

QUANTITATIVE METHODS FOR CALIBRATED SPATIAL  
MEASUREMENTS OF LARYNGEAL PHONATORY  
MECHANISMS

By

Hamzeh Ghasemzadeh

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Communicative Sciences and Disorders - Doctor of Philosophy  
Computational Mathematics Science and Engineering - Dual Major

2020

## **ABSTRACT**

### **QUANTITATIVE METHODS FOR CALIBRATED SPATIAL MEASUREMENTS OF LARYNGEAL PHONATORY MECHANISMS**

By

Hamzeh Ghasemzadeh

The ability to perform measurements is an important cornerstone and the prerequisite of any quantitative research. Measurements allow us to quantify inputs and outputs of a system, and then to express their relationships using concise mathematical expressions and models. Those models would then enable us to understand how a target system works and to predict its output for changes in the system parameters. Conversely, models would enable us to determine the proper parameters of a system for achieving a certain output. Putting these in the context of voice science research, variations in the parameters of the phonatory system could be attributed to individual differences. Thus, accurate models would enable us to account for individual differences during the diagnosis and to make reliable predictions about the likely outcome of different treatment options. Analysis of vibration of the vocal folds using high-speed videoendoscopy (HSV) could be an ideal candidate for constructing computational models. However, conventional images are not spatially calibrated and cannot be used for absolute spatial measurements. This dissertation is focused on developing the required methodologies for calibrated spatial measurements from in-vivo HSV recordings. Specifically, two different approaches for calibrated horizontal measurements of HSV images are presented. The first approach is called the indirect approach, and it is based on the registration of a specific attribute of a common object (*e.g.* size of a lesion) from a calibrated intraoperative still image to its corresponding non-calibrated in-vivo HSV recording. This approach does not require specialized instruments and can be implemented in many clinical settings. However, its validity

depends on a couple of assumptions. Violation of those assumptions could lead to significant measurement errors. The second approach is called the direct approach, and it is based on a laser-projection flexible fiberoptic endoscope. This approach would enable us to make accurate calibrated spatial measurements. This dissertation evaluates the accuracy of the first approach indirectly, and by studying its underlying fundamental assumptions. However, the accuracy of the second approach is evaluated directly, and using benchtop experiments with different surfaces, different working distances, and different imaging angles. The main significances and contributions of this dissertation are the following: (1) a formal treatment of indirect horizontal calibration is presented, and the assumptions governing its validity and reliability are discussed. A battery of tests is presented that can indirectly assess the validity of those assumptions in laryngeal imaging applications; (2) recordings from pre- and post-surgery from patients with vocal fold mass lesions are used as a testbench for the developed indirect calibration approach. In that regard, a full solution is developed for measuring the calibrated velocity of the vocal folds. The developed solution is then used to investigate post-surgery changes in the closing velocity of the vocal folds from patients with vocal fold mass lesions; (3) the method for calibrated vertical measurement from a laser-projection fiberoptic flexible endoscope is developed. The developed method is evaluated at different working distances, different imaging angles, and on a 3D surface; (4) a detailed analysis and investigation of non-linear image distortion of a fiberoptic flexible endoscope is presented. The effect of imaging angle and spatial location of an object on the magnitude of that distortion is studied and quantified; (5) the method for calibrated horizontal measurement from a laser-projection fiberoptic flexible endoscope is developed. The developed method is evaluated at different working distances, different imaging angles, and on a 3D surface.

I would like to dedicate this dissertation to the sparking light of my heart, Marjan, who has always been a true source of strength, passion, and support for me. The peace and comfort that I have found in her, have always given me the power to overcome all barriers and difficulties that I have been facing.

I would like to dedicate this dissertation to my lovely parents, whom I haven't seen for a long time, and miss deeply.



## **ACKNOWLEDGEMENTS**

I would like to express my gratitude to all who have contributed to my academic and personal development. In the first place, I am deeply grateful to my Ph.D. advisor, Dr. Dimitar Deliyski. He helped me to have a smooth and pleasant transitioning from engineering into science. I appreciate the flexibility that he offered, and his welcoming attitude toward new ideas. His approach allowed me to feel confident to combine my engineering skills with the scientific knowledge that I learned. Without his help I would not feel confident for starting my independent scholarship and academic career. I would also like to express my great appreciation to my committee members, Dr. Eric Hunter, Dr. Adam Alessio, Dr. Maryam Naghibolhosseini, and Dr. Dirk Colbry for their persistent help and guidance.

This dissertation was partially supported by the Michigan State University Foundation, the Council of Academic Programs in Communication Sciences and Disorders (CAPCSD) 2020 Ph.D. Scholarship, and the National Institutes of Health (NIH) - National Institute on Deafness and Other Communication Disorders (grants: R01 DC017923, R01 DC007640, P50 DC01546).

## TABLE OF CONTENTS

<b>LIST OF TABLES .....</b>	<b>x</b>
<b>LIST OF FIGURES .....</b>	<b>xiii</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
1.1. Background .....	1
1.2. Significance and rational.....	9
1.3. Structure of the dissertation and the research questions .....	14
1.4. Recordings setup and characteristics.....	24
1.4.1. Benchtop recording setup .....	24
1.4.2. Recording protocol .....	25
<b>CHAPTER 2: INDIRECT HORIZONTAL CALIBRATION OF IN-VIVO HSV</b>	
<b>RECORDINGS .....</b>	<b>28</b>
2.1. Introduction .....	29
2.2. Aim and hypothesis.....	32
2.3. Material and method.....	33
2.3.1. Participants and data acquisition .....	33
2.3.2. Indirect calibration principles and assumptions .....	34
2.3.2.1. Indirect calibration for between-subject size comparison .....	35
2.3.2.2. Indirect calibration for within-subject size comparison.....	37
2.3.3. Evaluation of indirect calibration .....	40
2.3.3.1. Registration uncertainty test .....	41
2.4. Experiments and results .....	42
2.4.1. Experiment1: Efficacy of registration uncertainty test.....	43
2.4.1.1. Database.....	43
2.4.1.2. Method .....	43
2.4.1.3. Results.....	44
2.4.2. Experiment 2: Effect of phonatory configuration on the calibrated length.....	45
2.4.2.1. Experiment 2.a: Vocal fold length attributes .....	46
2.4.2.1.1. Database.....	46
2.4.2.1.2. Method .....	47
2.4.2.1.3. Results.....	48
2.4.2.2. Experiment 2.b: Vocal fold width.....	49
2.4.2.2.1. Database.....	49
2.4.2.2.2. Method .....	49
2.4.2.2.3. Result .....	49
2.4.2.3. Experiment 2.c: Blood vessel on a vocal fold.....	52
2.4.2.3.1. Database.....	52
2.4.2.3.2. Method .....	53
2.4.2.3.3. Results.....	53
2.4.2.4. Experiment 2.d: Blood vessel on a nearby tissue .....	54

2.4.2.4.1. Database.....	54
2.4.2.4.2. Method.....	54
2.4.2.4.3. Results.....	55
2.4.3. Experiment 3: Selecting the most suitable common attribute.....	56
2.4.3.1. Experiment 3a: Registration uncertainty of different common attributes .....	56
2.4.3.2. Experiment 3b: Size consistency of different common attribute.....	57
2.4.3.2.1. Method.....	57
2.4.3.2.2. Results.....	58
2.5. Discussions.....	59
2.6. Conclusions.....	63

### **CHAPTER 3: APPLICATION OF INDIRECT HORIZONTAL CALIBRATION TO KINEMATIC MEASUREMENTS FROM IN-VIVO HSV RECORDINGS .....**

3.1. Introduction.....	66
3.2. Aim and hypothesis.....	67
3.3. Material and Method.....	71
3.3.1. Participants and data acquisition .....	71
3.3.2. Approach and measurements.....	73
3.3.2.1. Temporal segmentation.....	73
3.3.2.2. Motion compensation.....	75
3.3.2.3. Rotation correction.....	77
3.3.2.4. Spatial segmentation .....	80
3.3.2.5. Horizontal calibration .....	84
3.3.2.6. Velocity measurements.....	85
3.4. Experiments and results .....	87
3.4.1. Experiment1: Post-surgery changes in closing velocity.....	88
3.4.2. Experiment2: Post-surgery similarity between the two vocal folds.....	93
3.4.3. Experiment3: Effect of lesion size on post-surgery changes.....	95
3.5. Discussions.....	97
3.6. Conclusions .....	100

### **CHAPTER 4: DIRECT VERTICAL CALIBRATION OF HSV RECORDINGS .....**

4.1. Introduction.....	103
4.2. Aim and hypothesis.....	109
4.3. Material and method.....	110
4.3.1. Laser-projection endoscope.....	110
4.3.2. Calibration protocol and recordings .....	111
4.3.3. Measuring vertical distance.....	114
4.3.3.1. Compensating for the lens-coupler parameters.....	114
4.3.3.1.1. Recording model.....	115
4.3.3.1.2. Automatic estimation of the mapping.....	117
4.3.3.2. Algorithm for distance estimation .....	120
4.3.3.2.1. Automatic detection of laser points .....	120
4.3.3.2.2. Vertical distance decoding.....	122
4.4. Experiments and results .....	124
4.4.1. Experiment1: Evaluation of preprocessing components .....	124

4.4.1.1. Experiment1a: Evaluation of FOV and the fiducial finder modules .....	125
4.4.1.2. Experiment1b: Evaluation of the laser finder module .....	125
4.4.2. Experiment2: Displacement analysis and vertical resolution of the system .....	127
4.4.3. Experiment3: Evaluation of vertical distance measurements.....	129
4.5. Discussions.....	132
4.6. Conclusion.....	135
<b>CHAPTER 5: NON-LINEAR IMAGE DISTORTIONS IN FLEXIBLE FIBEROPTIC</b>	
<b>ENDOSCOPES .....</b>	<b>136</b>
5.1. Introduction .....	137
5.2. Aim and hypothesis.....	140
5.3. Optical principles of image formation .....	141
5.4. Material and method.....	145
5.4.1. Recording instrumentation and setup .....	145
5.4.2. Datasets.....	146
5.4.3. Automatic detection of grid lines .....	148
5.4.4. Pixel size.....	149
5.5. Experiments and results .....	150
5.5.1. Experiment 1: Differences between grid sizes .....	150
5.5.2. Experiment 2: Effect of spatial location.....	152
5.5.3. Experiment 3: Effect of the tilting angle .....	157
5.6. Discussions.....	164
5.7. Conclusions .....	167
<b>CHAPTER 6: DIRECT HORIZONTAL CALIBRATION OF HSV RECORDINGS .....</b>	<b>169</b>
6.1. Introduction .....	170
6.2. Aim and hypothesis.....	173
6.3. Material and method.....	174
6.3.1. Datasets.....	176
6.3.2. Segmentation and preprocessing .....	178
6.3.3. Horizontal calibration method .....	180
6.3.4. Horizontal measurement method.....	182
6.3.5. Estimation of the working distance .....	183
6.4. Experiments and results .....	186
6.4.1. Experiment 1: Accuracy of vertical measurements.....	186
6.4.2. Experiment 2: Performance of radial horizontal measurements .....	187
6.4.3. Experiment 3: Performance of central angle estimation .....	190
6.4.4. Experiment 4: Performance of general horizontal measurements.....	191
6.5. Discussion.....	194
6.6. Conclusion .....	196
<b>CHAPTER 7: VALIDITY AND ACCURACY OF HORIZONTAL AND VERTICAL</b>	
<b>MEASUREMENTS BASED ON DIRECT CALIBRATION.....</b>	<b>198</b>
7.1. Introduction .....	199
7.2. Aim and hypothesis.....	201
7.3. Material and method.....	203

7.3.1. Material and method for the effect of the imaging angle .....	203
7.3.1.1. Data acquisition .....	203
7.3.1.2. Database.....	204
7.3.1.2.1. Database for vertical measurements .....	206
7.3.1.2.2. Database for horizontal measurements .....	208
7.3.1.3. Analysis and measurements from a tilted surface.....	210
7.3.1.3.1. Vertical measurements from a tilted surface .....	210
7.3.1.3.2. Horizontal measurements from a tilted surface .....	213
7.3.2. Material and method for the effect of the 3D surface.....	213
7.3.2.1. Data acquisition .....	213
7.3.2.2. Analysis and measurements from a 3D surface.....	215
7.3.2.2.1. Vertical measurements from a 3D surface.....	215
7.3.2.2.2. Horizontal measurements from a 3D surface.....	217
7.4. Experiments and results .....	218
7.4.1. Experiment1: effect of the imaging angle .....	218
7.4.1.1. Experiment1a: effect of imaging angle on calibrated vertical measurements ....	219
7.4.1.2. Experiment1b: effect of imaging angle on calibrated horizontal measurements	222
7.4.2. Experiment2: effect of a 3D surface.....	225
7.4.2.1. Experiment2a: effect of a 3D surface on calibrated vertical measurements.....	225
7.4.2.2. Experiment2b: effect of a 3D surface on calibrated horizontal measurements ..	228
7.5. Discussions.....	230
7.6. Conclusions .....	235
<b>CHAPTER 8: SUMMARY OF THE FINDINGS.....</b>	<b>236</b>
8.1. Specific contributions of each dissertation chapter.....	238
8.2. Directions for further investigations.....	242
<b>REFERENCES.....</b>	<b>245</b>

## LIST OF TABLES

Table 1.1. Summary of different chapters of the dissertation.....	23
Table 2.1. Descriptive statistics of intra-sample registration variability. ....	45
Table 2.2. Descriptive statistics of the <i>mm</i> size of attributes of vocal fold length. ....	48
Table 2.3. Descriptive statistics of the <i>mm</i> width of the vocal fold.....	52
Table 2.4. Descriptive statistics of the <i>mm</i> size of attributes of a blood vessel on the vocal fold.. .....	54
Table 2.5. Descriptive statistics of the <i>mm</i> size of attributes of the blood vessel on a nearby tissue. .....	55
Table 2.6. Descriptive statistics of registration uncertainty for different selections of the common attribute. ....	56
Table 2.7. Individual differences in registration uncertainty of each common attribute.....	57
Table 2.8. Descriptive statistics of $\gamma$ for different selections of the common attribute. ....	58
Table 2.9. Individual trends regarding the size consistency of different common attributes. ....	59
Table 2.10. Comparing suitability of different common attributes for indirect calibration of vocal folds. ....	62
Table 3.1. Demographic and diagnosis information of the included subjects.....	73
Table 3.2. Descriptive statistics of closing velocity at different scanning lines (mean $\pm$ std).....	91
Table 3.3. Results of the paired-sample <i>t</i> -test for the closing velocity at different scanning lines. .....	91
Table 3.4. Descriptive statistics of closing velocity at different scanning lines (mean $\pm$ std).....	92
Table 3.5. Results of the paired-sample <i>t</i> -test for the closing velocity of the vocal fold with a lesion at different scanning lines. ....	92
Table 3.6. Results of the paired-sample <i>t</i> -test for the closing velocity of the vocal fold without a lesion at different scanning lines.....	93
Table 3.7. Results of paired-sample <i>t</i> -test for pre- and post-surgery recordings.....	94

Table 3.8. Correlation between post-surgery changes in the closing velocity and the area of the lesion. ....	97
Table 4.1. Literature-based taxonomy of different imaging systems with laser projection. These abbreviations were used in the table: VSB (videostroboscopy), HSV (high-speed videoendoscopy), 3D (three-dimensional reconstruction), <i>nm</i> (nanometer), mW (milli Watt). ....	108
Table 4.2. Statistics of the measurement error. All measurements have the unit of <i>mm</i> and the number in parentheses signifies the number of functions that were used in the measurements.....	131
Table 4.3. Results of correlation test for vertical measurement errors. The symbol $\epsilon$ means $p < 0.00001$ . ....	132
Table 5.1. Actual values of working distance and tilting angle for each target group. The first number represents the actual working distance in <i>mm</i> , and the second number the actual tilting angle in degree. ....	147
Table 5.2. Results of 2×2 robust ANOVA.....	151
Table 5.3. Descriptive statistics of pixel sizes. ....	152
Table 5.4. Results of 2×4 robust ANOVA.....	153
Table 5.5. Estimated values of pixel size.....	156
Table 5.6. Results of 7×4×3 ANOVA for trimmed means. ....	159
Table 5.7. The percentage of difference at the back and front peripheries from different working distances and tilting angles. ....	163
Table 5.8. Estimated uncalibrated length ( <i>i.e.</i> pixel length) of a 2 <i>mm</i> object at different locations of the FOV and different tilting angles. ....	164
Table 6.1. Correlation coefficients of the uniform model for radial measurement error. The symbol $\epsilon$ denotes a $p < 0.0001$ .....	188
Table 6.2. Correlation coefficients of the non-uniform model for radial measurement error. The symbol $\epsilon$ denotes a $p < 0.0001$ .....	189
Table 6.3. Accuracy of radial measurements from the uniform and the non-uniform models in different ranges of working distance.....	190
Table 7.1. The estimated working distance from the 3D surface .....	215

Table 7.2. Results of multiple linear regression for the JOV vertical measurement model. The symbols wd, a, and $\epsilon$ stands for the working distance, the imaging angle, and $p < 0.00001$ . .....	221
Table 7.3. Results of multiple linear regression for the PCA vertical measurement model. The symbols wd, a, and $\epsilon$ stands for the working distance, the imaging angle, and $p < 0.00001$ . .....	221
Table 7.4. Results of multiple linear regression for the uniform model for horizontal measurements. The symbols wd, a, and $\epsilon$ stands for the working distance, the imaging angle, and $p < 0.00001$ .....	224
Table 7.5. Results of multiple linear regression for the non-uniform model for horizontal measurements. The symbols wd, a, and $\epsilon$ stands for the working distance, the imaging angle, and $p < 0.00001$ .....	224
Table 7.6. Results of $2 \times 5$ ANOVA for vertical measurement errors .....	227
Table 7.7. Mean percent error and mean percent magnitude of error for vertical measurement.. .....	228
Table 7.8. Results of two-way ANOVA for horizontal measurement errors .....	229
Table 7.9. Mean percent error and mean percent magnitude of error for horizontal measurement. .....	230



## LIST OF FIGURES

Figure 1.1. Illustration of a horizontal plane and the vertical direction.....	15
Figure 1.2. The employed setup for benchtop recordings. ....	25
Figure 1.3. Examples of incorrect placements of the FOV in the image frame. ....	25
Figure 1.4. Some examples of the FOV with unclear edges.....	26
Figure 1.5. Some examples of the inadequate border between the FOV and the image frame.....	26
Figure 1.6. An example image with non-visible fiducial marker. ....	27
Figure 2.1. Two examples of intraoperative calibrated images, taken from references <sup>190</sup> and <sup>93</sup> . .....	31
Figure 2.2. Results of registration uncertainty test: (A) values of interquartile range for different patients and (C)omfortable and (H)igh pitch phonations, (B) estimated pdf of interquartile range over all recordings. ....	44
Figure 2.3. Boxplot of <i>mm</i> size of vocal fold length attribute of each subject for (C)omfortable and (H)igh pitch phonations. ....	48
Figure 2.4. Measurement of the vocal fold width: (A) the reference image with designated vocal fold and the target anchor point, (B) the measurement steps.....	51
Figure 2.5. Boxplot of <i>mm</i> size of vocal fold width of each subject for (C)omfortable and (H)igh pitch phonations. ....	51
Figure 2.6. Boxplot of <i>mm</i> size of an attribute of blood vessels on the vocal fold of each subject for (C)omfortable and (H)igh pitch phonations. ....	53
Figure 2.7. Boxplot of <i>mm</i> size of an attribute of blood vessels on a nearby tissue of each subject for (C)omfortable and (H)igh pitch phonations. ....	55
Figure 3.1. Result of registration uncertainty test for included subjects. ....	72
Figure 3.2. Intraoperative images from subjects with high uncertainty registration. ....	72
Figure 3.3. An example of temporal segmentation outcome. ....	75
Figure 3.4. An example of motion compensation: (A) kymogram before motion compensation, (B) kymogram after motion compensation. ....	76

Figure 3.5. Effect of endoscopic rotation on the kymogram: (A) kymogram before rotation compensation, (B) kymogram after rotation compensation.....	77
Figure 3.6. Estimation of the GAW: (A) pdf of the red channel, and the computed black threshold, (B) GAW estimate after applying the black threshold.....	79
Figure 3.7. Rotation correction for a frame of data: (A) before correction, (B) segmented glottis with the fitted line on the first moment of inertia from each row, (C) after correction. ....	80
Figure 3.8. Temporal curve fitting results: (A) local black reference estimation, the red window shows the search window, (B) ROI segmentation, (C) detection of vocal fold edges. ....	82
Figure 3.9. Spatial curve fitting results: (A) outlier removal step, (B and C) segmented edges of the vocal fold for two different timepoints. ....	83
Figure 3.10. Selection of the data: (A) the least stable portion of a phonation, (B) the most stable portion of a phonation. ....	86
Figure 3.11. Boxplot of closing phase maximum velocity for different subjects pre- and post-surgery: (A) box plot of $v_{\max}^L$ , (B) box plot of $v_{\max}^R$ .....	89
Figure 3.12. Boxplot of closing phase maximum velocity for different subjects pre- and post-surgery: (A) box plot of $v_{\text{lesion}}^L$ , (B) box plot of $v_{\text{lesion}}^R$ .....	89
Figure 3.13. Boxplot of closing phase maximum velocity for different subjects pre- and post-surgery: (A) box plot of $v_{\text{mid}}^L$ , (B) box plot of $v_{\text{mid}}^R$ . ....	90
Figure 3.14. Boxplot of closing phase maximum velocity for the vocal fold with the lesion and the (cont)ralateral side for different subjects: (A) pre-surgery condition, (B) post-surgery condition. ....	95
Figure 3.15. The relationship between area of a lesion and its post-surgery improvement: (A) The blue region shows the lesion (B) Scatter plot of post-surgery changes in $v_{\text{lesion}}^{\text{with}}$ vs. area of the lesion. The outliers are marked by a red circle. ....	96
Figure 4.1. Schematics of different laser projection techniques with the principle of encoding the vertical and/or horizontal distances: (A) laser triangulation method, (B) structured light projection, (C) a combined technique. Green and red dots depict hypothetical positions of the laser pattern at two different vertical distances. ....	106
Figure 4.2. The calibrated flexible endoscope with an insertion tube diameter of 4.9 mm and its main components. ....	111
Figure 4.3. A diagram of the recording conditions. ....	112

Figure 4.4. Calibration setup: (A) measuring the distance to the tip of the endoscope, (B) measuring the distance to the fixture. ....	113
Figure 4.5. Model for compensating the recording parameters of the system. ....	117
Figure 4.6. The intensity of the laser points: (A) sum of the intensity of pixels on the rows, (B) original image, (C) sum of the intensity of pixels on the columns. ....	121
Figure 4.7. Position of each laser point as a function of working distance where each color shows a different laser point: (A) $x$ - $y$ coordinates as a function of working distance, (B) $x$ -coordinate as a function of working distance, (C) $y$ -coordinate as a function of working distance. ....	123
Figure 4.8. Distribution of the variability in the output of FOV and the fiducial finder modules: (A) distribution of the centralized coordinates of the FOV center, (B) distribution of the centralized radius of FOV, (C) distribution of the centralized fiducial angle. ....	126
Figure 4.9. Distribution of the variability in the output of the laser finder module. ....	126
Figure 4.10. Displacement analysis of the laser points as the working distance is changing: (A) the magnitude of variation in the position of the laser points as the working distance is changing from 35 $mm$ to a new distance, (B) the magnitude of variation in the position of the laser points for 1 $mm$ decrement at different working distances. ....	128
Figure 4.11. The behavior of different laser points: (A) indexing used in this chapter, (B) the average magnitude of displacement of each laser point. ....	128
Figure 4.12. The average magnitude of displacement of each laser point. ....	129
Figure 4.13. Boxplot of vertical measurement errors at different working distances: (A) results from all functions, (B) results when the functions from the top row are discarded. ..	130
Figure 5.1. Optical principles of image formation: (A) parameters of the Snell's law, (B) image formation in the Gaussian optics model. ....	142
Figure 5.2. Effects of tilting the target surface on the geometry of the acquired images. ....	144
Figure 5.3. A schematic for measuring the tilting angle. ....	147
Figure 5.4. Automatic detection of the grid lines: (A) recording from 1 $mm$ grids at the working distance of 10 $mm$ , (B) the binary image showing the locations of the minima, (C) fitted second-order polynomials on the locations of the minima. ....	149
Figure 5.5. Groupings for experiments 1 and 2: (A) the solid red blocks and the patterned blue blocks denote the center and the periphery groups, (B) the selected sides of an example image. The Center of the image-FOV is denoted by a green cross mark. ....	151

Figure 5.6. Variation in pixel size for different working distances and groups.....	153
Figure 5.7. Boxplots of the pixel size for different groups and working distances. ....	154
Figure 5.8. Estimation of the dependence of pixel size on its spatial location, (A) selected line segments are shown in green dashed line, and the center of the image-FOV is denoted with a red cross mark, (B) dependence of pixel size on its distance from the center of the image-FOV and the working distance. The negative distance means blocks that were below the center of the image-FOV.....	155
Figure 5.9. Groupings for experiment 3. Solid red lines denote the back group, dotted green lines denote the middle group, and dashed blue lines denote the front group, (A) groupings at the working distance of 5 <i>mm</i> , (B) groupings at the working distance of 15 <i>mm</i> ...	158
Figure 5.10. Values of the mean and standard deviation of pixel size: (A) working distance of 5 <i>mm</i> , (B) working distance of 10 <i>mm</i> , (C) working distance of 15 <i>mm</i> , (D) working distance of 20 <i>mm</i> . ....	159
Figure 5.11. (A) The selected line segments are shown in green dashed lines, and the center of the image-FOV is denoted with a red cross mark. (B) Dependence of pixel size on its distance from the center of the image-FOV and the tilting angle at the working distance of 15 <i>mm</i> . ....	161
Figure 5.12. Dependence of location with the highest spatial resolution on the tilting angle. ....	161
Figure 6.1. Relationship between the length of an object ( $h_o$ ) and its image ( $h_i$ ) in an axially symmetrical optical system.....	174
Figure 6.2. Effects of working distance and spatial location on horizontal measurements: (A) working distance of 2.87 <i>mm</i> , (B) working distance of 2.24 <i>mm</i> . ....	176
Figure 6.3. The data for evaluation of central angle measurement: (A) the custom-designed grid, (B) segmented radial lines. ....	178
Figure 6.4. Segmentation of a circular grid: (A) horizontal and vertical strips with their respective summations, (B) final segmented circles after the fine-tuning stage.....	179
Figure 6.5. Models for horizontal measurements: (A) non-uniform model, (B) uniform model. ....	181
Figure 6.6. Expressing a general measurement in terms of radial measurements. ....	183
Figure 6.7. Mean absolute error (MAE) of original and the proposed PCA method for different values of the standard angle.....	184

Figure 6.8. Performance of estimating the working distance: (A) indexing of the laser points, (B) measurement accuracy of different laser points, (C) effect of working distance. ....	187
Figure 6.9. Performance of uniform model for radial measurements: (A) effect of object length, (B) effect of working distance. ....	188
Figure 6.10. Performance of non-uniform model for radial measurements: (A) effect of object length, (B) effect of working distance. ....	189
Figure 6.11. Boxplot of angle estimation error computed from set3.....	191
Figure 6.12. Performance of uniform model for general measurements: (A) effect of working distance, (B) effect of object length.....	192
Figure 6.13. Performance of non-uniform model for general measurements: (A) effect of working distance, (B) effect of object length.....	193
Figure 7.1. Imaging from a tilted surface: (A) effect of tilting the target surface on different objects within the FOV, (B) effect of tilting the target surface on the geometry of the FOV... ..	202
Figure 7.2. The effect of tilting the target surface vs. changing the imaging angle. ....	204
Figure 7.3. Recordings from a circular grid at the working distance of 8.66 <i>mm</i> : (A) the tilting angle of 15°, (B) the tilting angle of -15°, (C) tilting angle of 0° after making the endoscopic tip perpendicular to the target surface.....	205
Figure 7.4. The setup that allowed precise adjustment of the distal tip of the endoscope.....	206
Figure 7.5. A diagram of the recording conditions. Different colors correspond to the FOV cone at different working distances. To simplify the visualization, the target surface is kept fixed and the camera is displaced. However, in the experiments it was the other way around. ....	207
Figure 7.6. Placement of the 5- <i>mm</i> line segment inside the FOV for horizontal measurements. ....	209
Figure 7.7. A schematic for estimation of the true vertical distance of the laser point B.....	210
Figure 7.8. An example of computing the <i>mm</i> distance between two laser points B and R.....	212
Figure 7.9. The data used for investigating the effect of 3D shape: (A) the 3D model, (B) fiducial markers, (C) the printed composite model .....	214

Figure 7.10. The outcome of the registration process: (A) a composite image before the registration. Centers of the fiducial markers are marked with a red dot. Centers of the laser points are marked with a green cross mark, (B) the registration outcome for the composite image. ....	216
Figure 7.11. Boxplots of vertical measurement error using the JOV model at different working distances and imaging angles. ....	219
Figure 7.12. Boxplots of vertical measurement error using the PCA model at different working distances and imaging angles. ....	220
Figure 7.13. Boxplots of horizontal measurement error from the uniform model at different working distances and imaging angles. ....	222
Figure 7.14. Boxplots of horizontal measurement error from the non-uniform model at different working distances and imaging angles. ....	223
Figure 7.15. Performance of the PCA model on a flat surface: (A) vertical measurement errors, (B): magnitude of vertical measurement errors. ....	226
Figure 7.16. Performance of the vertical measurement errors on a 3D surface: (A) boxplot of error, (B) boxplot of the magnitude of error.....	226
Figure 7.17. Performance of the horizontal measurement errors on a 3D surface: (A) boxplot of error, (B) boxplot of the magnitude of error. ....	229
Figure 8.1. Graphical representation of the relationships among the chapters of this dissertation. ....	238

## CHAPTER 1: INTRODUCTION

### 1.1. Background

Voice and speech are the main communication channels for expressing our ideas, thoughts, and emotions. Furthermore, we use them for artistic creations (*e.g.* singing). Therefore, it is not very surprising that degradation in speech production and voice quality would lead to serious problems in communication. Several studies have shown that degraded voice quality is associated with significant negative bias and attitude of the society, and hence could have a negative effect on the social life of people with voice disorders.<sup>1-4</sup> Previous studies have also confirmed that the psychological and emotional burden of degraded voice quality could be very high and it could lead to a serious deterioration in the perceived quality of life.<sup>5-7</sup> Furthermore, voice and speech have a significant role in the career of professional voice users (*i.e.* news anchors, teachers, singers, etc.), and maintaining the voice quality becomes even more important for a large population. Considering the high prevalence of voice and speech problems in the general population (5 million school-age annually<sup>8</sup> and between 3% to 9% of the whole population<sup>9,10</sup>) and even higher incidence rate for professional voice users (*e.g.* prevalence between 50% to 80% in lifetime of teachers<sup>11</sup>), a very large population would benefit from quantitative research in this area.

The voice production system can be modeled as a dynamic system that takes a well-regulated air stream as the input and modifies it in a certain way to produce a specific acoustic signal in the output. Glottis in the larynx is the first and the primary place for conversion of airflow into the acoustic signal. Therefore, determining and understanding the behavior of vocal folds and their vibratory characteristics is important for the advancement of voice science research and clinical

applications. Considering the position of the larynx in the airway, direct assessment and observation of the larynx and the vocal folds have been challenging. Consequently, for a long-time people were relying on the output of the phonatory system (*i.e.* the acoustic signal and the aerodynamic measurements) for studying the behavior and functional assessment of the vocal folds. These methods are called output-based approaches in the rest of this dissertation. However, advancements in technology have made it possible to directly observe and study different parts of the phonatory system, including the vibration of the vocal folds. These approaches are called internal-based approaches in the rest of this dissertation, as they provide a direct means for studying the internal states and function of the phonatory system.

The output of the voice production system is an aerodynamic-acoustic phenomenon. Therefore, it is possible to analyze the aerodynamic and acoustic signals of the voice and derive certain information regarding the vibration of the vocal folds. Considering the interaction of the airflow with the cyclic abduction (*i.e.* opening) and adduction (*i.e.* closing) of the vocal folds, airflow could provide a very good means for studying the vibration of the vocal folds. In normal voice production, the glottal flow starts with the opening of the glottis and during a specific time, it reaches the maximum (positive slope) at this point the vocal folds are fully abducted. Then, the vocal folds start to move toward the midline and hence the flow starts to decline (negative slope) until the vocal folds are fully adducted and hence the flow stops (or reaches its minimum). The vocal folds remain adducted for a specific time and then this cycle is repeated again.<sup>12,13</sup> Several important characteristics from this cyclic behavior have been defined for differentiation between different modes of phonation and diagnostic purposes.<sup>12,13</sup> Open quotient is defined as the portion of a cycle that the glottis is open. Additionally, the duration of the positive slope is longer than the negative slope in normal phonation<sup>14,15</sup> which results in right-skewed flow measurements. This



characteristic has been associated with the inertia of the air in the vocal tract<sup>14,15</sup> and is an important feature of the glottal flow. Skewing quotient can capture and quantify this feature and is defined as the ratio of the duration of the increasing flow to the duration of the decreasing flow.<sup>12</sup> These two measurements provide a rough picture from the shape of the glottal flow and provide information about the underlying mechanism of the voice production system and the produced voice quality. For example, high values of open quotient have been associated with a breathy quality<sup>16–19</sup> and low values may indicate a pressed phonation.<sup>20–22</sup> The maximum flow declination rate is another important output-based measurement which is defined as the maximum value of the derivative of glottal airflow.<sup>23</sup> This measurement has been used as an indirect approach for estimation of the closing velocity of the vocal folds.<sup>24</sup> Maximum flow declination rate is closely related to collision forces of the vocal folds<sup>25–27</sup> and the produced acoustic output.<sup>23,28,29</sup>

Acoustic assessment is another example of output-based approaches and it accounts for the majority of the studies in voice research.<sup>30</sup> Two main approaches of auditory perceptual assessment and acoustic measurement can be identified for the evaluation of the acoustic data. Auditory perceptual assessment is considered as the gold standard<sup>31–34</sup>, and the most commonly used technique in the clinical settings.<sup>35</sup> This approach does not require additional investment, additional equipment, and technical knowledge. In perceptual assessment, the quality of the voice has been evaluated using qualitative terms such as wet<sup>36,37</sup>, gurgly<sup>38,39</sup>, breathy<sup>40,41</sup>, hoarse<sup>41,42</sup>, harsh<sup>43,44</sup>, rough<sup>40,41</sup>, creaky<sup>45,46</sup>, strained<sup>40,41</sup>, and many more. These terms refer to qualitative features of a voice that are clear enough for the general population such that almost everyone understands them.<sup>34</sup> Consequently, they are very useful for communication between different people and hence would facilitate client-clinician communications. On the other hand, researches have indicated low reliability of auditory perceptual evaluations.<sup>31,47,48</sup> More recent studies have

tried to remove some of the subjectivities from the evaluation, and hence to reduce the inter-rater and intra-rater variabilities by providing standard anchors or using matching tasks.<sup>49–52</sup> Variabilities in the evaluation terms and protocols were another big issue with the perceptual approaches. Efforts have been made to standardize the routine and the scales for evaluation of the voice quality. CAPEV<sup>40</sup> and GRBAS<sup>41</sup> are two widely used instruments for this purpose.

Acoustic measurements are objective approaches that have been designed based on signal processing techniques, and hence can alleviate some of the issues associated with the perceptual methods. Robustness to factors such as bias and variability is the primary advantage of acoustic measurements. Also, once these methods are developed, they provide fast and low-cost tools for assessing the voice in an automatic and repeatable fashion. Finally, since the procedures behind these measurements are known, their resolutions and sensitivities can be evaluated. Current objective measurements of voice quality can be grouped in four main categories of perturbation measurements (*e.g.* jitter<sup>53,54</sup>, pitch perturbation<sup>55</sup>, pitch perturbation quotient<sup>56</sup>, shimmer<sup>57</sup>, amplitude perturbation<sup>58</sup>, and amplitude perturbation quotient<sup>56</sup>), noise measurements (*e.g.* signal to noise ratio (SNR)<sup>59</sup>, harmonic to noise ratio (HNR)<sup>60</sup>, frequency domain HNR<sup>61</sup>, normalized noise energy<sup>62</sup>, glottal-to-noise excitation ratio<sup>63</sup>, the energy of the noise from a filter bank<sup>64</sup>), spectral and cepstral measurements (*e.g.* spectral slope<sup>65,66</sup>, cepstrum peak prominence (CPP)<sup>67</sup>, Mel-frequency cepstral coefficients<sup>68</sup>, energy and entropy of wavelet sub-bands<sup>69</sup>, and temporal and spectral dynamics of the speech<sup>70</sup>), and non-linear measurements (*e.g.* largest Lyapunov exponent<sup>71</sup>, correlation dimension<sup>72</sup>, and parameter of the phase-space<sup>73</sup>).

Using the output-based approaches for studying the voice production system is similar to reverse engineering. Output-based data are easier to collect; however, it is not an easy task to relate them back to their underlying mechanisms. Often, this step requires the assumption of a model

that describes the system very well. Additionally, there are several other factors that make things even more complicated. Based on control theory and mathematical analysis of dynamic systems, it is well-known that only under certain assumptions, internal states of a system can be inferred from its output.<sup>74,75</sup> Additionally, several characteristics of the speech production system (*e.g.* multiple to one mapping) may lead to an ambiguous interpretation of the underlying mechanism from the output. For example, researchers have shown that multiple significantly different articulatory configurations, can lead to the same acoustic measurements and output.<sup>76-78</sup> Considering the complex structure and interaction between intrinsic and extrinsic laryngeal muscles, and also their agonist-antagonist roles, a similar characteristic can be expected for the phonatory mechanism, too. Quantal and saturation effects are non-linear properties of the speech production system that describe stable regions such that changes within that region do not lead to a change in the acoustic output.<sup>79-84</sup> Interestingly, these characteristics are not unique to the articulatory system and also exist in the phonatory mechanism. For example, recent studies on the biomechanics of the larynx have indicated the existence of rich quantal regions in the larynx.<sup>85,86</sup> These characteristics facilitate motor planning and help with the production of stable sounds.<sup>87</sup> However, these features create ambiguity in determining the internal states of the system from its output. Based on these arguments, internal-based approaches are more favorable for studying the underlying mechanism of voice production and voice disorder.

Imaging techniques are probably the most important and popular internal-based approach for studying the voice. Imaging techniques can provide a wealth of information regarding the underlying mechanisms of voice production, their configuration, and their kinematics. Considering that vocal folds are vibrating at relatively high frequencies --with the typical range of 85-196 Hz for males, 155-334 Hz for females, and 208-440 Hz for children during normal speaking<sup>88</sup>--

imaging techniques should be able to track such frequencies. In fact, research has recommended a minimum of 4000 frames per second (fps) for a reliable functional assessment of the voice.<sup>89</sup> The existing imaging system can be classified based on different criteria. One important distinction can be made based on the imaging modality and how this frame-rate requirement is addressed. In that regard videostroboscopy (VSB), videokymography (VKG), and high-speed videoendoscopy (HSV) could be identified as the most common modalities for visualization of the vocal folds. Using a different criterion, imaging systems can be classified based on the type of endoscope that gets connected to the imaging system. Using this criterion two types of rigid and flexible systems could be identified. It is noteworthy that each of these factors would lead to different functionalities and applications for the acquired images. For example, the imaging modality (*i.e.* VSB, VKG, HSV) determines the type of phenomenon that can be captured and studied using the imaging system. However, the type of endoscopic instrument (*i.e.* flexible vs. rigid) determines the type of stimuli that can be elicited. Regardless of the employed imaging modality and the endoscopic instrument, acquired images can be evaluated with subjective visual assessment approaches<sup>89–96</sup> or objective measurements.<sup>97–103</sup>

To elaborate more on the effect of the imaging modality on the type of phenomenon that can be studied a brief introduction on principles of each imaging modality is presented. VSB system typically flashes a strobe light at specific phases of consecutive glottal cycles, and in this manner creates an illusion of slow-motion from vocal folds vibration.<sup>104,105</sup> Clearly, this technique requires a precise mechanism for estimation of the fundamental frequency and synchronization with it. Therefore, two important conditions for the correct functionality of VSB can be determined. First, the target phenomenon should be cyclic, and therefore it is not applicable to transient phenomena such as voice onset, voice offset, voice break, etc. Second, the target phenomenon should be nearly

periodic. Considering that fundamental frequency could be ambiguous in type2,3<sup>12</sup> and 4<sup>106</sup> voices --which correspond to many cases of disordered voices-- VSB does not represent a correct slow motion from highly dysphonic phonations. Additionally, it is a well-known phenomenon that depending on the sampling frequency, the strobe slow-motioned picture may appear freezing or even backward playing.<sup>107</sup> On the positive side, VSB can provide audio-synchronized visualization of the vocal folds which is very important for clinical evaluations.<sup>108</sup> Additionally, the distal-chip VSB systems can provide very high-quality images.<sup>109</sup> Therefore, despite its inherent flaws, it is still the gold-standard method for clinical evaluations.<sup>108,110,111</sup> VKG uses a different approach for visualization of the fast vibration of the vocal folds. The idea is to capture high-speed images from a single line of the vocal folds along its posterior-anterior axis, and then to stack them up and create a composite image.<sup>112</sup> VKG images can capture up to 8,000 images from the target section.<sup>112</sup> VKG captures the true behavior of the vocal fold and then show it in real-time; therefore, it is very appropriate for clinical evaluations.<sup>113</sup> VKG can demonstrate the existence of many vibratory characteristics including, subharmonics (*i.e.* type2 phonation<sup>12</sup>), left-right asymmetries, propagation of the mucosal waves, and open quotient.<sup>112,113</sup> However, limited spatial resolution (*i.e.* single line scanning) is the biggest limitation of VKG. HSV can provide full images at the rate of 20,000 fps or even higher<sup>89</sup>, and therefore can provide recordings with high temporal and spatial resolutions from vocal folds vibration. In comparison to VSB, HSV has a better temporal resolution and therefore can be used for studying aperiodic vibrations, as well as, transient phenomena. In comparison to VKG, HSV has a higher spatial resolution. This feature is necessary for studying the spatial aspects of the vibration such as spatial variations in the kinematics of the vocal folds. In summary, HSV captures vibration of the vocal folds as it is happening, and hence it could be the gold standard. More specifically, it can be used for validation of other

measurements<sup>100,114</sup> and also the validation of computational models<sup>115</sup>, which other imaging techniques cannot do as accurately. Finally, both VKG and VSB can be simulated from HSV recordings.<sup>111</sup> These characteristics make HSV the ideal tool for studying normal and disordered phonations. However, these significant benefits come at a price. Considering the huge data generated from HSV systems, manual analysis is not a viable solution and automated methods should be developed for the analysis of HSV recordings. Processing of HSV recordings typically consists of multiple steps including segmentation, motion compensation, and measurement. Segmentation is the first step in the analysis of the HSV recordings, where the phenomenon of interest is extracted.<sup>111</sup> Depending on the desired phenomenon segmentation can be performed in temporal<sup>111,116</sup>, spatial<sup>98,117–122</sup>, and spatial-temporal<sup>123,124</sup> domains. Motion compensation is another important step that could remove artifacts introduced by movements of the camera or the endoscope.<sup>124,125</sup>

The endoscopic instrument also has significant impacts on the application of the acquired images. Rigid endoscopes provide images with better spatial resolutions and visual qualities.<sup>126,127</sup> They have minimum image distortion<sup>127</sup> and can provide significantly more diagnosis information for a wide variety of voice disorders including vocal fold lesions<sup>127,128</sup> and laryngopharyngeal reflux.<sup>127,129</sup> Therefore, rigid endoscopes are considered the “gold standard” for awake imaging conditions.<sup>127,129</sup> On the other hand, due to transoral insertion, the rigid endoscopes affect the voice and speech production systems. For example, to get a decent view from the larynx the tongue should be retracted unnaturally.<sup>130</sup> This means that only limited types of stimuli can be elicited. Additionally, the altered voice production system could raise some concerns regarding the validity of the acquired data. For example, a previous study has shown that the presence of a rigid endoscope could significantly change the fundamental frequency and the quality of the produced

voice.<sup>131</sup> The changes in the fundamental frequency may indicate the altered functionality of the phonatory mechanism in the presence of the scope. Also, the changes in the voice quality may indicate issues regarding the validity of subsequent measurements.

Flexible endoscopy does not interfere with articulators and speech can be produced with minimal interference; therefore, it could be more ecologically valid. Additionally, there are fewer restrictions on the type of stimuli that could be produced. Thus, flexible endoscopes could be used for analysis and studying of vibratory patterns of the vocal folds during connected speech.<sup>116</sup> This feature has made flexible scopes the instrument of choice for diagnosis and evaluation of most neurological voice disorders.<sup>126</sup> Flexible endoscopes can also provide the possibility of simultaneous aerodynamic measurements.<sup>132–134</sup> This characteristic could provide significant information about the complex interactions between kinematics, aerodynamics, and the produced acoustic of the phonatory system. Additionally, flexible endoscopes allow the complete visual examination and evaluation of the vocal tract.<sup>135</sup> Last but not least, flexible endoscopes have been associated with higher success rates in adult<sup>127</sup> and especially pediatric<sup>136,137</sup> populations. On the other hand, flexible endoscopes are more invasive and have been associated with more pain and discomfort even among adult subjects.<sup>138</sup> Additionally, flexible endoscopes have inferior image quality and spatial resolution.

## **1.2. Significance and rational**

The ability to perform measurements is an important cornerstone and the prerequisite of any quantitative research. Measurements allow us to quantify inputs and outputs of a system, and then to express their relationships using concise mathematical expressions and models. These models could help us then, to understand how that system works. Additionally, measurements could enable us to make intelligent and accurate predictions about the output of a system if certain

characteristics of that system are changed. Conversely, they could enable us to determine the proper parameters of the system for achieving a certain output. Obviously, quantitative research from the phonatory system is not an exception. Moreover, the goal of *predicting the output of the system for changes in the parameters of the system* has significant and practical implications for people with dysphonia. Specifically, dissimilar intervention outcomes in different patients could be due to their individual differences. In this sense, models could improve our ability to account for individual differences during the diagnosis and to improve the likelihood of reliable predictions about the outcome of different treatment options. In this sense, the likely outcome of different interventions could be predicted, and the best one could be selected. The existence of computational models that could link the input, parameters, and the output of the phonatory system together are important components for developing *precision-medicine* and *personalized* approaches to diagnosis and treatment of voice disorder.

The voice production mechanism can be modeled as a dynamic system with specific input, system parameters, and output. Interestingly, the required methodology for measuring the input and the output of this system on calibrated scales has been around for a long time.<sup>13</sup> Specifically, the air stream is the input of the phonatory system which can be measured on calibrated scales using airflow and air pressure measurements.<sup>23,28,139</sup> The acoustic signal is the output of this system, which can also be measured on a calibrated scale using sound pressure level.<sup>140–146</sup> However, this is not the case for the system parameters. That is, the required methodology for kinematic measurements of the vocal folds and spatial measurements from the larynx on calibrated scales are missing.<sup>122</sup> Such measurements are necessary for a wide range of computational approaches to study and understand the biomechanics and aerodynamics of the phonatory mechanism.<sup>147–152</sup> Additionally, the calibrated spatial measurement could be very valuable for



studying the developmental aspects of vocal fold vibration.<sup>153</sup> The primary goal of this dissertation is to fill this significant gap and to present methods for calibrated spatial measurements from in-vivo laryngeal HSV images.

Another significance of this research is its application in the advancement of evidence-based practice in the field of laryngology and speech-language pathology. Specifically, the efficacy and outcome of voice therapy are usually evaluated using auditory perceptual changes between pre- and post-therapy conditions.<sup>154-157</sup> A survey study from experienced speech-language pathologists showed that perceptual assessments were the most likely evaluation tool in the field.<sup>35</sup> Additionally, some researchers have used acoustic measurements as an objective alternative for quantification of the efficacy of the intervention.<sup>158-162</sup> However, both auditory perceptual methods, and acoustic measurements are based on the output of the phonatory system and have all of the mentioned limitations of the output-based measurements. Most importantly, it is not trivial how to infer physiological changes from changes in the acoustic. A more straightforward alternative would be direct measurements of the physiological changes due to the intervention. For example, the efficacy of a therapy on a nodule could be evaluated and quantified in terms of changes in the lesion size, or changes in kinematics and vibratory patterns of the vocal folds. Obviously, this approach could provide significant information and provide the required evidence on the efficacy of different therapies. However, measuring the lesion size and the computation of kinematic measures (*e.g.* velocity measures) require horizontally calibrated images. That is, we need to compare the lesion size pre- and post-intervention, or correlate changes in the lesion size to changes in the kinematics of the vocal folds post-intervention. The proposed research presents methods for calibrated horizontal measurements and has the potential of addressing this need, and therefore is very significant.

Velocity measures are important kinematic features that can capture the dynamics of vocal folds' vibration. Velocity measures can relate different aspects of the phonatory system together, and therefore are significant for voice science and clinical applications. For example, the closing velocity of the vocal folds relates to their collision forces.<sup>25-27</sup> It also relates to the maximum flow declination rate<sup>26,163</sup> and the maximum area declination rate<sup>102,164,165</sup>, which their effect on the average produced acoustic output<sup>29</sup> and the vocal intensity<sup>23,28</sup> have been established. Finally, higher closing velocity increases the energy of high-frequency components of the voice, which in turn may improve the speech intelligibility.<sup>166</sup> However, velocity is the calibrated displacement of an object with respect to time. Consequently, the computation of the velocity depends on calibrated temporal and spatial measurements. Time is already calibrated in cameras. Therefore, calibration of the spatial domain would pave the path for the computation of velocity measures. This dissertation presents different approaches for calibrated horizontal measurement from HSV images and hence has the potential of addressing this need.

Another significance of this dissertation is its possible application in the quantification of the vertical movements and displacements of the phonatory mechanism. Specifically, imaging techniques provide a direct method for observation and assessment of the larynx and hence are important parts of diagnosis and functional assessment of the voice production system.<sup>167-171</sup> However, images are two-dimensional representations of the real world. Considering that, the real world is happening in three-dimensional (3D) space, images would not be a true representation of the actual phenomena that are being captured. In other words, the vertical dimension is lost during the data acquisition process, and we could not measure the distance of an object from the camera. This lack of vertical component means that the vertical motion of the larynx and the vertical component of the vibration of the vocal folds could not be measured and studied. Multiple

modeling studies have predicted the significance and the role of the vertical component of the vibration of the vocal folds on phonation.<sup>164,172-175</sup> For example, the mucosal wave is a surface wave that propagates along the medial surface (*i.e.* from the lower to upper margin) of vocal folds and in the direction of the airflow.<sup>176,177</sup> Mucosal wave can also be expressed as a phase difference between the upper and lower margins of the vocal folds.<sup>174,177</sup> Several important aspects of the voice production system have been attributed to mucosal wave. For example, mucosal wave velocity has been associated with the phonation threshold pressure (PTP)<sup>152,178-180</sup> in the sense that, a larger vertical phase difference may lead to a lower PTP which may indicate an easier phonation.<sup>174</sup> Mucosal wave has also been associated with voice quality.<sup>91,177,181</sup> Last but not least, subjective evaluations of the magnitude of the mucosal wave from in-vivo recordings have been used for diagnosis<sup>95,113,135,182-184</sup> or measuring the efficacy of an intervention.<sup>185-189</sup> However, the mucosal wave is a vertical aspect of the phonatory mechanism and therefore the capability of vertical measurements is the prerequisite for its objective quantification. This dissertation uses a laser-projection endoscope and presents the method for vertical measurements that can address these needs.

In summary, the main significances and contributions of this dissertation are the following: (1) a formal treatment of indirect horizontal calibration is presented, and the principles governing its validity and reliability are discussed. A battery of tests is presented that can indirectly assess the validity of those assumptions in laryngeal imaging applications; (2) recordings from pre- and post-surgery from patients with vocal fold mass lesions are used as a testbench for the developed indirect calibration approach. In that regard, a full solution is developed for measuring the calibrated velocity of the vocal folds. The developed solution is then used to investigate post-surgery changes in the closing velocity of the vocal folds from patients with vocal fold mass

lesions; (3) the method for calibrated vertical measurement from a laser-projection fiberoptic flexible endoscope is developed. The developed method is evaluated at different working distances, different imaging angles, and on a 3D surface; (4) a detailed analysis and investigation of non-linear image distortion of a fiberoptic flexible endoscope is presented. The effect of imaging angle and spatial location of an object on the magnitude of that distortion is studied and quantified; (5) the method for calibrated horizontal measurement from a laser-projection fiberoptic flexible endoscope is developed. The developed method is evaluated at different working distances, different imaging angles, and on a 3D surface.

### **1.3. Structure of the dissertation and the research questions**

This dissertation is focused on developing the required methodologies for calibrated spatial measurements from in-vivo HSV recordings. To that end, two projects are conducted in this framework. This section provides a brief overview of each project, with discussions on how different chapters are connected to each other. In that regard, this section connects different pieces of this dissertation together and describe how they fit into a single framework.

Considering that laryngeal HSV recordings are typically performed in an upright position, the following directions are defined for the rest of this dissertation. A horizontal plane is an imaginary plane that splits the body into the superior (*i.e.* above) and the inferior (*i.e.* below) sections. The vector normal (*i.e.* perpendicular) to that plane is called the vertical direction. Figure 1.1 presents an illustration of these terms.

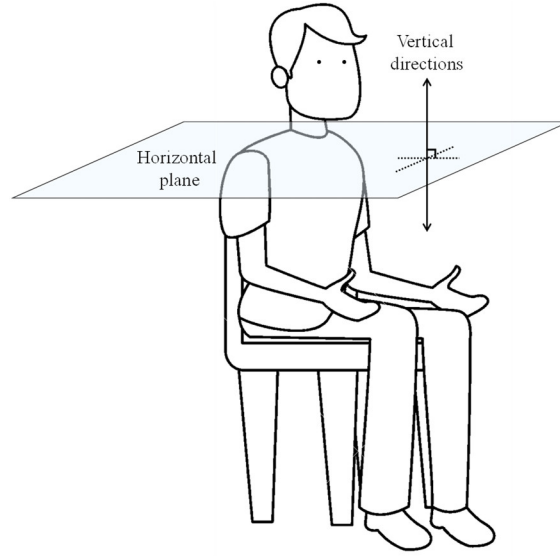


Figure 1.1. Illustration of a horizontal plane and the vertical direction.

Based on our daily experiences, we know that the size of an object in an image depends on its distance from the camera. This means that we could measure the pixel size of an object in an image, but we could not relate it to its actual size (*e.g.* the *mm* size). In this regard, we could not perform absolute *mm* measurements on typical images and hence we say that images are not spatially calibrated. However, if we have some specific *auxiliary information*, we could make *mm* measurements from the images. The pixel-to-*mm* conversion scale is the required *auxiliary* information. The procedure that allows us to compute the pixel-to-*mm* conversion scale is called calibration.

Based on figure 1.1 two different *types* of spatial measurements could be identified. They include horizontal and vertical measurements. Additionally, the *auxiliary information* could come from different sources, and depending on that, two different *methods* of direct and indirect calibrations can be distinguished and are defined here. The *indirect* calibration approach is defined as a method that its *auxiliary* information comes from a *different* image (possibly taken from a different imaging modality). Using the intraoperative calibrated measurement of the lesion

size<sup>93,190,191</sup> for horizontal calibration of its corresponding HSV recording is an example of the indirect approach. Conversely, the *direct* calibration approach is defined as a method that its *auxiliary* information comes from the *same* image that we want to make measurements from. Laser-calibrated endoscopes are the most common example of the direct approach in voice science research.<sup>191–195</sup>

The main goal of this dissertation is to devise the methods for performing calibrated spatial measurements from in-vivo HSV recordings. Therefore, the central hypothesis of this dissertation is:

H: Absolute spatial measurements from in-vivo HSV recordings using indirect and direct calibration approaches, are feasible.

In order to test H, several research questions and sub-hypotheses were formed which are presented in the rest of this section.

The indirect calibration approach does not require any specialized instruments and can be performed using the conventional and existing laryngeal imaging systems. Additionally, an image could be printed and a simple caliper would be enough for doing the horizontal measurement.<sup>93</sup> Consequently, indirect methods are very simple and could be used in many clinical settings. Chapter 2 of this proposal taps into this potential and proposes a method for horizontal calibration of an HSV recording using its corresponding calibrated intraoperative image. The main idea is to find a proper common attribute (*e.g.* lesion size in the pre-surgery recording) between the calibrated image and the HSV recording, and then to register that attribute (*i.e.* aligning the common attribute) on the HSV data. This project has external funding, and it is tightly related to a recently funded NIH R01 grant R01 DC017923 (PI: Verdolini Abbott) with a subcontract to Michigan State University (sub-award PI: Deliyski). The main research question of chapter 2 is:

Q1: How could calibrated intraoperative images be used for spatial calibration of pre- and post-surgery HSV recordings?

To answer this central research question, it is broken into three sub-questions.

Q1a: What are the main assumptions behind the validity of the indirect calibration approach?

Q1b: How can validity of the registration step be evaluated?

Q1c: Multiple common attributes can be identified for calibration of the post-surgery HSV recordings. How could we select the most appropriate one?

Associated with these research questions the following hypotheses were formed:

H1a: The proposed registration uncertainty test can detect instances of common attributes that have high registration uncertainties.

H1b: A proper object could be selected from the calibrated intraoperative image, such that its registration on the pre- and post-surgery HSV recordings would lead to the horizontal calibration of the HSV images.

Chapter 3 uses a pre-existing set of HSV recordings as a test bench to demonstrate the feasibility of the indirect calibration method. To that end, the required methods for computation of the calibrated velocity measures of the vocal folds are presented in chapter 3. The outcomes of chapter 3 have also significant scientific values for voice science and clinical applications. Perceptual evaluation and acoustic measurement studies have shown that the presence of a lesion on a vocal fold often changes the produced acoustic signal.<sup>60,61,196–201</sup> VSB, VKG, and HSV studies have made some connections between changes in the physiology and the vibratory characteristics of the vocal folds.<sup>93–95,108,183,184,202,203</sup> However, we do not know exactly how the kinematics of vocal folds changes in the presence of a lesion, and how removing the lesion improves it. The

closing velocity of the vocal folds is an important kinematic measure that relates to collision forces of the vocal folds<sup>25–27</sup> and the produced acoustic output.<sup>23,28,29</sup> Hence, this measure could link biomechanics of the phonation to the output of the system. Considering that time is already calibrated in the HSV recordings, its horizontal calibration would be the prerequisite of estimating the closing velocity of the vocal folds. Chapter 3 is aimed at studying the post-surgery changes in the closing velocity of the vocal folds. To this end, the following research question is answered in chapter 3.

Q2: How does the removal of a lesion from a vocal fold affects its kinematics?

Associated with this research question the following hypotheses were formed:

H2a: The closing phase maximum velocity will significantly increase after phonomicrosurgery.

H2b: For unilateral mass lesions, the closing phase maximum velocity of the two vocal folds will become more similar after the surgery.

H2c: Post-operative change in the closing phase maximum velocity will be positively correlated with the area of the lesion.

The indirect calibration approach has its own limitations. For example, it only provides the horizontal information and could not be used for vertical measurements. Additionally, it is based on some important implicit assumptions, which could not be validated or evaluated, directly. For example, a common attribute (*e.g.* lesion size) should be present in images from both modalities, and we should be able to register it accurately. Additionally, the actual size of the common attribute (*i.e.* its *mm* length) should be constant and does not change between different imaging sessions and imaging modalities. Also, the relationship between the length of the common attribute and the rest of the image in different imaging modalities should be a linear transformation of each other



(this is discussed in more detail in chapter 2). To put these conditions into context, if the lesion does not have a clear boundary, the first condition may be violated. If the size of the lesion changes (e.g. due to pliability of the tissue, different gravitation forces between supine and upright positions, etc.) the second condition may not hold. The last condition would be violated if the imaging angle is changed, or if the vertical distance between the common object and the rest of the image is changed between different imaging sessions. Consequently, the indirect approach could be prone to significant errors. The direct calibration approach could remedy this, at the expense of more sophisticated hardware (imaging instrument) and software (measurement algorithm). The remaining chapters of this dissertation are devoted to the development and evaluation of methodologies for direct calibration of in-vivo HSV recordings using a laser-projection transnasal fiberoptic endoscope.<sup>195</sup> Due to the optical design of this laser-calibrated endoscope, the horizontal distance between each pair of laser points is a function of the distance between the tip of the endoscope and the target surface (*i.e.* the working distance). Consequently, the horizontal measurement from this new system relies on the estimation of the working distance which is a vertical measurement. Therefore, the vertical measurement is presented before the horizontal measurement in this dissertation.

Chapter 4 presents the developed methodology for direct vertical measurements. Besides providing the required information for the horizontal measurement, multiple modeling studies have predicted a significant role for the vertical component of vibration of the vocal folds.<sup>164,172–174</sup> Therefore, the vertical measurements could be a significant source of information for improving our knowledge from the normal and disordered phonatory mechanisms. To achieve these goals, a system with the capability of vertical measurements should be developed first. The main research question of chapter 4 is:

Q3: How could we use a structured laser projection system for measuring the vertical distance between the distal tip of a flexible endoscope and the target surface?

Associated with this research question the following hypotheses were formed:

H3a: The position of each laser point will be a unique and deterministic function of the vertical distance between the distal tip of the flexible endoscope and the target surface, once the confounding factors are accounted for.

H3b: Vertical measurement error will be positively correlated to working distance.

Different parameters could be a confounding factor for calibrated vertical measurements. The effect of the focal distance of the lens coupler, the rotation of the endoscopic eyepiece inside the lens coupler, and the displacement of the eyepiece within the lens coupler are accounted for in the proposed method. Additionally, the effects of working distance, optical differences between different laser points, the imaging angle, and imaging from a non-flat surface on measurement errors are evaluated. Finally, there are other factors including the intensity of the light source, the frame rate of the camera, the exposure time of the camera, the sensitivity of the chip of the camera, the spatial resolution of the chip of the camera, the format of the images (*e.g.* raw data vs. avi), the intensity of the laser source, differences between different makes of the endoscope, the curvature of the target surface, the reflective properties of the target surface, the color of the target surface, and absorption properties of the target surface that are not investigated in this dissertation and should be investigated in future works.

Calibrated horizontal measurement depends on devising a scheme that could convert the pixel length of an object (*i.e.* its length on the image) to its true length (*i.e.* its *mm* length). Achieving this goal requires a precise knowledge of the confounding factors affecting the relationship between a pixel length and its *mm* length. The main aim of chapter 5 is to study two main

confounding factors of horizontal measurements. Additionally, the outcomes of this chapter could help us better understand possible confounding factors in subjective assessments and objective measurements from flexible endoscopy images. The main research questions of chapter 5 are:

Q4a: How much the *mm* size of a pixel depends on its spatial location?

Q4b: How much the imaging angle affects the *mm* size of a pixel?

Associated with these research questions the following hypotheses were formed:

H4b: Pixel size is significantly smaller in the center group than the periphery group.

H4c: Pixel size is significantly different between back, middle, and front groups when the target surface gets tilted.

Chapter 6 builds on the results of chapter 5 and presents the required methodology for calibrated horizontal measurements using the laser-calibrated endoscope. Horizontal measurements could provide a better and direct means for studying the developmental aspects of vocal folds<sup>153</sup> and laryngeal tissues, quantifying the relationship between an intervention and its resulting physiological changes (*e.g.* post-intervention changes in the lesion size), staging and grading of relevant laryngeal diseases<sup>191</sup>, and providing calibrated spatial measurements for patient-specific models. To achieve these goals, a system with the capability of calibrated horizontal measurements should be developed first. The aim of this chapter is to develop a method that could address this need. The main research question of chapter 6 is:

Q5: How could we use a structured laser projection system for measuring the horizontal distance between two points on a target surface?

Associated with this research question the following hypothesis was formed:

H5a: Horizontal measurement error from the laser-projection system significantly increases if the nonlinear distortion is not properly compensated for.

H5b: Horizontal measurement error will be positively correlated to working distance.

Different factors could affect the accuracy of calibrated horizontal measurements. The effects of vertical distance and the spatial location of the object inside the field of view (FOV) are accounted for in the proposed method. Additionally, the effect of working distance, the imaging angle, and imaging from a non-flat surface on the measurement errors are quantified and reported.

Chapters 4, and 6 of this dissertation are aimed at developing methods for calibrated vertical and horizontal measurements using a laser-projection endoscope. The methods are developed in a very controlled setting and using benchtop recordings from flat surfaces. In order to validate the methods in a more complex and realistic setting, chapter 7 is devoted to the validation of the developed methods. Specifically, based on our daily experiences we know that the angle of a camera relative to a scene affects the way that scene is recorded. Therefore, it is expected for the imaging angle to affect the accuracy of horizontal and vertical measurements. However, this topic has received very limited attention in the field of voice.<sup>204,205</sup> One possibility for this gap, could be a lack of quantitative values regarding the effect of imaging angle on the accuracy of measurements. Additionally, a 3D surface is used to evaluate the accuracy of the developed methods on non-flat surfaces.

The main research questions of chapter 7 are:

Q6a: How the imaging angle affects the performance of the vertical and horizontal measurements?

Q6b: How the topology of a 3D surface affects the vertical and horizontal measurements?

Associated with these research questions the following hypotheses were formed:

H6a: The tilting angle of the target surface and the working distance will be good predictors of the vertical measurement error.

H6b: The tilting angle of the target surface and the working distance will be good predictors of the horizontal measurement error.

H6c: The vertical measurement errors from a non-flat surface will be higher than those from a flat surface positioned at the same estimated average vertical distance.

H6d: The horizontal measurement errors from a non-flat surface will be higher than a flat surface positioned at the same estimated average vertical distance.

Finally, chapter 8 presents a summary of the findings.

Table 1.1 presents a summary of the chapters of the dissertation. Specifically, for each chapter the type of endoscope, the relevant dataset, and its primary goals are presented.

Table 1.1. Summary of different chapters of the dissertation.

Chapter	Endoscope	Dataset	Outcome
<b>Chapter2</b>	Rigid	Pre- and post-surgery in-vivo HSV recordings. Calibrated intraoperative still images.	Indirect calibrated horizontal measurements
<b>Chapter3</b>	Rigid	Pre- and post-surgery in-vivo HSV recordings. Calibrated intraoperative still images.	Indirect calibrated closing phase maximum velocity
<b>Chapter4</b>	Flexible	Benchtop recordings from white papers.	Direct calibrated vertical measurements
<b>Chapter5</b>	Flexible	Benchtop recordings from rectangular grid papers.	Quantification of non-linear distortion of a fiberoptic flexible endoscope.
<b>Chapter6</b>	Flexible	Benchtop recordings from white papers. Benchtop recordings from circular grid papers. Benchtop recordings from circle sectors. Benchtop recordings from line segments.	Direct calibrated horizontal measurements
<b>Chapter7</b>	Flexible	Benchtop recordings from line segments. Benchtop recordings from a 3D printed surface.	Validation of developed direct horizontal and vertical methods.

## **1.4. Recordings setup and characteristics**

The proposed research is based on different sets of video recordings. Specifically, two different types of in-vivo and benchtop recordings are used in this dissertation. Considering the extensive use of benchtop recordings, this section presents the employed setup, as well as, the protocol that was followed for the benchtop data collections.

### **1.4.1. Benchtop recording setup**

This dissertation uses benchtop recordings for the development of methods for vertical and horizontal measurements using a laser-calibrated fiberoptic flexible endoscope<sup>195</sup>, as well as, investigation of the effect of different factors on the accuracy of measurements. Therefore, a setup that allows precise variations in the working distance and the imaging angle was developed. The setup consisted of a vertical pillar that was connected to a horizontal surface. A high-speed monochrome camera Phantom v7.1 (Vision Research Inc., Wayne, NJ) was connected to the pillar such that it was perpendicular to the horizontal surface. A 45-*mm* lens coupler was used to connect the flexible endoscope to the camera. The distal tip of the endoscope was passed through two fixtures with small holes to keep the distal end of the endoscope fixed. The target surface was attached to an adjustable arm with two degrees of freedom. Specifically, the vertical adjustment of the arm allowed us to regulate the distance between the target surface and the distal end of the endoscope, accurately. Additionally, the setup allowed us to regulate the angle between the target surface and the imaging axis of the endoscope, accurately. The first parameter is called the working distance, and the second parameter is called the tilting angle for the rest of this dissertation. Figure 1.2 shows the employed setup for benchtop recordings.

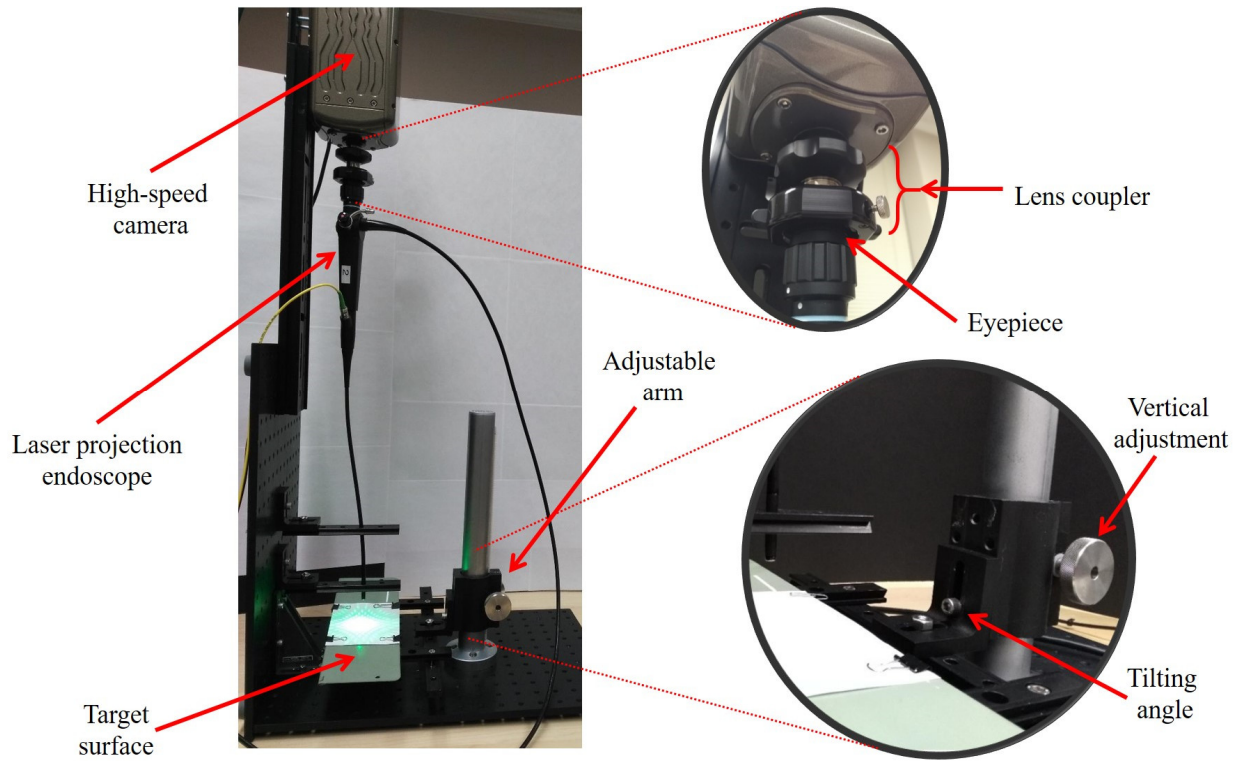


Figure 1.2. The employed setup for benchtop recordings.

#### 1.4.2. Recording protocol

Based on our preliminary studies and analyses a recording protocol was developed. Benchtop recordings were acquired using the following protocol.

- (1) We made sure that the FOV was completely inside the image frame and a border of at least five pixels was present on all four sides of the FOV. Figure 1.3 depicts this condition.

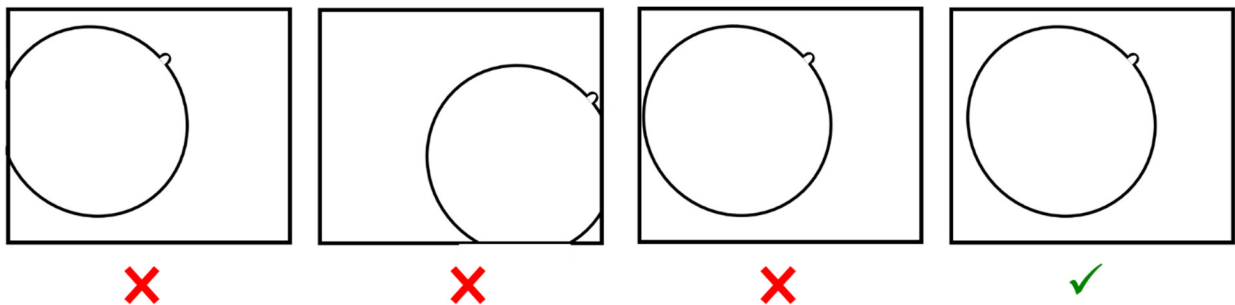


Figure 1.3. Examples of incorrect placements of the FOV in the image frame.

(2) We made sure that the FOV was quite visible and had a sharp contrast with the black background. Figure 1.4 depicts this condition.

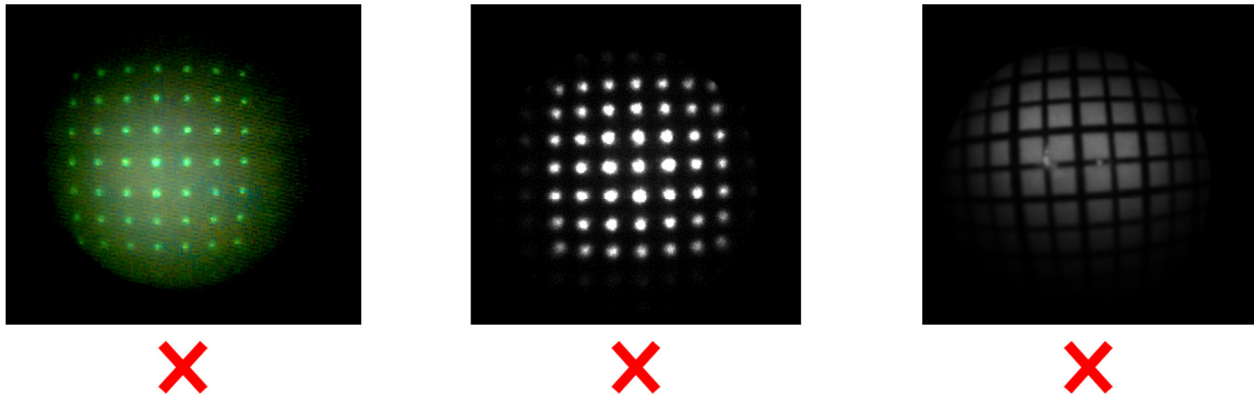


Figure 1.4. Some examples of the FOV with unclear edges.

(3) We made sure that the fiducial marker was completely inside the image frame and it had a border of at least five pixels. Figure 1.5 depicts this condition.

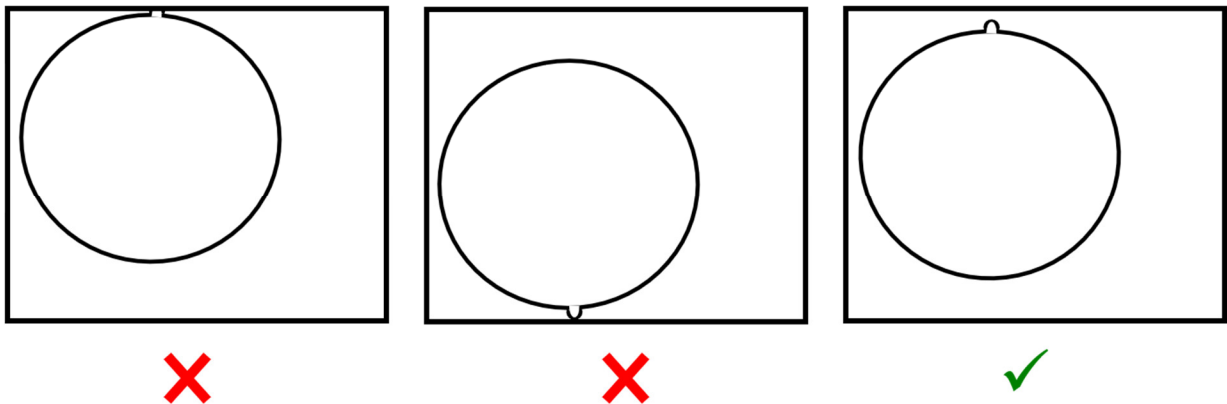


Figure 1.5. Some examples of the inadequate border between the FOV and the image frame.

(4) We made sure that the fiducial marker was visible and had a sharp contrast with the black background. Figure 1.6 shows an unacceptable example.



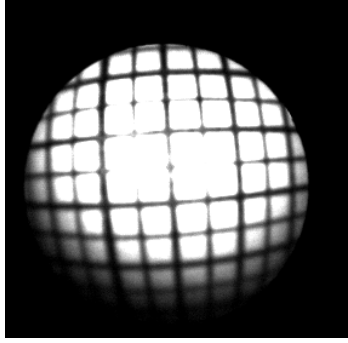


Figure 1.6. An example image with non-visible fiducial marker.

- (5) We used similar recording parameters for all recordings. The only parameters that were allowed to vary were the working distance, the tilting angle of the target surface (only in chapter 6), the illumination intensity of the light source, power of the laser source, and the exposure time of the camera.
- (6) The xenon light is essential for recording at high frame rates; however, it adds high-intensity divergence to the image. This divergence could interfere with the accuracy of the calibration protocols and add unnecessary complexities to the image processing steps. Therefore, all benchtop recordings used low frame rates and an external light source was used instead of the xenon light. For this purpose, we placed a studying lamp near to the target surface such that no shadow was projected on the target surface.

## CHAPTER 2: INDIRECT HORIZONTAL CALIBRATION OF IN-VIVO HSV RECORDINGS

*Based on:*

Ghasemzadeh H., Deliyski D. D., Hillman R. E., Mehta D. D. Indirect horizontal calibration of high-speed videoendoscopy recordings. How to do it, and what to look for? *in Preparation*.

Summary: Calibrated horizontal measurements from high-speed videoendoscopy recordings could offer significant advantage to precision medicine, patient-specific modeling, and evidence-based practice in the field of speech-language pathology and laryngology. Recently laser-projection systems have been developed for achieving the calibrated measurement goals. However, such systems are still in their infancy, and also only available to very few research labs. This chapter presents an alternative approach for achieving the horizontal calibration. The main idea of this alternative approach is to find a *proper common object* and then normalize lengths of other spatial measures to it. The underlying assumptions behind the validity of this approach are studied and it is shown that three main conditions should hold. First, the registration of the common object should be with negligible error. Second, the true length of the common object should be fixed. Finally, the common object and the target object should be on the same vertical distance. Two tests are proposed that could detect significant violation of the first and the second assumptions. In this study, a pre-existing dataset is used to demonstrate the feasibility of this approach.

## 2.1. Introduction

The uncalibrated size of a region of interest (ROI) (*e.g.* length, width, area) in any image can be measured by counting its corresponding number of pixels. Assuming a similar distance between all objects in the scene and the camera (*i.e.* the same working distance), one can compare the size of different objects within that image. Under this assumption, typical images can be utilized for within-image size comparisons. The main difficulty arises if we want to compare the size of an object from an image, with the size of a different object (or even the same object) from a different image (*i.e.* between-image size comparison). Considering the simplest scenario, we know that the size of an object in an image depends on its working distance. Therefore, differences between pixel lengths in different images could be attributed to either difference in their working distances or difference in their actual sizes. Obviously, this issue arises because we do not have a standard basis for comparison between the two images. This issue would be resolved if we could map both images to a fixed and standard basis. This task is called the horizontal calibration and it makes between-image size comparison possible. Considering that the meter is the base unit of the length in the International System of Units (SI), it is very typical to use it as a standard basis. In that regard, horizontal calibration is the process that one determines the size of a pixel in a metric unit (typically millimeter (*mm*) in voice science). This number then serves as a scale for conversion from pixel into *mm*, hence it is called the pixel size and the pixel-to-*mm* conversion scale in the rest of this dissertation. These two terms are used interchangeably in this dissertation. Different approaches are possible for the computation of this conversion scale. This chapter is devoted to what we have called the indirect computation of the conversion scale.

Let us consider a set of images from different scenes, all containing one common object. Between-image size comparisons can be made within this set if the pixel length of that common

object is used for calibration (*i.e.* the pixel length of the desired object is normalized using the pixel length of the common object). Now, if we know the metric size of that common object, images from that set can be mapped into a fixed and standard basis and between-image size comparisons can be made across different image sets. This approach is the basis of indirect horizontal calibration. In summary, we have a set of images with some *auxiliary* information regarding the metric size of a specific object or a specific spatial attribute from that set. In the context of laryngeal imaging, this *auxiliary* information could be for example the *mm* size of a lesion<sup>93</sup>, *mm* length of the vocal folds, or some *mm* features of a blood vessel.<sup>206</sup>

Indirect calibration is based on some *auxiliary* information that has come from a *different* image, than the one we would like to perform measurement from. One typical example is the intraoperative calibrated measurement of the lesion size.<sup>93,190,191</sup> Figure 2.1 shows some examples of calibrated intraoperative images. In this approach, a surgical instrument with a miniaturized ruler or a known *mm* length is positioned next to a lesion and the whole scene is recorded on a picture. The image then can be printed and the pixel lengths of the lesion and the surgical instrument can be measured using a high-precision caliper.<sup>93</sup> One can determine the pixel-to-*mm* conversion scale from this information. The computed scale could then be used for calibration of the HSV recording from the same patient. Another possible source of such *auxiliary* information could be a laser-calibrated VSB recording.<sup>133,207</sup> Considering that laser-calibrated VSB systems have been around longer than their HSV counterparts<sup>122</sup> and the simplicity of their optical design (typically parallel laser projections<sup>122</sup>), the required methodology for analysis and calibrated measurements of the laser-calibrated VSB images is already available. Additionally, the significantly shorter integration time of HSV systems and significantly brighter illumination sources of HSV systems in comparison to VSB, add extra requirements to the optical design of the

laser projection system. Therefore, in some instances, it may be more convenient and practical to use a combination of laser-calibrated VSB and a non-calibrated HSV system. One example would be a recently funded NIH R01 grant (R01 DC017923, PI: Verdolini Abbott) that is very tightly related to this project and the next chapter. In that grant proposal, VSB images will be used for calibrated measurement purposes, whereas HSV images will be used for the studying of the temporal and spatial vibratory characteristics of the vocal folds. Assuming a similar length for the vocal folds (or another spatial attribute) between VSB and HSV recordings, the *mm* length of the vocal folds (or another spatial attribute) from laser-calibrated VSB could be utilized for calibration of the HSV images. Previous studies have suggested that pitch of phonation depends (among other factors) on the length of the vocal folds and the subglottal air pressure.<sup>12</sup> Additionally, subglottal air pressure is a good predictor of the loudness.<sup>12</sup> Hence, if both recordings are done using the same pitch and loudness (*e.g.* habitual pitch, habitual loudness) similar attributes of the vocal folds across the recordings could be assumed to some extent.

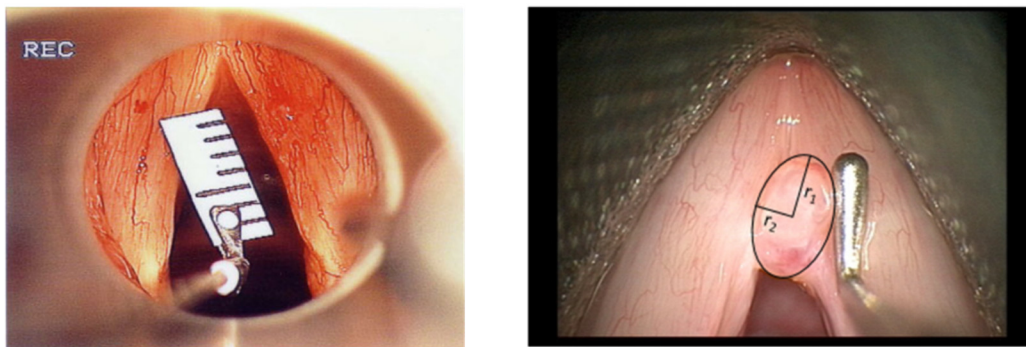


Figure 2.1. Two examples of intraoperative calibrated images, taken from references <sup>190</sup> and <sup>93</sup>.

This chapter presents a method for indirect horizontal calibration of HSV recordings from their corresponding intraoperative still calibrated images. The developed method could be used for different applications including, investigation of the kinematics of disordered phonation, studying the developmental aspects of vocal folds' vibration<sup>153</sup> and laryngeal tissues, the post-intervention

physiological changes (e.g. lesion size), staging and grading of relevant laryngeal disease<sup>191</sup>, and providing calibrated horizontal measurements for patient-specific models.

## 2.2. Aim and hypothesis

The project of this chapter has external funding, and it is tightly related to a recently approved NIH R01 grant R01 DC017923 (PI: Verdolini Abbott) with a subcontract to Michigan State University (sub-award PI: Deliyski). The second aim of that grant proposal requires calibrated horizontal measurements from in-vivo HSV recordings. The imaging will be conducted using a rigid endoscope, and the subjects will be pediatric patients diagnosed with nodules. Additionally, a parallel laser projection VSB system<sup>153</sup> will be utilized for measuring the size of the nodule pre- and post-therapy.

The idea of indirect horizontal calibration has already been used in multiple studies.<sup>93,190,191,206,207</sup> However, this notion has not yet received a formal treatment. Specifically, the conditions and assumptions behind the validity of this approach are yet to be studied. The main aims of this chapter are *to investigate the possible sources of error in the indirect calibration approach*, and then *to use intraoperative images for indirect calibration of pre- and post-surgery HSV recordings*. The central research question of this chapter is,

Q1: How could calibrated intraoperative images be used for spatial calibration of pre- and post-surgery HSV recordings?

To answer this central research question, it was broken into three sub-questions.

Q1a: What are the main assumptions behind the validity of the indirect calibration approach?

Q1b: How can validity of the registration step be evaluated?

Q1c: Multiple common attributes can be identified for calibration of the post-surgery HSV recordings. How could we select the most appropriate one?

The following hypotheses were formed in this chapter:

H1a: The proposed registration uncertainty test can detect instances of common attributes that have high registration uncertainties.

H1b: A proper object could be selected from the calibrated intraoperative image, such that its registration on the pre- and post-surgery HSV recordings would lead to the horizontal calibration of the HSV images.

## **2.3. Material and method**

### **2.3.1. Participants and data acquisition**

The aim of this chapter is pursued using retrospective data. Calibrate intraoperative images and HSV recordings were obtained from 26 adults with vocal fold mass lesions at Massachusetts General Hospital. Subjects were recorded using a custom-built HSV system over two different sessions. The first session was before the surgery, and the second recording was carried out on average 3.5 weeks after the surgery. The HSV system consisted of the following components, a color Phantom v7.3 camera (Vision Research, Inc., Wayne, New Jersey), a 300-Watt xenon light (Model 7152A, PENTAX Medical Company Montvale, New Jersey), and a 70° 10-mm rigid laryngoscope (Model 49-4072, JEDMED Instrument Co, St. Louis, Missouri). The recordings were done at a sampling rate of 6,250 fps with the maximum integration time and a spatial resolution of 320×352 pixels. The surgery was performed using cold instruments and/or a 532-nm pulsed potassium titanyl phosphate laser photoablation under general anesthesia. Before the

operation, a surgical instrument with a known *mm* length was placed next to the lesion and an intraoperative image was recorded.

Considering the aims of this chapter, the full temporal resolution of HSV recordings is not required. Therefore, a hamming window with a size of 5 was used for temporal smoothing of the data. To that end, every five consecutive frames of HSV data were weighted by a hamming window, and then they were averaged. This process significantly reduced the noise of the recordings.

### 2.3.2. Indirect calibration principles and assumptions

Let  $l_{mm}$  and  $l_{pixel}$  denote calibrated length and uncalibrated length of an object in an image. We can define the pixel size ( $\lambda$ ) as,

$$\lambda = \frac{l_{mm}}{l_{pixel}} \quad (2-1)$$

The validity of any uncalibrated size comparison depends on certain implicit assumptions, regarding the parameter  $\lambda$ . For example, the percentage change in the pixel length (*i.e.* the uncalibrated length in the image) of a lesion pre- and post-intervention could be used as a direct evaluation criterion for measuring and comparing the efficacy of different interventions. In this group-comparison scenario, an implicit assumption is that for each subject the measurement from the pre and post conditions are on the same scale, and hence can be compared with each other. Being on the same scale means that if the pixel length of the lesion in the post-intervention image is reduced by 20%, in reality, the true length in *mm* (*i.e.* the calibrated length) of that lesion has also been reduced by 20%. More precisely, the implicit assumption is that the *mm* size of a pixel (*i.e.* the pixel size) in the pre and post conditions are the same for each subject. We call this the *within-subject* size comparison assumption. It is noteworthy that most image-based group



comparison studies in voice (both objective and subjective) are based on this assumption. On the other hand, if the purpose of the research is to compare different groups or to relate post-intervention changes in the lesion size to some outcomes of the phonatory mechanisms (*e.g.* some acoustic measurements), a more strict assumption should hold. More precisely, not only the *mm* size of pixels in the pre- and post-conditions for each subject should be the same, but also the *mm* size of pixels in different subjects should be the same. We call this the *between-subject* size comparison assumption. This implicit assumption is present in most (if not all) image-based regression and modeling studies in voice. It is noteworthy that the *between-subject* size comparison assumption satisfies the *within-subject* size comparison assumption; however, the other direction does not necessarily hold. Therefore, the conditions and assumptions behind the validity of each approach are studied in different sections.

### 2.3.2.1. Indirect calibration for between-subject size comparison

The between-subject size comparison requires the mapping of all measurements into a standard and fixed basis. This often requires the knowledge of the *mm* length of the common attribute. Let  $l_{C,mm}^1$  denotes the calibrated length of the common attribute (*e.g.* the lesion size) in the first image (*e.g.* the intraoperative image). Additionally, let  $l_{C,pixel}^2$  and  $l_{T,pixel}^2$  denote the uncalibrated length of the common attribute and the target object (*e.g.* length of the vocal folds) in the second image (*e.g.* a frame of HSV recording). The aim of indirect calibration is to estimate the calibrated length of the target object in the second image ( $l_{T,mm}^2$ ) from the knowledge of  $l_{C,mm}^1$ . Obviously, if we have the pixel size of the target object in the second image ( $\lambda_T^2$ ) we can compute  $l_{T,mm}^2$  by,

$$l_{T,mm}^2 = \lambda_T^2 \cdot l_{T,pixel}^2 \quad (2-2)$$

Given that the second image is not calibrated, the  $\lambda_T^2$  is not known. Let  $\lambda_C^2$  denotes the pixel size of the common attribute in the second image. Now, assuming  $\lambda_T^2 = \lambda_C^2$  and  $l_{C,mm}^1 = l_{C,mm}^2$  we can compute  $l_{T,mm}^2$  as follows,

$$\lambda_C^2 = \frac{l_{C,mm}^2}{l_{C,pixel}^2} = \frac{l_{C,mm}^1}{l_{C,pixel}^2} \quad (2-3)$$

$$l_{T,mm}^2 = \lambda_C^2 \cdot l_{T,pixel}^2 \quad (2-4)$$

The value of  $\lambda$  could depend on different parameters of recording including the vertical distance, the imaging angle, and the spatial location of the object inside the field of view.<sup>208,209</sup> For simplicity we assume a case where  $\lambda$  only depends on the vertical distance. In that case, the  $\lambda_T^2 = \lambda_C^2$  assumption translates into the common attribute and the target object being on the same vertical distance from the endoscope. Based on the presented arguments, the main assumptions behind the validity of indirect calibration for between-subject size comparison applications are as follows. First, the common attribute can be registered accurately on the second image. Second, the common attribute and the target object are at the same vertical distance from the endoscope. Third,  $l_{C,mm}^1 = l_{C,mm}^2$  which means that the calibrated length of the common attribute should not change between the first and the second images. These three conditions will be referred to as the registration accuracy assumption, the similarity in the vertical distance assumption, and the consistency of the common attribute assumption in the rest of this dissertation. As a final note, the similarity in vertical distance assumption was derived based on the assumption that pixel size only depends on the working distance. However, spatial location in the fiberoptic endoscopes and the imaging angle are also significant factors for the value of the pixel size.<sup>208</sup> Therefore, the similarity in vertical distance assumption would become much more complicated during the fiberoptic endoscopy, or if the optical axis is not perpendicular during the imaging sessions.

### 2.3.2.2. Indirect calibration for within-subject size comparison

Performing the within-subject size comparison requires less information and in that regard is more practical. However, it is very likely that the outcome of calibration could not be used for absolute measurements, but rather the percent change in the size (*e.g.* percent change in the lesion size post-therapy). Assuming the availability of calibrated lengths of the target object in the first and the second image, the calibrated (*i.e.* the true) percent change ( $PC_{mm}$ ) of a target object is equal to,

$$PC_{mm} = \frac{l_{T,mm}^2 - l_{T,mm}^1}{l_{T,mm}^1} \times 100\% = \left( \frac{l_{T,mm}^2}{l_{T,mm}^1} - 1 \right) \times 100\% \quad (2-5)$$

Now, we could use the value of pixel size and re-write Equation 2-5 using uncalibrated lengths,

$$PC_{mm} = \left( \frac{l_{T,pixel}^2 \cdot \lambda_T^2}{l_{T,pixel}^1 \cdot \lambda_T^1} - 1 \right) \times 100\% \quad (2-6)$$

We could divide the numerator and denominator of Equation 2-6 with the same number,

$$PC_{mm} = \left( \frac{l_{T,pixel}^2 \cdot \lambda_T^2}{l_{T,pixel}^1 \cdot \lambda_T^1} \times \frac{l_{C,mm}^1}{l_{C,mm}^1} - 1 \right) \times 100\% \quad (2-7)$$

Assuming  $l_{C,mm}^2 = l_{C,mm}^1$  we can rewrite Equation 2-7 using the uncalibrated lengths of the common attribute,

$$PC_{mm} = \left( \frac{l_{T,pixel}^2 \cdot \lambda_T^2}{l_{T,pixel}^1 \cdot \lambda_T^1} \times \frac{l_{C,pixel}^1 \cdot \lambda_C^1}{l_{C,pixel}^2 \cdot \lambda_C^2} - 1 \right) \times 100\% \quad (2-8)$$

Doing some re-arrangements, we would have,

$$PC_{mm} = \left( \frac{l_{T,pixel}^2 / l_{C,pixel}^2}{l_{T,pixel}^1 / l_{C,pixel}^1} \times \frac{\lambda_C^1 \cdot \lambda_T^2}{\lambda_C^2 \cdot \lambda_T^1} - 1 \right) \times 100\% \quad (2-9)$$

Assuming  $\frac{\lambda_C^1 \lambda_T^2}{\lambda_C^2 \lambda_T^1} = 1$  (this assumption would be discussed shortly), then *calibrated* percentage change can be computed from *uncalibrated* lengths as follows. Pick a suitable common attribute and measure its uncalibrated length in the first ( $l_{C,pixel}^1$ ) and the second image ( $l_{C,pixel}^2$ ). Also, measure the uncalibrated lengths of the target object in the first ( $l_{T,pixel}^1$ ) and the second image ( $l_{T,pixel}^2$ ). The calibrated percentage change can be computed as:

$$PC_{mm} = \left( \frac{l_{T,pixel}^2 / l_{C,pixel}^2}{l_{T,pixel}^1 / l_{C,pixel}^1} - 1 \right) \times 100\% \quad (2-10)$$

Now, let us investigate the condition for  $\frac{\lambda_C^1 \lambda_T^2}{\lambda_C^2 \lambda_T^1} = 1$ . The value of  $\lambda$  could depend on different parameters of recording including the vertical distance, the imaging angle, and the spatial location of the object inside the field of view.<sup>208,209</sup> For simplicity we assume a case where  $\lambda$  only depends on the vertical distance. In that case, if we have the vertical distance we can compute the true value of  $\frac{\lambda_C^1 \lambda_T^2}{\lambda_C^2 \lambda_T^1}$  term and plug it in Equation 2-9. However, the vertical information is often (if not always)

lost during the image acquisition. Therefore, we need to find conditions that govern  $\frac{\lambda_C^1 \lambda_T^2}{\lambda_C^2 \lambda_T^1} = 1$ .

There are two trivial solutions to this. Either  $\lambda_C^1 = \lambda_C^2$  and  $\lambda_T^1 = \lambda_T^2$ , or  $\lambda_C^1 = \lambda_T^1$  and  $\lambda_C^2 = \lambda_T^2$ . The first solution means that the vertical distance between the endoscope and the target object, and the vertical distance between the endoscope and the common attribute should not change between the first and the second images. The second solution means that the target object and the common attribute are on the same vertical levels. Considering that the larynx can move in the vertical direction, the first condition cannot be controlled. Therefore, the second condition would be more feasible and practical case for laryngeal imaging applications. In summary, the validity of indirect calibration for within-subject size comparison depends on the same three main assumptions of:

registration accuracy assumption, the similarity in the vertical distance assumption, and the consistency of the common attribute assumption in the rest of this dissertation. As a final note, the similarity in the vertical distance assumption was derived based on the assumption that pixel size only depends on the working distance. However, spatial location in the fiberoptic endoscope and the imaging angle are also significant factors of the pixel size.<sup>208</sup> Therefore, the similarity in vertical distance assumption would become much more complicated in fiberoptic endoscopy or if the optical axis is not perpendicular in the two imaging sessions.

To provide more insights into these assumptions, some scenarios that violate each assumption are provided. For the first assumption, let us consider a common attribute that is blurry, or does not have a clear and sharp boundary. In that case, the common attribute cannot be registered accurately. For the second assumption, let the target object be some spatial features of the vocal folds and the common attribute be some spatial feature in the supraglottic region. Considering that the common attribute is closer to the camera it will have a smaller pixel size in comparison to the target object, which is a violation of the similarity in the vertical distance assumption. For the third assumption, let us consider a soft and pliable common attribute attached to the vocal which gets deformed easily. Now, if the vocal folds stretch, the common attribute will also elongate. Obviously, this is a contradiction of the consistency of the common attribute assumption.

It is noteworthy that, the second and the third assumptions often could be contradicting each other. Specifically, if the target object is on the vocal fold, using a common attribute that is not on the vocal fold would possibly satisfy the third assumption to the maximum extent. Conversely, the second assumption requires the selection of a common attribute that is as close to the target object as possible. In practice, we need to make a tradeoff between these two assumptions and other considerations and select a common attribute that is more suitable. A final word regarding the

second assumption, if the target object and the common attribute are not on the same vertical distance, the indirect calibration approach will introduce some errors into the measurement. The magnitude of this error will depend on the vertical distance between the common attribute and the target object and also the vertical distance between the imaging component (*i.e.* the endoscope) and the closer object. Additionally, a higher vertical difference between the common attribute and the target object will lead to higher error. Conversely, keeping the vertical difference between the common attribute and the target object fixed and increasing the vertical distance between the imaging component (*i.e.* the endoscope) and the closer object would decrease the error. This is especially important during flexible endoscopy, where vertical distance can be varied in a large range. As a practical guide, we need to keep the vertical distance between the common attribute and the target object as small as possible, especially when the imaging is done at a close working distance (*e.g.* during flexible endoscopy).

### **2.3.3. Evaluation of indirect calibration**

The indirect calibration lacks the existence of a universal and standard basis of comparison (*e.g.* metric scale) on the target data (*i.e.* HSV recording). However, it can offer some functionality of calibration by registering a proper common attribute on the HSV data. The validity and accuracy of any measurement following the indirect calibration would depend on correct and successful registration of the common attribute, as well as, the existence of the two other fundamental characteristics of the common attribute discussed in the previous section. While due to the lack of a universal and standard basis these conditions could not be checked directly, a test is presented in the next section that can indirectly evaluate the registration accuracy assumption.

### 2.3.3.1. Registration uncertainty test

Accuracy of indirect calibration relies heavily on correct registration of the common attribute. Additionally, not all common attributes may have similar registration accuracies. For example, a common attribute could be blurry or may lack a clear and sharp boundary. A test is developed here that indirectly provides an estimation of the accuracy of the registration process. The computed value provides a higher bound for the registration accuracy and therefore puts a lower bound on subsequent measurement uncertainty. That is, the subsequent measurements would *at least* have that amount of uncertainty. Finally, this test has a very high positive predictive value. Meaning that for a common attribute with a high score we could be quite confident that the amount of uncertainty is high, and therefore that common attribute would not be suitable.

The test assumes that a total number of  $k$  different HSV recordings from multiple subjects are being calibrated at the same time. Additionally, it assumes that for each subject a single image with the common attribute (*e.g.* the intraoperative image) is available. We will call this, the fixed image in the rest of this chapter. The test consists of three steps of data selection, data registration, and data analysis.

In the data selection step, a time point is selected randomly from one of the HSV recordings, and then frames within that glottal cycles are evaluated subjectively. The frame with the best visual appearance of the common attribute is selected. This process is repeated  $n$  times for the current recording. The whole process is repeated for all  $k$  recordings. This will lead to a total number of  $k \times n$  selected frames. These images will be referred to as the moving images in the rest of this chapter. Moving images are randomized, and then they are presented for the registration step.

In the registration phase, a two-panel graphical user interface (GUI) is used where one panel shows the fixed image, and the other panel shows the moving image. The GUI is equipped with

the zooming capability, to make the registration more accurate. The user clicks the boundary of the common attribute (*e.g.* the start and the endpoint of a lesion in the anterior-posterior direction) on the fixed image, and then do the same thing for the moving image. The software then computes the uncalibrated pixel size of the common attribute in the fixed and the moving image and then records them. If the lesion size is used for calibration, the size is computed as the Euclidean distance between the two click points. This process is repeated until all images are processed.

In the analysis phase, the data is de-randomized and then the ratios between sizes of the common attribute in the fixed and the moving images are computed for each recording. This will lead to  $n$  such ratio values. The interquartile range of these  $n$  ratio values is computed. This process is repeated for all  $k$  recordings. Any recordings with an interquartile range larger than a threshold would be an instance of calibration with a high level of uncertainty and should be removed from later analysis. A method for selecting a proper value of the threshold is presented in section 2.4.1.3.

It is noteworthy that, the registration uncertainty test can be extended to select the most suitable common attribute (*e.g.* anterior-posterior length of a lesion, medial-lateral length of a lesion, etc.) for each recording. For that purpose, different common attributes may be determined for each recording. The common attribute with the lowest interquartile range would be the most suitable common attribute for that recording.

## 2.4. Experiments and results

Three experiments were conducted to answer the research questions of this chapter. Experiment 1 demonstrates the efficacy of the proposed registration uncertainty test. Experiment 2 investigates the consistency of different common objects between different phonatory configurations. Experiment 3 demonstrates how the most appropriate common object may be



selected from a couple candidate ones. This section presents details of each experiment, followed by results and the related discussions.

#### **2.4.1. Experiment1: Efficacy of registration uncertainty test**

This experiment was conducted to demonstrate the performance and efficacy of the registration uncertainty test. The following hypothesis was formed for this experiment.

H1a: The proposed registration uncertainty test can detect instances of common attributes that have high registration uncertainties.

##### **2.4.1.1. Database**

From the 26 subjects with calibrated intraoperative images, lesions were not visible in three recordings, and one recording was blurry. These samples were excluded from the rest of the analysis. The included 22 subjects had an HSV recording from their comfortable pitch. Additionally, 14 subjects had an HSV recording from their higher pitch. Considering that the higher pitch requires a different glottal configuration and could also be an instance of interest for calibration, they were also included in the analysis. In summary, 36 different recordings were used for this experiment.

##### **2.4.1.2. Method**

Following the data selection step from the registration uncertainty test, 10 frames were randomly selected from each recording and were saved as images. Often, the frame with the maximum abduction resulted in the best visual exposure of the lesions. Additionally, from the saved 10 images, two images were randomly selected and were added again to the pool of saved

images (20% redundancy). This resulted in  $36 \times 12 = 432$  files to be registered. Registration and analysis followed the steps described in section 2.3.3.1.

### 2.4.1.3. Results

First, the redundant samples were excluded from the analysis. Figure 2.2(A) shows the computed score (the interquartile range) for each recording.

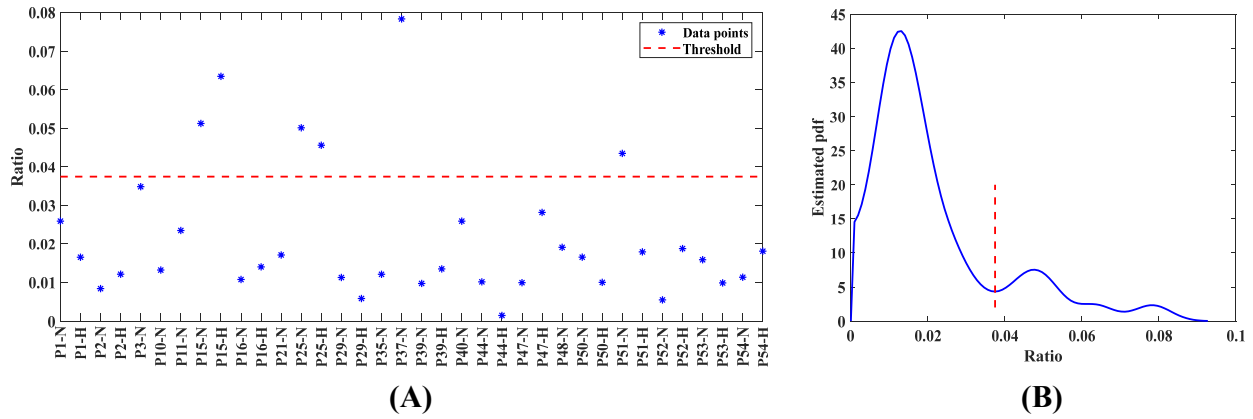


Figure 2.2. Results of registration uncertainty test: (A) values of interquartile range for different patients and (C)omfortable and (H)igh pitch phonations, (B) estimated pdf of interquartile range over all recordings.

A statistical approach was used to determine the appropriate value of the threshold. To that end, the probability density function (pdf) of the interquartile range over all 36 recordings was estimated using a Gaussian kernel. Figure 2.2(B) shows the result. Based on this figure the interquartile range can be attributed to three different classes. The first class has a very high value of the interquartile range, corresponding to calibration with a high level of uncertainty and hence high error. The second class has a moderate value of the interquartile range. And the last class corresponds to common attributes that can be registered with a low level of uncertainty. The value of the threshold was computed as the minimum value of the estimated pdf between the second and the third classes.

To test hypothesis H1a, the computed threshold was used to split the data into two groups, one with a high level of uncertainty (6 samples) and one with a low level of uncertainty (30 samples). Then, intra-sample registration variability was computed using the redundant samples that were excluded from the previous analysis (two per each recording). The absolute difference between the computed ratios for each pair of redundant data was computed. This led to two measurements per each recording. Intra-sample registration variability was computed as the average of those two values. Table 2.1 presents the descriptive statistics of each group.

Table 2.1. Descriptive statistics of intra-sample registration variability.

Group	Mean (pixel/pixel)	std (pixel/pixel)
High uncertainty	0.0453	0.0288
Low uncertainty	0.0096	0.0068

The dependent variable for H1 was the intra-sample registration variability. The independent variable was groupings (high vs. low uncertainty). A two-sample t-test was used to check H1. The test rejected the null hypothesis ( $p < .00001$ ,  $t = -6.28$ ). Based on this result and the values in table 2.1 we can conclude that the high uncertainty group had significantly higher intra-sample registration variability. This confirms the efficacy of the proposed test for detecting instances with high levels of registration uncertainty.

#### 2.4.2. Experiment 2: Effect of phonatory configuration on the calibrated length

The lesion is not present in post-surgery HSV recordings. Therefore, lesion size cannot be used for calibration of the post-surgery recordings. However, it is possible to find a common attribute between pre- and post-surgery recording and then use it for indirect calibration of the post-surgery data. In that sense, the lesion size would be used for the indirect calibration of the pre-surgery HSV data. Then the calibrated pre-surgery data would be used for indirect calibration of the post-surgery

HSV data. Considering the availability of the *mm* size of the lesion, the outcomes of both calibrations could be used for between-subject size comparison applications.

Going back to the consistency of the common attribute assumption, the *mm* size of the common attribute should be the same between the different imaging sessions, or imaging modalities. That is, the *mm* length of the lesion should be similar during the intraoperative imaging and the pre-surgery HSV recording. Additionally, the *mm* length of the object selected for calibration of the post-surgery HSV data should be similar in pre- and post-surgery conditions. Unfortunately, these conditions cannot be checked directly. However, we may use the information from different phonatory configurations in the pre-surgery recordings and check the robustness of the selected common attribute for calibration of the post-surgery recording. Experiment 2 presents the results of this analysis for different common attributes.

#### **2.4.2.1. Experiment 2.a: Vocal fold length attributes**

The length of the vocal folds (or some part of it) may be used as a common attribute. This idea has been used in several studies.<sup>103,203,206</sup> Considering the dependence of the fundamental frequency on the vocal fold length<sup>12</sup>, the following hypothesis was formed.

H1c: Calibrated vocal fold length during high-pitch phonation is significantly larger than its length during comfortable pitch phonation.

##### **2.4.2.1.1. Database**

From the 26 subjects with calibrated intraoperative images, 14 had recordings from both comfortable and the high-pitch phonations. Based on the result of experiment1 three subjects had high registration uncertainties, and hence were excluded from this analysis. Therefore, 11 subjects were used in this experiment.

#### 2.4.2.1.2. Method

11 glottal cycles were selected randomly from each recording. The frames within each glottal cycle were visually inspected and the frame with the highest glottal opening along the anterior-posterior direction was saved as an image. From the selected 11 images, one image was designated as the fixed image, and the remaining 10 images were designated as the moving image. The anterior commissure and the posterior part of the vocal folds were not visible in some of the images. Therefore, for each subject, the fixed images from both phonation tasks (*i.e.* comfortable and high pitch) were visually inspected and two suitable anchor points (one in the posterior and one in the anterior) were selected and marked on both fixed images. Some example anchors include the anterior commissure, a blood vessel on the vocal fold or a nearby tissue, or the midline of the lesion. Following the methodology of the registration uncertainty test, moving images from different recordings were randomized. The rationale for this randomization will be discussed in section 2.4.3.1.

A GUI with two panels (one showing the fixed image superimposed with the anchor points, and one showing the moving image) was developed for measuring the length of the vocal folds between the two anchor points. Due to occlusion and cropping of the recording, the measured value may only be part of the vocal fold length, hence it is named the vocal fold length attribute. The GUI had zooming capability for improved visual inspection and enabled marking of the anchor points on the target image. The pixel size of the vocal fold length attribute was measured as the Euclidian distance between the selected two anchor points on the moving image. Additionally, the pixel-to-*mm* conversion scale of each recording was computed from the known *mm* size of the lesion. Finally, the calibration was achieved by multiplying the pixel size of the attribute by the pixel-to-*mm* conversion scale.

### 2.4.2.1.3. Results

Figure 2.3 shows boxplot of the *mm* size of the vocal fold length attribute for each patient and each phonation task.

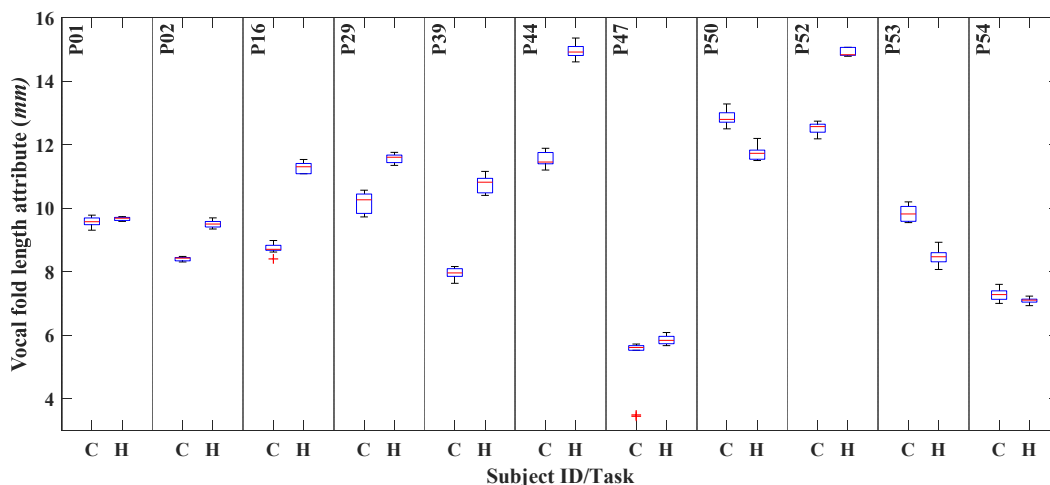


Figure 2.3. Boxplot of *mm* size of vocal fold length attribute of each subject for (C)omfortable and (H)igh pitch phonations.

The hypothesis H1c is based on the comparison of the *mm* sizes of the vocal fold length between the two phonation tasks. The *mm* size of the vocal fold attribute for each recording was computed as the median of the measurements from the 10 images. Table 2.2 presents the descriptive statistics of each phonatory group.

Table 2.2. Descriptive statistics of the *mm* size of attributes of vocal fold length.

Group	mean ( <i>mm</i> )	std ( <i>mm</i> )
Comfortable pitch	9.5	2.21
High pitch	10.53	2.84

To test the hypothesis H1c, a one-sided paired-samples *t*-test was used. The independent variable was the phonation task (comfortable vs. high pitch) and the dependent variable was the *mm* size of the vocal fold length attribute. The test detected a significant difference ( $p = 0.03$ ,  $t = -2.11$ ) between the two conditions. Therefore, the vocal fold length attribute is not a robust common

attribute for indirect calibration. This is especially important given that, the intraoperative image is taken under the resting state of the vocal fold with low tension, while the HSV recordings are captured during the phonation where the vocal folds have higher tension. Consequently, using vocal fold length attributes could lead to a violation of the consistency of the common attribute assumption, unless from the domain knowledge we know that the *visible* length of the vocal fold was not changing between the two imaging sessions, or imaging modalities.

#### **2.4.2.2. Experiment 2.b: Vocal fold width**

Vocal fold width is another spatial feature that can be used for calibration. For this experiment, the following hypothesis was formed.

H1d: Calibrated vocal fold width during high-pitch phonation is significantly different from a comfortable pitch.

##### **2.4.2.2.1. Database**

The data were similar to the experiment 2.a.

##### **2.4.2.2.2. Method**

11 glottal cycles were selected randomly from each recording. The frames within each glottal cycle were visually inspected for finding a frame where glottis would become very narrow, but not fully closed. Figure 2.4 shows an example image. The selected frames were saved as images. Considering that the left and the right vocal folds could have different widths, and also that width may be calculated at different locations along the anterior-posterior axis, the registration step could become inconsistent and hence susceptible to error. To remedy this, one image from the selected 11 frames was designated as the fixed image, and the rest were designated as the moving image.

On the fixed image, the target side of measurement (*i.e.* left or right vocal fold) was marked. Additionally, for each subject, the fixed images from both high and comfortable pitches were visually inspected for finding a proper anchor point (*i.e.* a point with a clear visual appearance in both phonation conditions) along the anterior-posterior direction. Some examples of the used anchor points were branching of a blood vessel on the vocal fold or a nearby tissue or specific topological attributes of the lesion. Figure 2.4 shows a fixed image with an anchor point selected based on a blood vessel on a nearby tissue. Following the methodology of registration uncertainty test images from different recordings were randomized. The rationale for this randomization will be discussed in section 2.4.3.1.

A GUI with two panels (one showing the fixed image superimposed with the anchor point, and one showing the moving image) was developed for measuring the width of the vocal fold, using the following procedure. A line was fitted to the target edge of the vocal fold on the moving image (solid red line in figure 2.4). Then the anchor point was marked on the moving image (symbol  $\times$  in figure 2.4). A line perpendicular to the line fitted to the edge of the vocal fold was passed through the selected anchor point (dashed blue line in figure 2.4). Then, the intersection of this new line with the periphery of the vocal fold was marked using the mouse (symbol O in figure 2.4). To reduce the inaccuracy of this selection, the selected point was analytically projected on the dashed line (point B in figure 2.4). The uncalibrated pixel width of the vocal fold was computed as the Euclidian distance between points A and B, where point A was the intersection of the two lines described above (figure 2.4). Finally, calibration was achieved by multiplying the pixel width of the vocal fold by pixel-to-*mm* conversion scale.



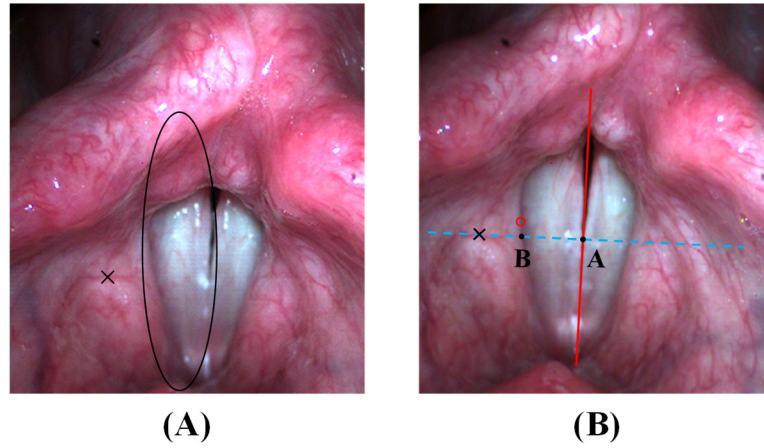


Figure 2.4. Measurement of the vocal fold width: (A) the reference image with designated vocal fold and the target anchor point, (B) the measurement steps.

### 2.4.2.2.3. Results

Figure 2.5 shows boxplot of the *mm* width of the vocal fold for each patient and each phonation task.

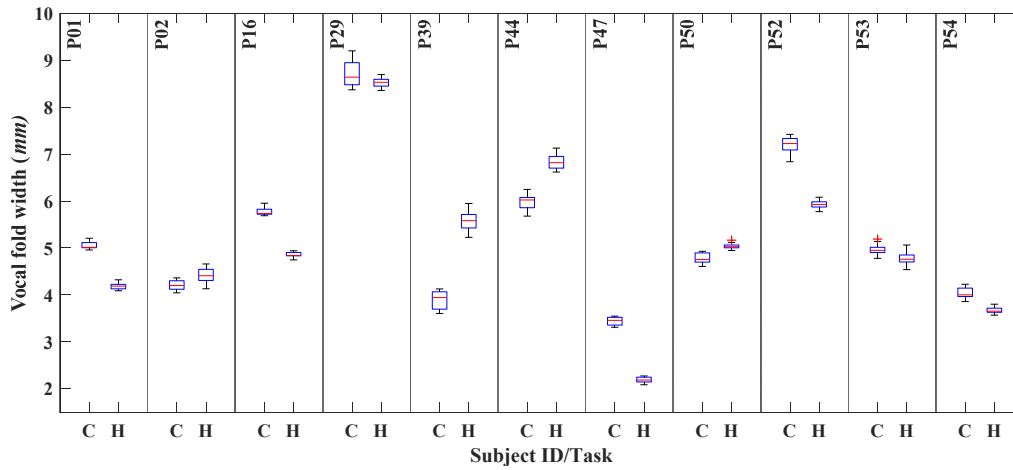


Figure 2.5. Boxplot of *mm* size of vocal fold width of each subject for (C)omfortable and (H)igh pitch phonations.

The hypothesis H1d is based on a comparison of the *mm* size of the vocal fold width between two phonation tasks. The *mm* width of the vocal fold for each recording was computed as the

median of the measurements from the 10 images. Table 2.3 presents the descriptive statistics of each phonatory group.

Table 2.3. Descriptive statistics of the *mm* width of the vocal fold.

Group	mean ( <i>mm</i> )	std ( <i>mm</i> )
Comfortable pitch	5.27	1.56
High pitch	5.08	1.67

To test hypothesis H1d, a two-sided paired-samples *t*-test was used. The independent variable was the phonation task (comfortable vs. high pitch) and the dependent variable was the *mm* width of the vocal fold. The test did not detect a significant difference ( $p=0.51$ ,  $t=0.68$ ) between the two conditions. Therefore, vocal fold width could be a robust common attribute for indirect calibration, and it may be used for indirect calibration.

#### 2.4.2.3. Experiment 2.c: Blood vessel on a vocal fold

The length of a blood vessel is another spatial feature that can be used for indirect calibration.<sup>206,207</sup> This experiment explores the suitability of a blood vessel on the vocal fold. For this experiment, the following hypothesis was formed.

H1e: Calibrated attribute of a blood vessel on the vocal fold during high-pitch phonation is significantly different from a comfortable pitch.

##### 2.4.2.3.1. Database

From the data included in experiments 2.a and 2.b, seven subjects had a visible blood vessel on their vocal folds. Recordings from comfortable and high-pitch phonations of these subjects were used for this experiment.

### 2.4.2.3.2. Method

The method was similar to the one described in experiment 2.a, but instead frames with the best visual appearance of the blood vessels were selected. Additionally, anchor points were selected based on unique features of each blood vessel, including their branching, or looping.

### 2.4.2.3.3. Results

Figure 2.6 shows boxplot of the calibrated length of a blood vessel on the vocal fold of each patient for both phonation tasks.

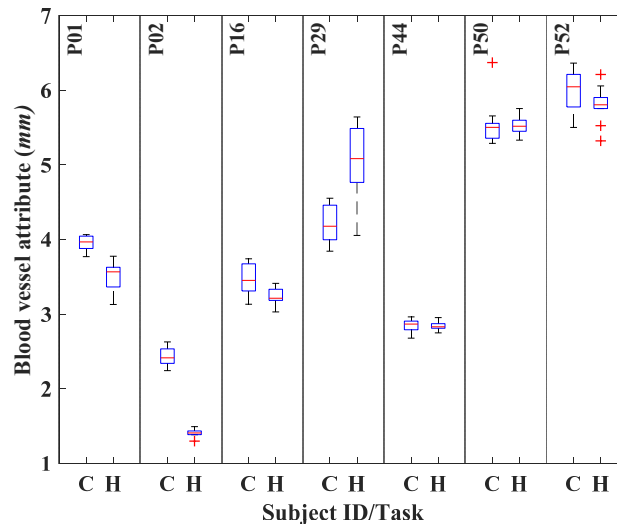


Figure 2.6. Boxplot of *mm* size of an attribute of blood vessels on the vocal fold of each subject for (C)omfortable and (H)igh pitch phonations.

The hypothesis H1e is based on a comparison of the *mm* size of a blood vessel on the vocal fold between two phonation tasks. The *mm* size of the blood vessel for each recording was computed as the median of the measurements from the 10 images. Table 2.4 presents the descriptive statistics of each phonatory group.

Table 2.4. Descriptive statistics of the *mm* size of attributes of a blood vessel on the vocal fold.

Group	mean ( <i>mm</i> )	std ( <i>mm</i> )
Comfortable pitch	4.06	1.33
High pitch	3.92	1.61

To test hypothesis H1e, a two-sided paired-samples *t*-test was used. The independent variable was the phonation task (comfortable vs. high pitch) and the dependent variable was the *mm* size of some attribute of the blood vessel on the vocal fold. The test did not detect a significant difference ( $p=0.53$ ,  $t=0.66$ ) between the two conditions. Therefore, attributes of the blood vessel on the vocal fold is a robust common attribute for indirect calibration, and it could be used for indirect calibration.

#### 2.4.2.4. Experiment 2.d: Blood vessel on a nearby tissue

The length of a blood vessel is a spatial feature that can be used for indirect calibration.<sup>206,207</sup> This experiment explores the suitability of a blood vessel on a nearby tissue. For this experiment, the following hypothesis was formed.

H1f: Calibrated attribute of a blood vessel on a nearby tissue during high-pitch phonation is significantly different from a comfortable pitch.

##### 2.4.2.4.1. Database

From the data included in experiments 2.a and 2.b, seven subjects had a visible blood vessel on a tissue near the vocal folds. Recordings from comfortable and high-pitch phonations of these subjects were used for this experiment.

##### 2.4.2.4.2. Method

The method was similar to the one described in experiment 2.c.

### 2.4.2.4.3. Results

Figure 2.7 shows boxplot of the calibrated length of a blood vessel on a nearby tissue of each patient for both phonation tasks.

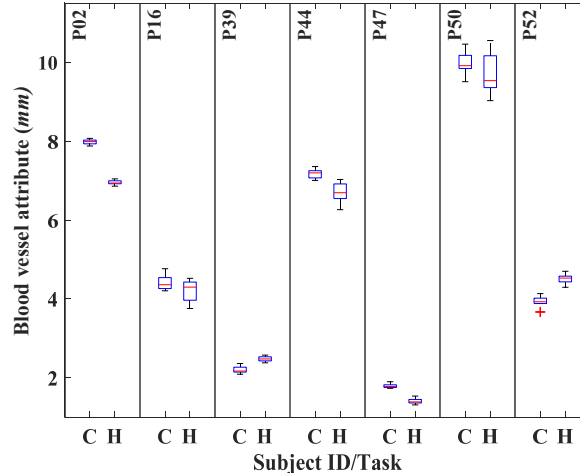


Figure 2.7. Boxplot of *mm* size of an attribute of blood vessels on a nearby tissue of each subject for (C)omfortable and (H)igh pitch phonations.

The hypothesis H1f is based on a comparison of the *mm* size of a blood vessel on a nearby tissue between the two phonation tasks. The *mm* size of the blood vessel for each recording was computed as the median of the measurements from the 10 images. Table 2.5 presents the descriptive statistics of each phonatory group.

Table 2.5. Descriptive statistics of the *mm* size of attributes of the blood vessel on a nearby tissue.

Group	mean ( <i>mm</i> )	std ( <i>mm</i> )
Comfortable pitch	5.33	3.09
High pitch	5.12	2.81

To test hypothesis H1f, a two-sided paired-samples *t*-test was used. The independent variable was the phonation task (comfortable vs. high pitch) and the dependent variable was the *mm* size of some attribute of the blood vessel on a tissue near to the vocal fold. The test did not detect a significant difference ( $p = 0.35$ ,  $t = 1.02$ ) between the two conditions. Therefore, attributes of the

blood vessel on a nearby tissue is a robust common attribute for indirect calibration, and it could be used for indirect calibration.

### 2.4.3. Experiment 3: Selecting the most suitable common attribute

Often, we could select several common attributes for performing indirect calibration. Then, an important question is how to select the most suitable one. The aim of experiment 3 is to answer this question.

#### 2.4.3.1. Experiment 3a: Registration uncertainty of different common attributes

Experiments 2.a-2.d were based on randomized measurements of different target objects from multiple recordings, and in that sense resemble the registration uncertainty test. Therefore, we could use a similar approach and estimate the registration uncertainty of each common attribute. To that end, the data was de-randomized and then the ratios between the 10 measurements per each recording and their median was computed. The interquartile range of the computed ratios was used as an estimate of the registration uncertainty. Table 2.6 presents the descriptive statistics of the registration uncertainty for different common attributes.

Table 2.6. Descriptive statistics of registration uncertainty for different selections of the common attribute.

Common attribute	mean ( <i>mm</i> )	std ( <i>mm</i> )
Vocal fold length attribute	0.027	0.012
Vocal fold width	0.036	0.019
Blood vessel on a vocal fold	0.061	0.037
Blood vessel on a nearby tissue	0.047	0.027

Based on table 2.6 the vocal fold length attribute may lead to the lowest registration uncertainty, followed by the vocal fold width, blood vessel on a nearby tissue, and finally blood vessel on a vocal fold. Additionally, for each subject, we could compare the registration

uncertainty for different common attributes and determine the best one. This was computed as the average of the registration uncertainty of the two phonation tasks per each subject. Table 2.7 shows this, where the lowest value is presented in bold face letters.

Table 2.7. Individual differences in registration uncertainty of each common attribute.

Subject ID	Common attribute			
	VF length	VF width	VF blood	Nearby blood
P01	<b>0.016</b>	0.022	0.058	×
P02	0.015	0.048	0.058	<b>0.011</b>
P16	0.023	<b>0.017</b>	0.076	0.086
P29	0.04	<b>0.036</b>	0.126	×
P39	<b>0.036</b>	0.073	×	0.049
P44	<b>0.025</b>	0.036	0.031	0.04
P47	<b>0.033</b>	0.047	×	0.051
P50	<b>0.024</b>	0.027	0.032	0.059
P52	<b>0.018</b>	0.026	0.049	0.033
P53	0.041	<b>0.027</b>	×	×
P54	<b>0.025</b>	0.034	×	×

### 2.4.3.2. Experiment 3b: Size consistency of different common attribute

#### 2.4.3.2.1. Method

Calibrated (*i.e.* the true) percent change ( $PC_{mm}$ ) of a target object was defined in section 2.3.2.2 as,

$$PC_{mm} = \frac{l_{T,mm}^2 - l_{T,mm}^1}{l_{T,mm}^1} \times 100\% \quad (2-11)$$

Using the common attribute consistency assumption (*i.e.*  $l_{C,mm}^1 = l_{C,mm}^2$ ), Equation 2-11 can be written as,

$$PC_{mm} = \frac{\frac{l_{T,mm}^2}{l_{C,mm}^2} - \frac{l_{T,mm}^1}{l_{C,mm}^1}}{\frac{l_{T,mm}^1}{l_{C,mm}^1}} \quad (2-12)$$

If the common attribute consistency assumption is violated, we would have  $l_{C,mm}^2 = l_{C,mm}^1 + \Delta$ , which leads to,

$$\widetilde{PC}_{mm} = \frac{l_{T,mm}^2 / l_{C,mm}^2 - l_{T,mm}^1 / l_{C,mm}^1}{l_{T,mm}^1 / l_{C,mm}^1} \quad (2-13)$$

Obviously, if  $\Delta = 0$ , then  $\widetilde{PC}_{mm} = PC_{mm}$ . However, if the size of the common attribute changes by  $\Delta mm$  between the two imaging sessions or imaging modalities,  $\widetilde{PC}_{mm} \neq PC_{mm}$  and the indirect approach will introduce some error into measurements. Assuming a similar pixel-to- $mm$  conversion scales for the target and the common object, we could simplify 2-13 into,

$$\widetilde{PC}_{mm} = \frac{l_{T,pixel}^2}{l_{T,pixel}^1} \times \frac{l_{C,mm}^1}{l_{C,mm}^2} - 1 \quad (2-14)$$

If  $\frac{l_{C,mm}^1}{l_{C,mm}^2} = 1$  there is no error ( $\widetilde{PC}_{mm} = PC_{mm}$ ), but as  $\frac{l_{C,mm}^1}{l_{C,mm}^2}$  deviates from the value 1, the magnitude of error increases. Consequently, we could use  $\gamma = \frac{l_{C,mm}^1}{l_{C,mm}^2}$  for comparing the consistency of different common attributes.

#### 2.4.3.2.2. Results

The value of  $\gamma$  was computed for the four common attributes presented in experiment2. Table 2.8 presents descriptive statistics of  $\gamma$  for different common attributes.

Table 2.8. Descriptive statistics of  $\gamma$  for different selections of the common attribute.

Common attribute	mean (mm/mm)	std (mm/mm)
Vocal fold length attribute	1.11	0.165
Vocal fold width	0.968	0.21
Blood vessel on a vocal fold	0.94	0.188
Blood vessel on a nearby tissue	0.974	0.136



Based on table 2.8, a blood vessel on nearby tissue may lead to the lowest measurement error, followed by the vocal fold width, blood vessel on a vocal fold, and finally the vocal fold length. Additionally, for each subject, we could compare values of  $\gamma$  for different common attributes and determine the best one. Table 2.9 shows individual trends regarding the size consistency of different common attributes.

Table 2.9. Individual trends regarding the size consistency of different common attributes.

Subject ID	Common attribute			
	VF length	VF width	VF blood	Nearby blood
P01	<b>1.012</b>	0.835	0.899	×
P02	1.127	<b>1.05</b>	0.583	0.868
P16	1.299	0.844	0.931	<b>0.986</b>
P29	1.131	<b>0.987</b>	1.217	×
P39	1.359	1.416	×	<b>1.141</b>
P44	1.302	1.132	<b>0.987</b>	0.93
P47	<b>1.04</b>	0.632	×	0.78
P50	0.916	1.058	<b>1.003</b>	0.962
P52	1.18	0.82	<b>0.96</b>	1.153
P53	0.863	<b>0.962</b>	×	×
P54	<b>0.976</b>	0.913	×	×

## 2.5. Discussions

This chapter presented a formal treatment of indirect spatial calibration. The purpose of indirect calibration is to use images taken from *different* imaging modalities or, imaging sessions, to account for confounding factors of horizontal measurements. Depending on the type of available information, the outcome of indirect calibration could be used for within-subject or between-subject size comparisons. Specifically, if one of the images is spatially calibrated (*i.e.* *mm* measurement can be achieved) the outcome of indirect calibration could be utilized for between-subject size comparison. However, if neither of the images are spatially calibrated, the outcome of indirect calibration could be utilized only for within-subject size comparison (*e.g.* pre/post changes

in the same person). The indirect calibration approach is based on identifying a proper common attribute between different recordings of each subject. The size of this common attribute is then used as a scale for spatial calibration. If the *mm* size of this common attribute is available, the calibration will map all images into a standard basis, and obviously, between-subject size comparison can be achieved. This chapter identified three conditions that govern the validity of the indirect calibration. These conditions were the registration accuracy assumption, the similarity in the vertical distance assumption, and the consistency of the common attribute assumption

The registration accuracy stipulates that the common attribute can be identified and registered accurately for each subject across all images. For example, if images are taken from different angles, the common attribute would be recorded differently and hence some error will be introduced during the registration step. Additionally, common attributes without a sharp contrast with the background would be another example of low registration consistency. A test was proposed in this chapter that can estimate the magnitude of the registration uncertainty. In that regard, a high value of registration uncertainty may indicate a serious violation of the registration accuracy assumption, which means the presence of significant errors in the calibration outcome. As a rule of thumb, a common attribute that is larger and has a sharper contrast with the background would lead a lower registration uncertainty value.

Another assumption of the indirect calibration was the existence of a certain relationship between the pixel-to-*mm* conversion scales of the common attribute and the target object. Pixel-to-*mm* conversion scale may depend on the vertical distance, the spatial location of the object, and the imaging angle.<sup>208,209</sup> However, we assumed that the pixel-to-*mm* conversion scale only depended on the vertical distance, and derived the similarity in the vertical distance assumption. Unfortunately, the vertical distance is lost during the imaging and consequently, evaluation of the

similarity in the vertical distance assumption is not an easy task (if possible, at all). In sections 2.3.2.1 and 2.3.2.2 we derived conditions satisfying the vertical distance assumption. One solution assumed that the vertical distances from the endoscope to the common attribute, and from the endoscope to the target object are similar between different imaging sessions or imaging modalities. Achieving this condition is very hard in practice. The second solution assumed that the common attribute and the target object have the same vertical distance from the endoscope. This condition may be achieved by selecting a common attribute that is part of the region of interest.

The consistency of the common attribute assumption stipulates that the actual size of the common attribute (*i.e.* its *mm* length) is fixed and does not change between different imaging sessions, or imaging modalities. This assumption is quite fundamental in the indirect calibration method and is the basis of its validity. Unfortunately, checking this assumption requires the existence of spatially calibrated images (which obviously is not available), and hence cannot be done directly. However, for laryngeal images, a method was developed that could evaluate this assumption indirectly. The method was based on the comparison of sizes of the common attribute during different vocal behaviors and phonation tasks.

This chapter used laryngeal images as a test bench for indirect calibration. The images were acquired from two different imaging modalities of intraoperative and HSV recordings. The intraoperative recordings were still images and they provided the calibrated *mm* measurements of the lesion. Conversely, HSV data were not calibrated but they provided the motion and the vibration of the vocal folds. The indirect calibration was aimed to achieve a valid comparison of some spatial (or tempo-spatial) attributes of the vocal fold between different recordings (*e.g.* vocal fold length, vocal fold velocity, etc.). HSV data were recorded pre- and post-surgery. Assuming that size of the lesion is consistent between the intraoperative and HSV sessions, the pre-surgery

HSV recordings can be calibrated using the lesion. However, the lesion is not present in the post-surgery recording and a different common attribute should be used. To that end, four common attributes of vocal fold length, vocal fold width, size of a blood vessel on the vocal fold, and size of a blood vessel on a nearby tissue were identified. The registration test was used to compare the registration accuracy of these four objects. Experiment 3.a showed that vocal fold length had the lowest registration uncertainty. This low registration uncertainty may stem from the fact that the vocal fold length was significantly longer than the other three common attributes. Additionally, the dark glottis provides a very sharp contrast for accurate detection of the vocal folds and measurement of its length. Experiment 3.b compared the consistency of the four common attributes. Interestingly, the vocal fold length had the lowest consistency. Additionally, in experiments 2.a-2.d we saw that vocal fold length was the only attribute that was significantly different between different phonation tasks. Therefore, using vocal fold length for calibration could add significant error into calibration and subsequent measurements. Table 2.10 presents a summary of the three assumptions for the indirect calibration using different common attributes. Based on table 2.10, the vocal fold width may provide the best trade-off between the three assumptions of indirect calibration. Another significant advantage of vocal fold width is the lack of ambiguity in its measurement. Specifically, some parts of the vocal fold may be occluded during the data collection which makes measurement of the vocal fold length ambiguous.

Table 2.10. Comparing suitability of different common attributes for indirect calibration of vocal folds.

Common attribute	Registration consistency	Size consistency	Vertical distance consistency
<b>Vocal fold length attribute</b>	Highest	Lowest	High
<b>Vocal fold width</b>	High	High	High
<b>Blood vessel on vocal fold</b>	Lowest	Low	High
<b>Blood vessel on a nearby tissue</b>	Low	Highest	Low

This work had several limitations that should be mentioned. The main assumption of this work was that the calibrated size of the lesion was not changing between the intraoperative and HSV recordings. The majority of the subjects included in this study were diagnosed with vocal fold polyps. Vocal fold polyp has been associated with increased stiffness<sup>210</sup>, which provides some evidence regarding the validity of this assumption. The small sample size was another limitation of this study. Specifically, the results regarding the vocal fold length and the vocal fold width were based on measurements from 11 subjects, and the results for blood vessel attributes were based on measurements from 7 subjects.

## **2.6. Conclusions**

Calibrated spatial measurements from laryngeal images could provide significant benefits for voice science research and clinical practice. However, the calibration of endoscopic images requires the existence of some auxiliary information. Recent advancements in laser-calibrated scopes may provide the required auxiliary information.<sup>122</sup> However, that technology is still in its infancy and requires significant effort and investment to become fully developed. Additionally, the functionality of the laser-calibrated system depends on specialized hardware and software, which will not be widely available in a near future. Meanwhile, an alternative calibration approach that is more accessible is needed. The indirect calibration approach could be an answer to this need. The indirect calibration depends on identifying a *proper* object that is common in different images for achieving the calibration. This chapter presented a formal treatment of this problem and identified three fundamental assumptions behind the validity of the indirect calibration. These conditions were the registration accuracy assumption, the similarity in the vertical distance assumption, and the consistency of the common attribute assumption. The registration accuracy stipulates that the common attribute can be registered with a small error on different images. The

similarity in the vertical distance stipulates that the common attribute and the region of interest are on the same vertical distances from the endoscope. The consistency of the common attribute stipulates that the calibrated size of the common attribute does not change between different images. A test was developed for evaluating the registration accuracy assumption. The similarity in the vertical distance and consistency of the common attribute could come from the domain knowledge and the anatomy of the larynx. Calibrated intraoperative images were used for calibration of pre- and post-surgery HSV recordings. Considering the absence of the lesion on post-surgery HSV recordings, four common attributes of the vocal fold length, the vocal fold width, the length of a blood vessel on a vocal fold, and the length of a blood vessel on a nearby tissue were identified. The three assumptions of indirect calibration were tested on these four attributes and it was concluded that the vocal fold width may provide the best trade-off.

### **CHAPTER 3: APPLICATION OF INDIRECT HORIZONTAL CALIBRATION TO KINEMATIC MEASUREMENTS FROM IN-VIVO HSV RECORDINGS**

*Based on:*

Ghasemzadeh H., Deliyski D. D., et al. Spatial segmentation of high-speed videoendoscopy with sub-pixel resolution using adaptive-thresholding and double curve fitting, *in Preparation*.

Ghasemzadeh H., Deliyski D. D., Hillman R. E., Mehta D. D., Verdolini K. A. Post-surgery changes in vocal fold closing velocity in patients with mass lesions, *in Preparation*.

Summary: Vocal fold kinematic measures are important features that can aid in modeling the input, output, and parameters of the phonatory system. This chapter investigates the post-surgical changes in the closing velocity of the vocal folds during phonation in patients with VF mass lesions. Transoral rigid high-speed videoendoscopy from habitual pitch/loudness of sustained phonation from 16 subjects with benign vocal fold mass lesions were recorded pre- and post-surgery, along with spatially calibrated intraoperative images. HSV recordings underwent temporal segmentation, motion compensation, spatial segmentation and spatial calibration processes. The pre-surgical HSV images were spatially calibrated by registering the lesions from the intraoperative images. The vocal fold width from each calibrated pre-surgical HSV was selected, then registered to its corresponding post-surgical HSV to provide indirect spatial calibration. Three different experiments were conducted to investigate the: (1) post-surgical changes in closing velocity, (2) differences in pre- and post-surgical left-right closing velocity symmetry, and (3) association between post-surgical changes in closing velocity and lesion size. Significant post-surgery increases were found in the closing velocity of the surgically-treated vocal

fold at multiple points throughout its length. The contralateral vocal fold showed a small insignificant improvement in the lesion contact area. (2) Closing velocity of the two vocal folds became more symmetric after surgery. (3) Post-surgical changes in closing velocity and lesion size were not significantly correlated.

### 3.1. Introduction

The closing velocity is an important kinematic feature of vocal folds' vibration which relates to their collision forces.<sup>25-27</sup> The closing velocity also correlates with the maximum flow declination rate<sup>26,163</sup> and the maximum area declination rate<sup>102,164,165</sup>, which have established association with the average produced acoustic output<sup>29</sup> and the vocal intensity.<sup>23,28</sup> Additionally, based on the time-frequency duality of the Fourier transform<sup>211</sup>, a faster phenomenon results in high-frequency components. Consequently, it is expected for higher closing velocity to lead to an increase in the energy of high-frequency components of the voice, which in turn may improve the speech intelligibility.<sup>166</sup> In summary, investigation of the closing velocity of the vocal folds could provide significant information about the phonatory mechanism and could link the input (*i.e.* airflow measurements), the output (*i.e.* the produced acoustic signal), and parameters of the phonatory system together. However, velocity is the calibrated displacement of an object with respect to time. Consequently, the computation of any velocity (including the closing velocity) depends on temporal and spatial measurements that are calibrated. In cameras, time is already calibrated. Therefore, only the spatial component should be calibrated, which can be done using the method developed in chapter 2.

Spatial calibration of in-vivo laryngeal images is a challenging task, and therefore the number of studies on velocity measures of the vocal folds is very limited, and also limited to normal



subjects. Using the color Doppler imaging technique the vocal fold velocity of  $68 \pm 10$  cm/s was reported for comfortable pitch and loudness of a sustained phonation from 10 healthy male subjects.<sup>212</sup> Another study employed the photoglottography recordings from 32 healthy subjects covering a wide range of sound pressure levels (65.46- 86.89 dB<sub>A</sub>) and reported the vocal fold maximum closing velocity of  $112 \pm 53$  cm/s.<sup>213</sup> Finally, using a parallel-laser project endoscope the average value of 100 cm/s was reported for the maximum velocity of the vocal folds from 9 healthy male subjects.<sup>25</sup> A different study reported the values of 9 to 110 cm/s during talking and phonation from 20 normal subjects.<sup>214</sup> Finally, a different study divided the uncalibrated velocity of the vocal folds by the vocal fold length, for achieving the spatial calibration.<sup>103</sup> Considering the results and discussions of chapter 2 this approach could be prone to significant errors, especially for between-subject comparison applications.

### 3.2. Aim and hypothesis

The project of this chapter has external funding, and it is tightly related to a recently approved NIH R01 grant R01 DC017923 (PI: Verdolini Abbott) with a subcontract to Michigan State University (sub-award PI: Deliyski). The second aim of that grant proposal is “*to investigate the influence of children’s **physical development** on their biological **response to voice therapy***”. Where the physical development will be quantified using velocity measures from HSV recordings. Also, the response to voice therapy will be measured using a laser-calibrated VSB system. Therefore, this project could constitute an example of indirect horizontal calibration which was developed in chapter 2 of this dissertation. In that case, the *auxiliary* information would come from the calibrated VSB recordings.

This chapter is aimed at developing a method for computation of the closing velocity of the vocal folds and studying the post-surgery changes in the closing velocity of patients with vocal fold mass lesions. To this end, the following research question is answered in this chapter.

Q2: How does the removal of a lesion from a vocal fold affects its kinematics?

To answer this research question three hypotheses were formed that are presented in this section.

Let  $m$  and  $a(t)$  denote mass and the instantaneous acceleration of a lumped model of a vocal fold, based on Newton's second law of motion we have,

$$F(t) = m \cdot a(t) \quad (3-1)$$

where  $F(t)$  denotes the net exerted external force on the vocal fold. Let  $t_{abd}$  and  $t_{add}$  denote the timestamps of the maximum abduction and the maximum adduction of the vocal fold from the same glottal cycle. The time window between  $t_{abd}$  and  $t_{add}$  is defined as the closing phase of the vocal folds. Additionally, the vocal folds are at rest at  $t_{abd}$  and  $t_{add}$  and therefore, vocal fold velocity would be equal to zero at these timepoints. Let  $t$  denotes a time during the closing phase ( $t_{abd} < t < t_{add}$ ), the closing velocity of the vocal fold at  $t$  can be computed as,

$$v(t) = \int_{t_{abd}}^t a(\tau) \cdot d\tau \quad (3-2)$$

Let,  $t_{max}$  be the time point during the closing phase ( $t_{abd} < t_{max} < t_{add}$ ) that the magnitude of Equation 3-2 becomes maximum, then  $|v(t_{max})|$  is called the closing phase maximum velocity and is the main dependent variable of this chapter.

Assuming similar sub-glottal air pressures and aerodynamic characteristics between the pre- and post-surgery conditions, and also similar interactions between the airflow and vocal folds, similar forces would be exerted on the vocal folds in both conditions. Therefore, based on Equation 3-1 we would have,

$$m_{pre} \cdot a_{pre}(t) = m_{post} \cdot a_{post}(t) \quad (3-3)$$

We can rearrange Equation 3-3 and derive,

$$a_{post}(t) = \frac{m_{pre}}{m_{post}} \cdot a_{pre}(t) \quad (3-4)$$

In vocal fold mass lesions, some extra mass is accumulated on the vocal fold, which is removed during the surgery, thus  $\frac{m_{pre}}{m_{post}} > 1$ . With everything else being equal, it is logical to hypothesize that the magnitude of the acceleration of the vocal fold would increase in the post-surgery recording. Assuming similar pitches between the pre- and post-surgery conditions, we can expect similar timings between the two recordings. Finally, by plugging 3-4 into 3-2 and using  $t_{max}$  for the upper bound of the integral we get,

$$v(t_{max})_{post} = \frac{m_{pre}}{m_{post}} \cdot \int_{t_{abd}}^t a_{pre}(\tau) \cdot d\tau \quad (3-5)$$

Obviously  $v(t_{max})_{pre} = \int_{t_{abd}}^t a_{pre}(\tau) \cdot d\tau$ , therefore,

$$v(t_{max})_{post} = \frac{m_{pre}}{m_{post}} \cdot v(t_{max})_{pre} \quad (3-6)$$

Using the  $\frac{m_{pre}}{m_{post}} > 1$ , we get,

$$|v(t_{max})_{post}| > |v(t_{max})_{pre}| \quad (3-7)$$

Consequently, it is expected for the closing phase maximum velocity of the vocal fold to increase after the surgery. Another reason behind the expected increase in the velocity of the vocal folds is the improved post-surgery glottal closure. Specifically, vocal fold lesions have been associated with an incomplete glottal closure.<sup>168,210,215–219</sup> The increased glottal gap has been associated with increased PTP.<sup>220</sup> This may indicate reduced energy transfer from the air stream into the vocal

folds. In that regards, it is quite possible that  $|F_{post}(t)| > |F_{pre}(t)|$  which would further increase the post-operative changes in the velocity.

Based on the presented rationales, the following hypothesis is made,

H2a: The closing phase maximum velocity will significantly increase after phonomicrosurgery.

Phonomicrosurgery probably would leave a scar on the vocal fold with the lesion<sup>221</sup>, which in turn, leads to changes in the biomechanical properties of the scarred-vocal fold, including increased stiffness.<sup>222</sup> However, after the phonomicrosurgery mass and morphology of the vocal fold with the lesion and the contralateral vocal fold would become more similar. Considering that the goal of the surgery is to improve the voice, it is expected for the positive changes to outweigh the negative side effects of the surgery. Therefore, it is hypothesized that,

H2b: For unilateral mass lesions, the closing phase maximum velocity of the two vocal folds will become more similar after the surgery.

A lesion with a larger area probably indicates a larger accumulation of the extra mass on the vocal fold. It is expected that removing a larger mass leads to larger post-surgery changes in the velocity of the vocal fold. Additionally, a previous subjective study based on visual evaluation of HSV recordings has suggested that the area of a lesion is a better predictor for qualitative changes in the vibratory characteristics of the vocal folds (*e.g.* left-right phase asymmetry) than its length.<sup>93</sup>

Based on these rational, the following hypothesis is made,

H2c: Post-operative change in the closing phase maximum velocity will be positively correlated with the area of the lesion.

### 3.3. Material and Method

#### 3.3.1. Participants and data acquisition

The aims of this chapter were pursued using retrospective data. Calibrate intraoperative images and HSV recordings were obtained from 26 adults with vocal fold mass lesions at Massachusetts General Hospital. Subjects were recorded using a custom-built HSV system over two different sessions. The first session was before the surgery and the second recording was carried out on average 3.5 weeks after the surgery. The HSV system consisted of the following components, a color Phantom v7.3 camera (Vision Research, Inc., Wayne, New Jersey), a 300-Watt xenon light (Model 7152A, PENTAX Medical Company Montvale, New Jersey), and a 70° 10-*mm* rigid laryngoscope (Model 49-4072, JEDMED Instrument Co, St. Louis, Missouri). The recordings were done at a sampling rate of 6,250 fps with the maximum integration time and at a spatial resolution of 320×352 pixels. The surgery was performed using cold instruments and/or a 532-*nm* pulsed potassium titanyl phosphate laser photoablation under general anesthesia. Before the operation, a surgical instrument with a known *mm* length was placed next to the lesion and an intraoperative image was recorded.

Reviewing the HSV data showed that 6 subjects (p1, p24, p39, p43, p50, p51) did not have the post-surgery HSV data, recordings from the comfortable pitch of two subjects were missing (p40, p44), the pre-surgery recording from one subject was quite blurry (p34), and glottis in the pre-surgery recording of one subject was not visible (p46). These subjects were excluded, and the rest of the analyses were carried out using recordings from the comfortable pitch and comfortable loudness sustained phonations of the remaining 16 subjects.

The registration uncertainty test described in section 2.3.3.1 was applied to all of the data including the redundant samples described in section 2.4.1. Based on this analysis, the value of

0.0357 was used as the threshold. Figure 3.1 shows a scatter plot of the registration uncertainty of the included subjects. The red dashed line represents the value of the threshold.

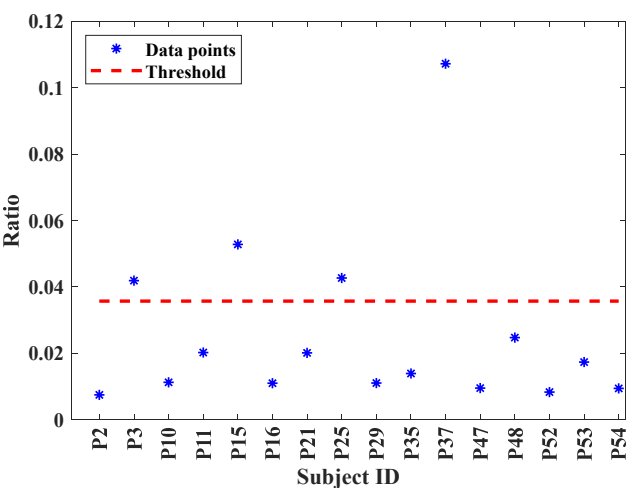


Figure 3.1. Result of registration uncertainty test for included subjects.

Based on figure 3.1, subjects p3, p15, p25, and p37 had high registration uncertainty, and hence they were excluded from the rest of the analysis. Figure 3.2 shows the intraoperative images of these subjects.

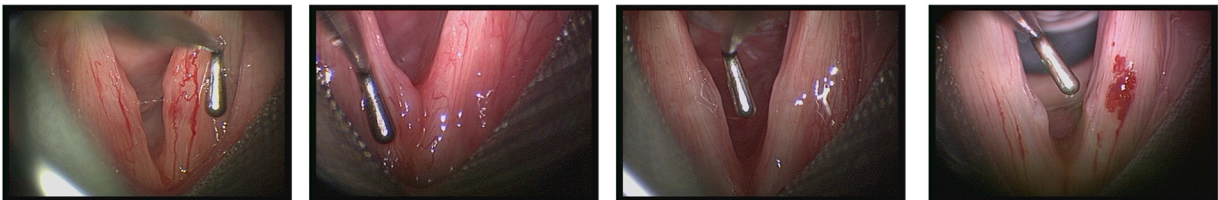


Figure 3.2. Intraoperative images from subjects with high uncertainty registration.

Table 3.1 reflects demographic and diagnosis information from the included subject.

Table 3.1. Demographic and diagnosis information of the included subjects.

Subject ID	Gender	Age	Diagnosis
P2	F	24	Right vocal fold mucoid polyp
P10	F	23	Left vocal fold hemmorhagic polyp
P11	F	50	Left vocal fold hemmorhagic polyp
P16	F	40	Left vocal fold polyp
P21	M	42	Hemmorhagic cyst on anterior aspect of left vocal fold
P29	M	52	Hemmorhagic polyp on left vocal fold
P35	M	63	Left vocal fold polyp
P47	F	17	Bilateral phonotraumatic vocal fold lesions
P48	F	50	Left vocal fold hemorrhagic polyp and a fibrovascular contact lesion on the right vocal fold
P52	M	27	Keratin cyst, sessile fibrovascular polyp, and residual sulcus on right vocal fold
P53	M	45	Right vocal fold hemmorhagic polyp
P54	M	40	Left vocal fold hemmorhagic polyp

### 3.3.2. Approach and measurements

To measure the velocity of the vocal folds, a series of pre-processing steps should be performed. These steps include temporal segmentation, motion compensation, rotation correction, spatial segmentation, and horizontal calibration. These steps are described in the following.

#### 3.3.2.1. Temporal segmentation

The act of phonation is a complex phenomenon and requires accurate timing between different body organs and depends on specific laryngeal posture and glottal configuration.<sup>223–225</sup> The phonation starts with the pre-phonatory adjustment phase, where the vocal folds take the appropriate posturing.<sup>226</sup> Additionally, there is a time lag between the first vibration of the vocal folds to the first glottal contact, also known as glottal attack time.<sup>116</sup> Voice offset and voice break are other temporal characteristics of the phonation. Recording from a full phonation cycle includes most of these temporal features. Additionally, a single recording could include multiple repetitions

of the phonatory cycle. Considering the aim of this chapter, we need to find timestamps corresponding to the onset and offset of phonation. The purpose of temporal segmentation is to address this need.

Temporal segmentation can be automated based on different glottal features. We adopted the method based on the fundamental frequency ( $f_0$ ) contour.<sup>111</sup> The fundamental frequency was estimated based on the glottal area waveform (GAW) estimate.<sup>111</sup> Main steps of the method are described here shortly. A more detailed description of the algorithm can be found in <sup>111,116,227</sup>. First, the temporal difference between consecutive frames of the video recording was computed. Large temporal differences, corresponding to the movements of the vocal fold edges, were detected using a thresholding technique, and then they were summed over time. Following this process, a mask was created that contained the region of interest (*i.e.* all possible spatial locations of the edges of the vocal folds). Let  $f(x,y,t)$  denotes the result of applying the mask on the frame  $t$  of the recording. The  $y$ -direction second central moment of inertia ( $I_2(y,t)$ ) of  $f$  was computed as,

$$I_2(y,t) = \sqrt{\frac{\int f(x,y,t)x^2dx}{\int f(x,y,t)dx}} - I_1^2(y,t) \quad (3-8)$$

where  $I_1(y,t)$  was equal to,

$$I_1(y,t) = \frac{\int f(x,y,t)x dx}{\int f(x,y,t)dx} \quad (3-9)$$

The  $y$ -direction estimated of the GAW ( $GAW_y(t)$ ) was computed as the integral of  $I_2(y,t)$  over all rows of the image,

$$GAW_y(t) = \int I_2(y,t)dy \quad (3-10)$$

The  $x$ -direction estimated of the GAW ( $GAW_x(t)$ ) was computed similarly. The final estimate of the GAW was computed as the root mean square of  $x$ - and  $y$ -direction GAW estimates,



$$GAW(t) = \sqrt{(GAW_x(t)^2 + GAW_y(t)^2)/2} \quad (3-11)$$

Finally, the fundamental frequency was computed based on windowing and autocorrelation analysis of the GAW estimate. Temporal segmentation was achieved based on the analysis of the fundamental frequency contour. Figure 3.3 shows an example of temporal segmentation outcome.

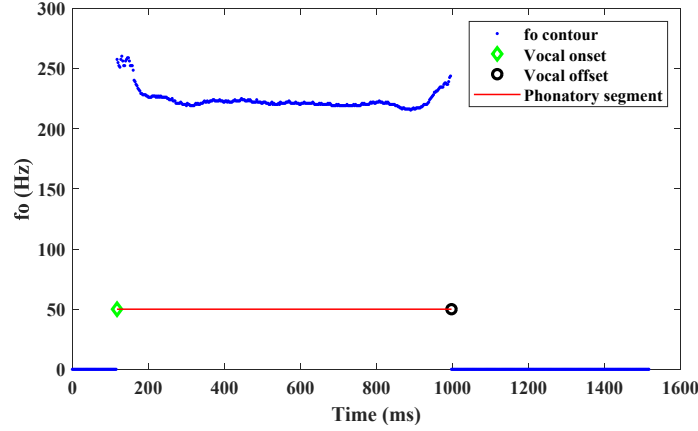


Figure 3.3. An example of temporal segmentation outcome.

### 3.3.2.2. Motion compensation

The position of the endoscope could change during the HSV data collection. Such movements will lead to changes in the spatial location of the vocal folds and could impact the performance and accuracy of subsequent measurements or analysis. Motion compensation can be employed to account for endoscopic movements. Motion compensation is an image registration process that maps vocal folds from different frames of the recording into a fixed and constant coordinate. Depending on the type of endoscopic movement, different types of motion compensations may be needed.<sup>227</sup> We assumed a motion that leads to anterior-posterior and left-right displacement of the vocal fold in the HSV frames. The method proposed in<sup>125,227</sup> was adopted to compensate for this factor. The main steps of the method are described here shortly, but a more detailed description of the algorithm can be found in<sup>125,227</sup>.

An intensity-based registration method was used for motion compensation.<sup>125</sup> The key idea of the method is based on the fact that motions of vocal folds are happening much faster (70-400 Hz) than the motion of the endoscope.<sup>125</sup> Therefore, we could use a low pass filter and separate the two components from each other. The method starts with computing the temporal difference of consecutive frames of the data. Then, the high-frequency components of the motions are filtered out. This step leads to the removal of the vibration of the vocal folds, and hence only the gross movements of the vocal folds will remain. Then, the region containing the vocal fold (ROI) is determined. This is achieved by applying a thresholding technique and only retaining pixels with high intensities. Next, the motion vector between two frames of the data is determined. The translation vector that minimized the least absolute difference ( $L_1$  norm) between intensities of the ROI from the two frames was selected as the best estimate. Finally, the registration task was achieved by applying the motion vector on the data. Figure 3.4 shows kymogram of a data before and after motion compensation. The scanning lines of both kymograms were matched based on the location of a blood vessel from the first frame of the selected portion of each video data.

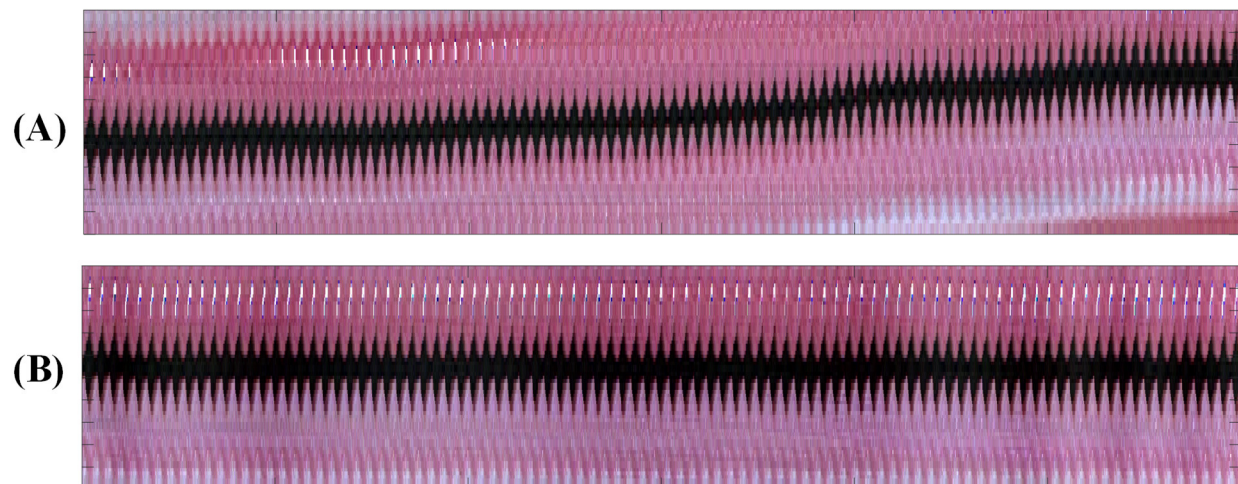


Figure 3.4. An example of motion compensation: (A) kymogram before motion compensation, (B) kymogram after motion compensation.

### 3.3.2.3. Rotation correction

The endoscope could have an angle relative to the vocal folds. This would result in vocal folds that are rotated in the image. More precisely, under such circumstances, the glottal midline would have an angle with the  $y$ -axis of the image. This rotation could change the kymogram and the subsequent velocity measurements. Figure 3.5(A) depicts a kymogram from a recording with a  $30^\circ$  rotation. Our measurements showed a maximum excursion of 12 pixels for the vocal fold at each cycle. Figure 3.5(B) depicts the kymogram of the same recording after the correction. Our measurements showed a maximum excursion of 10 pixels for the vocal fold at each cycle. Comparing these two conditions shows a significant error in uncorrected data. Specifically, uncorrected data shows 20% higher excursion, which translates into higher velocity. It is noteworthy that, the scanning lines of both kymograms were matched based on the location of a blood vessel from the first frame of the corresponding video data. Another problem with uncorrected data is that measurements from the left and right vocal folds would not be comparable. Because the left and right edges of the vocal folds in the uncorrected kymogram do not belong to the same section along the anterior-posterior axis.

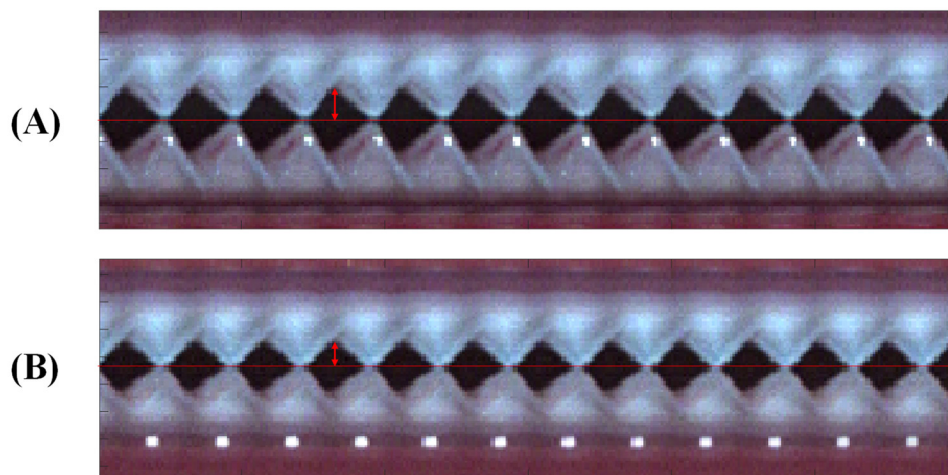


Figure 3.5. Effect of endoscopic rotation on the kymogram: (A) kymogram before rotation compensation, (B) kymogram after rotation compensation.

An automated method is presented here that can account for this factor. The method consists of four steps.

**Step1:** Estimation of the GAW

GAW was estimated based on an adaptive thresholding method. The method assumes that the location of the anterior commissure and the posterior end of the vocal folds is known. The user can provide these parameters by clicking the two ends of the vocal folds. A box with a width of 100 pixels around the clicked points is selected from the recording. This box will enclose the vocal folds. The probability density function (pdf) of the red channel from the box is estimated using a Gaussian kernel. Considering the high number of data points, this step can be sped up by random sampling. Figure 3.6(A) depicts the estimated pdf of data for 100000 randomly selected samples. The pdf can often be modeled as a mixture of three different distributions. The first distribution would be an estimate of pdf of pixels inside the glottis. The second distribution would be an estimate of pdf of pixels on the vocal folds or the nearby tissues. The third distribution would be an estimate of the pdf of reflection lights. The black reference was defined as the bin corresponding to the deep between the first two peaks. Figure 3.6(A) illustrates this. The black reference was used for the thresholding of the data. GAW for each frame was computed as the number of black pixels.

**Step2:** Finding frames with the maximum abduction from each glottal cycle.

First, the ripples of GAW were removed by applying a Hanning window with a size of 5. Figure 3.6(B) presents the smoothed GAW of the data. Timepoints of all local maxima of the smoothed GAW were detected, and their corresponding frames were extracted from the data.

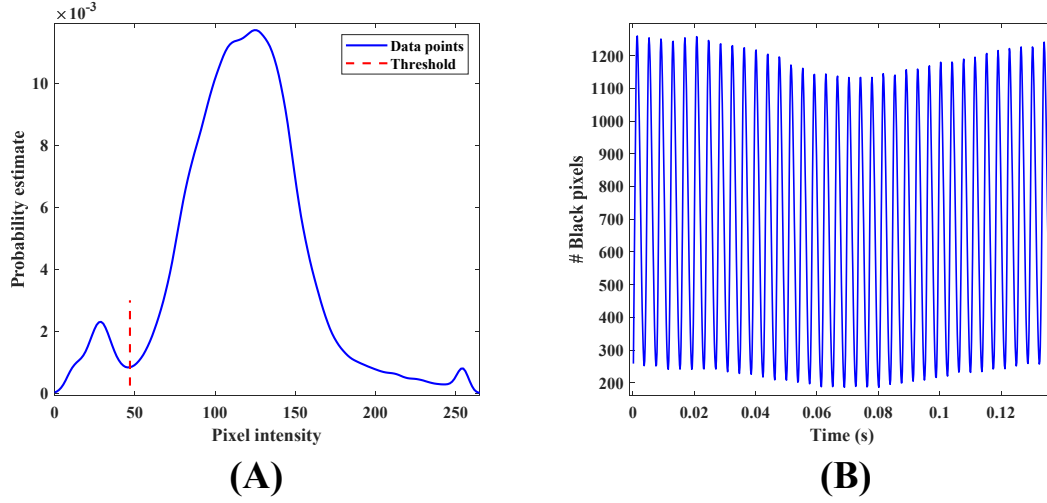


Figure 3.6. Estimation of the GAW: (A) pdf of the red channel, and the computed black threshold, (B) GAW estimate after applying the black threshold.

### Step3: Detection of the glottal midline.

The following process was repeated for all extracted frames. The frame was thresholded using the black reference and it was converted into a binary image. The object with the largest area was selected and then it underwent the morphological operation of closing with a circular structuring element with a radius of 2 pixels. The first moment of inertia (corresponding to the center of glottis at each row) was computed for each row of the image. Let  $J(x,y)$  denotes the binary image. Equation 3-12 shows the formula for the computation of the first moment of inertia for row  $y$  ( $I_y$ ).

$$I_y = \frac{\int J(x,y)x dx}{\int J(x,y) dx} \quad (3-12)$$

A linear line was fitted on the computed centers of the glottis. The angle between this line and the  $x$ -axis was computed and stored for further analysis. Figure 3.7(B) shows the outcome of this step.

### Step4: Rotation correction

Assuming a constant rotation angle throughout the recording, the correction angle was estimated as the mean of values computed from all frames, after removing the top and bottom 5% of the data (trim mean with 0.1 level). This approach makes estimation of the angle robust to the

presence of outliers. Finally, all frames were rotated by this value. Figure 3.7(C) shows the outcome for a frame of data. It is noteworthy that the method can easily be adapted to conditions where the rotation angle is changing throughout the recording.

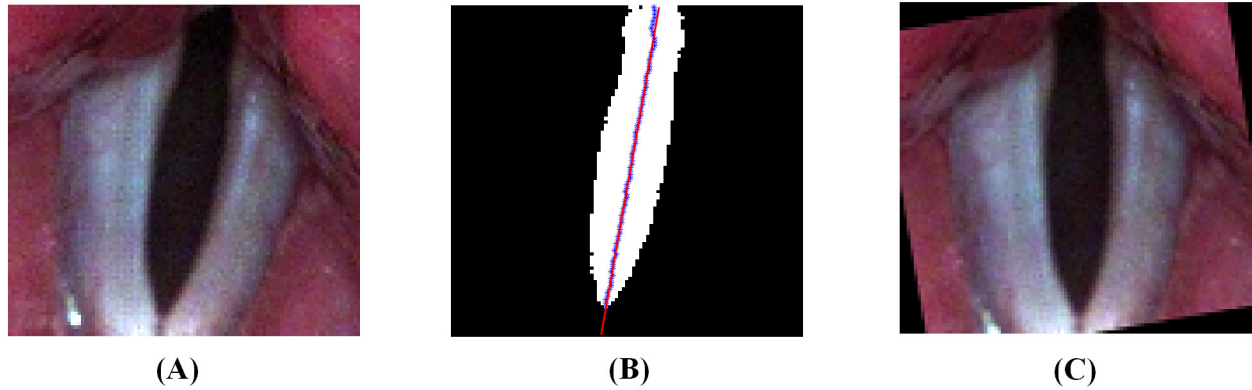


Figure 3.7. Rotation correction for a frame of data: (A) before correction, (B) segmented glottis with the fitted line on the first moment of inertia from each row, (C) after correction.

#### 3.3.2.4. Spatial segmentation

Computation of the velocity of vocal folds depends on the accurate detection of the edges of the vocal folds. Spatial segmentation is the process that achieves this. Different methods have been proposed in the literature for this purpose, including intensity thresholding<sup>98</sup>, level set segmentation<sup>120</sup>, active contours<sup>123,124</sup>, and region growing<sup>118</sup>. A new method for spatial segmentation is presented here that takes full advantage of the temporal and spatial redundancy of the vocal fold edges and can achieve a sub-pixel resolution. The method assumes that recordings are motion- and rotation-compensated. Additionally, the method assumes that the location of the anterior commissure and the posterior end of the vocal folds is known. The user can provide these parameters by clicking the two ends of the vocal folds. While this information can be estimated automatically (*e.g.* processing the temporal difference of frames), the user can provide it very accurately and without too much effort. This information is used as an initial estimation of the

glottal midline and the two ends of the glottis in a recording. The proposed algorithm consists of 3 steps.

**Step1:** Temporal curve fitting

The spatial location of a certain point on a vocal fold edge cannot abruptly change from one frame to the next one. More precisely, the function determining the coordinate of a specific point on a vocal fold edge should be continuous in time. This step exploits this temporal redundancy of the data. To that end, kymograms of the recording between the two user-selected points were created. Then, the following processes were done on each kymogram.

1.a: The local black threshold (*i.e.* the threshold for a specific scanning line along the anterior-posterior axis) was computed. The 20<sup>th</sup> percentile of each row of the red channel of the kymogram was computed. A 5<sup>th</sup> order Hanning window was used to remove the ripples and to make the result smooth. A window with a size of 31, centered at the glottal midline was selected from the result (figure 3.8(A)), and its minimum was selected as the local black threshold.

1.b: The ROI corresponding with the glottis was segmented. To that end, the red channel of the data was thresholded with the computed black reference. The clutters were removed for the computed binary image. This was achieved by computing the area of all objects, and then constructing their pdf using a Gaussian kernel. For multimodal distributions, the maximum size of the clutter was determined as the minimum between the first two peaks (refer to figure 3.6 for an example), and the value of 4 was used otherwise. The binary image underwent a closing operation with a circular structuring element with a radius of 1 pixel. A window with a size of 21, centered at the glottal midline was retained from the binary image, and the rest was set to zero.

1.c Two different curves (one per each vocal fold edges) were fitted on the data. First, the ROI mask was summed on all columns. The result was smoothed with a 3<sup>rd</sup> order Hanning window.

The location of the maximum was recorded as the current midline estimate. Let  $M(:, i)$  denotes column  $i$  of the ROI. The row index of the first non-zero element of  $M(:, i)$  was stored in a variable called  $u(i)$ . If all elements of  $M(:, i)$  were zero, the current midline estimate was stored in  $u(i)$ . Using a similar approach, the last non-zero element of  $M(:, i)$  was stored in a variable called  $l(i)$ . If all elements of  $M(:, i)$  were zero, the current midline estimate was stored in  $l(i)$ . In that regard,  $u(i)$  and  $l(i)$  stored the initial estimate of  $y$ -coordinates of the two edges of the vocal folds from the kymogram. Separate curves were fitted on vectors  $\mathbf{u}$  and  $\mathbf{l}$ . Depending on the vibrating characteristics of the vocal folds different types of curves may be employed at this step. If the kymogram has clear periodicity, using Fourier curves offers more robustness to noise and outliers. Otherwise, spline lines may be used. Due to the presence of lesions, some of our kymograms were not fully periodic, hence the spline curves with a smoothing factor of 0.1 were used. Figure 3.8 depicts different stages of step1.

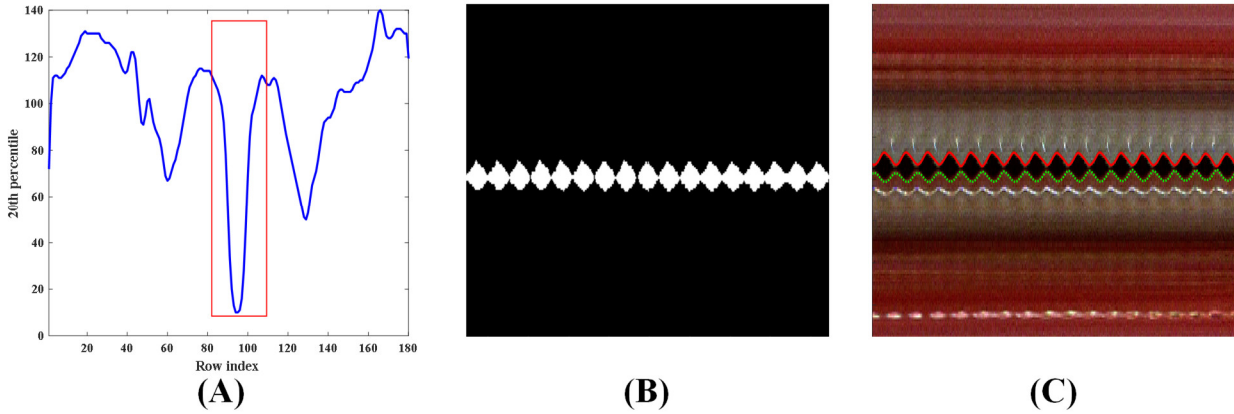


Figure 3.8. Temporal curve fitting results: (A) local black reference estimation, the red window shows the search window, (B) ROI segmentation, (C) detection of vocal fold edges.

## Step2: Outlier removal

The step1 only exploited the temporal redundancy of the data. That is, each scanning line in the anterior-posterior direction was segmented independently. Therefore, two points adjacent to each other on a vocal fold can show very strong and abrupt changes. Often, this phenomenon was



observed on the lesion site or the two ends of the vocal folds. This step takes care of such instances and prepares the data for the next stage.

Executing the step1 results in two vectors per each scanning line. Each vector stores the  $x$ -coordinates of one of the edges of the vocal folds for different time points. Therefore, the information from step1 may be concatenated into L and R matrices. Let  $L(:, i)$  denotes the column  $i$  of matrix L, where it will store the  $x$ -coordinate of all points on the edge of the left vocal fold at time point  $i$ . A 9<sup>th</sup> order polynomial with the least absolute residuals (LAR) cost function was fitted on  $L(:, i)$ . Rows corresponding to the absolute value of residual greater than 2 were designated as outliers and excluded from further analysis. Matrix R was processed similarly. Figure 3.9(A) shows this step.

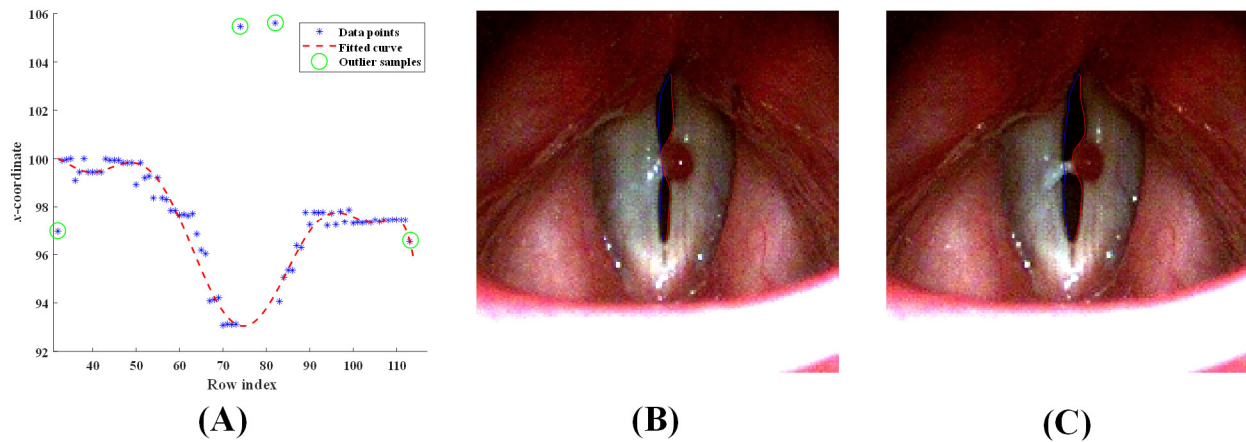


Figure 3.9. Spatial curve fitting results: (A) outlier removal step, (B and C) segmented edges of the vocal fold for two different timepoints.

### Step3: Spatial curve fitting

The  $x$ -coordinate of two adjacent points on a vocal fold edge cannot abruptly change in each frame of the data. More precisely, the function determining the edges of the vocal fold at each time point should be continuous in space. This step exploits this spatial redundancy of the data. To that end, a spline curve with a smoothing factor of 0.06 was fitted on every column of matrices  $L$  and

R. These curves will be the output of the spatial segmentation process and they will constitute the edges of the vocal folds for different time points. Figures 3.9(B-C) show the result.

### **3.3.2.5. Horizontal calibration**

Computation of the velocity of the vocal folds depends on tracking *mm* displacements of the edges of the vocal folds which are horizontal measurements. This task can be achieved by computing the pixel displacements of the edges of the vocal folds and then converting them into *mm* displacement using the indirect calibration method developed in chapter 2. To that end, a *proper* common attribute should be determined. Three steps were followed for horizontal calibration of HSV recordings.

**Step1:** Computing the *mm* length of the lesion from the intraoperative images

The pixel lengths of the lesion and the surgical instrument were measured from the intraoperative image of each subject. This task was repeated 10 times and then their median was recorded. Considering the known *mm* length of the surgical instrument, the pixel-to-*mm* conversion scale of the intraoperative image was computed. This value was then multiplied with the computed median of pixel length of the lesion to compute the *mm* length of the lesion.

**Step2:** Calibration of pre-surgery HSV recording

Ten timepoints were selected randomly from each HSV recording and then frames within their corresponding glottal cycles were evaluated subjectively. The frame with the best visual appearance of the lesion was selected. The pixel length of the lesion was computed from each selected frame. The median of these 10 measurements was used as the final estimate of the pixel length of the lesion. Considering the known *mm* length of the lesion (Step1), the pixel-to-*mm* conversion scale of the pre-surgery HSV data was computed.

**Step3:** Calibration of post-surgery HSV recording

In chapter 2 we showed that the vocal fold width was a robust attribute for calibration of HSV recordings. Considering that the lesions are not present in the post-surgery recordings, the vocal fold width was used for calibration of the post-surgery data. Following the method described in section 2.4.2.2.2 pre- and post-surgery recordings of each subject were investigated for an appropriate anchor point. Ten frames from the pre- and post-surgery recordings of each subject were selected. Following the method described in section 2.4.2.2.2, the pixel width of the vocal fold was measured from all selected frames. The medians of the measurements from the pre- and post-surgery recordings were computed for each subject. Based on the outcome of the step2, the *mm* width of the vocal fold in pre-surgery data was computed. This value in combination with the median of pixel width of the vocal fold from the post-surgery recording was used for computation of the pixel-to-*mm* conversion scale of post-surgery HSV data.

### **3.3.2.6. Velocity measurements**

Reviewing the data showed that each recording contained different numbers of glottal cycles. Additionally, some of the recordings did not include the onset or offset. To make this factor uniform across all recordings, the most stable portion of each recording was detected and used for further analysis. The selection strategy was as follows. The uncalibrated GAW was computed based on detected edges. GAW was smoothed using a 5<sup>th</sup> order Hanning window. Indexes of the maximum (corresponding to the maximum abduction) were computed and used as timestamps for different glottal cycles. The vocal fold velocity depends on the magnitude of the lateral excursion of the vocal folds; therefore, the most stable region of phonation was determined based on the dynamics of the excursion of the vocal folds. Specifically, the average value of GAW in each glottal cycle was computed. The fifty consecutive cycles that showed the lowest value of the interquartile range for the mean of GAW were used for the rest of the analysis. This approach also

ensures that any possible occlusion of the vocal folds remains relatively constant. Figure 3.10 presents a comparison between GAW from the least and the most stable portions of a recording.

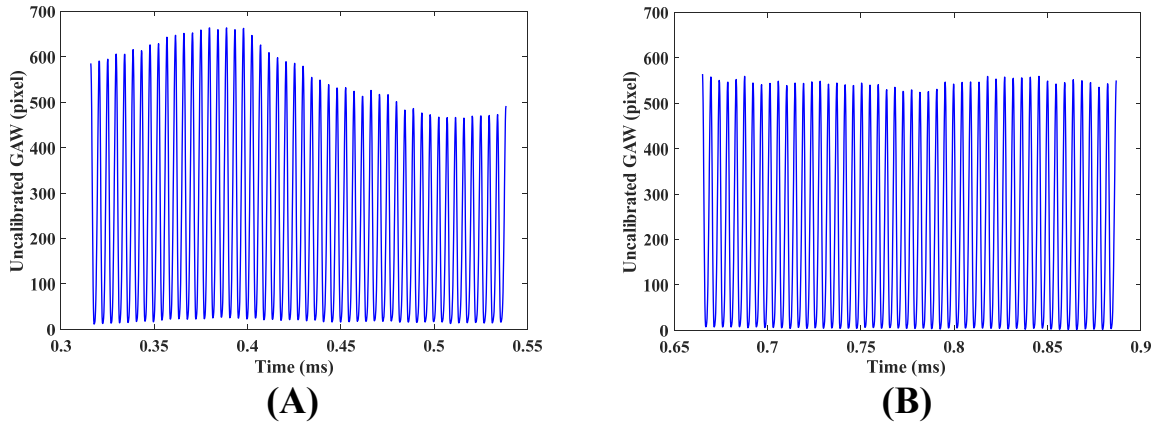


Figure 3.10. Selection of the data: (A) the least stable portion of a phonation, (B) the most stable portion of a phonation.

The pixel displacements of the estimated edges between consecutive frames were measured and then converted into the *mm* displacements, using the appropriate pixel-to-*mm* conversion scale (section 3.3.2.5). Finally, the velocity of each vocal fold at scanning line  $y$  (corresponding to location  $y$  along the anterior-posterior axis) was computed according to Equation 3-13.

$$v_y(t) = \frac{d_y(t)}{\tau} \quad (3-13)$$

where  $d_y(t)$  denotes the *mm* displacement of point  $y$  along the anterior-posterior axis on a vocal fold edge between frames  $t$  and  $t+1$ , and  $\tau$  denotes the time-difference between consecutive frames.  $\tau$  can be computed based on the known frame rate of the recording. Our investigation showed that the computed velocities near the two ends of the vocal folds, and near the lesion site sometimes had very sharp discontinuity. To remedy this, each  $v_y(t)$  was smoothed using a Hanning window.

The investigation of hypotheses of this chapter depends on inter- and intra-subject comparisons of vocal folds velocities. For a vocal fold with the length of  $l$  pixels,  $l$  different velocity time-sequences can be computed, per each vocal fold. However, meaningful inter- and

intra-subject comparisons depend on selecting comparable points on the vocal folds. This selection was subjected to multiple complications. First, the true length (*i.e. mm*) of the vocal fold would be dissimilar in different subjects. Second, investigating the data showed that the full vocal length was not visible in some of the recordings. This was primarily due to arytenoid hooding, epiglottis obstruction, or accumulation of significant mucous on the anterior commissure. Third, the recording from different subjects was done at different working distances. In summary, the number of measured velocity time-sequences depends on the true length of the vocal fold, the imaging working distance, and the magnitude of the vocal fold occlusion. To tackle this problem, three different strategies were taken. Each strategy provides a scanning line  $y$  (corresponding to location  $y$  along the anterior-posterior axis) for computation of the velocity.

The first strategy was based on finding the scanning line  $y$  that led to the maximum velocity measure. To that end, the scanning line with the maximum velocity measure was determined from each glottal cycle. This process led to 50 values. The mode of the computed value was used as the scanning line  $y$ . The second strategy was based on the scanning line  $y$  that passed through the middle of the lesion. The value of  $y$  was determined from each pre-surgery recording. A proper anchor point was selected for determining the comparable point on the post-surgery recording. The third strategy was based on the scanning line  $y$  that passed through the middle of the visible vocal fold. Regardless of the strategy taken, for each selected scanning line  $y$ , the velocity time-sequences at lines  $[y-2, y-1, y, y+1, y+2]$  were computed, and then they were averaged (over the  $y$ -direction) for the analysis. This step was taken to remove some of measurement errors.

### **3.4. Experiments and results**

Three experiments were conducted to answer the research questions of this chapter. Experiment 1 investigates changes in the closing velocity of the vocal folds following the surgery.

Experiment 2 presents the analysis on similarity between the closing velocity of the left and the right vocal folds in pre- and post-surgery conditions. Experiment 3 studies the association between the area of the lesion and the post-surgery changes in the closing velocity of the vocal fold with a lesion. This section presents details of each experiment, followed by results and related discussions.

### 3.4.1. Experiment1: Post-surgery changes in closing velocity

This experiment investigates the intra-subject changes in the closing velocity of the vocal fold following the surgery. The following hypothesis was formed for this experiment.

H2a: The closing phase maximum velocity will significantly increase after phonomicrosurgery.

To investigate H2a, timestamps of the closing phrases of the vocal folds should be determined. GAW was computed based on the detected edges, and then it was smoothed using a 5<sup>th</sup> order Hanning window. Indexes of the local maxima (corresponding to the maximum abduction) and the local minima (corresponding to the maximum adduction) were computed. The time window between a minimum and its preceding maximum was defined as a closing phase. All closing phases were determined for each token.

Following the discussion of the previous section, three different locations for measuring the closing velocities were used. The measurement from the scanning line that led to the maximum value will be represented as  $v_{max}^L$  and  $v_{max}^R$  for the left and right vocal folds. The measurement from the scanning line passing through the middle of the lesion will be represented as  $v_{lesion}^L$  and  $v_{lesion}^R$  for the left and right vocal folds. The measurement from the scanning line passing through the middle of vocal fold length will be represented as  $v_{mid}^L$  and  $v_{mid}^R$  for the left and right vocal folds.

Figure 3.11 shows the boxplots of  $v_{max}^L$  and  $v_{max}^R$  for different subjects pre- and post-surgery. The most immediate observation is that different subjects have dissimilar behaviors. For example, subjects p16, p35, p47, p53, p54 showed a decrease in  $v_{max}^L$  following the surgery. Additionally, the left and right vocal folds could show dissimilar trends following the surgery (e.g. p47 and p53).

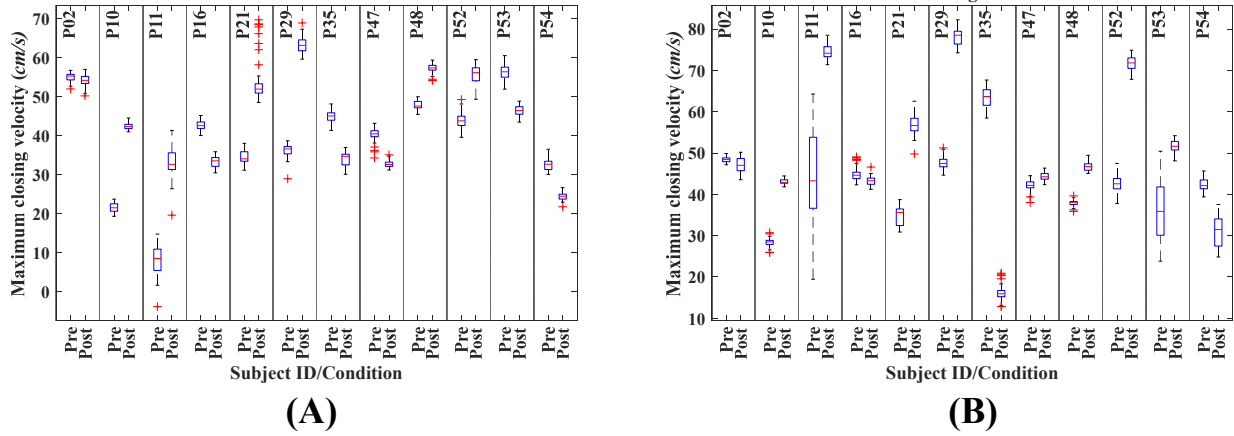


Figure 3.11. Boxplot of closing phase maximum velocity for different subjects pre- and post-surgery: (A) box plot of  $v_{max}^L$ , (B) box plot of  $v_{max}^R$ .

A similar analysis was done for  $v_{lesion}^L$  and  $v_{lesion}^R$ . Figure 3.12 shows the result. Based on this figure we see that most subjects had an increase of closing velocity at the lesion site.

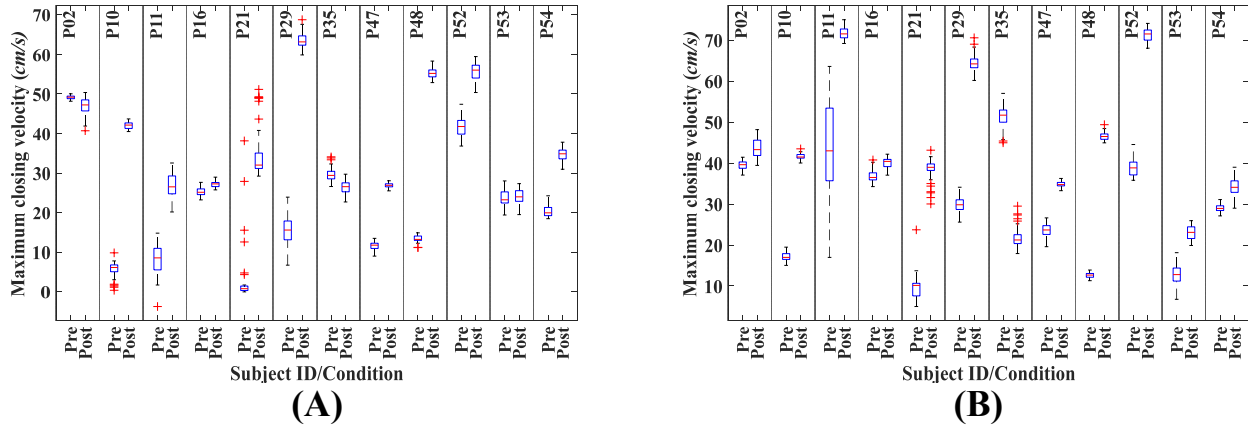


Figure 3.12. Boxplot of closing phase maximum velocity for different subjects pre- and post-surgery: (A) box plot of  $v_{lesion}^L$ , (B) box plot of  $v_{lesion}^R$ .

A similar analysis was done for  $v_{mid}^L$  and  $v_{mid}^R$ . Figure 3.13 shows the result. Based on this figure we see that most subjects had an increase of closing velocity in the middle of the vocal fold.

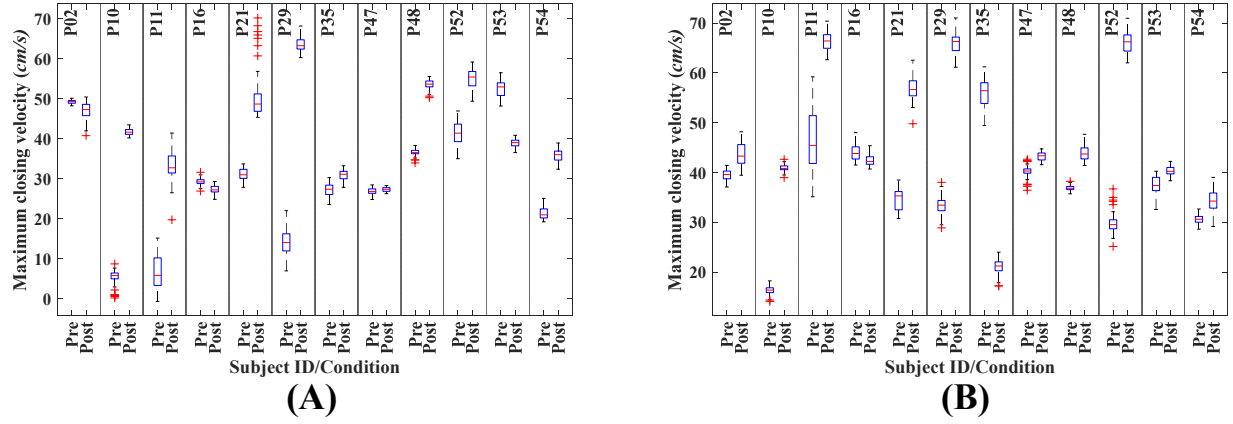


Figure 3.13. Boxplot of closing phase maximum velocity for different subjects pre- and post-surgery: (A) box plot of  $v_{mid}^L$ , (B) box plot of  $v_{mid}^R$ .

Finally, investigating figures 3.11-3.13 reveals a peculiar trend for subject p35. Specifically, closing velocity for this subject shows a consistent decrease (with the exception of  $v_{mid}^L$ ) in the closing velocity post-surgery.

To quantify the qualitative trends observed in boxplots and to test H2a, a paired-sample  $t$ -test was adopted. The independent variable was the recording condition (pre/post) and the dependent variable was the maximum closing velocity at different scanning lines. The closing velocity for each subject was computed as the median of measurements from the 50 cycles. The Bonferroni correction was used to address the issue of the increased likelihood of type I error due to multiple testing. Table 3.2 shows the descriptive statistics of each measurement. Table 3.3 shows the result of  $t$ -tests.



Table 3.2. Descriptive statistics of closing velocity at different scanning lines (mean $\pm$ std).

Scanning location	Pre (cm/s)	Post (cm/s)
$v_{max}^L$	38.72 $\pm$ 13.53	44.13 $\pm$ 12.43
$v_{max}^R$	42.7 $\pm$ 8.63	50.37 $\pm$ 18.01
$v_{lesion}^L$	20.39 $\pm$ 14.39	38.5 $\pm$ 13.81
$v_{lesion}^R$	28.72 $\pm$ 13.6	44.26 $\pm$ 16.83
$v_{mid}^L$	28.36 $\pm$ 15.27	41.87 $\pm$ 11.77
$v_{mid}^R$	37.1 $\pm$ 9.77	47.06 $\pm$ 14.07

Table 3.3. Results of the paired-sample  $t$ -test for the closing velocity at different scanning lines.

Scanning location	$t$	$p$
$v_{max}^L$	1.28	0.23
$v_{max}^R$	1.19	0.26
$v_{lesion}^L$	3.58	0.004
$v_{lesion}^R$	2.84	0.016
$v_{mid}^L$	2.63	0.02
$v_{mid}^R$	1.8	0.09

Using the Bonferroni correction and the significance level of 0.05, only the closing velocity of the left vocal fold at the lesion site ( $v_{lesion}^L$ ) shows a significant change after the surgery. Referring to table 3.3 we see a positive  $t$ -value for this variable and therefore, we could conclude that the closing phase maximum velocity at the lesion site has increased significantly after the surgery. It is noteworthy that  $v_{lesion}^R$ ,  $v_{mid}^L$ , and even  $v_{mid}^R$  had also low  $p$ -values, but they did not reach the significant level. It is quite possible for these variables, to become significant if we had a bigger sample size. Finally, we could see a consistent and interesting trend in tables 3.2 and 3.3. Specifically, the right vocal fold on average shows a higher closing velocity for all variables in both conditions (pre/post) than the left vocal fold. Investigation of table 3.1 shows that the majority of the subjects had a lesion on the left vocal fold. If this is correct, we may expect to see a bigger improvement in the closing velocity of the left vocal fold following the surgery. Referring to the  $t$

column in table 3.3 we see a bigger  $t$ -statistic for measurements from the left vocal fold, which supports this expectation. A different analysis was used to test this subjective observation.

The post-surgery changes in the closing phase maximum velocities of patients with unilateral lesions at different scanning lines were investigated. Table 3.4 shows the descriptive statistics, where  $v_{max}^{with}$ ,  $v_{lesion}^{with}$ ,  $v_{mid}^{with}$  correspond with the scanning line producing the maximum closing velocity, the line passing through the middle of the lesion, and the line passing through the middle of the vocal fold with the lesion, respectively. Table 3.5 shows the results of the  $t$ -test for these variables.

Table 3.4. Descriptive statistics of closing velocity at different scanning lines (mean $\pm$ std).

Scanning location	Pre (cm/s)	Post (cm/s)	Improvement (cm/s) %
$v_{max}^{with}$	34.8 $\pm$ 11.95	45.33 $\pm$ 14.75	30%
$v_{lesion}^{with}$	19.68 $\pm$ 13.41	39.04 $\pm$ 16.46	98%
$v_{mid}^{with}$	24.02 $\pm$ 12.09	42.98 $\pm$ 13.08	79%

Table 3.5. Results of the paired-sample  $t$ -test for the closing velocity of the vocal fold with a lesion at different scanning lines.

Scanning location	$t$	$p$
$v_{max}^{with}$	2.09	0.07
$v_{lesion}^{with}$	3.65	0.005
$v_{mid}^{with}$	3.43	0.008

Comparing tables 3.3 with 3.5 shows a consistent improvement (a lower  $p$ -value and a higher  $t$ -statistic) in the later analysis for all measures. Finally, a similar analysis was done for the contralateral vocal fold. Table 3.6 shows the results of the  $t$ -tests. This table shows the opposite behavior of table 3.5, where the  $t$ -statics show lower values (hence smaller effect sizes), and the  $p$ -value show higher values.

Table 3.6. Results of the paired-sample  $t$ -test for the closing velocity of the (cont)ralateral vocal fold at different scanning lines.

Scanning location	$t$	$p$
$v_{max}^{cont}$	0.52	0.62
$v_{lesion}^{cont}$	1.75	0.11
$v_{mid}^{cont}$	0.99	0.35

In summary, we could make the following conclusion regarding the closing phase maximum velocity of the vocal folds following the surgery. The closing phase maximum velocity of the vocal fold with a lesion improves (at least) at multiple points following the surgery. The closing phase maximum velocity of the contralateral side only shows a small improvement at the location of the lesion (this improvement did not reach the significance level due to the small sample size). Finally, the closing phase maximum velocity could be computed from different scanning lines along the anterior-posterior axis. The result of this experiment suggests that the selection of the scanning line could have a significant effect on the potency of the measure for explaining the intervention outcome. For example, the scanning line producing the maximum velocity is the easiest approach to implement, as it does not need a registration step (*i.e.* finding a comparable scanning line in different recordings). However, it may produce a significantly inferior outcome (none of the  $p$ -values even reached the 0.05). Conversely, employing the scanning line passing through the middle of the lesion seems to be the most promising location for computing the velocity measures. The  $p$ -values for this measure from the lesioned-vocal fold and the contralateral side produced the smallest  $p$ -values and the largest  $t$ -statistic (hence a larger effect size).

### 3.4.2. Experiment2: Post-surgery similarity between the two vocal folds

Phonomicrosurgery probably leaves a scar on the vocal fold with the lesion<sup>221</sup>, which in turn, leads to changes in the biomechanical properties of the scarred-vocal fold, including increased

stiffness.<sup>222</sup> However, after the phonomicrosurgery mass and morphology of the vocal fold with the lesion and the contralateral side would become more similar. Considering that the goal of the surgery is to improve the voice, it is expected for positive changes to outweigh the negative side effects of the surgery. Additionally, based on tables 3.5 and 3.6 the vocal fold with a lesion showed a higher improvement following the surgery, comparing to the vocal fold without a lesion. This higher improvement may compensate for the small value for the vocal fold with the lesion at the baseline (table 3.4). Therefore, we may expect for kinematics of the two vocal folds to become more similar following the surgery. Therefore, it is hypothesized that,

H2b: For unilateral mass lesions, the closing phase maximum velocity of the two vocal folds will become more similar after the surgery.

To test H2b two separate paired-sample  $t$ -tests were used per each scanning line. First, the pre-surgery differences in the closing phase maximum velocities between the vocal fold with the lesion and the contralateral side were investigated. Then, the same process was repeated for the post-surgery recording. Table 3.7 shows the results for  $v_{max}$ ,  $v_{lesion}$ , and  $v_{mid}$ .

Table 3.7. Results of paired-sample  $t$ -test for pre- and post-surgery recordings.

Scanning location	Pre-surgery		Post-surgery	
	$t$	$p$	$t$	$p$
$v_{max}$	3.36	0.008	0.86	0.41
$v_{lesion}$	4.82	0.0009	0.97	0.36
$v_{mid}$	4.92	0.0008	0.96	0.36

Based on the results of table 3.7 we see that the vocal fold with the lesion has a significantly lower closing phase maximum velocity (positive value of  $t$ ) than the contralateral side in pre-surgery condition for all scanning locations. However, none of the tests were significant for post-surgery condition. Therefore, we could conclude that the two vocal folds were significantly

dissimilar in pre-surgery data, but they become similar following the surgery. Figure 3.14 shows the individual behavior of  $v_{lesion}$  for both conditions, which corroborates the findings of table 3.7.

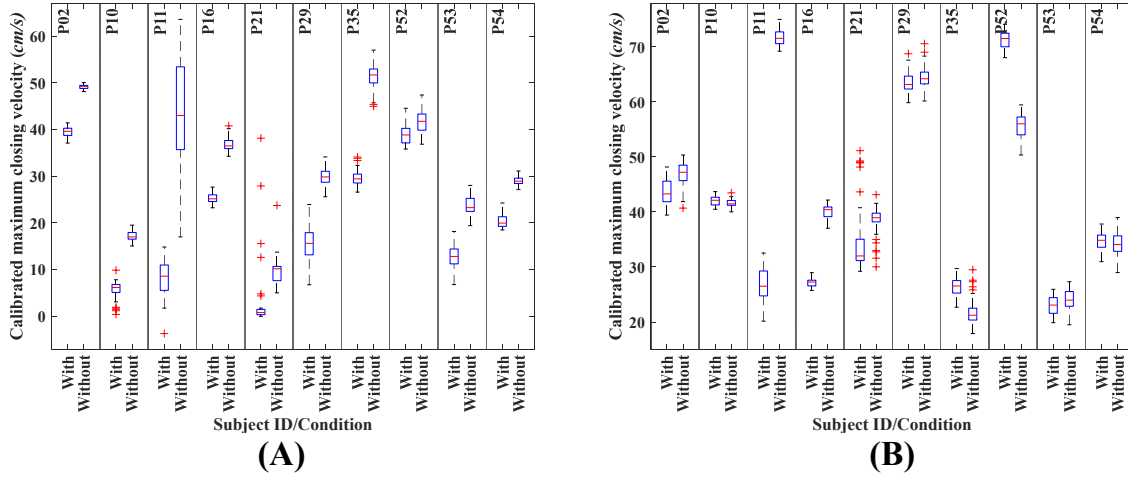


Figure 3.14. Boxplot of closing phase maximum velocity for the vocal fold with the lesion and the (cont)ralateral side for different subjects: (A) pre-surgery condition, (B) post-surgery condition.

### 3.4.3. Experiment3: Effect of lesion size on post-surgery changes

A lesion with a larger area probably indicates a larger accumulation of the extra mass on the vocal fold. It is expected for the removal of a larger mass to lead to larger post-surgery increase in the velocity of the vocal fold. Additionally, subjective visual evaluation of HSV recordings has suggested that the area of a lesion is a better predictor for qualitative changes in the vibratory characteristics of the vocal folds (*e.g.* left-right phase asymmetry) than its length.<sup>93</sup> Based on these rational, the following hypothesis is made,

H2c: Post-operative change in the closing phase maximum velocity will be positively correlated with the area of the lesion.

The intraoperative images were imported into an image editing software and then the area of the lesion was painted blue (figure 3.15(A)). The edited images were imported into Matlab and the numbers of solid blue pixels (*i.e.* red =0, green =0, and blue =255) were counted. This number corresponds with the uncalibrated (*i.e.* pixel) areas of the lesions. Calibration was done by

multiplying the uncalibrated area with the square of the pixel-to-*mm* conversion scale computed from the corresponding intraoperative image. Figure 3.15(B) shows a scatter plot of post-surgery changes in  $v_{lesion}$  computed from the vocal fold with the lesion.

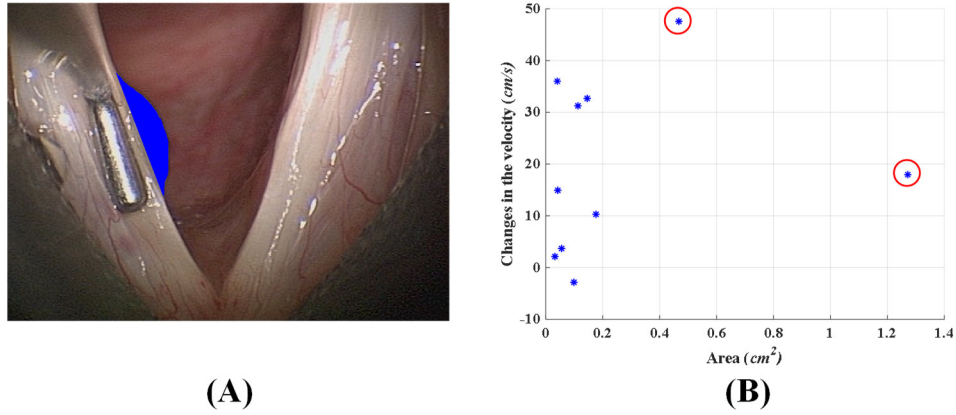


Figure 3.15. The relationship between area of a lesion and its post-surgery improvement: (A) The blue region shows the lesion (B) Scatter plot of post-surgery changes in  $v_{lesion}^{with}$  vs. area of the lesion. The outliers are marked by a red circle.

The scatter plot shows that two data points were not follow the trend. These data points belonged to subjects p11 and p29. Investigation of the recordings from these two subjects showed that they share two common characteristics. First, the size of the lesion was very big. Second, the lesion site was near the anterior commissure. Edges near the two ends of the vocal fold have low excursions and hence small maximum velocities. Therefore, removing a very big lesion from these locations could have a smaller impact on the velocity. To test hypothesis H2c the correlation coefficient between the post-surgery changes in the  $v_{lesion}^{with}$  and the calibrated area of the lesion was computed. Considering the above-mentioned differences for subjects p11 and p29, two different cases were tested. First, these samples were included in the analysis (N=10). In the second analysis, these outliers were excluded (N=8). Table 3.8 shows the results.

Table 3.8. Correlation between post-surgery changes in the closing velocity and the area of the lesion.

Sample size (N)	$r$	$p$
10	0.17	0.63
8	0.16	0.71

Based on the results of table 3.8 the null hypothesis cannot be rejected, which could be due to the small sample size.

### 3.5. Discussions

The closing velocity is an important kinematic measure of the vocal folds' vibratory motion. For example, the closing velocity relates to collision forces between the two vocal folds<sup>25–27</sup>, as well as, to the average produced acoustic output<sup>29</sup> and vocal intensity.<sup>23,28</sup> Additionally, previous studies have suggested that closing velocity could be a predictor for tissue elasticity<sup>102,103</sup>, and hence a predictor of physical development of the vocal folds.<sup>103</sup> Therefore, accurate velocity measures could significantly improve our understanding of the normal and disordered phonatory mechanisms. There is also a general agreement in the association between the phonotrauma and the collision forces between the vocal folds.<sup>169,210,216,228</sup> Therefore, clinical diagnosis and treatment could significantly benefit from further research into velocity measures.

This chapter provided the required methodology for accurate measurements of calibrated horizontal (*i.e.* medial-lateral direction) velocity of the vocal fold edges. Two primary steps were performed to achieve this. First, a method with a sub-pixel resolution was developed for the segmentation of the edges of the vocal folds. This was done using an adaptive thresholding technique, followed by fitting proper curves on temporal and spatial domains. Second, calibrated velocity measures require the existence of calibrated time and space. Fortunately, HSV videos are temporally calibrated, that is, the time difference between consecutive frames is known. However,

spatial information is not readily calibrated. The spatial calibration was done using the method developed in chapter 2. Specifically, the intraoperative images were used to determine the *mm* lengths of the lesions, which in turn were used for spatial calibration of the pre-surgery HSV recordings. The *mm* widths of vocal folds at specific locations along the anterior-posterior axis were measured from the pre-surgery HSV data, and then they were used for calibration of the post-surgery HSV data. The employed method was used to measure the closing phase maximum velocity of subjects with mass lesions pre- and post-surgery. Based on table 3.2 the closing phase maximum velocity of the pre-surgery condition was on average between 28.36 *cm/s* and 42.7 *cm/s*, depending on where the measurements were computed from. A similar measurement from the post-surgery condition was on average between 38.5 *cm/s* and 50.37 *cm/s*, depending on where the measurements were computed from. We may compare these values with the velocity reported in other studies. Using the color Doppler imaging technique the vocal fold velocity of  $68 \pm 10$  *cm/s* was reported for comfortable pitch and loudness of a sustained phonation from 10 healthy male subjects.<sup>212</sup> Using the photoglottography the vocal fold maximum closing velocity of  $112 \pm 53$  *cm/s* was reported for 32 healthy subjects covering a wide range of sound pressure levels (65.46- 86.89 dBA).<sup>213</sup> Finally, using a parallel-laser project endoscope the average value of 100 *cm/s* was reported for the maximum velocity of the vocal folds for 9 healthy male subjects.<sup>25</sup> Considering that our subjects had voice disorders, the computed values seem to be in a sensible range.

Assuming a vocal fold with the length of  $l$  pixel, we could compute  $2l$  time-sequences describing the velocity of every point on the edges of the two vocal folds at every time. Obviously, this high number of measurements has a lot of redundancy and should be reduced. Such reduction should have two parts. First, each velocity time-sequence should be represented by a limited number of attributes. This step is a temporal reduction. Next, computed attributes from the  $2l$  points



on the edges of the two vocal fold should be represented by a limited number of features. This step is a spatial reduction. In this chapter, the temporal reduction was achieved by selecting time points in the closing phase that led to the maximum velocity (*i.e.* the closing phase maximum velocity). The spatial reduction was achieved by just selecting certain points on the edges of the vocal folds. These were: the point leading to the maximum value, the midpoint of the vocal fold length, and the midpoint of the vocal fold lesion. Results from experiment1 indicated a significant effect for the spatial reduction operation. Specifically, the measurement from the point with the maximum value showed the least discriminative power (*i.e.* the lowest effect size between pre/post), followed by the measurement from the middle of the vocal folds, and then the measurement from the middle of the lesion. This outcome suggests that future studies should consider this factor during their experiment designs.

Referring to table 3.4 we see for patients with unilateral lesions the closing phase maximum velocity of the vocal fold with the lesion on average improves by 98%, 79%, and 30% at the midpoint of the lesion, midpoint of the vocal fold length, and the point with the maximum closing velocity, respectively. Referring to table 3.5 we can conclude that the closing velocity of the vocal fold with the lesion improves significantly, at the midpoint of the lesion and the midpoint of the vocal fold length. The improvement for the point with the maximum closing velocity also showed a promising improvement, but due to the small sample size, it did not reach the significance level ( $p=0.07>0.05/3$ ). However, this was not the case for the contralateral side. Specifically, table 3.6 did not establish a significant improvement for the contralateral side. The line passing through the midpoint of the lesion (based on the other vocal fold) was the only location that showed some level of improvement. However, due to the small sample size, it did not reach the significance level ( $p = 0.11 > \frac{0.05}{3}$ ). In summary, the finding from experiment1 suggests that the closing velocity of

the vocal fold with a lesion improves, at least, at multiple points along the length of the vocal fold following the surgery. However, the improvement of the contralateral side is more local and probably more limited to the area in direct contact with the lesion.

Experiment2 provided some evidence regarding the similarity of the vibration of the two vocal folds following the surgery. Specifically, table 3.7 showed that the closing velocities of the two vocal folds during the pre-surgery phonation were significantly different, at least, at multiple locations. However, the closing velocities of the two vocal folds after the surgery were not significantly different. This finding suggests that kinematics of the two vocal folds become more similar after the surgery. Finally, experiment3 investigated the association between the area of the lesion and the post-surgery improvement in the closing velocity of the vocal fold with the lesion. Table 3.8 indicated a very weak association between the two, and the correlation failed to reach the significance level. Considering the small sample size, a firm conclusion cannot be made. However, this result suggests that the area of the lesion is not a good predictor for closing velocity improvement.

### **3.6. Conclusions**

This chapter was motivated by the importance and the relevance of closing velocity of the vocal fold for clinical applications, and voice science research. The computation of the calibrated velocity measures depends on two primary steps. The accurate segmentation of the edges of the vocal folds with sub-pixel resolution, and the spatial calibration of the recording. A new segmentation method based on an adaptive thresholding technique, followed by fitting proper curves on temporal and spatial domains was presented. An indirect approach based on intraoperative images was employed for calibration of the pre- and post-surgery HSV recordings. Investigation of post-surgery changes revealed a significant effect for the location that the velocity

is computed from. The line passing through middle of the lesion showed the highest improvement (an average improvement of 98%). Additionally, the analysis suggested that the closing velocity of the vocal fold with a lesion improves, at least, at multiple points following the surgery. However, the improvement of the contralateral side was more local and probably more limited to the area in direct contact with the lesion. Furthermore, the result showed that the closing velocity of the two vocal folds become more similar following the surgery. This study also investigated the association between the size of the lesion and the post-surgery closing velocity improvements. The result showed a very weak association between the two ( $r=0.17$ ), which did not reach the significant level ( $p=0.63$ ).

## CHAPTER 4: DIRECT VERTICAL CALIBRATION OF HSV RECORDINGS

*Based on:*

Ghasemzadeh H., Deliyski D. D., Ford D. S., Kobler J., Hillman R. E., Mehta D. D. Method for Vertical Calibration of Laser-Projection Transnasal Fiberoptic High-Speed Videoendoscopy. *Journal of Voice*. 2020 Nov;34(6):847-861. doi: 10.1016/j.jvoice.2019.04.015. PMID: 31151853; PMCID: PMC6883161.

Summary: The ability to provide absolute calibrated measurement of the laryngeal structures during phonation is of paramount importance to voice science and clinical practice. Calibrated three-dimensional measurement could provide essential information for modeling purposes, for studying the developmental aspects of vocal fold vibration, for refining functional voice assessment and treatment outcomes evaluation, and for more accurate staging and grading of laryngeal disease. Recently, a laser-calibrated transnasal fiberoptic endoscope compatible with high-speed videoendoscopy (HSV) and capable of providing three-dimensional measurements was developed. The optical principle employed is to project a grid of  $7 \times 7$  green-laser points across the field of view (FOV) at an angle relative to the imaging axis, such that (after calibration) the position of each laser point within the FOV encodes the vertical distance from the tip of the endoscope to the laryngeal tissues. The purpose of this chapter was to develop a precise method for vertical calibration of the endoscope. Investigating the position of the laser points showed that, besides the vertical distance, they also depend on the parameters of the lens coupler, including the FOV position within the image frame and the rotation angle of the endoscope. The presented automatic calibration method was developed to compensate for the effect of these parameters. Statistical image processing and pattern recognition were used to detect the FOV, the center of FOV, and the

fiducial marker. This step normalizes the HSV frames to a standard coordinate system and removes the dependence of the laser-point positions on the parameters of the lens coupler. Then, using a statistical learning technique, a calibration protocol was developed to model the trajectories of all laser points as the working distance was varied. Finally, a set of experiments was conducted to measure the accuracy and reliability of every step of the procedure. The system was able to measure absolute vertical distance with mean percent error in the range of 1.7% to 4.7%, depending on the working distance.

#### **4.1. Introduction**

Typical images are two-dimensional representations of the real world. Considering that the real world has a three-dimensional (3D) spatial structure, images are not a true representation of the actual phenomena that are being captured. Basically, for any pixel of an image, we could construct a hypothetical square pyramid such that its tip is on the sensor of the camera and its base is toward the front of the camera. Anything inside this pyramid would be represented by the same pixel, or equivalently, all the space inside that pyramid is squeezed into a single point on the image. This model predicts several important features for an image. If the pyramid contains several objects at different distances, the closest one gets recorded. Also, based on this model the height of the pyramid is lost during the imaging. Finally, the size of an object in the image depends on its distance from the camera. The main aim of this chapter is to devise a method that can estimate the height of this hypothetical pyramid for laryngeal endoscopy. Assuming an upright position for the patient during the laryngeal imaging, this height would correspond to the vertical distance between the tip of the endoscope and different points on the superior view of the larynx. Therefore, the term vertical distance is used for the rest of this chapter.

The larynx has a 3D structure, and its different components reside at different vertical distances. Extrinsic laryngeal muscles could also elevate or depress the larynx<sup>12,229</sup> which would lead to changes in the vertical distance of the larynx from the endoscope. Additionally, vocal folds have 3D morphology and in fact, their vibration is happening in both horizontal and vertical planes. Multiple studies have predicted the significant role of the vertical component of the vibration on the phonation.<sup>164,172–174</sup> Therefore, measuring the vertical movements of the larynx and the vertical component of the vibration of vocal folds could provide significant amount of information regarding the mechanism of normal and disordered phonations. At the same time, the ability to obtain absolute horizontal measurements from laryngeal tissues and structures may depend on estimating their distances from the endoscope. It is expected for accurate horizontal and vertical measurements from in-vivo laryngeal images to provide essential information for modeling of the vocal fold behavior<sup>151,230</sup>, studying the developmental aspects of vocal fold vibration<sup>153</sup> and laryngeal tissues, better evaluation of treatment outcome of voice disorders, and more accurate grading of relevant laryngeal diseases.<sup>191</sup> To achieve these goals laryngeal imaging systems should provide calibrated measurement capabilities.

Researchers have been working on augmenting the laryngeal imaging systems with absolute measurements and/or 3D reconstruction capabilities for more than two decades.<sup>25,153,190–192,194,231–234</sup> Most often, these goals are achieved by projecting a laser pattern with certain topological properties on the field of view (FOV) and then using the information from the position and displacement of the laser pattern for achieving absolute measurement or 3D reconstruction.<sup>191,192,194,231,235,236</sup> Three main components can be identified in (almost) all systems that have been designed for this purpose: the laser projection component, the imaging component,

and the endoscopic instrument. These three components determine the functionality, characteristics, and capabilities of the final imaging system.

Considering the underlying principles for creating the laser pattern three main categories may be distinguished. Systems in the first category use the well-known laser triangulation principle for performing measurements.<sup>237</sup> The main idea behind systems in this category is to project a laser point (or line) on the target surface and then record the scene from a different angle. The angle difference between the laser projection and the imaging axes captures the vertical displacement of the target surface. The single-point<sup>231,232</sup> and single-line<sup>233</sup> laser projection systems fall under this category. Systems in the second category have been developed based on the projection of structured laser lights. These systems project a set of (commonly two) parallel laser beams with known horizontal distance on the target surface. Then, the distance between the parallel laser patterns on the image acts as a scale for converting pixel into *mm*. Two-point<sup>25,190,192</sup>, two-parallel-line<sup>234</sup>, and multiple-parallel-line<sup>153</sup> projection systems are examples from this category. Finally, systems in the third category have combined structured light projection with the laser triangulation technique for achieving the desired measurement goals. The multiple-point laser projection systems are examples of this category.<sup>191,194,235</sup> It is noteworthy that systems from each category have different functionalities. Systems from the first category could only capture the vertical movements of the target surface, whereas systems from the second category are typically used for absolute measurements on the horizontal plane. The systems in the third category are by far the most flexible approach and, depending on the design, can provide detailed information regarding vertical movements and absolute measurements on the horizontal plane. This wealth of information comes at the cost of more complex hardware (optical) and software (algorithm) design.

Figure 4.1 presents a schematic of different laser projection systems. Fig 4.1(A) shows the projection of a single laser beam on a target surface ( $S_1$ ). When the surface  $S_1$  moves  $h$  mm in the vertical direction, the laser point moves  $\Delta$  mm in the horizontal plane. This horizontal component is captured on the image as a  $\delta$ -pixels displacement. Fig 4.1(B) shows a projection of two parallel laser points on a target surface ( $S_1$ ). The actual distance between laser points ( $d$  mm), is reflected by a  $\delta$ -pixel distance on the image. Fig 4.1(C) shows a schematic image of the combined approach. Specifically, hypothetical positions of laser points on the image for two different vertical distances are shown in red and green colors. Change in the vertical distance leads to the displacement of the laser pattern by  $D$  pixels. Additionally, the distance between pairs of laser points ( $d_1, d_2$ ) could be used for horizontal measurements.

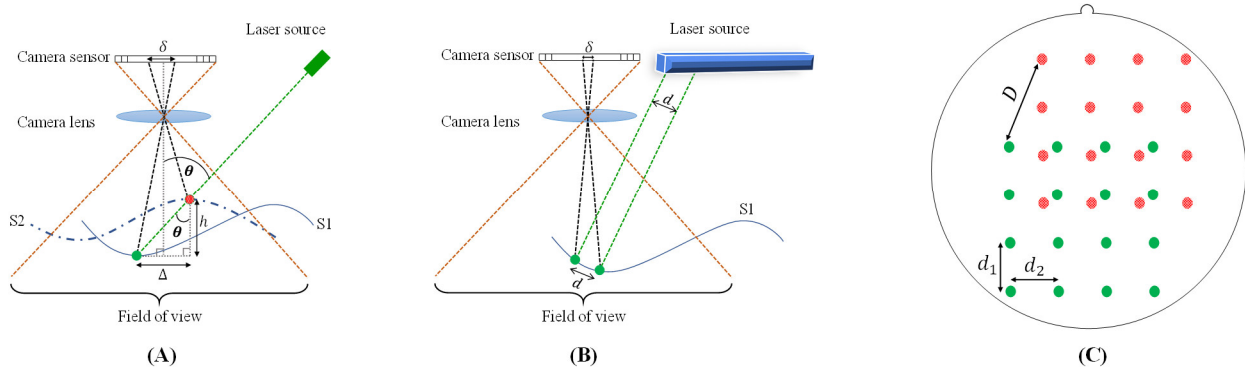


Figure 4.1. Schematics of different laser projection techniques with the principle of encoding the vertical and/or horizontal distances: (A) laser triangulation method, (B) structured light projection, (C) a combined technique. Green and red dots depict hypothetical positions of the laser pattern at two different vertical distances.

Considering the optical imaging component, two main technologies of VSB and HSV can be differentiated. VSB has been the “gold standard” approach for clinical voice evaluations<sup>108,110,111</sup> and it “provides real-time audiovisual feedback and continues to be the imaging modality of choice by voice clinicians.”<sup>108</sup> This technique uses very short flashes of light and takes a sequence of pictures from different glottal cycles and then assembles them into a motion picture. An external trigger based on the vibratory phase of the acoustic or electroglottographic signal determines the



time of the flashes. In this fashion, the assembled images represent a slow motion of the true vibration of the vocal folds.<sup>104,105</sup> Consequently, VSB does not present the actual vibratory patterns of the vocal fold, and its captured images would substantially deviate from the true pattern as the vibration becomes irregular and aperiodic.<sup>93,108</sup> On the other hand, HSV systems capture the true vibratory patterns of the vocal fold, and therefore it is more appropriate when studying the intra-cycle characteristics of vocal fold vibration.<sup>105,111</sup> In summary, the imaging component would determine the temporal resolution of the captured images and consequently, it has a significant role in the type of phenomena that can be captured and studied. Systems based on stroboscopy are applicable to stationary phenomena, whereas HSV systems can be used for capturing non-stationary behaviors such as onset and offset of phonation and also aperiodic phonation.

Considering the type of endoscopic instrument, two categories of rigid and flexible endoscopes are available. The rigid endoscope provides images with better spatial resolution and visual quality but at the same time it affects the voice and speech production due to transoral insertion that requires unnatural retraction of the tongue for adequate laryngeal exposure, thus, only limited types of stimuli can be elicited. On the other hand, flexible endoscopy does not interfere with articulators and speech can be produced with minimal interference, therefore, it could be more ecologically valid. Additionally, there are fewer restrictions on the type of stimuli that could be produced, thus, it could be used for analysis and studying of the vibratory pattern of vocal folds during connected speech.<sup>116</sup> Finally, flexible endoscopes provide the possibility of simultaneous recordings of the aerodynamic measurements.<sup>132–134</sup>

Table 4.1 summarizes the taxonomy of different systems with laser projection capabilities in the literature. Recently, we developed a new flexible, fiberoptic endoscope with laser-projection capabilities.<sup>195</sup> The new system uses a flexible endoscope for accessing the superior view of the

larynx, which allows eliciting a wide range of stimuli, while at the same time the optical characteristics of the laser projection system were designed to be compatible with HSV systems and provide good visual contrast between laser points and the background. The system was designed so that absolute measurements in both horizontal and vertical planes are possible. Combining these characteristics, the new system could provide 3D information regarding the vocal fold vibratory pattern and the laryngeal configuration during laryngeal maneuvers, phonation, and connected speech.

Table 4.1. Literature-based taxonomy of different imaging systems with laser projection. These abbreviations were used in the table: VSB (videostroboscopy), HSV (high-speed videoendoscopy), 3D (three-dimensional reconstruction), *nm* (nanometer), mW (milli Watt).

Year	Ref.	Laser pattern	Projection technique	Imaging	Endoscope	Functionality	Other notes
1997	[192]	2-point	parallel beams	VSB	90°, rigid	horizontal	red laser, 670 <i>nm</i> , 3 mW power per laser point
2001	[193]	1-point	triangulation	VSB	70°, rigid	vertical	red laser, 643 <i>nm</i>
2002	[25]	2-point	parallel beams	HSV	90°, rigid	horizontal	red laser, 633 <i>nm</i>
2004	[233]	1-point	triangulation	HSV	70°, rigid	vertical	1 mW power
2004	[190]	2-point	parallel beams	VSB	70°, rigid	horizontal	green laser, 1 mW power at source
2006	[191]	23-point	structured light+ triangulation	VSB	flexible	horizontal	green laser, 150 mW power at source
2008	[234]	1-line	triangulation	HSV	90°, rigid	vertical+horizontal along a single line	red laser, 653 <i>nm</i> , irradiance of 1800 W/m <sup>2</sup>
2008	[235]	2-line	parallel lines	HSV	90°, rigid	vertical+horizontal along two lines	red laser, 635 <i>nm</i> , 22 mW power
2010	[194]	196-point	structured light+ triangulation	HSV	70°, rigid	vertical+ horizontal+3D	-
2013	[153]	21-line	parallel lines	HSV	70°, rigid	horizontal	green laser, 300 mW power at source, irradiance of 1100 W/m <sup>2</sup> at working distance of 30 <i>mm</i>
2016	[236]	324-point	structured light+ triangulation	HSV	70°, rigid	vertical+ horizontal+3D	532 <i>nm</i> , 150 mW power at source, 80 mW at the tip of the endoscope, irradiance of 1000 W/m <sup>2</sup> at working distance of 60 <i>mm</i>
2019	[195]	49-point	structured light+triangulation	HSV	flexible	vertical+ horizontal+3D	green laser, 520 <i>nm</i> , 55 mW at source, 20 mW at the tip of the endoscope, irradiance of 372 W/m <sup>2</sup> at working distance of 20 <i>mm</i>

To achieve the above-mentioned measurement goals, the laser-projection endoscope should be calibrated first. This chapter presents the methodology for vertical calibration and subsequent measurements.

## 4.2. Aim and hypothesis

The main aim of this chapter is to develop the methodology of vertical calibration and subsequent vertical measurement for a laser-projection transnasal fiberoptic HSV system. The main research question of this chapter is:

Q3: How could we use a structured laser projection system for measuring the vertical distance between the distal tip of a flexible endoscope and the target surface?

To answer this research question and to pursue the aim of this chapter, two hypotheses were formed that are presented in this section.

Referring to figure 4.1(A) horizontal and vertical displacements of each laser point could be related using trigonometric rules. Equation 4-1 shows this.

$$\Delta = h \cdot \tan(\theta) \quad (4-1)$$

At the same time, the horizontal displacement component ( $\Delta$ ) and the pixel displacement ( $\delta$ ) are related through the magnification factor of the camera ( $m$ ).

$$\delta = m \cdot \Delta \quad (4-2)$$

Combining Equations 4-1 and 4-2 we would have,

$$h = \frac{\delta}{m \cdot \tan(\theta)} \quad (4-3)$$

Based on Equation 4-3 the *mm* vertical displacement ( $h$ ) is a function of the pixel displacement ( $\delta$ ), magnification of the camera ( $m$ ), and the angle difference between imaging and laser projection axes ( $\theta$ ). Additionally, magnification of the camera depends on the focal length of its

lens ( $f$ ) and the vertical distance between the object and the focal point of the lens ( $x$ ).<sup>238</sup> Equation 4-4 shows this,

$$m = \frac{f}{x} \quad (4-4)$$

Based on the presented Equations, it is hypothesized that,

H3a: The position of each laser point will be a unique and deterministic function of the vertical distance between the distal tip of the flexible endoscope and the target surface, once the confounding factors are accounted for.

Considering the hypothetical pyramid of introduction, the resolution of the system decreases as the working distance increases. It means that for every  $mm$  increase in  $\Delta$ , the number of pixel displacement on the image ( $\delta$ ) would be a decreasing function of the working distance. Therefore, it is hypothesized that

H3b: Vertical measurement error will be positively correlated to working distance.

### 4.3. Material and method

#### 4.3.1. Laser-projection endoscope

A surgical flexible endoscope, Fiber Naso Pharyngo Laryngoscope Model FNL-15RP3 (PENTAX Medical, Montvale, NJ), with three channels (surgical, imaging, and light-delivery channels) was used for developing the laser-projection endoscope with absolute measurement capabilities.<sup>195</sup> The surgical channel is used for delivering a green laser light with a wavelength of  $520\text{ nm}$  to the distal tip of the endoscope, where a diffraction-based system splits it into a mesh-pattern of  $7 \times 7$  laser points. The size of the laser pattern is  $16 \times 16\text{ mm}$  at a working distance of  $20\text{ mm}$ . The imaging channel of the endoscope allows for coupling the endoscope with a color/monochrome high-speed digital camera and recording of the superior view of the larynx with

the projected laser pattern at distance ranging from 5 mm to 35 mm. The third channel utilizes a fiberoptic light-delivery system that can be coupled with a xenon light source with power up to 300 W. Figure 4.2 depicts the calibrated endoscope with its main components.

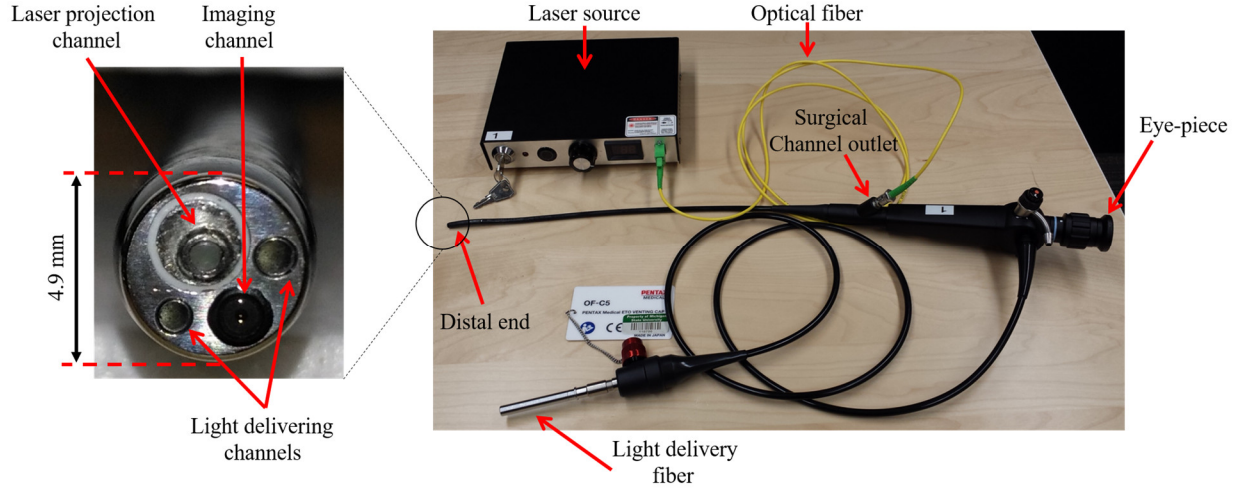


Figure 4.2. The calibrated flexible endoscope with an insertion tube diameter of 4.9 mm and its main components.

#### 4.3.2. Calibration protocol and recordings

To achieve the absolute measurements in the vertical plane the endoscope should be calibrated first. More specifically, the position of the laser points in the FOV is a non-linear function of the lens-coupler parameters and the working distance. Calibration is the process that accounts for these factors and finds the mathematical function for decoding the desired measurements from the positions of the laser points. To find that function, a data-driven approach based on statistical pattern recognition and statistical learning techniques were adopted.

The setup presented in chapter 1 (figure 1.2) with one degree of freedom was used in this chapter. Specifically, the tilting angle was kept fixed and at zero angle (*i.e.* perpendicular imaging angle) and only the working distance was varied. The laser-projection endoscope was connected to a high-speed monochrome camera Phantom v7.1 (Vision Research Inc., Wayne, NJ) using a 45-

*mm* lens coupler and a 300-Watt xenon light source. Considering that calibrations are typically done under a controlled environment and the best possible settings, a monochrome camera was used for this phase. Monochrome cameras have higher sensitivity comparing to their color versions and they don't use the Bayer-decomposition filters.<sup>239</sup> These characteristics result in a sharper image with better-defined edges. It is noteworthy that using the monochrome camera does not impose restrictions on the application of the system, and the calibrated endoscope can be used with a color camera after calibration. The camera and the endoscope were mounted on a vertical plane perpendicular to the target surface and FOV was recorded at the speed of 7000 frames per second with a spatial resolution of  $288 \times 280$  pixels. The target surface was attached to an adjustable arm that allowed to regulate with high precision the working distance to the distal end of the endoscope. The working distance was varied from 5 *mm* to 35 *mm* using a 1-*mm* step and it was measured using a digital height gauge with an accuracy of 0.001'' (0.03 *mm*). Figure 4.3 presents a diagram of the recording conditions.

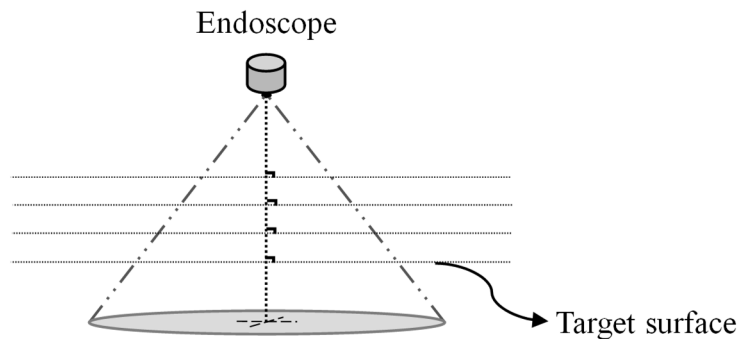
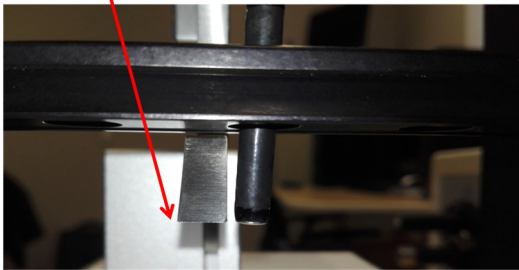


Figure 4.3. A diagram of the recording conditions.

Accurate measurement of working distance depended on accurate leveling of the arm of the gauge with the distal end of the endoscope (figure 4.4(A)), which should be determined visually and therefore time-consuming and subject to variability. Therefore, in the setup, a fixture was

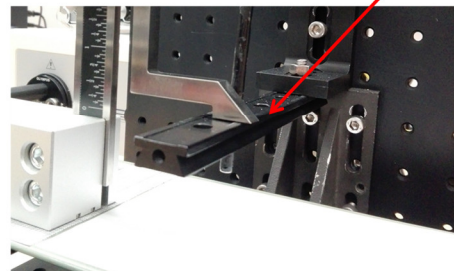
placed about 2 *cm* above the distal end of the endoscope, and the following procedure for measuring the distance between the tip of the endoscope and the top surface of the fixture was implemented. The measurement arm of the gauge was positioned on the top surface of the fixture and the height was recorded (figure 4.4(B)). Then, the measurement arm was positioned parallel to the tip of the endoscope and the height was recorded again. To check the leveling of the two surfaces, a 13-megapixel smartphone camera was positioned on the same vertical level as the tip of the endoscope and the digital magnification feature of the camera was used to fine-tune the position of the adjustable arm (figure 4.4(A)). These steps were repeated ten times and then the results were averaged. The average distance to the fixture was  $45.01 \pm 0.03$  *mm* and the average distance to the tip of the endoscope was  $21.85 \pm 0.11$  *mm*. The measurement of the distance to the fixture shows a lower value of standard deviation, supporting better accuracy of measurement when the fixture is used as the reference point. From these measurements, the distance between the tip of the endoscope and the top surface of the fixture was estimated to be 23.16 *mm*.

Leveling the arm and  
the tip of endoscope



(A)

Positioning the arm on the  
top surface of the fixture



(B)

Figure 4.4. Calibration setup: (A) measuring the distance to the tip of the endoscope, (B) measuring the distance to the fixture.

Two different recordings were made at each working distance. In the first recording, a white piece of paper was used, the xenon light was turned off, and the laser projection system was turned

on with maximum power. In the second recording, a multi-resolution grid paper (1-*mm*, 2-*mm*, and 10-*mm* boxes) was used, the laser projection system was turned off, and the xenon light was turned on. Throughout this chapter, these two recordings will be referred to as laser recordings and grid recordings, respectively. Each of these two sets of recordings serves a different purpose in the calibration procedure. The laser recordings are used for finding the accurate position of laser points in the FOV, whereas the grid recordings are used for estimating the parameters of recordings. The grid recordings are also necessary for the horizontal calibration of the system, which is the topic of the next two chapters. It is noteworthy that these two recording conditions are only used to remove confounding factors from different calibration processes and to maximize the accuracy, but they don't impose any restrictions on the application of the system, and they don't need to be replicated during clinical data collection. Finally, since the intensity of pixels increases at shorter working distances leading to possible saturation of the image, the exposure time of the camera and the power of the light source were adjusted at each step to prevent image saturation.

#### **4.3.3. Measuring vertical distance**

The position of the laser points in the captured image is a deterministic function of the vertical distance between the distal tip of the endoscope, the target surface, and the lens-coupler parameters. This section presents the automatic approach for compensating the effect of lens-coupler parameters and for decoding the vertical distances from the positions of laser points.

##### **4.3.3.1. Compensating for the lens-coupler parameters**

Some of the lens-coupler parameters change the position of the laser points in the FOV even if the working distance is kept constant. Those parameters include the focal distance of the lens coupler connecting the endoscope to the camera and the position and angle of the endoscopic



eyepiece relative to the lens coupler. To decode the vertical displacements, first, these parameters should be estimated from the recordings and then compensated for. After that, the positions of the laser points become only a function of the vertical distance and could be used for the measurements. The effects of different lens-coupler parameters and the corresponding compensation approaches are presented as follows.

#### **4.3.3.1.1. Recording model**

The focal distance of the lens coupler determines the magnification of the camera. Using higher magnification results in an image where everything is larger. Therefore, the number of pixels between certain laser points (equivalently  $x$ - $y$  coordinates of the laser points in the image) would depend on the magnification of the camera. The second variability comes from the rotation of the endoscopic eyepiece inside the lens coupler attached to the camera. Because the camera is fixed, the recording frame would remain constant, but the FOV with everything inside of it would undergo a rotation transformation. Therefore, when the endoscope gets rotated, the projected laser pattern would also get rotated. This means that the  $x$ - $y$  coordinates of the laser points in the image would depend on the endoscope rotation. The last variability stems from the displacement of the eyepiece within the lens coupler. More specifically, the position of the eyepiece inside the lens coupler is not fixed and it can move in the horizontal plane. When the eyepiece is displaced, the whole FOV is displaced within the image frame. Consequently, the  $x$ - $y$  coordinates of the laser points in the image would depend on the position of the eyepiece within the lens adapter.

To account for variations due to these lens-coupler parameters, first, we need to have a model that describes the effect of each parameter on the recorded images. The model that was used for this purpose consists of three main transformations of scaling (effect of magnification), rotation (effect of eyepiece rotation), and translation (eyepiece displacements). This model aims to map the

recordings with variable parameters into a fixed and standard coordinate system where  $x$ - $y$  coordinates of the laser points are independent of those lens-coupler parameters.

Let  $I(x,y)$  and  $i(x,y)$  denote the original image and a pixel from it and  $J(x',y')$  and  $j(x',y')$  denote the mapping of that image in the standard coordinate system and the corresponding pixel in the new image. Also, let  $T$ ,  $R$ ,  $I_2$ , and  $k$  denote a translation vector, a rotation matrix, an identity matrix with the size of two, and a scaling factor, respectively. Equation 4-5 shows the model.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = k \cdot I_2 \cdot R \cdot \begin{bmatrix} x \\ y \end{bmatrix} + T \quad (4-5)$$

Now, if values of  $T$ ,  $R$ , and  $k$  are determined, the mapping can be carried out. Considering the aim of mapping, a few considerations should be taken into account when determining these parameters. First, the parameters should be determined so that the new image ( $J(x,y)$ ) is invariant from the lens-coupler parameters. Second, the estimation of those parameters should be computationally efficient. Third, the estimated parameters should be relatively robust to different sources of noise.

The effect of the eyepiece displacements manifests itself as the position change of FOV in the image frame. Also, the effect of the focal length of the lens coupler is manifested through the FOV size. Therefore, both magnification and eyepiece displacement can be compensated by the parametrization of the FOV. Both visual inspection and objective assessment confirmed that FOV can be estimated with a circle. Fortunately, very efficient algorithms have been developed for the parametrization of circular objects.<sup>240,241</sup> Additionally, circles have very well-defined and smooth topological shapes, which makes estimation of their parameters robust to noise. Therefore, the translation transformation ( $T$ ) was defined so that the center of the new coordinate system coincides with the center of FOV. Also, the radius of FOV was used to account for the magnification effect making the size of pixels constant. Flexible endoscopes have a fixed fiducial marker on their distal end that remains fixed relative to FOV and the target surface (figure 4.5).

This fiducial marker helps with determining the orientation during flexible endoscopy. The laser projection optics are glued inside the surgical channel and their position relative to the fiducial marker is fixed, therefore, the position of the fiducial marker can be used as a reference for compensating the effect of rotation of the endoscope within the lens coupler. Figure 4.5 summarizes the model. Based on this model the recorded image ( $I(x,y)$ ) undergoes a series of transformations including a translation, a rotation, and a scaling, and gets converted into a new image ( $J(x,y)$ ) in a standard coordinate system. In the translation phase, the center of the coordinate system is shifted to the center of the FOV. Rotation transformation brings the fiducial marker to a predetermined position (*e.g.*, 0 degree in figure 4.5). Finally, the scaling transformation stretches or shrinks the FOV so that its radius gets equal to a predetermined value of the radius  $r$ .

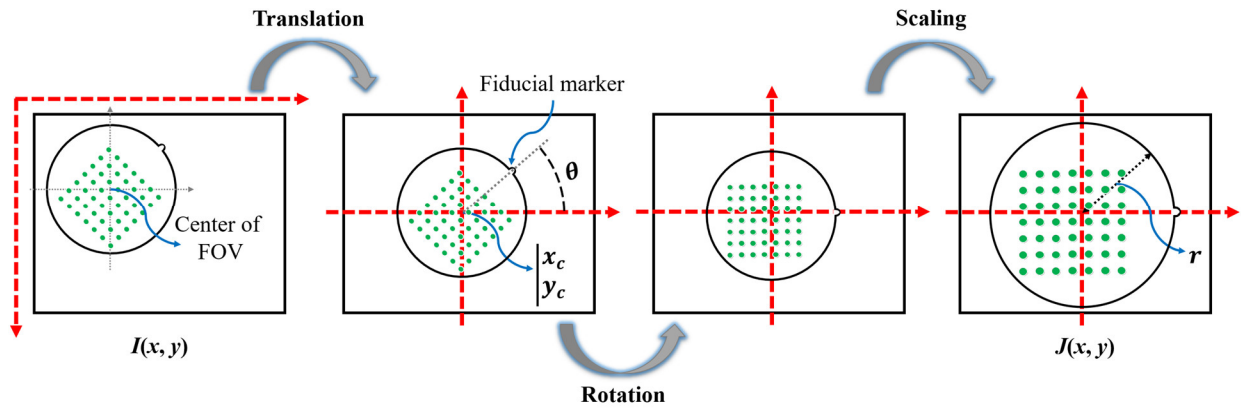


Figure 4.5. Model for compensating the recording parameters of the system.

#### 4.3.3.1.2. Automatic estimation of the mapping

Based on Equation 4-5, the mapping consists of three main transformations. As shown in Figure 4.5 the parameters of those transformations can be estimated based on two components of the image. That is, the center of FOV and its radius are used for estimating parameters of translation and scaling transformations, and the angle ( $\theta$ ) between the line connecting the center of FOV to the fiducial marker and the horizontal axis determines the parameter of the rotation

transformation. This section presents the algorithms and image processing techniques that were used for finding these two important landmarks.

First, the detection of FOV is investigated. The lighting channels of the endoscope provide illumination for the FOV, which is then trimmed by the field of view of the endoscope, leaving the pixels outside of the endoscopic circle quite dark. Therefore, it is possible to apply a thresholding technique and find a rough estimation of FOV. However, any error in estimating the center and radius of FOV would change the position of the laser points in the standard coordinate system, introducing an error in estimating the vertical distances. To find a more robust approach, the FOV finder module requires an additional source of information. Assuming the noise and distortions have linear effects, the geometrical shape of FOV would remain intact. Therefore, combining the geometrical information with the illumination differences inside and outside of FOV could help devise a robust method.

FOV has a circular shape and therefore the pixels on its boundary can be expressed using a precise mathematical Equation. Let  $c = \begin{bmatrix} x_c \\ y_c \end{bmatrix}$  denote the center of a circle with radius  $r$ , Equation 4-6 shows the locus of points on the perimeter of that circle.

$$\{(x, y) \in R^2 : (x - x_c)^2 + (y - y_c)^2 = r^2\} \quad (4-6)$$

The Hough transform is a very popular approach in the computer vision community, which initially was developed for the detection of lines and other analytically defined shapes (*e.g.* circles, ellipses)<sup>242</sup>, but later on, it was extended to other shapes.<sup>243</sup> Considering that FOV has an analytically well-defined shape, Hough transform can be used for capturing the geometrical information of FOV. In summary, the FOV finder module consists of two steps. In the first step, a thresholding technique is applied to the grayscale image. This step uses information from differences between the intensity of pixels inside and outside of FOV and converts the grayscale

image into a binary image. In the second step, the binary image is fed into the Hough transform algorithm, where it finds the center and radius of a circle that fits the binary image the best.

The second landmark in the image is the fiducial marker (figure 4.5). The position of this landmark relative to the center of the FOV and horizontal line determines the rotation transformation parameter. To make its detection as accurate and robust as possible, two different sources of information were identified and combined. First, the fiducial marker is fabricated through a physical notch in the FOV. Therefore, it is the most likely region outside the FOV to be bright and hence there would be differences between the intensity of the pixels within the fiducial marker comparing to other regions outside of FOV. Second, the fiducial marker is attached to the exterior of FOV, and therefore there is no need to check all pixels outside of FOV. Using this spatial information would remove some incorrect candidates and improve the performance of the fiducial finder module.

The fiducial finder module has two main steps. First, a torus mask centered at the center of FOV with an inner radius of  $r+1$  and outer radius  $r+8$  was applied to the image. This step incorporates the spatial information into the method. Next, a threshold was applied to the remaining pixels. Due to the imperfect circular shape of FOV and the leakage of light to the outside of FOV, a very thin arc could be present at this step. To remove those artifacts, the binary image at this stage underwent a morphological opening operation with a disk-shaped structuring element.<sup>244</sup> The final step is to quantify the position of the detected fiducial marker. It is known that the centroid of an object is relatively insensitive to noise and therefore is a robust estimation of the location of that object inside the image. Therefore, the centroid of the biggest element was computed as the location of the fiducial marker. Let  $B(x, y)$  denote a binary image with a size of  $m \times n$  pixels. Equation 4-7 shows how its centroid can be computed.

$$\mu_x = \frac{1}{|A|} \cdot \sum_{x=1}^n \sum_{y=1}^m x \cdot B(x, y) \quad (4-7)$$

$$\mu_y = \frac{1}{|A|} \cdot \sum_{x=1}^n \sum_{y=1}^m y \cdot B(x, y)$$

Where  $|A|$  denotes area of image  $B$  and it can be computed from Equation 4-8.

$$|A| = \sum_{x=1}^n \sum_{y=1}^m B(x, y) \quad (4-8)$$

After finding the centroid of the fiducial marker, the rotation angle is computed using Equation 4-9.

$$\theta = \tan^{-1} \left( \frac{\mu_y}{\mu_x} \right) \quad (4-9)$$

#### 4.3.3.2. Algorithm for distance estimation

After mapping a frame into the standard coordinate system, the position of each laser point on the new image ( $J(x, y)$ ) depends only on the vertical distance between the distal tip of the endoscope and the target surface. This section presents details of the algorithm for the automatic detection of laser points and the decoding of vertical distances from those positions.

##### 4.3.3.2.1. Automatic detection of laser points

The accuracy of the laser point detection module would have a significant effect on the accuracy of vertical distance estimation. Any error in the detection of the laser points, or in the quantification of their positions, would translate into vertical distance inaccuracies. To devise a robust detection algorithm and an accurate calibration method, the characteristics of the projected laser points should be known. Figure 4.6 shows a frame from one of the laser recordings data. As shown, the energy of the laser source is not uniformly divided between the laser points, where the points in the middle are significantly brighter than the points in the periphery. Additionally, as

shown in figure 4.6(A, C), sums of the image on rows and columns indicate that the intensity of each laser point has a bell-shaped spatial distribution, with the highest intensity at the center followed by a fast decay toward the distal pixels.

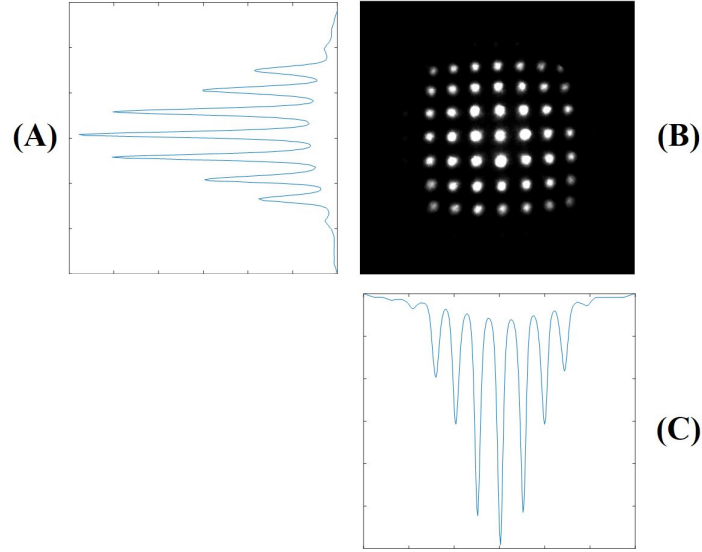


Figure 4.6. The intensity of the laser points: (A) sum of the intensity of pixels on the rows, (B) original image, (C) sum of the intensity of pixels on the columns.

Different characteristics and sources of information were taken into account during the design of the laser detection module. First, the difference between the intensity of the laser points and the background was exploited through an adaptive thresholding approach. Considering that intensity of pixels is a function of working distance, using the adaptive approach was inevitable. For that purpose, the histogram of the intensity of pixels was constructed with 200 bins and the first bin was considered as the black reference and was discarded. The cumulative distribution function (CDF) of the logarithms of the remaining bins was estimated, and the value corresponding to 0.4 was selected as the intensity threshold. Second, referring to Figures 4.6(A) and (C), a very large magnitude of gradient around the laser points is expected; therefore, an adaptive thresholding approach was used for exploiting this information, as well. To that end, the histogram of the magnitude of the gradient of the image was constructed with 200 bins and the value of the sixth

bin was used as the gradient threshold. The two thresholding values were applied to the image followed by a morphological opening operation with a disk-shaped structuring element. At this point, every laser point would be represented by a blob and its centroid can be computed. Considering that the intensities of the laser points have a bell-shaped spatial distribution, this information was used too. This information was incorporated by using the weighted centroid instead. Third, the laser points should have circular shapes, but most of the time the extracted blobs don't have that characteristic; therefore, the weighted centroid would be affected by those artifacts. To remedy that and also to incorporate the morphological information of laser points, a disk with a radius of 7 pixels was constructed around every centroid, and then the final position of laser points was computed as a weighted centroid of the pixels within those disks.

#### **4.3.3.2.2. Vertical distance decoding**

Figure 4.7 shows how the  $x$ - $y$  coordinates of each laser point vary depending on the working distance. Based on this figure, each laser point travels along a unique and well-defined trajectory, and hence its position within that trajectory can be used for decoding the vertical distance from that point to the tip of the endoscope. Additionally, it is evident that each laser point has some idiosyncratic characteristics. As seen in Figure 4.7(B), the behaviors of the laser points are different, where some of the laser points travel along a line (almost) perpendicular to the  $x$ -axis, indicating very small variations in the  $x$ -coordinate of these points; while other points travel along non-linear trajectories and show significant variations in the  $x$ -coordinate. Interestingly, some of these points have deflection to the right and some of them have deflection to the left. Considering that each laser point has a slightly different projection angle, these variations are to be expected. It is desirable to have trajectories with variation only along one axis, but these observations show that such characteristics cannot be achieved perfectly.



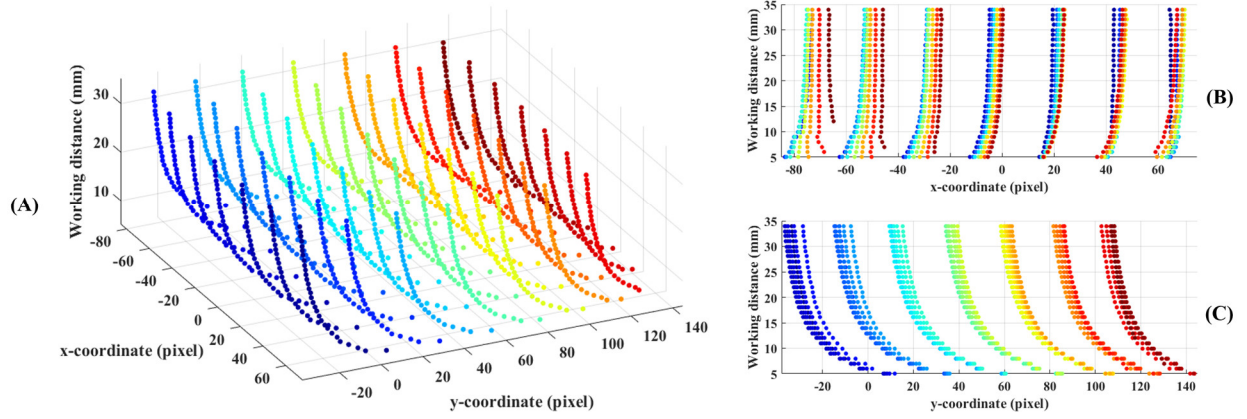


Figure 4.7. Position of each laser point as a function of working distance where each color shows a different laser point: (A)  $x$ - $y$  coordinates as a function of working distance, (B)  $x$ -coordinate as a function of working distance, (C)  $y$ -coordinate as a function of working distance.

To make the decoding process efficient and fast, the trajectory of each laser point was modeled using a function. Let  $d_p$  be a specific working distance and  $p = \begin{bmatrix} x_p \\ y_p \end{bmatrix}$  denote the position of a certain laser point corresponding to that working distance. This point can be converted into the polar coordinate system using Equation 4-10.

$$r_p = \sqrt{x_p^2 + y_p^2}, \quad \varphi_p = \tan^{-1}\left(\frac{y_p}{x_p}\right) \quad (4-10)$$

Now, the goal is to find a family of parametrized function  $\mathcal{F}$  and their proper parameters  $\boldsymbol{\beta}$  such that on average the estimated distances ( $\tilde{d}_p$ ) and the true distances ( $d_p$ ) are near each other based on a properly defined distance function ( $\mathcal{D}$ ). Equations 4-11 and 4-12 show these,

$$\tilde{d}_p = \mathcal{F}_{\boldsymbol{\beta}}(r_p) \quad (4-11)$$

$$\arg \min_{\boldsymbol{\beta}} \mathcal{D}(r_p, d_p) \quad (4-12)$$

The parameter  $\boldsymbol{\beta}$  could be determined using optimization Equation 4-12 and a set of data points (training phase). After that, the trained function could be used for decoding the working distances of new data points. As shown in figure 4.7(B-C), different laser points follow relatively similar

semi-exponential trajectories; thus, the same family of curves ( $\mathcal{F}_\beta$ ) was used for decoding purposes. But, to capture the idiosyncratic characteristics of each trajectory, the training phase was done separately for each laser point. Therefore, a total number of 49 different curves were trained during this phase. Equation 4-13 shows the family of curves there were used.

$$\mathcal{F}_\beta(r_p) = \beta_1 \cdot e^{\beta_2 \cdot r_p} + \beta_3 \cdot e^{\beta_4 \cdot r_p} \quad (4-13)$$

Finally, most often normalization of data points improves the performance of machine learning algorithms.<sup>73</sup> Let  $m_i$  and  $\sigma_i$  denote the mean and standard deviation of the radius of all laser points from trajectory  $i$ , Equation 4-14 shows the employed normalization process. The normalized values were then used for the training purpose.

$$r'_p = (r_p - m_i)/\sigma_i \quad (4-14)$$

#### 4.4. Experiments and results

Three experiments were conducted to answer the research questions of this chapter. Experiment 1 tests the performance of different preprocessing components of the proposed method. Experiment 2 presents displacement analysis and vertical resolution of the system. Experiment 3 investigates the performance of the proposed method for vertical measurements. This section presents details of each experiment, followed by results and related discussions.

##### 4.4.1. Experiment1: Evaluation of preprocessing components

The performance of the proposed method relies on accurate estimation of parameters of the mapping and also accurate detection of the laser points. Experiment1 was conducted to assess the performance of these components.

#### **4.4.1.1. Experiment1a: Evaluation of FOV and the fiducial finder modules**

The performance of the methods for compensating the effect of lens-coupler parameters was evaluated. Doing that requires a ground truth as a reference for comparison. Additionally, to measure the performance of each module separately, the standard deviation of the estimated parameters within a recording was used as the evaluation criterion. To that end, the videos from grid recordings were used. During those recordings, the configuration between the camera and endoscope was kept constant; therefore, the position of FOV, the radius of FOV, and the position of the fiducial marker should be the same for all of them. This observation was used for objective evaluation of the implemented algorithms. For that purpose, each recording was divided into batches with 200 frames. Then, frames within each batch were averaged and the result was fed into the algorithm for estimating the center of FOV, the radius of FOV, and the angle of the fiducial marker. Figure 4.8 shows the centralized distribution (the means were subtracted to make the plots more comparable) of estimated parameters over all batches and recordings. As seen in these figures, the centralized probability density functions are concentrated around zero with very sharp peaks. This supports that the proposed FOV and fiducial finder modules are quite robust and have very stable performances.

#### **4.4.1.2. Experiment1b: Evaluation of the laser finder module**

The performance of the laser detecting module was evaluated using the videos from the laser recordings. To that end, each recording at a specific working distance was divided without any overlaps into 11 batches of 200 frames. Then, the frames within each batch were averaged to remove the effect of additive noise. The positions of the laser points for each batch were estimated using the presented algorithm. Because all batches were recorded at the same working distance,

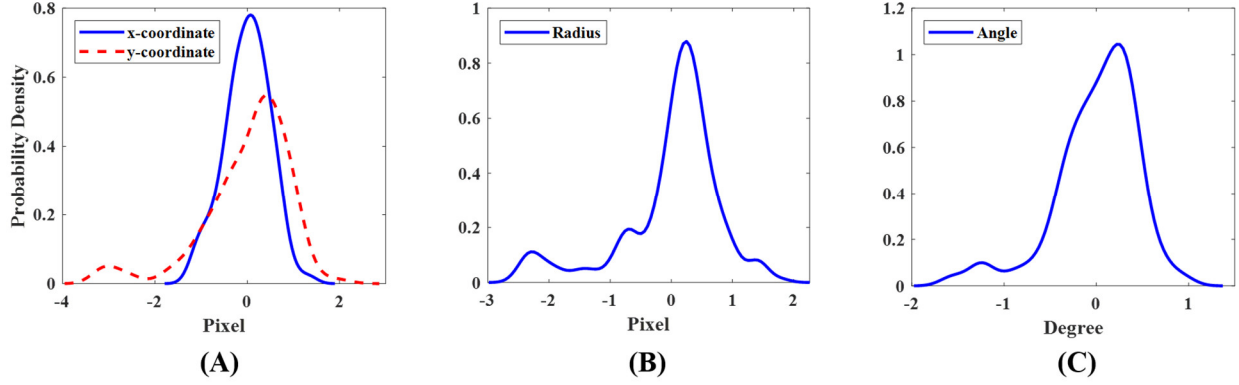


Figure 4.8. Distribution of the variability in the output of FOV and the fiducial finder modules: (A) distribution of the centralized coordinates of the FOV center, (B) distribution of the centralized radius of FOV, (C) distribution of the centralized fiducial angle.

the estimated position should have no variation in the ideal case. Therefore, the standard deviation of the  $(x,y)$  coordinates of each laser point over all batches could be used to evaluate the performance of the algorithm. The same approach was repeated for all working distances. The distribution of this evaluation criterion had a *mean* of 0.012 pixels and *std* of 0.0224 pixels. Figure 4.9 shows the distribution of this evaluation metric. The figure shows that the probability density function is concentrated around a small number near the zero with a very sharp peak. This supports that the employed approach for detection of the laser points is quite robust and has very stable performance.

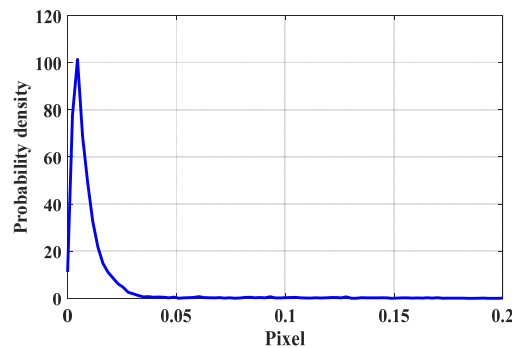


Figure 4.9. Distribution of the variability in the output of the laser finder module.

#### 4.4.2. Experiment2: Displacement analysis and vertical resolution of the system

The displacement of the laser points when the working distance was varied was analyzed. To that end, the positions of all laser points were computed for all working distances. Then, the magnitude of the displacement was plotted as a function of the variation in working distance. Figure 4.10(A) shows the magnitude of displacement when the working distance is changed from 35 *mm* to another target distance. This figure clearly shows a semi-exponential relationship between the working distance and the magnitude of displacement, where, at large working distances the displacement is small, but at small working distances the magnitude of displacement is much larger. To present this phenomenon better, the magnitude of displacement between two consecutive working distances was computed. In this fashion, the amount of decrement in the working distance is kept constant (around 1 *mm*), but the effect of different working distances can be studied. Figure 4.10(B) shows the result. Clearly, at large working distances (>20 *mm*) reducing the working distance by 1 *mm* leads to a small variation in the position of the laser points. On the other hand, as the working distance is reduced, a much larger variation in the position of the laser points for the same reduction in working distance is seen. Because the variation in the position of the laser points captures the vertical displacement of the target surface, these analyses show that the vertical resolution of the endoscope is a function of the working distance, where the vertical movements can be measured with higher resolution at shorter working distances.

Figure 4.10 indicates that different laser points at the same working distance exhibit different behaviors. More specifically, some laser points show a higher magnitude of displacement indicating a higher sensitivity to variation in working distance. To find whether those points have certain relationships with each other or not, another analysis was carried out. The average magnitude of displacement for a 1 *mm* decrement in different working distances (figure 4.10(B))

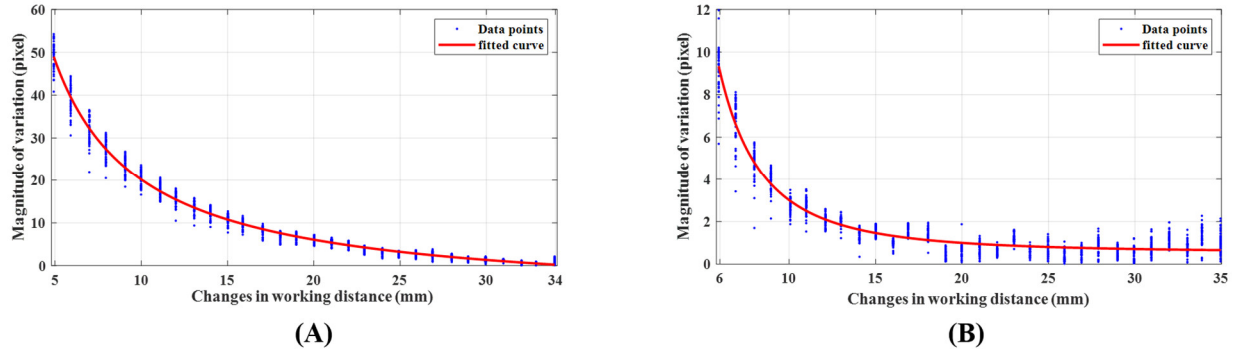


Figure 4.10. Displacement analysis of the laser points as the working distance is changing: (A) the magnitude of variation in the position of the laser points as the working distance is changing from 35 mm to a new distance, (B) the magnitude of variation in the position of the laser points for 1 mm decrement at different working distances.

was computed separately for each laser point and then the result was plotted. Figure 4.11 presents the employed indexing and the result. The result from figure 4.11(B) is significant in several regards. First, the figure has a specific pattern and it is not random. Therefore, the variability seen in figure 4.11 does not stem from the detection algorithm, but it is rather inherent to the characteristics of the system. Second, assuming that the square grid of  $7 \times 7$  points is parallel to the  $x$ - $y$  axes (figure 4.11 (A)), the points with the highest sensitivity to vertical displacement were the three middle rows and the first and last rows had the lowest sensitivity to vertical displacement. Therefore, the best reconstruction of vertical movements is achieved if the target region is covered with laser points from the three middle rows.

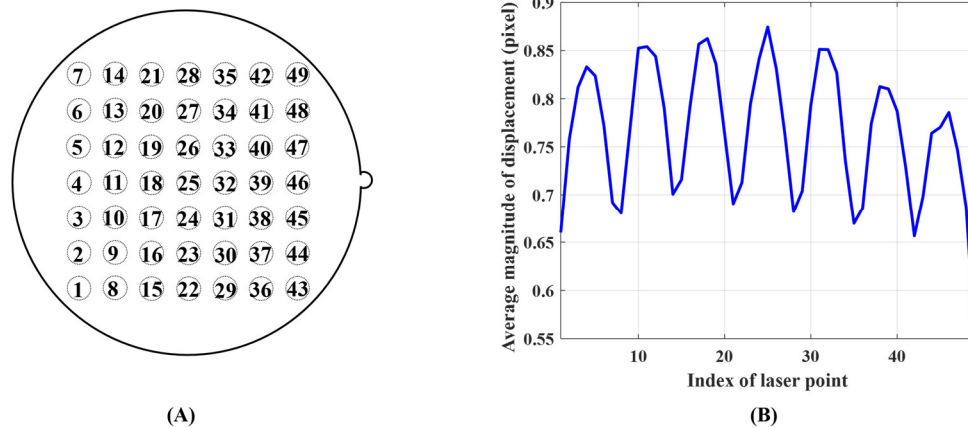


Figure 4.11. The behavior of different laser points: (A) indexing used in this chapter, (B) the average magnitude of displacement of each laser point.

#### 4.4.3. Experiment3: Evaluation of vertical distance measurements

This experiment was conducted to quantify the accuracy of estimated vertical measurements. The following hypothesis was formed for this experiment.

H3b: Vertical measurement error will be positively correlated to working distance.

The proposed method was evaluated using two different criteria. First, the goodness of fit of the functions during the training phase was analyzed. For that purpose, the values of root mean square error (RMSE) and adjusted  $r$ -squared were computed. Figure 4.12 shows these values for each individual function.

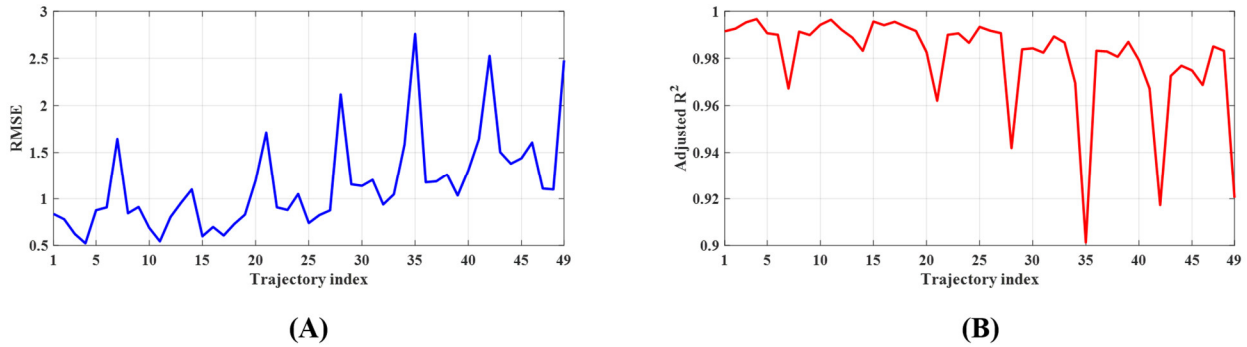


Figure 4.12. The average magnitude of displacement of each laser point.

Figure 4.12 shows that the training error has peaks for RMSE and dips for adjusted  $r$ -squared for laser point indices  $\{7, 14, 21, 28, 35, 42, 49\}$ . These trajectories correspond to the top row of the projection pattern. Therefore, for these points, high values of error in the testing phase are expected. That is, the points on the top row would have a higher vertical measurement error. Referring to the trajectories with the best performance, most of them were from middle rows of the projection pattern and hence those rows would have lower vertical measurement error. This last observation concurs with the results shown in figure 4.11.

Next, the performance of the system in the testing scenario was analyzed. To that end, the target surface was positioned at fifteen different new working distances, and positions of the laser

points were recorded. After finding the  $x$ - $y$  coordinates of the laser points from the above-described approach, they were mapped into the polar coordinate system using Equation 4-10. The radius in the polar coordinate system was then fed as the input to all 49 trained functions, and each function returned an estimated vertical distance. Figure 4.13 shows boxplot of the error at each working distance during the testing phase. Considering that the top row of the laser projection pattern had quite different performance in the training phase (figure 4.12), two different scenarios are reported, first results from all trajectories were used for finding the vertical distance (A), second, the functions corresponding to the top row in the projection pattern were excluded from the analysis (B).

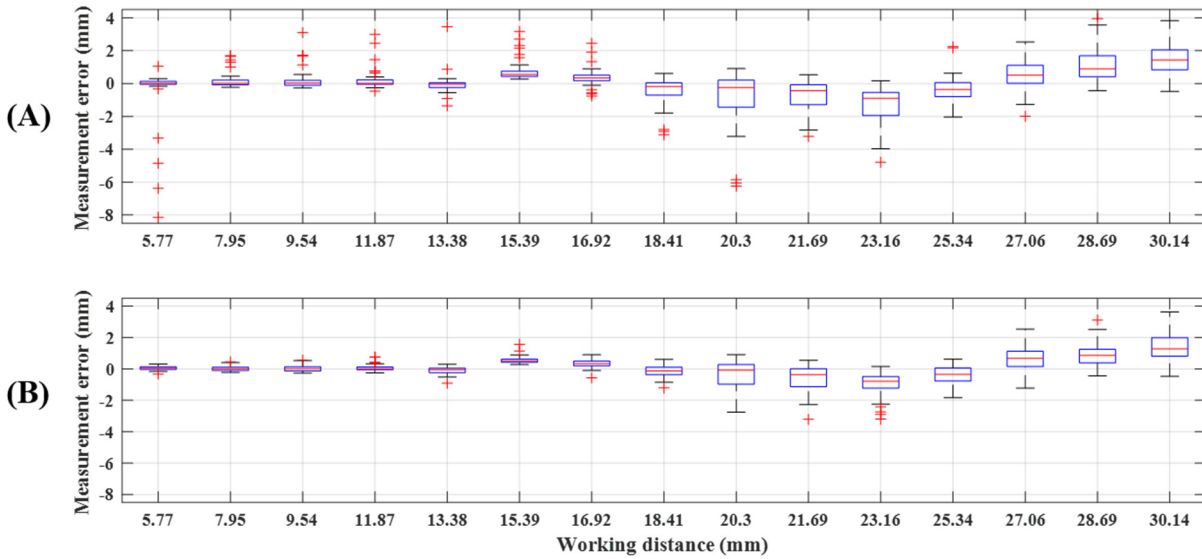


Figure 4.13. Boxplot of vertical measurement errors at different working distances: (A) results from all functions, (B) results when the functions from the top row are discarded.

Referring to figure 4.13, some observations can be made. First, the estimation error at short working distances ( $<20$  mm) is much lower than at large working distances. It is noteworthy that this observation agrees with the result and discussions of experiment2 and figure 4.10(B). Considering the working distance of flexible endoscopy and the fact that these endoscopes could



get near to the target tissue, this characteristic could be utilized very efficiently during the examination. As a rule of thumb, the proximity of the endoscope to the target tissue can be ensured by filling the image with the tissue of interest. Second, when the points on the top row are discarded, the estimation error is reduced considerably. Finally, table 4.2 reports the measurement error for each working distance, comparing the mean of the estimation error for the whole pattern (averaged over all 49 functions) to the case when the top row is discarded (averaged over 42 functions). This value is significant because if a flat and horizontal target surface can be assumed, averaging multiple measurements would remove significant amount of error from the measurements. It also provides the lower (upper) bound on error (accuracy) of the measurements from the device. Additionally, the mean percent error (mPE) defined as the average absolute value of error divided by the working distance, and the maximum percent error (MPE) defined as the maximum of the absolute value of the error divided by the working distance, are also computed and reported.

Table 4.2. Statistics of the measurement error. All measurements have the unit of mm and the number in parentheses signifies the number of functions that were used in the measurements.

<b>Distance</b>	<b>Mean (49)</b>	<b>mPE (49)</b>	<b>MPE (49)</b>	<b>Mean (42)</b>	<b>mPE (42)</b>	<b>MPE (42)</b>
<b>5.77</b>	-0.43	10.4%	141.1%	0.04	1.7%	5.7%
<b>7.95</b>	0.17	3.4%	21.3%	0.02	1.7%	5.6%
<b>9.54</b>	0.19	3.1%	32.5%	0.03	1.5%	5.8%
<b>11.87</b>	0.21	2.6%	25.2%	0.06	1.2%	6.4%
<b>13.38</b>	-0.03	2.1%	25.9%	-0.09	1.3%	6.8%
<b>15.39</b>	0.76	4.9%	20.6%	0.57	3.7%	10.2%
<b>16.92</b>	0.38	2.8%	14.5%	0.33	2.2%	5.3%
<b>18.41</b>	-0.42	3.1%	16.9%	-0.15	1.8%	6.6%
<b>20.3</b>	-0.84	5.5%	30.8%	-0.34	3.3%	13.6%
<b>21.69</b>	-0.78	4%	14.8%	-0.6	3.2%	14.8%
<b>23.16</b>	-1.3	5.6%	20.7%	-0.99	4.3%	13.8%
<b>25.34</b>	-0.4	2.7%	8.8%	-0.43	2.2%	7.2%
<b>27.06</b>	0.51	3.3%	9.3%	0.66	3.2%	9.3%
<b>28.69</b>	1.13	4.1%	13.8%	0.9	3.3%	10.8%
<b>30.14</b>	1.53	5.1%	12.7%	1.41	4.7%	12%

To test hypothesis H3b two correlation tests were used. Based on the results of table 4-2 the top-row laser points were omitted from this analysis. The first test establishes the dependence of measurement error on the working distance. Specifically, a correlation test with the working distance as the independent variable and the average measurement error as the dependent variable was conducted. The second test establishes the dependence of the magnitude of measurement error on the working distance. Specifically, a correlation test with the working distance as the independent variable and the average of absolute measurement error as the dependent variable was conducted. Table 4-3 reflects the results. Based on this analysis, the measurement error has a non-significant and weak correlation with the working distance. However, the magnitude of the error and working distance has a very strong, significant, and positive correlation.

Table 4.3. Results of correlation test for vertical measurement errors. The symbol  $\epsilon$  means  $p < 0.00001$ .

Error		Magnitude of error	
<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
0.28	0.32	0.9	$\epsilon$

## 4.5. Discussions

Speech and voice are the outcomes of intricate collaborative functions between different systems of the body. The pulmonary system provides the driving force for the voice and speech production system and its effect can be measured on a calibrated scale using air-flow and air-pressure measurements, which are used for modeling the underlying mechanisms. On the output, the intensity of the acoustic signal can also be measured on a calibrated scale using sound pressure level. The methodology and the required instrumentation for performing these measurements have been available to researchers for a long time.<sup>13</sup> One of the remaining pieces for developing a comprehensive model for voice and speech production is performing the kinematic measurements on the vocal folds and their vibratory pattern on a calibrated scale. Having access to a device with

absolute measurement capabilities along the horizontal and vertical planes would address this gap. Additionally, personalized medicine<sup>245</sup> and patient-specific modeling<sup>246</sup> are topics of high importance to medicine, because they allow taking into account the differences between individuals during diagnosis and treatment. In patient specific-modeling, such differences could be fed into computational models for improving the diagnosis and treatment of patients by making better predictions about the outcome of different therapeutic options and surgeries.<sup>246</sup> Considering that most current patient-specific modeling approaches rely on the geometry of the tissues derived from 3D imaging techniques, instrumentation with absolute measurement and 3D reconstruction capabilities would be beneficial for developing patient-specific models for populations with voice disorders. Finally, imaging techniques with absolute measurement capabilities can significantly enhance evidence-based practice, an important clinical topic in all fields, including laryngology and speech-language pathology.<sup>30</sup> More specifically, the ability to perform absolute measurements on tissues and to reconstruct the 3D vibratory patterns of the vocal folds would provide researchers and clinicians with means for measuring the size of lesions and performing quantitative analysis on the kinematics of the vocal folds. This information can be obtained before, and after therapy, and the comparison between the two would allow evaluating the efficacy of the therapy. Other important clinical applications of an imaging system with calibrated measurement capabilities include studying the developmental aspects of the laryngeal tissues and the resulting changes in vocal fold vibration<sup>153</sup>, and the more accurate grading of relevant laryngeal diseases.<sup>191</sup>

This chapter provided a detailed analysis of the calibration characteristics and procedures, which is the first step into developing an accurate instrument allowing absolute measurements of the vocal fold vibratory kinematics. Achieving the above-mentioned goals depends on a software solution that performs several additional tasks. Considering the end-user perspective, the laser

points should be first detected and tracked on in-vivo recordings. This module should handle efficiently the non-uniform intensity of the laser points, the high-intensity reflection points in the recorded images, and the non-uniform reflections of the tissues. Further, a second module would take the estimated position of the laser points as an input and perform the required measurements and the reconstruction of the 3D envelope of the vibratory pattern of the vocal folds. Establishing the relationship between the position of the laser points and the target measurements is the prerequisite for this second module. This process is known as calibration, where the calibration along the vertical dimension was the focus of this chapter. To that end, an automatic modular solution was proposed for performing vertical calibration. The modular solution allows the system to be broken into different components and has several important advantages. It makes objective analysis of each module possible, in that different sources of error can be distinguished and each of them can be quantified separately. Also, it provides flexibility in the design where each module may be replaced independently with a better solution in the future. Another feature of the proposed calibration method was using the data-driven approach. Considering that each of the laser points has idiosyncratic characteristics, the manufacturing of each endoscope, and the different endoscopic brands introduce differences, this approach adds significant flexibility to the system. In that regard, the calibration system was designed based on a set of parameters (a translation vector, a rotation matrix, a scaling factor, and parameters of the decoding functions  $\mathcal{F}$ ) where the parameters of  $\mathcal{F}$  are determined separately for each endoscope and the remaining parameters are computed per recording. Another distinctive feature of the data-driven approach is its robustness to measurement error. More specifically, all measurements have some inherent errors, and using statistical learning approaches can remove the random error component and hence, improve the

performance of the system. This feature may improve by increasing the number of training samples.

#### **4.6. Conclusion**

The ability to provide absolute calibrated measurements and to estimate the vertical vibratory pattern of the vocal folds would further advance the kinematic and aerodynamic modeling of voice production, enabling new clinically significant research approaches, such as patient-specific modeling and studying laryngeal development. With these goals in mind, this chapter presented an automatic and modular approach for calibration of a newly developed transnasal fiberoptic endoscope with absolute horizontal and vertical measurement capabilities. This was achieved by mapping the recorded image into a standard and fixed coordinate systems, where the position of the laser points was independent of the lens-coupler parameters such as the magnification of the camera, the rotation of the endoscope relative to the camera, and the displacement of the endoscope within the lens coupler. Consequently, the position of the laser points in this new coordinate system is only a function of working distance. The analysis showed that each laser point travels along a unique and deterministic trajectory, making the efficient decoding of the vertical distance possible. The decoder was implemented based on statistical learning techniques, where a different function was trained per each trajectory. The trained function produces the estimated vertical distance upon a given input. Each module of the system was tested separately, and the results were satisfactory. The system was able to measure absolute vertical distance with the mean percent error varying from 1.7% to 4.7%, depending on the working distance.

## **CHAPTER 5: NON-LINEAR IMAGE DISTORTIONS IN FLEXIBLE FIBEROPTIC ENDOSCOPES**

*Based on:*

Ghasemzadeh H., Deliyski D. D. Non-Linear Image Distortions in Flexible Fiberoptic Endoscopes and their Effects on Calibrated Horizontal Measurements Using High-Speed Videoendoscopy. *Journal of Voice*. 2020 Sep 18:S0892-1997(20)30331-3. doi: 10.1016/j.jvoice.2020.08.029. Epub ahead of print. PMID: 32958427.

Summary: Laryngeal images obtained via high-speed videoendoscopy are an invaluable source of information for the advancement of voice science because they can capture the true cycle-to-cycle vibratory characteristics of the vocal folds in addition to the transient behaviors of the phonatory mechanism, such as onset, offset, and breaks. This information is obtained through relating the spatial and temporal features from acquired images using objective measurements or subjective assessments. While these images are calibrated temporally, a great challenge is the lack of spatial calibration. Recently, a laser-projection system allowing for spatial calibration was developed. However, various sources of optical distortions deviate the images from reflecting the reality. The main purpose of this chapter was to evaluate the effect of the fiberoptic flexible endoscope distortions on the calibration of images acquired by the laser-projection system. Specifically, it is shown that two sources of non-linear distortions could deviate captured images from reality. The first distortion stems from the wide-angle lens used in flexible endoscopes. It is shown that endoscopic images have a significantly higher spatial resolution in the center of the field of view (FOV) than in its periphery. The difference between the two could lead to as high as 26.4% error in calibrated horizontal measurements. The second distortion stems from variation in the imaging

angle. It is shown that the disparity between spatial resolution in the center and periphery of endoscopic images increases as the imaging angle deviates from the perpendicular position. Furthermore, it is shown that when the imaging angle varies, the symmetry of the distortion is also affected significantly. The combined distortions could lead to calibrated horizontal measurement errors as high as 65.7%. The implications of the findings on objective measurements and subjective visual assessments are discussed. These findings can contribute to the refinement of the methods for clinical assessment of voice disorders. Considering that the studied phenomena are due to optical principles, the findings of this study, especially those related to the effects of the imaging angle, can provide further insights regarding other endoscopic instruments (*e.g.* distal-chip and rigid endoscopes) and procedures (*e.g.* gastroendoscopy and colonoscopy).

## 5.1. Introduction

Imaging techniques provide a direct method for observation, assessment, and precision measurement of characteristics of the laryngeal mechanisms. Therefore, they are important in voice research<sup>30,247</sup> and functional assessment of voice production.<sup>110,167,169</sup> Regardless of the imaging modality (*e.g.* VSB, HSV, or videokymography) the acquired images can be evaluated using two main approaches of visual-perceptual assessments<sup>30</sup> or image measurements. Visual-perceptual assessments and image measurements respectively lead to subjective and objective evaluations of some features of the phonatory mechanism. Using a different taxonomy, features from the acquired images may belong to spatial, temporal, or spatial-temporal domains. Some examples of spatial features would be the size of a lesion<sup>206</sup>, glottal closure pattern<sup>95</sup>, and glottic angle.<sup>248</sup> Some examples of spatial-temporal features would be velocity measures<sup>27,103</sup>, mucosal wave<sup>184</sup>, glottal area waveform<sup>102,202</sup>, and kymogram.<sup>113,249</sup> Objective measurements and

subjective assessments based on spatial and spatial-temporal features rely on some implicit but important assumptions. Those implicit assumptions may vary depending on the purpose of measurements or assessments. The notions of *within-* and *between-subject* size comparisons were defined in section 2.3.2, but they are repeated here shortly. Comparison between size of a feature (e.g. lesion size) in the same person but between two different imaging conditions (e.g. pre- post-intervention) is called the *within-subject* size comparison. The implicit assumption of this scenario is that for each subject the measurement from the two recording conditions are on the same scale, and hence can be compared with each other. More precisely, the implicit assumption is that the *mm* size of a pixel (*i.e.* pixel size) in the two conditions for each subject are the same. However, *between-subject* size comparison is the scenario that we want to compare size of a feature (possibly from two recording conditions) among different subjects. The implicit assumption of this scenario is stricter. More precisely, not only the *mm* size of pixels in two recording conditions for each subject should be the same, but also the *mm* size of pixels in different subjects should be the same. Obviously, the *between-subject* size comparison assumption satisfies the *within-subject* size comparison assumption; however, the other direction does not necessarily hold.

Different approaches are possible to satisfy the *between-subject* size comparison assumption. Regardless of the employed approach, all methods are based on the same principle. Basically, pixels are building blocks of images. Therefore, if we know the *mm* size of pixels, all objects in the image could be mapped in *mm* scale which is a universal and standard basis. Intraoperative calibrated images<sup>93,190</sup> and laser-calibrated imaging systems<sup>153,192,195,236</sup> are some possible approaches for determining the *mm* size of pixels. In the intraoperative calibration method, a surgical instrument is placed next to a target tissue and an image is recorded.<sup>93,190</sup> Considering the known *mm* length of the surgical instrument, the *mm* size of pixels in the image could be estimated.



On the other hand, laser-calibrated systems are based on well-designed laser patterns that are projected on the laryngeal tissues. The laser patterns often have specific topological characteristics that help with determining the *mm* size of pixels in the acquired images.

Deriving the *mm* size of pixels based on intraoperative calibrated images or parallel laser projection is based on an important condition. Basically in these approaches, the *mm* size of a pixel is computed from some specific part of the image --in the intraoperative approach this is the target tissue that the surgical instrument is placed next to, and in the laser projection is the part of the image that falls between the two laser points-- and then we assume that the same number is valid for other parts of the image too. Specifically, we assume that all pixels in the image have the same *mm* sizes and therefore the conversion from pixel to *mm* can be achieved using a constant number (*i.e.* independent from the spatial location of the pixel). This assumption is critical for both *within-subject* and *between-subject* size comparison applications, and its violation could lead to significant error in the measurements. To put this argument into perspective let us consider a hypothetical imaging system with a specific non-linear distortion where pixels in the right half of the image correspond to 1 *mm*, and pixels in the left half of the image correspond to 0.5 *mm*. Obviously, using a constant pixel size would lead to significant errors in *between-subject* size comparison applications, as well as, *within-subject* size comparison applications (*e.g.* if the lesion site in pre- and post-intervention are on different halves of the image). Consequently, studying the presence of image distortion is the prerequisite of reliable and accurate spatial measurements.

Reviewing the literature showed that the effect of non-linear distortions has found little attention in the field of voice. Hibi and colleagues investigated the effects of non-linear distortions in flexible endoscopes.<sup>204</sup> They showed that the magnitude of distortion increases with the deviation of the imaging axis from the perpendicular angle.<sup>204</sup> Distortion as high as 20% was

reported for a 30° deviation in the imaging angle. Considering that calibrated horizontal measurements were not possible at that time, that work was geared more toward practical recommendations for keeping the effect of distortions to a minimum. A different research aimed at studying the normative values of the glottic angle using flexible endoscopy acknowledged the significant effect of non-linear barrel distortion on the measurements.<sup>248</sup> However, the study neither provided details on how the distortion was compensated for nor reported the magnitude of errors in presence of the non-linear barrel distortion. Finally, a very recent work investigated the effects of parameters of HSV recordings on the estimation of the phonatory parameters of synthetic vocal folds.<sup>205</sup> This work suggested that the imaging angle was the most influential factor, where a 10° change in the imaging angle led to a 10% error in the estimation of the subglottal air pressure from the glottal area waveform.<sup>205</sup> However, none of these works were aimed at calibrated horizontal measurement and effects of barrel distortion or changes in the imaging angle on it.

## 5.2. Aim and hypothesis

The main aim of this chapter is to investigate if non-linear distortions are present in flexible HSV endoscopy and if so, to quantify their impacts on subsequent horizontal measurements. The main research questions of this chapter are:

Q4a: How much the *mm* size of a pixel depends on its spatial location?

Q4b: How much the imaging angle affects the *mm* size of a pixel?

To answer these research questions and to pursue the aim of this chapter two main hypotheses were formed that are presented in this section.

H4b: Pixel size is significantly smaller in the center group than the periphery group.

H4c: Pixel size is significantly different between back, middle, and front groups when the target surface gets tilted.

The outcomes of this chapter will help us develop a more accurate and reliable method for horizontal calibration and measurements from the laser-calibrated endoscope (to be presented in the next chapter). It is expected for the derived horizontal measurement to improve our understanding of the effect of individual differences on the function of the phonatory mechanism<sup>207</sup> and consequently advancement of personalized medicine in the field of laryngology and speech-language pathology. Additionally, the outcomes of this chapter could help us to better understand possible confounding factors in subjective assessments and objective measurements from flexible endoscopy images. It is noteworthy that the application of the outcomes is not limited to horizontal measurements from laser-calibrated endoscopes. For example, the outcomes could be utilized to increase the accuracy of horizontal measurements from intraoperative images, as well as, any other calibration approach. Additionally, the outcomes of this research would shed light on possible confounding factors affecting the accuracy and reliability of objective measurements and subjective evaluations on images recorded using distal-chip flexible endoscopes or rigid endoscopes. However, the exact effects in distal-chip flexible endoscopes and rigid endoscopes are not the purpose of this chapter and need to be investigated in a separate study.

### **5.3. Optical principles of image formation**

The formation of an image in a camera follows principles of optics. Snell's law is one of the main principles that govern image formation in the presence of a lens.<sup>238</sup> Based on Snell's law, the path of a ray of light changes, when it passes through the boundary of two different mediums. Specifically, let  $n_1$  and  $\theta_1$  denote the refractive index and the angle of incidence in the first

medium. Also, let  $n_2$  and  $\theta_2$  denote the refractive index and the refracted angle in the second medium. Figure 5.1(A) shows these symbols. Equation 5-1 shows the Snell's law.

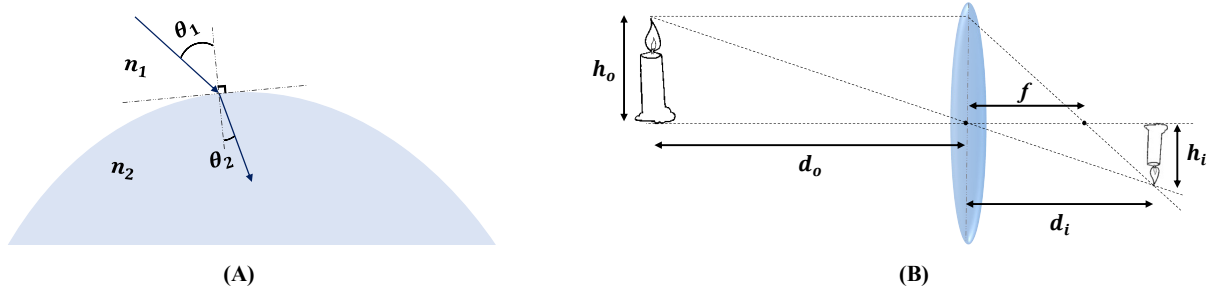


Figure 5.1. Optical principles of image formation: (A) parameters of the Snell's law, (B) image formation in the Gaussian optics model.

$$n_1 \cdot \sin \theta_1 = n_2 \cdot \sin \theta_2 \quad (5-1)$$

Snell's law could be utilized to trace rays of light as they insert and exit the lens, and hence properties of the resulting image could be estimated. However, Snell's law is based on trigonometric functions and hence involves complex computations. One solution is to use approximations to Snell's law. Specifically, using the thin lens assumption and small-angle approximation we can derive a simplified model known as the Gaussian optics<sup>238</sup> which is very easy to use. The small-angle approximation stipulates that the height (length in laryngeal images) of the object relative to its distance from the lens is small. More precisely, in Gaussian optics the object should be near to the optical axis of the imaging system, otherwise, a significant error will be introduced into the computation. Based on Gaussian optics, the properties of the image can be expressed in terms of simple measurements. Let  $d_o$  and  $d_i$  be distances from the lens to the object and its image, respectively (figure 5.1(B)). Also, let  $f$  denotes the focal distance of the lens and  $h_o$  and  $h_i$  be the actual size of the object (*i.e.* *mm* length) and its image size (*i.e.* pixel length), respectively. Equations 5-2 and 5-3 present the relationship between these variables, under the

Gaussian optics.<sup>238</sup> Also, in Equation 5-3  $m$  denotes the magnification of the imaging system and the negative sign is due to the inversion of the image.

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \quad (5-2)$$

$$m = \frac{h_i}{h_o} = -\frac{d_i}{d_o} \quad (5-3)$$

Referring to Equation 5-3,  $h_o$  can be measured in a metric unit (*e.g. mm*) and  $h_i$  can be measured in pixels. We can define the reciprocal of the magnification factor as the pixel size. The value of the pixel size could serve similarly to the scale printed on a map, and it could enable us to estimate the actual length (*i.e. mm* length) of an object from its uncalibrated image length (*i.e. pixel* length). Additionally, based on Equation 5-3 magnification of the camera only depends on  $d_o$  and  $d_i$ . Therefore, under Gaussian optics, all pixels of the image would have similar pixel sizes. However, Gaussian optics approximation is only valid under the small-angle assumption. The optical lens of the flexible endoscope gets very near to the target surface. In that case, a lens with a small FOV angle (and hence valid small-angle approximation) can only visualize a very small portion of the target surface. To remedy this and to increase the size of the FOV, flexible endoscopes are equipped with wide-angle lenses. Considering the significant deviation from the small-angle approximation in such lenses, we may expect significant errors in using the Gaussian optics approximation. In reality, the magnification of imaging systems equipped with wide-angle lenses could become a function of the spatial location of the object in the FOV. Such characteristics will lead to a non-linear distortion. Specifically, if the magnification of an imaging system decreases with the distance from the optical axis, it is called barrel distortion.<sup>250</sup> Conversely, if the magnification of an imaging system increases with the distance from the optical axis, it is called pincushion distortion.<sup>250</sup>

The second source of non-linear distortion could come from deviation in the imaging angle. This effect can be described clearly using the concept of field-of-view cone. A cone can be constructed for an imaging system with its apex on the center of the lens and its base toward the target scene. Sides of this cone denote the last ray of light that can reach the sensor of the camera. Using this concept, an imaging system only records objects that are inside its FOV cone. Figure 5.2 shows the intersections of the FOV cone with two different surfaces. Specifically, the line AC denotes the optical axis,  $S_1$  denotes a surface that is perpendicular to the optical axis, and  $S_2$  denotes a non-perpendicular surface. Intersection of  $S_1$  with the FOV cone creates the circle centered at point B (it is drawn as an ellipse due to perspective principles). However, the intersection of  $S_2$  with the FOV cone creates the ellipse centered at point D. Pictures are only a two-dimensional representation of the three-dimensional world, hence the height is lost during the imaging. Therefore, differences in heights of the left and right sides of the ellipse centered at D (*i.e.* the one located on  $S_2$ ) are lost and it is also mapped into a circle in the final picture. To differentiate between the intersection of a surface with the FOV cone and its recorded image, the former one is called the FOV while the latter one is called the image-FOV in the rest of this chapter.

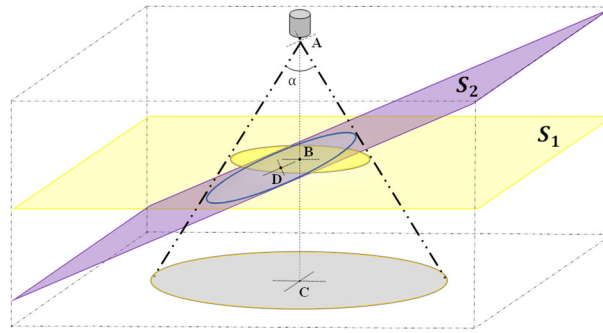


Figure 5.2. Effects of tilting the target surface on the geometry of the acquired images.

Assuming the small-angle approximation (*i.e.* small  $\alpha$  in figure 5.2), Equation 5-3 could be used for finding the magnification of the imaging system. Specifically, if two objects are on  $S_1$ ,

one to the left and one to the right side of point B, they would have similar distances from the lens ( $d_i$ ) and hence similar magnification factors. On the other hand, if the two objects are on  $S_2$ , one to the left and one to the right side of point D, they would have unequal distances from the lens. That is, the object on the right will be closer to the camera and hence will have a larger magnification factor comparing to the object on the left. This example indicates another case of the dependence of the magnification factor of an imaging system to the spatial location of the target object. Another interesting observation from figure 5.2 is that when the surface is perpendicular to the optical axis, the center of the image-FOV (*i.e.* point B) coincides with the intersection point of the optical axis, and the surface  $S_1$ . However, when the surface is tilted the center of the image-FOV (*i.e.* point D) moves away from the intersection point of the optical axis, and the surface  $S_2$ . Combining this observation with properties of the barrel distortion would lead to an interesting anticipation, which is tested in this chapter. We know in imaging systems with a barrel distortion the maximum magnification happens near the optical axis. Therefore, we could anticipate that if the surface is tilted, the point with the maximum magnification (*i.e.* the point with the smallest pixel size) would move from the center of the image toward the direction that gets closer to the imaging system.

## **5.4. Material and method**

### **5.4.1. Recording instrumentation and setup**

To answer the research questions of this study, different sets of benchtop recordings should be collected. Therefore, the setup presented in chapter 1 (figure 1.2) with both degrees of freedom was used. Considering that images were taken from static surfaces, high frame rates were not required and the moderately low speed of 200 frames per second was used for data collection. The

main benefit of reducing the frame rate is the increase in the integration time that we could get. Therefore, the target surface does not need to be very bright and instead of a xenon light, a conventional study incandescent lamp could be used for illuminating the target surface. The main problem with the xenon light was that it produced a spatially non-uniform illumination (*i.e.* the intensity of the light at different spatial locations was very different). This non-uniformity led to images with high-intensity divergence, which would unnecessarily complicate the required image processing algorithms. Therefore, a study lamp was employed as the light source for data collection.

#### 5.4.2. Datasets

This study used recordings from a target surface at multiple working distances and multiple tilting angles for answering the research questions. The working distance was varied from 5 *mm* to 20 *mm* in 5-*mm* increments. The working distance was measured using a digital height gauge with an accuracy of 0.001'' (approximately 0.03 *mm*). The tilting angle was varied from -15° to 15° in 5-degree increments. The following procedure was followed for measuring and adjusting the tilting angle. First, the target surface was leveled using a leveler. Then the distance between the front edge of the target surface (figure 5.3) and the desk was measured using the digital height gauge. The same measurement was carried out for the back edge of the target surface. Let  $D$  and  $l$  denote the difference between the back and front measurements and the length of the target surface, respectively. Additionally, for a desired tilting angle let  $h_b$  and  $h_f$  denote heights of the front and back edges of the target surface from the table. Figure 5.3 depicts definitions of these quantities. Now, the trigonometric functions could be employed for measuring the tilting angle of the target



surface ( $\gamma$ ). Equation 5-4 shows the formula. Based on Equation 5-4, a negative angle corresponds to the case where the front edge of the target surface is higher than the back edge.

$$\gamma = \arcsin\left(\frac{h_b - h_f - D}{l}\right) \quad (5-4)$$

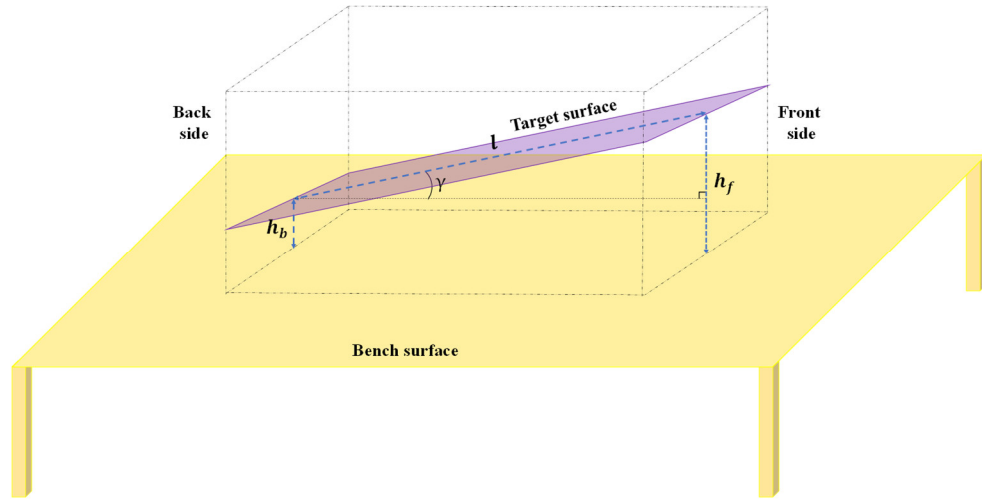


Figure 5.3. A schematic for measuring the tilting angle.

Finally, it is hard to adjust the setup for achieving the exact target working distances and tilting angles; therefore, the actual values deviated from the target values. Table 5.1 reflects the actual value of these parameters for each set of recordings. However, in the rest of this chapter groups will be referenced using their target values.

Table 5.1. Actual values of working distance and tilting angle for each target group. The first number represents the actual working distance in *mm*, and the second number the actual tilting angle in degree.

		Working distance group			
		5	10	15	20
Tilting angle	-15	5.12, -15.6	9.93, -15.6	15.05, -15.6	20.06, -15.6
	-10	5.06, -10.1	10.04, -10.1	15.27, -10.1	20.02, -10.1
	-5	5.12, -5.1	10.07, -5.1	15.18, -5.1	20.05, -5.1
	0	4.95, 0	10, 0	15.08, 0	20.12, 0
	5	5.14, 5	10.05, 5	15.29, 5	20.05, 5
	10	5.08, 10.3	10.08, 10.3	15.30, 10.3	20.15, 10.3
	15	5.26, 15.6	10, 15.6	15.07, 15.6	20.07, 15.6

Considering the aim of this chapter, square grid papers were attached to the target surface and they were recorded with the spatial resolution of  $288 \times 280$  pixels, the frame rate of 200, and exposure time of  $4900 \mu s$ . Subjective investigations showed that 1 mm grids were quite blurry and hard to detect at the working distance of 20 mm. Therefore, two different square grids with 1 mm and 2 mm spacings were used for data collection. Working distances of 5 mm, 10 mm, and 15 mm were recorded using 1 mm-spacing grids and working distances of 15 mm and 20 mm were recorded using 2 mm-spacing grids. The overlap between the two cases was used to investigate any possible effect of different grid sizes on the measurements. This is discussed in more detail in section 5.5.1.

### **5.4.3. Automatic detection of grid lines**

The main aim of this study was to investigate the effect of non-linear distortions in flexible endoscopy on horizontal measurements from the acquired images. Accurate detection of the grid lines from benchtop recordings was the prerequisite of that. Visual investigation of the recordings showed that grid lines in the images did not constitute straight lines but had some curvature. This characteristic is a classic case of barrel distortion. Figure 5.4(A) shows an example image taken from the 1 mm grid at the working distance of 10 mm. Therefore, an automatic algorithm based on statistical image processing was developed to account for possible curvature of the grid lines. Frames of each video recording were averaged over time and then a spatial averaging filter with the size of 2 pixels was applied. The following algorithm was then used for the detection of lines parallel to the y-axis. The filtered image was segmented in strips parallel to the x-axis with the width of 10 pixels and maximum overlaps (*i.e.* 9 pixels). The strip was averaged over the columns, and then locations of its local minima were detected. A zero-vector mask was created, and locations of the minima were set to 1. This procedure was repeated for all strips parallel to the x-

axis, and all masks were concatenated vertically to create a binary image. The binary image at this stage underwent two morphological operations of dilation and erosion<sup>244</sup> using rectangular structuring elements with the size of  $8 \times 2$  and  $3 \times 1$  pixels. Finally, second-order polynomials were fitted on the regions with large areas. Figure 5.4 shows the outputs of the algorithm at different stages. The procedure for the detection of the lines parallel to the  $x$ -axis followed similar steps. However, the filtered image was segmented in strips parallel to the  $y$ -axis instead. Also, the strips were averaged on the rows, zero-vector masks were concatenated horizontally, and rectangular structuring elements had the size of  $2 \times 8$  and  $1 \times 3$  pixels.

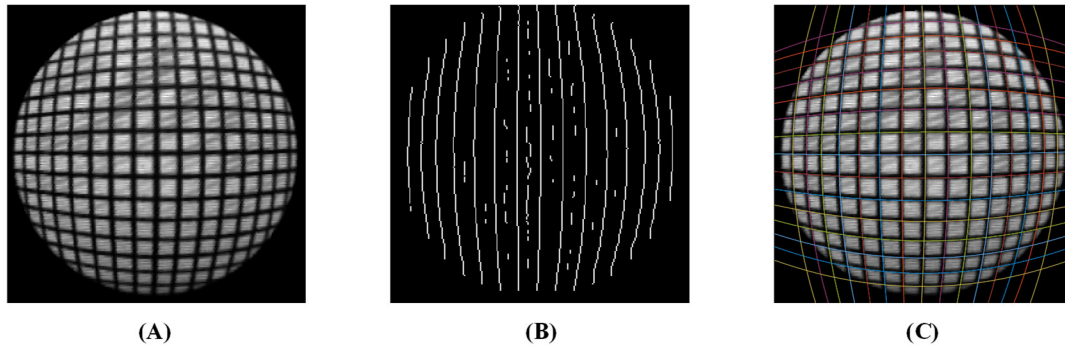


Figure 5.4. Automatic detection of the grid lines: (A) recording from 1  $mm$  grids at the working distance of 10  $mm$ , (B) the binary image showing the locations of the minima, (C) fitted second-order polynomials on the locations of the minima.

#### 5.4.4. Pixel size

This study relies on a variable called the pixel size. This quantity could play a similar role to a scale on a printed map. Basically, we can multiply the uncalibrated pixel length of an object with this quantity and estimate its calibrated  $mm$  length. This number can be estimated as the ratio of the  $mm$  length of a target object to its pixel length during the horizontal calibration process. In this study, the target surfaces were calibrated square grids; hence, the  $mm$  lengths of sides of all blocks were known. Therefore, we could measure pixel lengths of sides of blocks from the image and

then compute their corresponding values of pixel sizes. To that end, pixel lengths of sides of blocks were determined from the fitted curves (figure 5.4(C)). Specifically, coordinates of intersections of all curves were determined with the precision of 0.1 pixels. Then, the pixel length of a side was computed as the Euclidian distance between its corresponding intersection points.

## 5.5. Experiments and results

Three experiments were conducted to answer the research questions of this chapter. Experiment 1 investigates the existence of differences in pixel sizes computed from different grids. Experiment 2 presents the results on dependence of the pixel size on the spatial location of the target region. Experiment 3 tests the effect of imaging angle on pixel size. This section presents details of each experiment, followed by results and related discussions.

### 5.5.1. Experiment 1: Differences between grid sizes

We saw in section 5.4.2 that two different grids with 1 *mm* and 2 *mm* spacing were used for collecting data from different working distances. Before proceeding with further analysis, we need to make sure that measurements from 1 *mm* and 2 *mm* grids are comparable. The following hypothesis was formed to test this.

H4a: Pixel sizes computed from 1 *mm* grids are significantly different from 2 *mm* grids.

Rejection of H4a would indicate that measurements from 1 *mm* and 2 *mm* grids are comparable. The dataset for this experiment were images from 1 *mm* and 2 *mm* grids recorded at the working distance of 15 *mm*. Considering the possible effect of spatial location on the pixel size, two different groups of blocks were distinguished. The center group included all sides of blocks that were nearest to the center of the image-FOV. The periphery group included the farthest side of the

blocks that were farthest from the center of the image-FOV. Figure 5.5 depicts the two groupings with their corresponding selected sides.

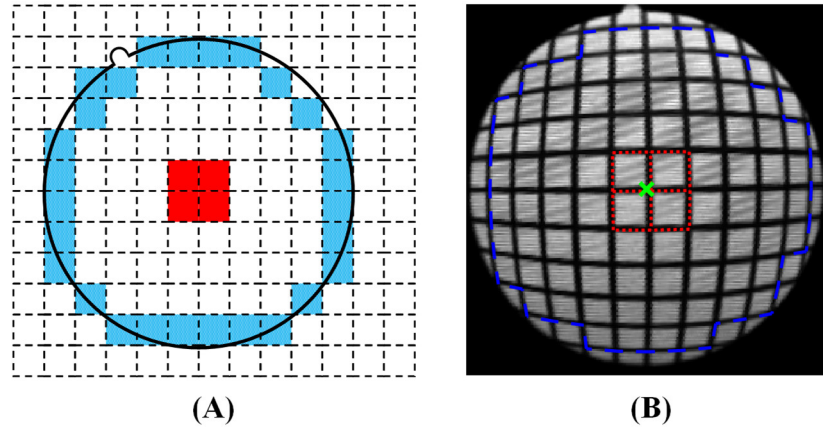


Figure 5.5. Groupings for experiments 1 and 2: (A) the solid red blocks and the patterned blue blocks denote the center and the periphery groups, (B) the selected sides of an example image. The Center of the image-FOV is denoted by a green cross mark.

The dependent variable for this experiment was the computed pixel size. The independent variables were grid sizes (1 *mm* vs. 2 *mm*) and groupings (center vs. periphery). A two-way *ANOVA* was used to test H4a. Since it is known that *ANOVA* is generally not robust to the violation of homogeneity of variance if groups have different sample sizes<sup>251</sup>, *Levene's* test was first employed to check the homogeneity of variance. The test rejected the null hypothesis ( $p < .00001$ ). Therefore, the analysis was carried out using M-estimators for the location with 1000 bootstrap, which provides ANOVA with a robust performance for non-homogeneous variance between groups.<sup>252</sup> Table 5.2 reflects the results of this analysis.

Table 5.2. Results of 2x2 robust ANOVA.

variable	<i>p</i>
Grouping ( <i>G</i> )	<0.00001
Grid size ( <i>S</i> )	0.12
<i>G</i> × <i>S</i>	0.13

Based on table 5.2, we see a non-significant effect of grid size on the pixel size. Therefore, we could conclude that measurements from 1 *mm* and 2 *mm* grids are comparable. Additionally, we see a significant effect for the grouping variable. It means that pixel sizes were significantly different between the center and the periphery groups. To better investigate this, experiment 2 was conducted.

### 5.5.2. Experiment 2: Effect of spatial location

The aims of this experiment were to establish the dependence of the pixel size on its spatial location and then to quantify that dependence. Specifically, the effects of different groups (center vs. periphery as depicted in figure 5.5) and different working distances on the pixel size were analyzed. Table 5.3 presents descriptive statistics of pixel size in different conditions.

Table 5.3. Descriptive statistics of pixel sizes.

Working distance ( <i>mm</i> )	Center		Periphery	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
	mean ( <i>mm</i> )	std ( <i>mm</i> )	mean ( <i>mm</i> )	std ( <i>mm</i> )
<b>5</b>	0.028	0.001	0.037	0.004
<b>10</b>	0.054	0.001	0.074	0.008
<b>15</b>	0.08	0.001	0.107	0.012
<b>20</b>	0.106	0.001	0.141	0.017

Figure 5.6 depicts how the pixel size changes between different groups and working distances.

Based on figure 5.6 we can hypothesize that,

H4b: Pixel size is significantly smaller in the center group than the periphery group.

To test this hypothesis a new dataset was compiled. The dataset consisted of images from 1 *mm* grids recorded at the working distances of 5 *mm* and 10 *mm* and from 2 *mm* grids recorded at the working distances of 15 *mm* and 20 *mm*. The dependent variable for this experiment was the pixel size. The independent variables were groups (center vs. periphery) and working distance. A

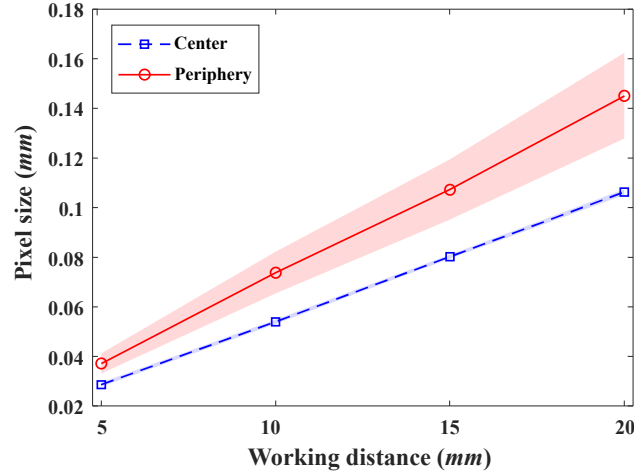


Figure 5.6. Variation in pixel size for different working distances and groups.

two-way *ANOVA* could be used to test H4b. It is known that *ANOVA* is not robust to the violation of homogeneity of variance if groups have different sample sizes<sup>251</sup>; therefore, *Levene's* test was used to check the homogeneity of variance. *Levene's* rejected the null hypothesis ( $p < .00001$ ) indicating non-homogeneity of variance between different groups. Consequently, the robust two-way *ANOVA* using M-estimators for the location with 1000 bootstrap samples was used instead.<sup>252</sup> Table 5.4 reflects the results of the analysis.

Table 5.4. Results of 2×4 robust ANOVA.

variable	<i>p</i>
Groups ( <i>G</i> )	<0.00001
Working distance ( <i>WD</i> )	<0.00001
<i>G</i> × <i>WD</i>	<0.00001

Based on table 5.4 we see a significant main effect of groups (center vs. periphery), a significant main effect of the working distance, and a significant interaction effect. In order to pinpoint differences, robust post hoc analysis with 1000 bootstrap samples was used.<sup>252</sup> The analysis showed significant differences between all contrasts. Figure 5.7 presents the boxplots of the pixel size for different groups and working distances.

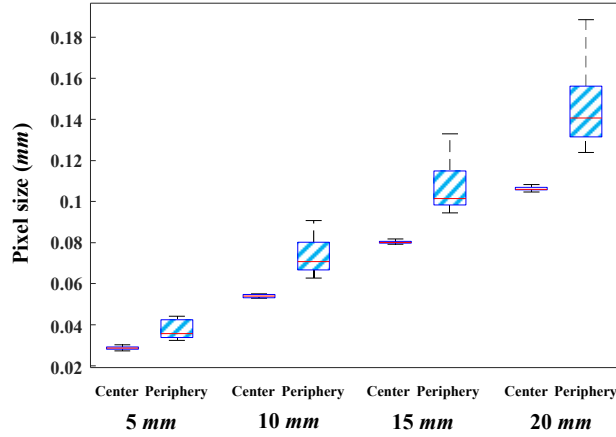
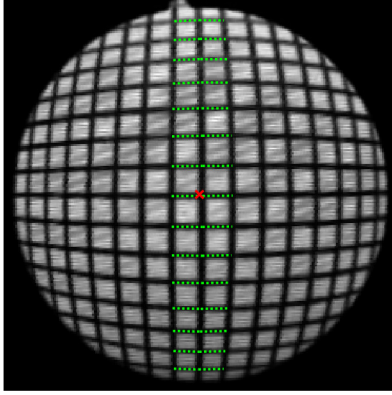


Figure 5.7. Boxplots of the pixel size for different groups and working distances.

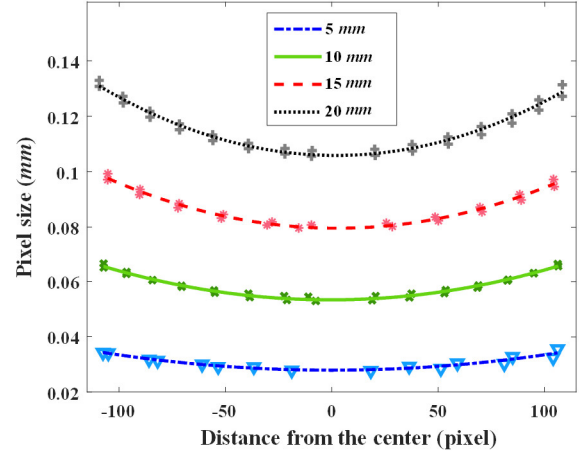
Based on figure 5.7 we could conclude that, at a fixed working distance, pixels from the center group have smaller pixel sizes than pixels from the periphery group. Additionally, the pixel size increases with the working distance, which was to be expected. Finally, as the working distance increases the disparity in the pixel size between the center and the periphery groups increases. This observation which concurs with the significant interaction effect presented in table 5.4 has practical implications. Specifically, measurement errors due to the usage of pixel length for comparing sizes of two objects, one in the center and one in the periphery, increases with the working distance.

To quantify the effect of the spatial location of a pixel on its pixel size a different analysis was carried out. The pixel size of line segments highlighted in figure 5.8(A) were computed. Also, the Euclidian distances between the center of all line segments and the center of the image-FOV were computed. Finally, a negative sign was assigned to the distance of blocks that were below the center of the image-FOV. Figure 5.8(B) presents a scatter plot of the pixel size for different distances from the center of the image-FOV. Second-order polynomials were fitted to measurements.





(A)



(B)

Figure 5.8. Estimation of the dependence of pixel size on its spatial location, (A) selected line segments are shown in green dashed line, and the center of the image-FOV is denoted with a red cross mark, (B) dependence of pixel size on its distance from the center of the image-FOV and the working distance. The negative distance means blocks that were below the center of the image-FOV.

Based on figure 5.8 the following conclusions can be made. First, the relationship between the pixel size and the distance from the center of the image-FOV is non-linear. Second, curves are symmetrical around the center (*i.e.* zero distance). This characteristic has practical implications. Basically, it means that pixel length cannot be used for *within* (and *between*) *subject* size comparison, unless the target objects have similar distances from the center of the image-FOV, in addition to similar working distances and zero tilting angle. For example, pixel length could not be used for comparing spatial features of a point on the left vocal fold to a similar point on the right vocal fold, unless those points have similar distances from the center of the image-FOV. Third, pixels in the center of the image have the smallest pixel size, and as we move toward the periphery the value of pixel size increases. This characteristic has important practical implications. Moving the target tissue to the center of the image-FOV provides better spatial resolution and details in the captured images. Fourth, the curvature of plots increases with the working distance. That is the difference between pixel size in the center and the periphery increases with the working

distance. This result concurs with the results and discussion of figure 5.7, and the significant interaction effect of table 5.4.

The fitted second-order polynomials could be used to quantify the magnitude of variations in the pixel size between the center and the periphery. Table 5.5 shows the estimated values of pixel size at the center and periphery of the image-FOV. Considering the dependence of the pixel size on its spatial location, a possible simplistic approach for computing the *mm* length of an object could be to compute the average values of all pixel sizes in the image-FOV and use it as the pixel size. The mean column in table 5.5 reflects this value. However, if this mean value is used for measuring the *mm* length of an object in the center and the periphery, some error will be introduced into the measurement. The percent value of this error for a center pixel was defined as the difference between the mean pixel size and the pixel size in the center divided by the mean value. A similar approach was followed for computing the percent difference of a periphery pixel. These values are presented in the last two columns of table 5.5.

Table 5.5. Estimated values of pixel size.

<b>Working distance (mm)</b>	<b>Center (mm)</b>	<b>Periphery (mm)</b>	<b>Mean (mm)</b>	<b>Center diff. %</b>	<b>Periphery diff. %</b>
<b>5</b>	0.028	0.035	0.03	8	-16
<b>10</b>	0.053	0.067	0.058	8.1	-16.4
<b>15</b>	0.079	0.099	0.086	7.9	-14.6
<b>20</b>	0.106	0.131	0.115	7.8	-14.6

Combining previous results with table 5.5 the following conclusions can be made. Despite the fact that the absolute value of difference increases with the working distance (more curvature in figure 5.8(B) at larger working distances), yet the percentage of error remains relatively constant. This characteristic means that the non-linear distortion mostly depends on the optical characteristics of the endoscope and it is relatively independent of the working distance. This

independence translates into a simpler method for compensating the effect of such non-linear distortions in horizontal measurements. This topic would fully be investigated in the next chapter.

To put the results of table 5.5 into perspective an extreme case of a *within-subject* size comparison scenario is presented. Let us consider the actual size of a lesion is reduced from 2 *mm* to 1.5 *mm* post an intervention. If the pre-intervention lesion is recorded at the working distance of 10 *mm* and on the periphery of the image, it would be presented by approximately 30 pixels. However, if the post-intervention lesion is recorded at the same working but on the center of the image, it would be presented by approximately 28 pixels. That is, despite a 25% reduction in the *mm* length of the lesion we would get only a 6.7% reduction in the pixel length. This reduced sensitivity requires a bigger sample size in scientific research in order to achieve a significant effect.

### 5.5.3. Experiment 3: Effect of the tilting angle

Experiments 1 and 2 were done at zero tilting angle (*i.e.* imaging axis was perpendicular to the target surface). However, changes in the tilting angle could also lead to non-linear distortions. The aim of this experiment was to study and quantify the effects of this parameter on horizontal measurements. Therefore, values of pixel size in three different groups at multiple working distances and multiple tilting angles were studied. Figure 5.9 shows the groupings that were used in this experiment. Recording at the working distance of 5 *mm* resulted in 14 line segments in the front and back groups and 18 line segments in the middle group. For all other working distances, all three groups had 22 line segments.

We saw in experiment2 that the pixel size increases with the working distance. Considering that tilting the target surface decreases the working distance of one side of the image and increases the working distance of the other side, the following hypothesis was formed.

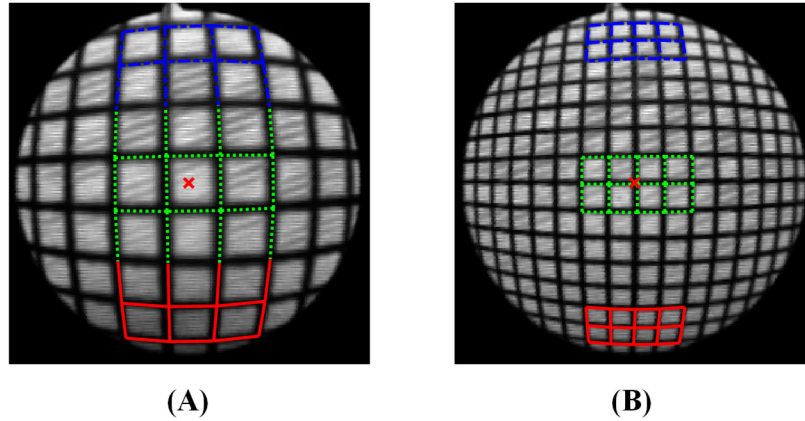


Figure 5.9. Groupings for experiment 3. Solid red lines denote the back group, dotted green lines denote the middle group, and dashed blue lines denote the front group, (A) groupings at the working distance of 5 *mm*, (B) groupings at the working distance of 15 *mm*.

H4c: Pixel size is significantly different between back, middle, and front groups when the target surface gets tilted.

To test this hypothesis a new dataset was compiled. The dataset consisted of images from 1 *mm* grids recorded at working distances of 5 *mm* and 10 *mm* and from 2 *mm* grids recorded at working distances of 15 *mm* and 20 *mm*. The dependent variable for this experiment was the pixel size. The independent variables were groups (back, middle, and front), the working distance, and the tilting angle. Figure 5.10 presents the mean and standard deviation of the pixel size for the three groups at different working distances and tilting angles.

A three-way *ANOVA* was used to test H4c. *Levene's* test rejected the null hypothesis ( $p < .00001$ ). Therefore, the analysis was carried out using trimmed means (0.2 trimming level), which provides ANOVA with a robust performance for non-homogeneous variance between groups.<sup>252</sup> Table 5.6 reflects the results of the analysis. Based on this table we see all main effects were significant. Additionally, except for the Angle $\times$ WD, all other interaction effects were significant.

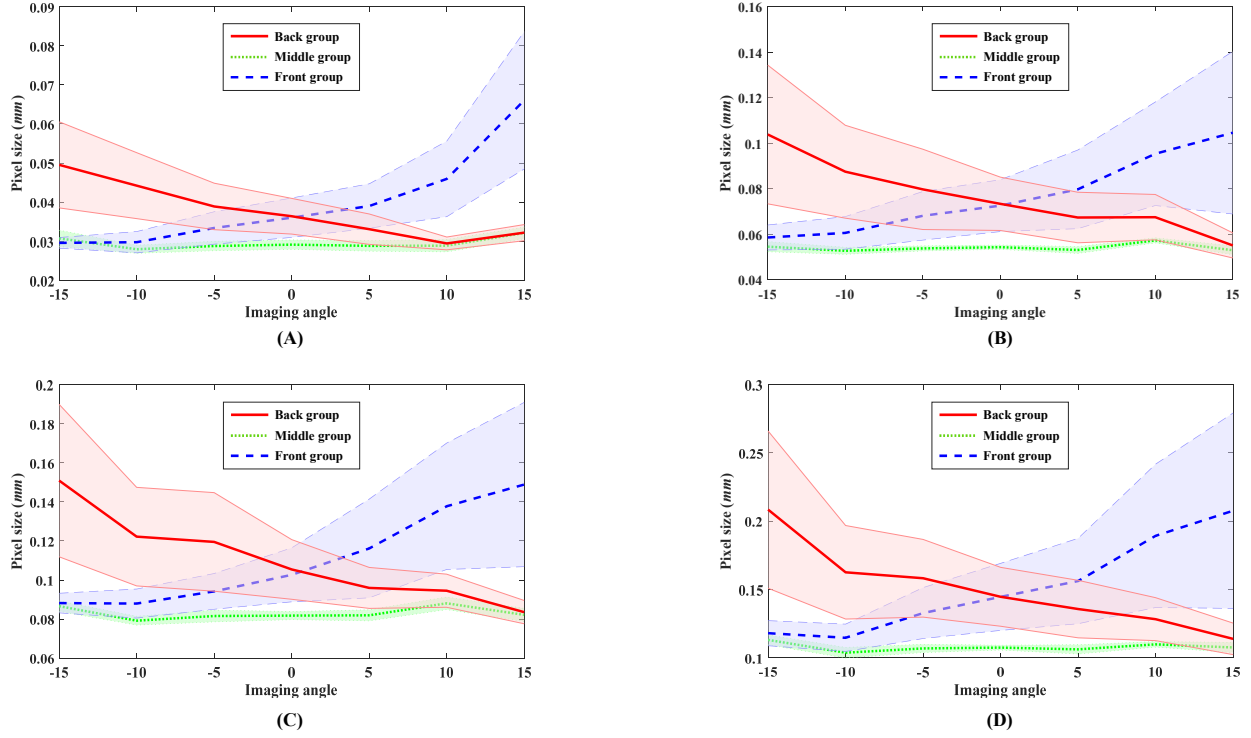


Figure 5.10. Values of the mean and standard deviation of pixel size: (A) working distance of 5 mm, (B) working distance of 10 mm, (C) working distance of 15 mm, (D) working distance of 20 mm.

Table 5.6. Results of  $7 \times 4 \times 3$  ANOVA for trimmed means.

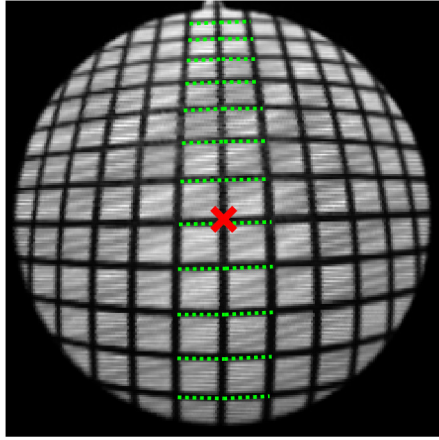
variable	<i>p</i>
Angle	0.0001
Working Distance ( <i>WD</i> )	0.001
Groups ( <i>G</i> )	0.0001
Angle $\times$ WD	0.86
Angle $\times$ G	0.001
G $\times$ WD	0.001
Angle $\times$ G $\times$ WD	0.001

Based on figure 5.10 the following conclusions can be made. First, when the tilting angle is zero, the back and front groups have similar pixel sizes. However, as the magnitude of the tilting angle increases the difference in the pixel size of the back and front group increases. Specifically, at positive angles (*i.e.* when the backside is higher) pixels in the back group have smaller pixel sizes than the front group (hence higher spatial resolution in the backside). Conversely, at negative

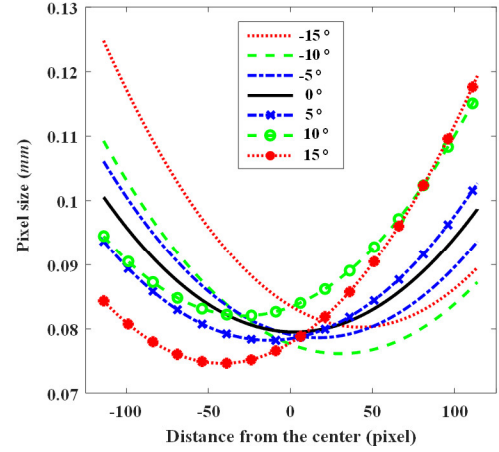
angles, pixels in the front group have smaller pixel sizes than the back group. Second, crudely speaking, the behavior of the front group at a negative angle is similar to the behavior of the back group at a similar but positive angle, and vice versa. This characteristic indicates the presence of a specific symmetry in the distortion. Third, the standard deviations of different groups show dissimilar trends. The middle group exhibits the least variations and its behavior remains relatively constant for different tilting angles. However, as the tilting angle goes from  $-15^\circ$  to  $15^\circ$  standard deviation of the pixel size in the front group (back) increases (decreases). This behavior may indicate a non-linear dependence of the pixel size on the tilting angle and the spatial location of a target pixel. To quantify this behavior a further analysis was carried out.

The pixel sizes for line segments highlighted in figure 5.11(A) were computed. Then, the Euclidian distance between the center of all line segments and the center of the image-FOV was computed. Then, a second-order polynomial curve was fitted for data points computed from each tilting angle. Figure 5.11(B) represents the result of this analysis for the working distance of 15 *mm*. It is noteworthy that, the negative sign denotes blocks that were below the center of the image-FOV.

Based on figure 5.11(B) we see significant differences between different curves. Specifically, when the tilting angle is zero, the minimum of the curve is near point zero (*i.e.* the minimum pixel size is at the center of the image-FOV). However, when the tilting angle becomes positive the minimum of the curve (*i.e.* position with the minimum pixel size and hence the highest spatial resolution) deviates from the center of the image-FOV and goes toward the negative direction (*i.e.* toward the back of the target surface). Additionally, the magnitude of this deviation is positively correlated with the magnitude of the tilting angle. Conversely, when the tilting angle becomes negative the minimum of the curve deviates from the center of the image-FOV and goes toward



(A)



(B)

Figure 5.11. (A) The selected line segments are shown in green dashed lines, and the center of the image-FOV is denoted with a red cross mark. (B) Dependence of pixel size on its distance from the center of the image-FOV and the tilting angle at the working distance of 15 mm.

the positive direction (moving toward the front of the target surface). Additionally, the magnitude of this deviation is positively correlated with the magnitude of the tilting angle. To quantify these qualitative observations, further analysis was carried out. The minimum of each curve was estimated using the analytical approach (*i.e.* equating the derivative to zero). Figure 5.12 shows the distance of the minimum pixel size from the center of the image-FOV.

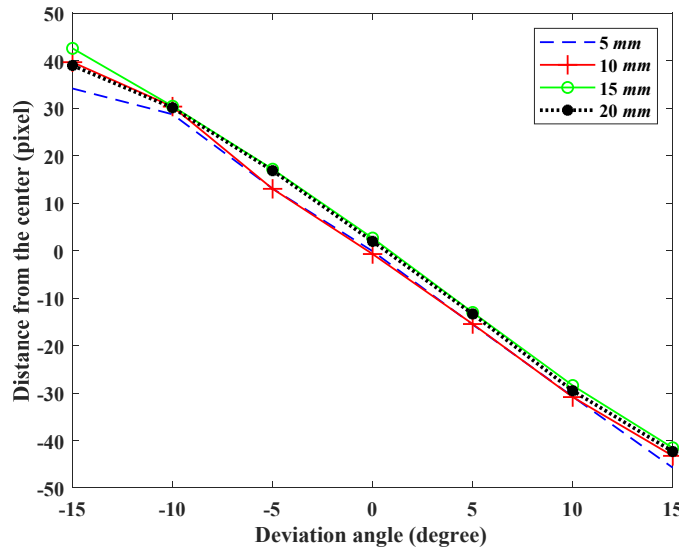


Figure 5.12. Dependence of location with the highest spatial resolution on the tilting angle.

Another significant observation from figure 5.11 is that at zero tilting angle the curve is symmetrical around the minimum point (which coincides with the center of the image-FOV). This means the points with similar distances from the center of the image-FOV would have similar pixel sizes. However, as the tilting angle starts to deviate, the curves become exceedingly asymmetric. That is, the dissimilarity between the two portions of the curves (left of the minimum and right of the minimum) increases with the magnitude of the tilting angle. To quantify these qualitative observations, further analysis was carried out. Let  $\lambda_{min}$  and  $\lambda_{front}$  denote the minimum pixel size and the pixel size at the front periphery of the image-FOV. Then, the percentage of difference at the front periphery ( $D_f$ ) was defined as follows.

$$D_f = \frac{\lambda_{front} - \lambda_{min}}{\lambda_{min}} \times 100\% \quad (5-5)$$

The percentage of difference at the back periphery ( $D_b$ ) was defined similarly. These values were computed for each working distance and tilting angle. Table 5.7 shows the results, which support the preceding qualitative discussions. Specifically, at negative tilting angles, pixel sizes are significantly larger at the back periphery (larger values of percentage of difference), and at positive tilting angles, pixel sizes are significantly larger at the front periphery (hence smaller spatial resolution). Additionally, as the magnitude of the tilting angle increases the percentage of difference from one side (the side that is getting away from the camera) increases while the other side (the side that is getting near to the camera) decreases. For example, at the working distance of 10 mm and the tilting angle of 15°, the pixel sizes at the front and back peripheries are 61.2% and 12.4% larger than the minimum pixel size. In summary, pixel sizes at the side of FOV that gets closer to the camera become more similar, whereas, the other side become more divergent.



Table 5.7. The percentage of difference at the back and front peripheries from different working distances and tilting angles.

		<b>5 mm</b>		<b>10 mm</b>		<b>15 mm</b>		<b>20 mm</b>	
		D <sub>b</sub> %	D <sub>f</sub> %	D <sub>b</sub> %	D <sub>f</sub> %	D <sub>b</sub> %	D <sub>f</sub> %	D <sub>b</sub> %	D <sub>f</sub> %
Tilting angle	-15°	61.5	17.9	59	13.8	55.5	11.5	59.6	14.3
	-10°	46.1	16.5	45.3	15.2	43.5	14.6	42.2	14.3
	-5°	28.7	18.1	29.4	18.6	35	19.1	34.1	18.8
	0°	25.7	25.8	25.7	26.3	26.5	24.2	25.8	24.1
	5°	18.1	31.4	19.5	33.7	19.8	31.3	20.5	32.8
	10°	16.8	50.9	14.5	43.9	15.2	42.1	15.1	43.6
	15°	12	65.7	12.4	61.2	13	59.9	12.7	60.3

The main goal of this chapter was to investigate the effect of two non-linear distortions on horizontal measurements. To that end, we simulated a situation where an object with the actual length of 2 mm was placed at different locations of the FOV (front periphery, center, back periphery, and the location with the highest spatial resolution). Then we used the estimated value of pixel size (figure 5.11(B)) for computing the pixel length of that object at the working distance of 10 mm and different tilting angles. Table 5.8 presents the results. The location with the highest spatial resolution (denoted as the maximum in table 5.8) was determined analytically (*i.e.* equating the derivative to zero) from curves in figure 5.11(B).

Table 5.8 clearly demonstrates the effect of spatial location and tilting angle on the uncalibrated size (*i.e.* pixel length) of an object in flexible endoscopy. For example, for a constant spatial location in the back periphery the uncalibrated size of the object could increase by 48% if the tilting angle changes from -15° to 15°. Also, for a constant tilting angle of 15° the uncalibrated size of the object could increase by 57.7% if the object moves from the minimum resolution location to the maximum resolution location in the FOV. Finally, we can see the interaction effect of grouping (front, center, back) and the tilting angle. Specifically, at zero tilting angle the uncalibrated size of an object on the front periphery increases by 26.7% if that object moves to the

center of the image-FOV. However, at the tilting angle of 15°, the increase could be as high as 53.8%.

Table 5.8. Estimated uncalibrated length (*i.e.* pixel length) of a 2 mm object at different locations of the FOV and different tilting angles.

Tilting angle		Front	Center	Back	Maximum
	-15°	36	39	25	40
	-10°	34	39	27	40
	-5°	32	38	29	38
	0°	30	38	30	38
	5°	29	38	32	38
	10°	25	36	32	37
	15°	26	40	37	41

## 5.6. Discussions

Imaging techniques are widely employed in clinical practice. The fields of speech-language pathology and laryngology are not an exception. However, the access for direct functional observation of the laryngeal tissues is not trivial, and therefore, the visualization is channeled through an endoscopic instrument. Hence, the functionality and characteristics of the endoscope determine the characteristics of the acquired images. For example, rigid endoscopy is based on transoral insertion, which limits the types of stimuli that can be elicited. Also, the unnatural retraction of the tongue<sup>130</sup> required for adequate laryngeal exposure may alter the voice production system and hence may not reflect the natural function of the phonatory system. For example, research has shown that the presence of a rigid endoscope could significantly change the fundamental frequency and quality of the produced voice<sup>131</sup>, which may support a modified function of the phonatory mechanism during the rigid endoscopy. Flexible endoscopy helps address some of these concerns. Also, flexible endoscopes provide the possibility of simultaneous

aerodynamic measurements.<sup>132–134</sup> This could provide significant information about the complex interactions between kinematics, aerodynamics, and the produced acoustic of the phonatory mechanism. Additionally, coupling a laser-calibrated flexible endoscope<sup>195</sup> to an HSV system and recording synchronized aerodynamic measurements could help us tease apart the effect of individual differences on the phonatory mechanism. Last but not least, flexible scopes have been associated with higher success rates in adult<sup>127</sup> and pediatric<sup>136,137</sup> populations. However, flexible endoscopes are associated with non-linear distortions.

The main aim of this chapter was to quantify the effects of two different sources of non-linear distortions in the images acquired from a fiberoptic flexible endoscope. The first source stems from the wide-angle lens that is used in the flexible endoscope in order to compensate for short working distances and hence maximizing the FOV. The second source of non-linearity stems from changes in the imaging angle. A significant error can be introduced into measurements if these distortions are not compensated for. Two different interpretations of the effects of these distortions are presented here. The first interpretation relates to the usage of uncalibrated measurements (*i.e.* pixel lengths) and quantifies the magnitude of error in comparing pixel length of objects from different locations of the image-FOV. Whereas, the second interpretation relates to calibrated measurements (*i.e.* estimating the *mm* lengths) in the absence of proper compensation methods. This interpretation quantifies the magnitude of error in estimating the *mm* length of objects from different locations of the image-FOV. Experiments 1 and 2 demonstrated the significant effect of spatial location of a pixel on its *mm* size. Based on results of table 5.5 pixels in the periphery could have about 26.4% lower spatial resolution than pixel in the center. This means that if pixel lengths are used for comparing two similar objects one in the center and one the periphery, length of the object in the center will be overestimated by 26.4%. Considering the *mm* measurement, a simplistic

solution could be to compute the average pixel size and then use it for conversion from pixel into *mm*. Based on results of table 5.5 this approach could lead up to 8.1% overestimation of the object in the center and up to 14% underestimation of the object in the periphery. Experiment 3 investigated the effect of tilting angle and showed its significant effect on measurements. Specifically, table 5.8 showed that pixel length of an object in the periphery of the image-FOV could change by 48% if the tilting angle goes from  $-15^{\circ}$  to  $15^{\circ}$ . If the average pixel size (table 5.5, column Mean) is used and the effect of tilting angle is not compensated for, calibrated *mm* measurements could have significant error. Specifically, at the tilting angle of  $15^{\circ}$  the *mm* length on one side of the periphery could be underestimated by 34%.

The focus of this study was on non-linear distortion from a laser-calibrated laryngeal fiberoptic flexible endoscope and their effect on the horizontal measurement. However, the results may provide insights and motivation for further analysis of other types of endoscopes, as well as, other endoscopic procedures (*e.g.* gastroendoscopy, colonoscopy). Specifically, the first non-linear distortion was due to the wide-angle lens of the fiberoptic flexible endoscopes. Considering that distal-chip flexible endoscopes, gastroendoscopy, and colonoscopy also use wide-angle lenses, one may expect to see some residual distortions. However, the exact magnitude of distortion would be different from this study and should be investigated in a separate study. Rigid endoscopes have a narrower angle of view and hence the small-angle approximation may be valid. Therefore, the effect of the first source on non-linear distortion could be minimal in rigid endoscopes. On the other hand, the effect of the imaging angle seems to be universal. Therefore, it is expected for accuracy of measurements from rigid endoscopy, distal-chip flexible endoscopes, gastroendoscopy, and colonoscopy to depend on the imaging angle. However, the exact magnitude of that distortion could be different from fiberoptic endoscopes and should be investigated in a

separate study. To address this need, we are planning to use a similar approach and evaluate the distortions of distal-chip videoendoscopy systems, which would quantify the effect of tilting angle and spatial location on the validity and reliability of horizontal measurements. Considering the popularity and widespread usage of distal chip videoendoscopy systems in clinical settings, such a study is warranted to provide more immediate clinical value.

Implications and findings from this study seem to extend beyond horizontal measurements. For example, in figure 5.4(C) we see that parallel lines exhibit a bowing effect in the captured images. This may indicate that subjective visual assessments of laryngeal images captured from fiberoptic flexible endoscopes for assessment of vocal fold bowing may get biased. Figure 5.11(A) shows that when the imaging angle is not perpendicular, parallel lines may result in divergent lines in the image. This may indicate that vocal folds that are in fact parallel may be captured as divergent ones in laryngeal images (regardless of the imaging modality) if the imaging angle is not perpendicular. Last but not least, the objective and subjective measurements of asymmetry have been used in previous literature.<sup>98,249</sup> However, the investigated non-linear distortions could significantly change the accuracy of those subjective assessments and objective measurements.

## 5.7. Conclusions

This study was motivated by performing calibrated (*i.e.* *mm*) horizontal measurement from a laser-calibrated HSV system. The system was designed based on a fiberoptic flexible endoscope. Two different sources of non-linear distortions in the fiberoptic flexible endoscope were investigated, the wide-angle lens used in flexible endoscopes, and the deviation in the imaging angle. It was shown that the first source of distortion, the wide-angle lens, results in a pixel size (*i.e.* the conversion scale from pixel into *mm*) that depends on the spatial location of that pixel. More precisely, it was shown that if the imaging axis is perpendicular, all pixels with similar

distances to the center of the image-FOV will have similar pixel sizes. Additionally, it was shown that as we move away from the center of the image-FOV the pixel size increases. A different interpretation of this observation would be that the spatial resolution of the image decreases as we move away from the center of the image-FOV toward its periphery. Therefore, keeping the region of interest in the center of the image-FOV would improve the details of the captured image. Studying the second source of non-linear distortion, the effect of imaging angle, showed that it disturbs the radial symmetry of the images. That is, the spatial resolution of points with similar distance to the center of the image-FOV become dissimilar, and also that dissimilarity increases with an increase in the tilting angle. Additionally, this distortion leads to the dislocation of the points with the highest spatial resolutions from the center of the image-FOV. The analysis showed that the combined non-linear distortions could result in calibrated horizontal measurement errors up to 65.7%.

## CHAPTER 6: DIRECT HORIZONTAL CALIBRATION OF HSV RECORDINGS

*Based on:*

Ghasemzadeh H., Deliyski D. D., Hillman R. E., Mehta D. D. Method for Horizontal Calibration of Laser-Projection Transnasal Fiberoptic High-Speed Videoendoscopy, *in Preparation*.

Summary: Calibrated horizontal measurements (e.g., *mm*) from the vibrating vocal folds and the surrounding laryngeal structures during phonation could improve our knowledge of the function of normal and disordered phonatory mechanisms. Additionally, it could be used for direct assessment of therapeutic outcomes, implementation of evidence-based practice, and advancement of personalized medicine in the fields of laryngology and speech-language pathology. However, the size of an object in laryngeal images is not routinely calibrated during endoscopic assessment and depends on a couple of factors, including the distance between the endoscope and the target surface. This chapter used a recently developed *in-vivo* laser-projection fiberoptic endoscope and proposes a method for calibrated spatial measurements. To that end, a set of circular grids were recorded at multiple working distances. A statistical model was trained that would map from pixel length of the object, the working distance, and the spatial location of the target object into its *mm* length. A detailed analysis of the performance of the proposed method is presented. The analyses have shown that the accuracy of the proposed method does not depend on the working distance and length of the target object. The estimated average magnitude of error was 0.27 *mm*, which is three times less than the existing approaches.

## 6.1. Introduction

The length of an object in an image depends on the magnification factor of the camera, which in turn depends on several factors including the distance of the object from the camera. Considering that most often, we do not know the distance of an object from the camera, measuring the calibrated (*i.e.* *mm*) lengths of an object from an image is not a trivial task unless some *auxiliary* information is provided. Providing a conversion scale probably is the most common approach, which is present in every printed map. Another less common approach would be to add an object with a known size (*e.g.* a penny) to the scene before taking the picture. Regardless of the employed approach, calibrated horizontal measurement on an image follows the same steps. Pixel lengths of the target object and the *auxiliary* information (*i.e.* the scale on a map or the penny) are measured. These measurements are then combined with the a-prior knowledge of the *mm* length of the *auxiliary* information, and the pixel-to-*mm* conversion scale is computed. Finally, the pixel-to-*mm* conversion scale can be used for measuring the *mm* length of any object in the image.

In chapter 1 we discussed that horizontal calibration approaches can be classified into direct and indirect methods, depending on the source of the auxiliary information. The *auxiliary* information for the indirect approach comes from a *different* image (or source), whereas the *auxiliary* information for a direct approach comes from the *same* image that we want to make measurements from. This subtle difference has very important consequences regarding the validity of the measurements. Specifically, indirect approaches have the following implicit assumptions. (1) The *auxiliary* information is exactly the same in both images. (2) The *auxiliary* information can be registered accurately on the target image. (3) Both images were captured under similar conditions (*e.g.* imaging angle, a similar vertical distance between the *auxiliary* information and the target surface, etc.). These conditions were discussed in detail in chapter 2.



Let us consider the length of a vocal fold for indirect horizontal calibration. Obviously, the phonation system is a moving mechanism and hence it is changing constantly. For example, the length of the vocal folds can change from one recording to the next one. Also, the larynx could move in the vertical plane (*i.e.* be elevated or depressed) which would change the working distance of the camera. Therefore, elevation-depression of the larynx could change the pixel length of a vocal fold between two recording sessions, even if *mm* lengths of the vocal fold was similar in both imaging sessions. Things get even more complicated during phonation. The relationship between different parameters of the phonatory system (*e.g.* activity of different intrinsic muscles and subglottal pressure) and the acoustic output (*e.g.* pitch, intensity) is very complex.<sup>12</sup> Consequently, using the measured (or the self-perceived) pitch and loudness could not necessarily warrant the assumption of a similar length of the vocal fold between different recordings. Finally, if different recordings are done pre- and post-surgery, then the system was changed between the two conditions. Obviously, using pitch and loudness could be even more problematic in such instances. On the other hand, direct calibration approaches do not have these important implicit assumptions. Thus, their measurements could be more accurate. It is noteworthy, that the improved accuracy is achieved at the expense of higher complexity. Specifically, indirect calibration approaches do not require specialized instruments and can be performed using existing laryngeal imaging systems. Additionally, images could be printed and a simple caliper would be enough for measurements.<sup>93</sup> Conversely, direct approaches rely on specialized and more sophisticated imaging instruments, and often any measurement requires complex calibration and processing steps.

The definition of the direct calibration approach stipulates the existence of some *auxiliary* information on the recorded images. This means that we need to add some properly designed

fiducial markers to the FOV. Two important problems should be addressed in order to realize this requirement. The first problem is to design and create fiducial markers with certain topological properties. The second problem is to deliver the created pattern to the laryngeal mechanism. Reviewing the literature on laryngeal imaging shows that researchers have been working on these problems for more than two decades.<sup>122</sup> Laser source emits spatially coherent light and therefore can be used for creating fiducial patterns with specific topological properties. The created pattern could then be delivered by clipping the laser projection component to the endoscope.<sup>25,153,190,232,236,253</sup> Obviously, this approach increases the insertion diameter of the endoscope which would exacerbate the discomfort level of the patient and hence reduce the success rate of the endoscopy. A more elegant approach is using a surgical endoscope<sup>195</sup> or employing some portions of the illumination fibers of a flexible endoscope.<sup>191</sup>

Two main approaches of parallel laser markers and multiple laser points have been used for creating the laser-fiducial markers in the field of voice.<sup>122</sup> The projection of the parallel laser markers is the simplest approach. Two-point laser projection<sup>25,190,192</sup>, two-line laser projection<sup>234</sup>, and multiple line laser projection<sup>153</sup> are some examples of this category. The multiple-laser-points projection is more sophisticated and involves the projection of many laser points on the FOV.<sup>191,194,195,235</sup> Each method has its own merits. The parallel laser projection category benefits from the simplicity of its optical design and subsequent measurement methodology. Detection of the laser markers on the image is the only required step for measurement in those systems. After that, the distance between the laser markers may be used similarly to a scale on a map and calibrated horizontal measurements may be achieved using a simple caliper. The main assumption of this method is that all pixels in the image have the same *mm* lengths which could be violated if different objects of the image have different distances from the camera, or if different locations of

the image have different pixel size representations.<sup>208</sup> Violation of these assumptions will lead to measurement errors. Conversely, multiple-laser projection systems benefit from the presence of laser points in all parts of the image. Not only this information helps with vertical measurements<sup>122</sup> and 3D reconstruction of the envelope of the FOV<sup>236</sup>, but it also means that with a high probability some laser points would be near to, or on the target surface. Therefore, the above-discussed problems would be resolved. However, systems from this category require more sophisticated optical hardware and processing software design. Horizontal measurements from these systems depend on a calibration step, where the confounding factors of the pixel-to-*mm* conversion scale are determined and accounted for.

## 6.2. Aim and hypothesis

The main aim of this chapter is to develop the methodology of horizontal calibration and subsequent horizontal measurement for a laser-projection transnasal fiberoptic HSV system.

The main research question of this chapter is:

Q5:       How could we use a structured laser projection system for measuring the horizontal distance between two points on a target surface?

In chapter 5 we saw that fiberoptic endoscopes have significant non-linear distortions. Therefore, the following hypothesis was formed for this chapter.

H5a:       Horizontal measurement error from the laser projection system significantly increases, if the nonlinear distortion is not properly compensated for.

Additionally, in chapter 4 we saw that vertical measurement error was positively correlated to working distance. Considering that horizontal measurements depend on the estimation of the working distance, the following hypothesis was formed.

H5b:       Horizontal measurement error will be positively correlated to working distance.

### 6.3. Material and method

The *mm* length of an object is its length perpendicular to the line of sight of the scope. Accurate estimation of the *mm* length of an object from the pixel length of the object's image is the primary goal of calibrated horizontal measurement. Assuming an optical system that is symmetrical around its optical axis, the relationship between an object and its image can be determined. Let  $h_o$  denote the *mm* length of an object and  $h_i$  denote the pixel length of the object's image. Also, let  $O$  be the intersection point between a ray of light from the object and the aperture of the camera (figure 6.1) expressed in the polar coordinates  $(\rho, \varphi)$ . We would have<sup>238</sup>,

$$h_i = A_1 \rho \cos \varphi + A_2 h_o + B_1 \rho^3 \cos \varphi + B_2 \rho^2 h_o (2 + \cos 2\varphi) + (3B_3 + B_4) \rho h_o^2 \cos \varphi + B_5 h_o^3 \quad (6-1)$$

where  $A_j$  and  $B_j$  are constants, and HOT represents the higher-order terms.

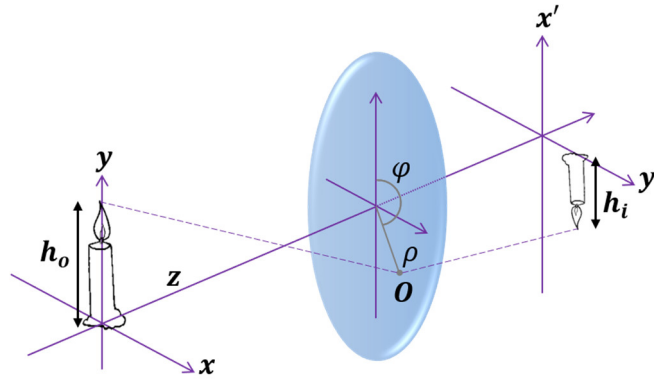


Figure 6.1. Relationship between the length of an object ( $h_o$ ) and its image ( $h_i$ ) in an axially symmetrical optical system.

Equation 6-1 shows a non-linear and complex relationship between pixel and *mm* lengths. Using the thin-lens assumption and small-angle approximation<sup>238</sup> Equation 6-1 can be approximated with a much simpler model known as the Gaussian optics.<sup>238</sup> In this model, the ratio of pixel length to *mm* length is a constant number, which is called the magnification factor of the system ( $m$ ). Equation 6-2 shows this:

$$m = \frac{h_i}{h_o} \quad (6-2)$$

In the Gaussian optics, the magnification factor and the working distance have an inverse relationship<sup>238</sup>, and therefore the working distance would be a confounding factor for calibrated horizontal measurements.

Flexible fiberoptic endoscopes employ wide-angle lenses to maximize their FOV sizes. However, wide-angle lenses violate the small-angle approximation of the Gaussian optics. This leads to a more complex relationship between pixel and *mm* lengths. Specifically, this deviation introduces significant non-linear distortion into recorded images. Distortion of a flexible laryngoscope was studied in chapter 5. We showed that when the imaging axis is perpendicular to the target surface, the distortion is symmetrical around the optical axis, and points with similar distances from the FOV center experience similar distortions. Considering this symmetry, Equation 6-1 may govern the image formation in flexible endoscopy. Additionally, we showed that the pixel length of an object significantly depends on its spatial location within the FOV.<sup>208</sup> Therefore, the spatial location of the target object is another confounding factor for horizontal measurements. Circular grids can exploit this symmetry efficiently; thus, the proposed method uses circular grids to account for the effect of working distance and the spatial location of the target object.

To demonstrate this, a circular grid with a spacing of 0.5 *mm* was recorded at working distances of 2.87 *mm* and 2.24 *mm*. Figure 6.2 shows the recorded images. The circles had a constant distance of 0.5 *mm* from each other. However, in figure 6.2(A) we see as we go from the center toward the periphery the distance between consecutive circles decreases from 30 pixels to 20 pixels. This clearly demonstrates the dependence of horizontal measurements on the spatial location. Comparing figures 6.2(A) and 6.2(B) we see the effect of working distance, where the

distance between the two smallest circles increases from 30 pixels to 35.5 pixels when the working distance decreases from 2.87 *mm* (figures 6.2(A)) to 2.24 *mm* (figures 6.2(B)).

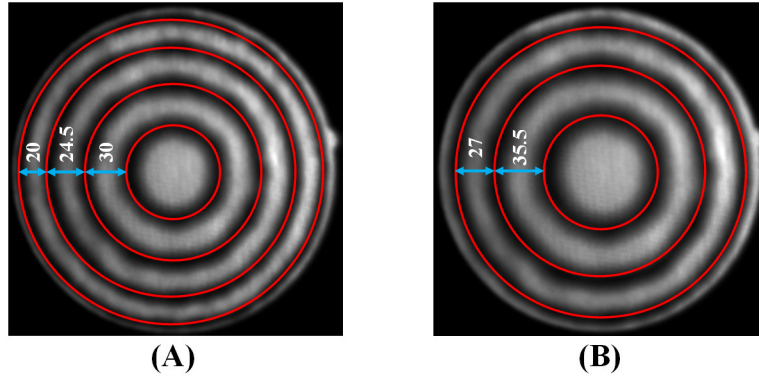


Figure 6.2. Effects of working distance and spatial location on horizontal measurements: (A) working distance of 2.87 *mm*, (B) working distance of 2.24 *mm*.

### 6.3.1. Datasets

The proposed calibration and subsequent horizontal measurement methods were developed and then evaluated based on different sets of benchtop recordings. The setup presented in section 1.4.1, figure 1.2 with only one degree of freedom was used for data collection. That is, the tilting angle was fixed (perpendicular to the imaging axis) and only the working distance was changed.

This study used four different sets of recordings. Set 1 contained 65 recordings from circular grids (figure 6.2) at different working distances. This set was used for training and testing of the model converting a pixel length to its *mm* length. The working distance was gradually increased from 2 *mm* to 32 *mm* and at each working distance, a recording was done. This process was repeated three times to reduce measurement error. For each recording, the grid was adjusted subjectively inside the FOV such that the largest visible circle had a uniform distance from the border of the FOV. Considering the limited spatial resolution, grids became significantly blurry after a certain working distance. Hence, three different circular grids with the spacing of 0.5 *mm*,

1 mm, and 2 mm were used for working distances in the range of [2, 10], [10, 20], [20, 32] mm. The laser source was turned off during these recordings.

The proposed method requires an accurate estimation of the distance between the tip of the endoscope and the target surface (*i.e.*, the working distance). We showed in chapter 4 that a statistical model can be trained to decode the working distance from locations of the laser points.<sup>122</sup> Set 2 had 72 recordings and was used for the training of this model. For this set, the laser source was turned on, and the light source was turned off and recordings were done from a white paper. The working distance was gradually increased from 2 mm to 35 mm and at each working distance, a recording was done. The recording process was repeated four times to reduce measurement error.

The proposed method relies on an accurate estimation of a central angle (*i.e.*, an angle that has its apex on the center of a circle). However, flexible endoscopy images exhibit significant nonlinear distortions.<sup>208</sup> Set 3 was recorded to investigate possible effects of the introduced nonlinear distortion on central angle measurements. This set was based on a custom-designed grid. A circular grid was divided into 24 equal sectors, which created 24 central angles in 15° increments (figure 6.3(A)). The grid was recorded at four working distances of 6.16 mm, 13.20 mm, 19.54 mm, and 26.44 mm. At each recording distance, the grid was adjusted subjectively inside the FOV such that the largest visible circle had a uniform distance from the border of the FOV. This process insured that the center of the grid was at the center of the FOV. This characteristic governs that estimated angles from the image are central angles. The laser source was turned off during these recordings.

Set 4 was recorded for evaluating the accuracy of the proposed method. Line segments with known mm lengths were recorded at fifteen arbitrary locations in the FOV with arbitrary rotations. To provide a comprehensive evaluation, a wide range of lengths and working distances were used.

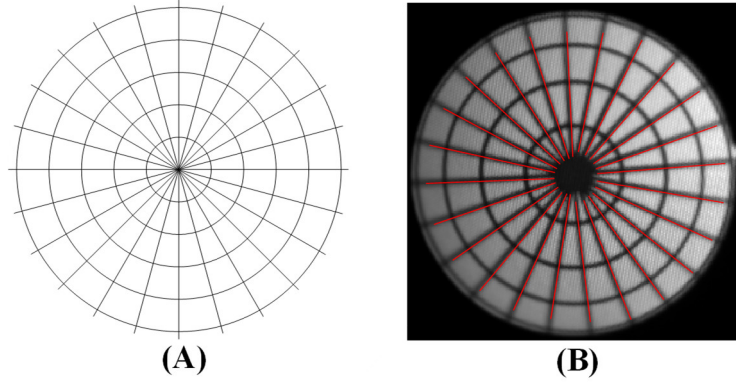


Figure 6.3. The data for evaluation of central angle measurement: (A) the custom-designed grid, (B) segmented radial lines.

Specifically, 5 mm, 10 mm, 15 mm, and 20 mm line segments were recorded at a working distance of 20.18 mm. These recordings were used to investigate the possible effect of object length on the accuracy of the method. Additionally, a 5 mm line segment was recorded at working distances of 5.12 mm, 9.98 mm, 14.98 mm, and 20.18 mm, which covers the common range of administration of fiberoptic laryngeal endoscopy. These recordings were used to investigate the possible effect of working distance on the accuracy of the method. The laser source was turned on during these recordings.

### 6.3.2. Segmentation and preprocessing

Accurate detection of circular grids is a prerequisite of the proposed calibration method. An automatic two-stage method was developed for the segmentation of the circles from Set 1. To take advantage of the full 72-dB dynamic range of the camera, recordings were imported into MATLAB directly in the native 12-bit format from the proprietary Vision-Research .cine files without any conversion or compression. Frames of the recordings were averaged over time and then a Gaussian filter with a size of 2 pixels was applied. The Center and radius of the FOV were estimated using the method described in<sup>122</sup>. A strip parallel to the  $x$ -axis centered at the center of the FOV with a width of 9 pixels was selected. The strip was averaged over the rows, and then



locations of its local minima were detected. Detected locations were paired based on their distances from the center of the FOV. The average of each pair was used as the coarse estimation of the  $x$ -coordinate of centers of circles. Half of the difference between each pair was used as the row-wise estimation of radii of circles. This process was repeated for a strip parallel to the  $y$ -axis averaged over the columns. The average of each pair was used as the coarse estimation of the  $y$ -coordinate of centers of circles. Half of the difference between each pair was used as the column-wise estimation of radii of circles. The final coarse estimation of the radius of each circle was computed as the average of its row- and column-wise radii. A grid search over all combinations of the three estimated parameters  $\pm 1$  pixel with the resolution of 0.25 pixels was used for fine-tuning of the estimated parameters. Specifically, for each case, the target parameters were used to create a ring mask with a width of 1 pixel. The mask was then applied to the gradient of the image, and the summation of the results was used as the cost function. The set of parameters that minimized the cost function was selected as the final estimation of the center and radius of each circle. Figure 6.4 shows the process, with the results on an example image.

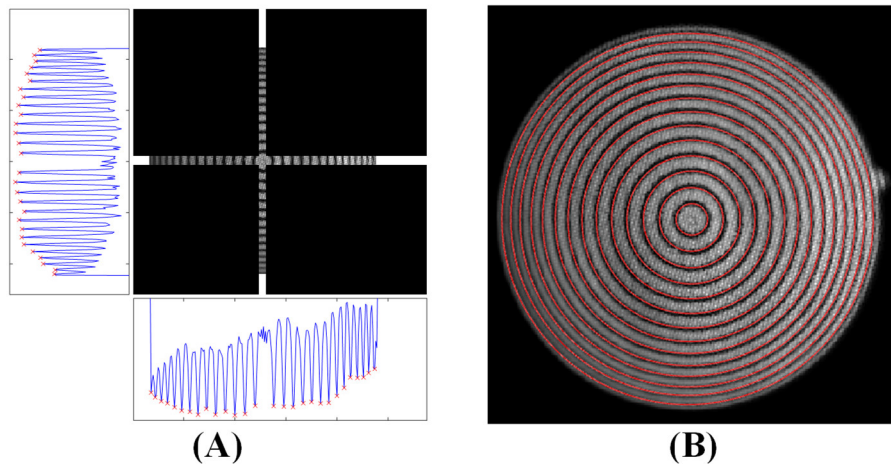


Figure 6.4. Segmentation of a circular grid: (A) horizontal and vertical strips with their respective summations, (B) final segmented circles after the fine-tuning stage.

The segmentation of the laser points for Set 2 was based on the method described in<sup>122</sup>. Target

objects in Set 3 were 24 radial lines, which were detected using the Hough transform<sup>242</sup>. Figure 6.3(B) shows the grid after segmentation. The actual horizontal measurements on laryngeal images will rely on the manual segmentation of the target object. To better reflect this characteristic during the evaluation, a graphical user interface was developed for manual segmentation of line segments from Set 4.

### 6.3.3. Horizontal calibration method

Working distance and spatial location of the target object are the main confounding factors of horizontal measurements. Circular grids provide an effective way for the spatial sampling of the location inside the FOV. This information can be utilized for determining the dependence of horizontal measurements on the spatial location. Additionally, the grids can be recorded at multiple working distances. This information may be utilized for determining the dependence of horizontal measurements on the working distance. To that end, all circles from Set 1 were segmented. This process led to 612 different data points. Let  $w$ ,  $r_p(w)$ , and  $r_{mm}(w)$  denote the working distance, pixel radius, and  $mm$  radius of a circle, respectively, recorded at  $w$   $mm$ . Then, a statistical model can be trained using  $w$  and  $r_p$  as the predictor variables and  $r_{mm}$  as the outcome variable. Let  $\mathcal{F}_{M,N}$  denote a polynomial model in two variables,  $w$  and  $r_p$ , with maximum degrees of  $M$  and  $N$ , respectively. Equations 6-3 and 6-4 show the model, where  $a_{l,k}$  are some constants determined during the training process:

$$r_{mm} = \mathcal{F}_{M,N}(w, r_p) \quad (6-3)$$

$$\mathcal{F}_{M,N}(w, r_p) = \sum_{k=1}^M \sum_{l=1}^N a_{l,k} \cdot w^k \cdot r_p^l \quad (6-4)$$

To select the best model, polynomial models with different degrees were evaluated using 10-fold

cross-validation. The cost function was defined as the mean absolute error (MAE) over all testing samples from all folds. The  $\mathcal{F}_{5,5}$  resulted in the MAE of  $0.025 \text{ mm}$ , which was the lowest value. This model will be referred to as the non-uniform model in the rest of this dissertation. Figure 6.5(A) presents the trained non-uniform model.

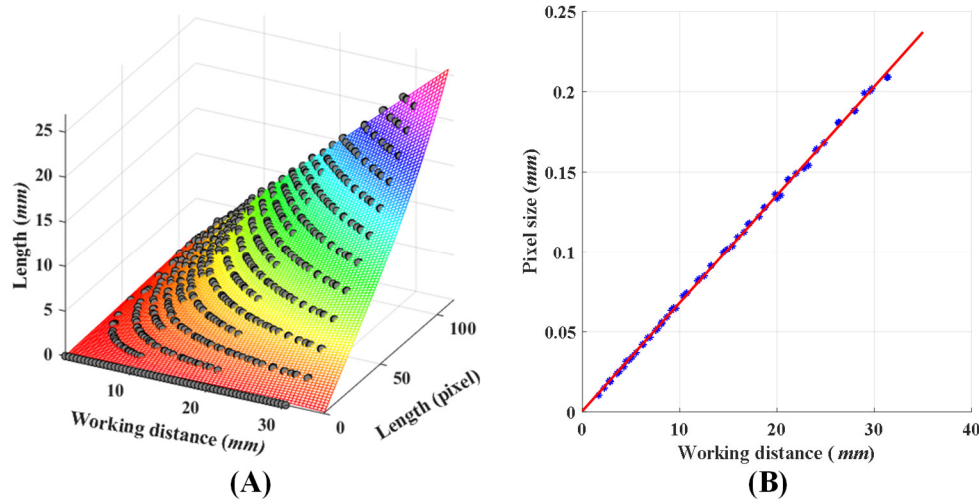


Figure 6.5. Models for horizontal measurements: (A) non-uniform model, (B) uniform model.

To highlight the effect of spatial location on horizontal measurements, and also to test H5a, a second model was trained where all pixels in the FOV had similar pixel sizes. This scenario mimics horizontal measurement from a parallel-laser projection system. Let,  $R_p(w)$  and  $R_{mm}(w)$  denote the pixel radius and  $mm$  radius of the largest circle visible in the FOV recorded at the working distance of  $w \text{ mm}$ . The uniform pixel size  $\gamma(w)$  is defined as

$$\gamma(w) = \frac{R_{mm}(w)}{R_p(w)}. \quad (6-5)$$

The uniform pixel size was computed for all recordings in Set 1. Then, a statistical model was trained using the working distance as the predictor variable and  $\gamma$  as the outcome variable. Investigating the relationship between working distance and  $\gamma$  revealed a linear model. This model is shown in figure 6.5(B) and it will be referred to as the uniform model in the rest of this

dissertation.

#### 6.3.4. Horizontal measurement method

The application of the uniform model is simple and quite similar to the estimation of a distance on a printed map. The pixel size ( $\gamma$ ) allows the conversion from pixel length into *mm* length. Considering the dependence of  $\gamma$  on the working distance, the following steps were followed for horizontal measurements using the uniform model. The working distance was estimated from the positions of the laser points<sup>122</sup>; then the appropriate value of the pixel size was computed from the uniform model. The pixel length of the target object was measured on the image; then the pixel length was multiplied with the multiplicative factor of pixel size to estimate its *mm* length.

The application of the non-uniform model is more involved, and it is described under two categories of radial and general measurements. A radial measurement is defined as the length of an object that has one of its ends on the FOV center. The non-uniform model was trained using circles centered at the FOV center. Therefore, the model can estimate the *mm* radius of a circle centered at the FOV center, which would be equivalent to a radial measurement. Thus, the following steps were followed for horizontal radial measurements using the non-uniform model. The working distance was estimated from the positions of the laser points.<sup>122</sup> Then, the pixel length of the target radial object was measured on the image. The values of working distance and pixel length were fed into the trained non-uniform model (Equation 6-3), and the *mm* length of the object was estimated.

A general measurement needs to be expressed in terms of radial measurements before the application of the non-uniform model. Figure 6.6 shows this process. The main goal is to determine the length of the line segment AB in *mm*. We can construct the triangle AOB on the image, where O is the FOV center. Referring to Figure 6.6, OA and OB each have one of their ends at the FOV

center, and hence they constitute radial measurements, and their *mm* lengths can be computed using the non-uniform model. At the same time, we can measure the angle  $\alpha$  from the image. Let  $A \begin{vmatrix} x_A \\ y_A \end{vmatrix}$  and  $B \begin{vmatrix} x_B \\ y_B \end{vmatrix}$  then, we can determine the angle between OA and the positive *x*-axis ( $\theta_A$ ) as follows,

$$\theta_A = \tan^{-1} \left( \frac{y_A}{x_A} \right) \quad (6-6)$$

where  $\tan^{-1}$  denotes the four-quadrant inverse tangent function. The angle between OB and the positive *x*-axis ( $\theta_B$ ) can also be measured, similarly. Finally, the angle  $\alpha$  can be computed as,

$$\alpha = |\theta_A - \theta_B| \quad (6-7)$$

Now, we can apply the law of cosines for determining the *mm* length of the line segment AB:

$$AB^2 = OB^2 + OA^2 - 2 \cdot OA \cdot OB \cdot \cos(\alpha) \quad (6-8)$$

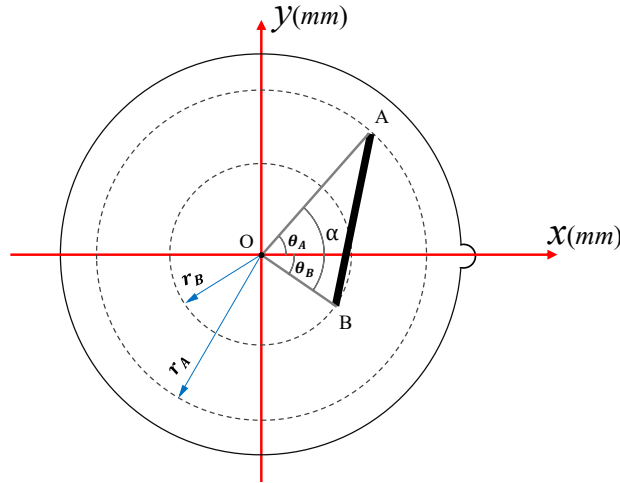


Figure 6.6. Expressing a general measurement in terms of radial measurements

### 6.3.5. Estimation of the working distance

Referring to Equations 6-3 and 6-5, we see that accurate estimation of the working distance is a prerequisite of both uniform and non-uniform methods. The method for estimating the working distance has been presented in chapter 4.<sup>122</sup> The model assumed that the data were mapped into a

standard template by applying a chain of rotation, translation, and scaling operations on the recorded images. The rotation operation was parametrized in terms of the angle between the positive  $x$ -axis and the line connecting the fiducial marker to the FOV center. The rotation operation brings this angle to a fixed and standard value across all recordings.<sup>122</sup> First, we show that the performance of the original method depends on the value of this angle; then we propose an improved version to alleviate this problem.

Ten-fold cross-validation over Set 2 was used to evaluate the effect of different standard angles on the accuracy of estimation of the working distance. To that end, the standard angle of the method presented in in<sup>122</sup> was varied between  $0^\circ$  and  $180^\circ$  in  $5^\circ$  increments, and then the segmented laser points from the training set were used to create the model. It has been shown that laser points from the top row degrade the accuracy of measurements<sup>122</sup>; therefore laser points from the top row were discarded for this analysis. The trained model was then applied to the testing set, and measurement errors from the remaining 42 laser points were computed. MAE over all folds is shown in figure 6.7.

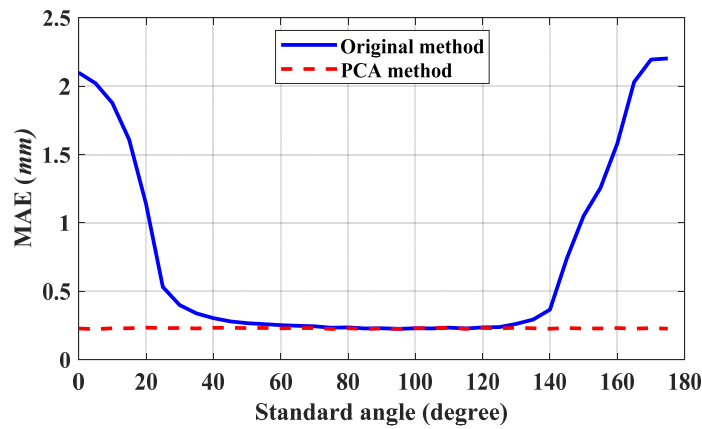


Figure 6.7. Mean absolute error (MAE) of original and the proposed PCA method for different values of the standard angle.

Investigation of figure 6.7 shows that the accuracy of the original method highly depends on the choice of the standard angle. Principal component analysis (PCA) is a mapping that is robust

to linear transformations of the data points, including their rotation. Consequently, we propose a slight improvement over the original method. Let  $p_{ij} \begin{smallmatrix} x_{ij} \\ y_{ij} \end{smallmatrix}$  be the cartesian coordinates of the laser point  $i$  ( $1 \leq i \leq 49$ ) at the working distance  $w_j$ . We can store  $p_{ij}$  for a specific value of  $i$  and all values of  $j$  ( $1 \leq j \leq n$ ) into a  $2 \times n$  data matrix  $P_i$ , where  $n$  is the number of working distances in the dataset. Let  $\mu_{ix}$  and  $\mu_{iy}$  denote the average values of  $P_i$  over the first and the second row, respectively. Now, we can center the data and construct the matrix  $Q_i$ . Equations 6-9 through 6-11 show these definitions.  $\mathbf{1}_n$  is a column vector containing 1 in all of its  $n$  rows.

$$\mu_{ix} = \frac{\sum_{j=1}^n x_{ij}}{n} \quad (6-9)$$

$$\mu_{iy} = \frac{\sum_{j=1}^n y_{ij}}{n} \quad (6-10)$$

$$Q_i = P_i - \begin{bmatrix} \mu_{ix} \\ \mu_{iy} \end{bmatrix} \cdot \mathbf{1}_n^T \quad (6-11)$$

Now the direction capturing most of the variance of the data ( $\mathbf{u}_i$ ) can be computed as,

$$\mathbf{u}_i = \arg \max_{\|\mathbf{u}\|=1} (\mathbf{u}^T \cdot Q_i Q_i^T \cdot \mathbf{u}) \quad (6-12)$$

The first principle component ( $\mathbf{v}_i$ ) would be the projection of the data points on the direction  $\mathbf{u}_i$  and is computed as,

$$\mathbf{v}_i = \mathbf{u}_i^T \cdot Q_i \quad (6-13)$$

Now, the first principal component may be used to train the vertical calibration model. Let  $v_{ij}$  denotes the  $j$  component of the vector  $\mathbf{v}_i$  (*i.e.* projection of the point  $p_{ij}$  in direction  $\mathbf{u}_i$ ).

Equations 6-14 and 6-15 are repeated for each laser point  $i$ .

$$w_j = \mathcal{G}_i(v_{ij}) \quad (6-14)$$

$$\mathcal{G}_i(v_{ij}) = \beta_{i1} e^{\beta_{i2} v_{ij}} + \beta_{i3} e^{\beta_{i4} v_{ij}} \quad (6-15)$$

Ten-fold cross-validation over Set 2 was used to evaluate the effect of different standard

angles on the accuracy of the improved model. The standard angle was varied between  $0^\circ$  and  $180^\circ$  in  $5^\circ$  increments and for each value. The training set was used to estimate  $\mu_{ix}$ ,  $\mu_{iy}$ ,  $\mathbf{u}_i$ , and parameters of the model ( $\beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4}$ ). The trained model was then applied to the testing set and measurement errors were computed. Figure 6.7 shows the computed MAE of the proposed method over all folds. This figure shows the robustness of the improved method to variations in standard angle. Experiment 1 in the next section presents the performance of the proposed improved method in more detail.

## 6.4. Experiments and results

Four experiments were conducted to answer the research questions of this chapter. Experiment 1 presents the performance of the vertical measurement. Experiment 2 quantifies the accuracy of horizontal radial measurements. Experiment 3 presents the performance of central angle estimation from recorded images. Experiment 4 tests the performance of the proposed method for general horizontal measurements. This section presents details of each experiment, followed by results and related discussions.

### 6.4.1. Experiment 1: Accuracy of vertical measurements

The accuracy of the improved vertical measurement model (Equations 6-14 and 6-15) was compared with the original method<sup>122</sup> using 10-fold cross-validation. The original method used the value of  $30^\circ$  for the standard angle. At this angle, the grid becomes a square that is parallel to the  $x$ - $y$  axis, which facilitates the labeling of the laser points. The same standard angle was also used for the proposed improved version. Recordings from Set 2 were split into training and testing sets. Both models were trained using the training set. The trained models were then applied to the testing set and measurement errors were computed. First, the effect of different laser points on the error



was investigated. MAE was computed for each laser point averaged over all working distances. Figure 6.8(B) shows the result. Based on this figure we see different laser points exhibit different performances in the original method, where the top-row laser points produce inferior results. Conversely, all laser points exhibit comparable performances in the improved PCA method. A second analysis was conducted to test the effect of working distance on the accuracy of both methods. For this analysis, laser points from the top row were discarded from the original method and only the remaining 42 laser points were used. However, all 49 laser points were used for the analysis of the improved PCA method. Figure 6.8(C) shows the results. The lines represent the linear model fitted on the individual data points. We can use the slope of regression lines to compare the magnitude of error of different methods with the working distance. Slopes of original and PCA methods were  $0.008 \text{ mm/mm}$  and  $0.001 \text{ mm/mm}$ , respectively. Therefore, we may conclude that the performance of the improved PCA method is less dependent on the working distance.

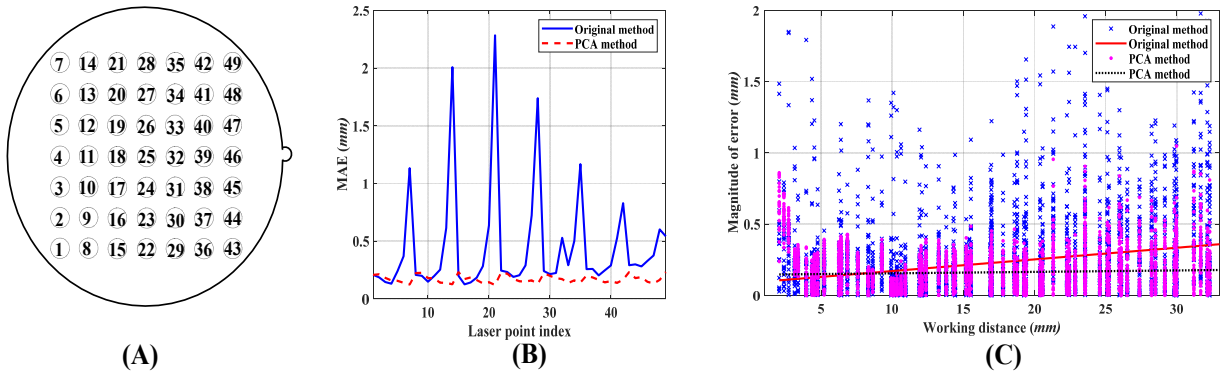


Figure 6.8. Performance of estimating the working distance: (A) indexing of the laser points, (B) measurement accuracy of different laser points, (C) effect of working distance.

#### 6.4.2. Experiment 2: Performance of radial horizontal measurements

The accuracy of the uniform model for radial measurement was evaluated using 10-fold cross-validation. To that end, Set 1 recordings were split into training and testing sets. The uniform

model was trained using the largest enclosed circles of the training set. The trained uniform model was then evaluated for estimating  $mm$  radii of all circles from the testing set, in addition to smaller circles (those that were not used during the training process) of the training set. Figure 6.9 presents scatter plots of absolute errors of all folds versus the radial length of the target circle and the working distance.

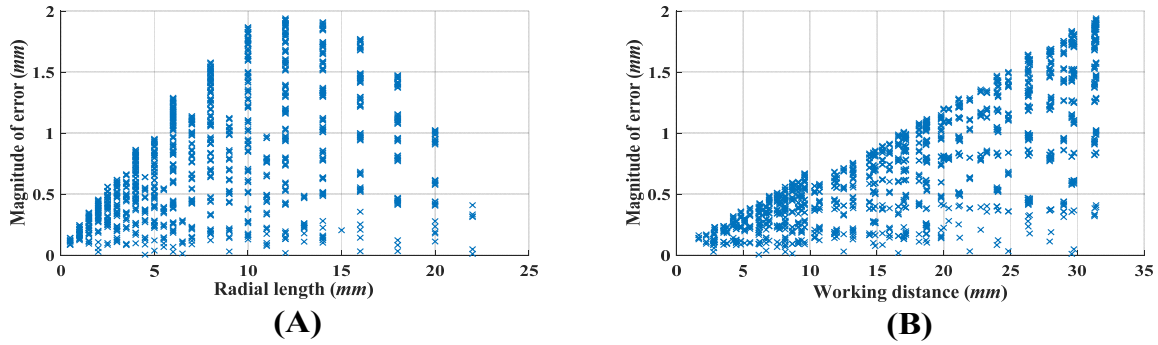


Figure 6.9. Performance of uniform model for radial measurements: (A) effect of object length, (B) effect of working distance.

Investigating scatter plots of figure 6.9 reveals that the measurement error of the uniform model depends on the working distance and the length of the target object. However, the relationship seems to be non-linear. Additionally, our analysis showed that neither of the variables had a normal distribution. Therefore, both parametric and non-parametric tests were used to quantify the effect of working distance and length of the object on the magnitude of the error. Table 6.1 reports the values of Pearson's  $r$ , Kendall's  $\tau$ , and Spearman's  $\rho$ .

Table 6.1. Correlation coefficients of the uniform model for radial measurement error. The symbol  $\varepsilon$  denotes a  $p < 0.0001$ .

Parameter	Pearson's		Kendall's		Spearman's	
	$r$	$p$	$\tau$	$p$	$\rho$	$p$
Radial length	0.59	$\varepsilon$	0.56	$\varepsilon$	0.69	$\varepsilon$
Working distance	0.76	$\varepsilon$	0.57	$\varepsilon$	0.74	$\varepsilon$

Based on Table 6.1, we see a moderate positive correlation between the magnitude of error and

length of the target object and a strong positive correlation between the magnitude of the error and the working distance.

The non-uniform model was trained using the training set, and then its performance for estimating the  $mm$  radii of circles was evaluated using the testing set. Figure 6.10 presents scatter plots of absolute errors of all folds versus the radial length of the target circle and the working distance.

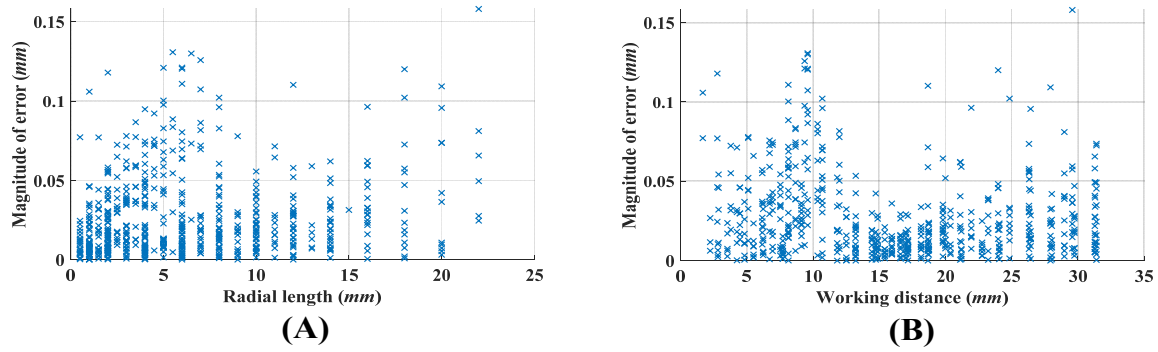


Figure 6.10. Performance of non-uniform model for radial measurements: (A) effect of object length, (B) effect of working distance.

Table 6.2 quantifies the effect of the radial length of the object and working distance on the magnitude of error from the non-uniform model. Based on Table 6.2, we see the magnitude of error has very weak associations with the working distance and length of the target object.

Table 6.2. Correlation coefficients of the non-uniform model for radial measurement error. The symbol  $\epsilon$  denotes a  $p < 0.0001$ .

Parameter	Pearson's		Kendall's		Spearman's	
	$r$	$p$	$\tau$	$p$	$\rho$	$p$
Radial length	0.16	$\epsilon$	0.12	$\epsilon$	0.17	$\epsilon$
Working distance	-0.14	$\epsilon$	-0.08	0.003	-0.13	0.001

Comparing the results of tables 6.1 and 6.2 highlights a primary advantage of the non-uniform method over its uniform counterpart. Specifically, the non-uniform method has a stable and

relatively constant error for a wide range of working distances and target lengths. Additionally, comparing figures 6.9 and 6.10, we see the non-linear method reduces measurement error significantly. To better quantify this, the range of working distance was divided into separate intervals. The average and standard deviation of error and magnitude of error for both uniform and non-uniform methods were calculated in each interval. Table 6.3 presents the results. Based on this table we see another advantage of the non-uniform approach. The average value of error in the non-uniform method is almost zero; therefore, measurement error using the non-uniform approach

Table 6.3. Accuracy of radial measurements from the uniform and the non-uniform models in different ranges of working distance.

Working distance interval (mm)	Non-uniform				Uniform			
	Error (mm)		Magnitude of error (mm)		Error (mm)		Magnitude of error (mm)	
	mean	std	mean	std	mean	std	mean	std
(0, 5)	0.003	0.039	0.029	0.026	-0.192	0.077	0.192	0.075
[5, 10)	-0.012	0.049	0.04	0.031	-0.351	0.151	0.352	0.15
[10, 15)	0.02	0.028	0.025	0.024	-0.489	0.217	0.492	0.21
[15, 20)	-0.005	0.02	0.015	0.015	-0.692	0.303	0.693	0.299
[20, 25)	0.001	0.031	0.022	0.022	-0.955	0.352	0.956	0.347
[25, 30)	-0.001	0.039	0.029	0.026	-1.159	0.476	1.162	0.47

has a random nature. Thus, multiple radial measurements can reduce the error significantly. Conversely, the average error in the uniform approach is not zero, indicating the systematic nature of the error. Finally, the error of the non-uniform method is several orders of magnitude smaller than the uniform approach. This result confirms a recent finding suggesting the presence of significant errors in horizontal measurements if nonlinear distortion of fiberoptic endoscopy is not compensated.<sup>208</sup>

### 6.4.3. Experiment 3: Performance of central angle estimation

Equation 6-8 is at the core of general calibrated measurements using the non-uniform model and relies on the angle  $\alpha$ . Experiment 3 was conducted to investigate the accuracy of the estimation of  $\alpha$  from an image. This experiment is especially important, given the presence of non-linear

distortion in flexible endoscopy.<sup>208</sup> Angle differences between adjacent lines from Set 3 (figure 6.3) were estimated, and then they were subtracted from their true value (*i.e.*  $15^\circ$ ). Figure 6.11 presents boxplots of this error for different working distances. Running a one-way analysis of variance (ANOVA) did not indicate any significant effect of working distance. Therefore, all measurement errors were combined into a single group. The overall angle estimation error had the value of  $-0.03^\circ \pm 0.6^\circ$  (average $\pm$ std). Consequently, central angles can accurately be estimated from acquired images. This result may seem contradictory with a previous finding, suggesting significant errors in the estimation of angles from flexible endoscopes<sup>248</sup>, and hence requires further explanation. The proposed method relies on central angles; however, the work of <sup>248</sup> was based on a general angle. Considering the radial nature of the non-linear distortion, lines passing through the center do not experience bending and curving. Therefore, the central angles can be measured very accurately.

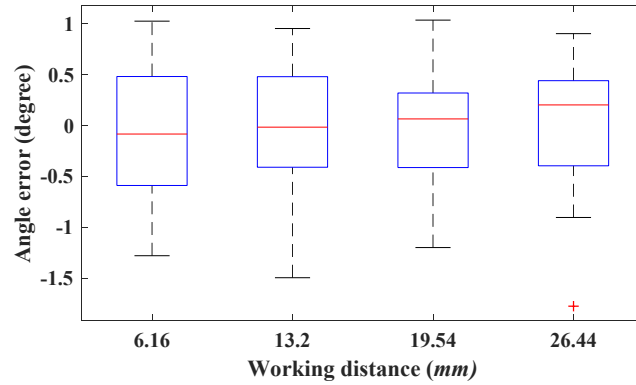


Figure 6.11. Boxplot of angle estimation error computed from set3.

#### 6.4.4. Experiment 4: Performance of general horizontal measurements

Set 4 was used to compare the accuracy of uniform and non-uniform models for general horizontal measurements. Both models were trained with all data points from Set 1. Additionally, Set 4 was recorded in the presence of laser points. Therefore, the required working distance was

estimated using the improved PCA method (Equations 6-14 and 6-15).

To investigate the effect of working distance on general horizontal measurement in the uniform model, measurement errors from a 10 mm line segment recorded at working distances of 6.16 mm, 13.20 mm, 19.54 mm, 26.44 mm were computed. One-way ANOVA with a trimming level of 0.2 and 1000 bootstrap samples<sup>252</sup> was non-significant ( $p=0.61$ ). Figure 6.12(A) presents boxplot of errors for different working distances. To investigate the effect of length of the target object on general horizontal measurement in the uniform model, measurement errors from 5 mm, 10 mm, 15 mm, 20 mm line segments recorded at the working distance of 26.44 mm were computed. One-way ANOVA with a trimming level of 0.2 and 1000 bootstrap samples<sup>252</sup> was non-significant ( $p=0.22$ ). Figure 6.12(B) presents boxplot of errors for target objects with different lengths. Considering these non-significant results, all measurement errors were combined into a single group. The overall measurement error was  $-0.8\pm0.69$  mm, and the magnitude of error was  $0.86\pm0.6$  mm for the uniform method.

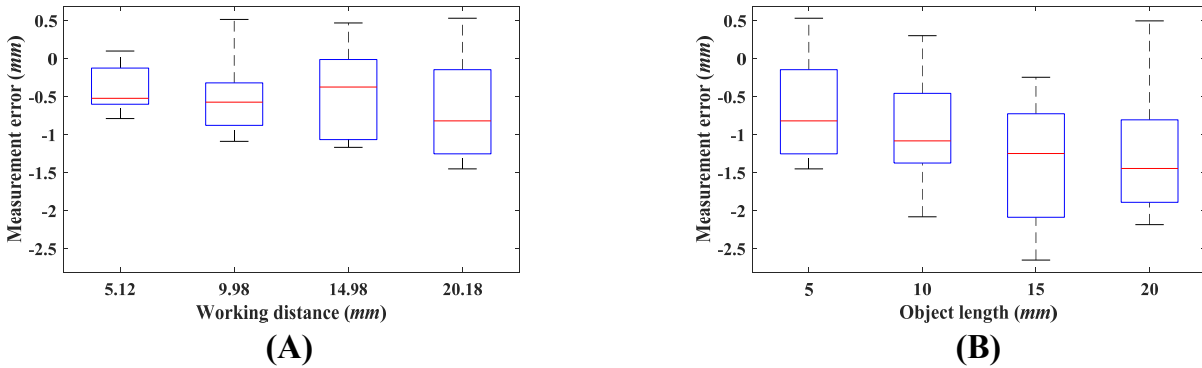


Figure 6.12. Performance of uniform model for general measurements: (A) effect of working distance, (B) effect of object length.

A similar approach was followed for the non-uniform method. Figure 6.13 presents boxplot of errors for this analysis. The effects of working distance ( $p=0.64$ ) and length of the target object ( $p=0.43$ ) were non-significant. Considering these non-significant results, all measurement errors

were combined into a single group. The overall measurement error for the non-uniform method was  $-0.2 \pm 0.29 \text{ mm}$ , and the magnitude of error was  $0.27 \pm 0.24 \text{ mm}$ .

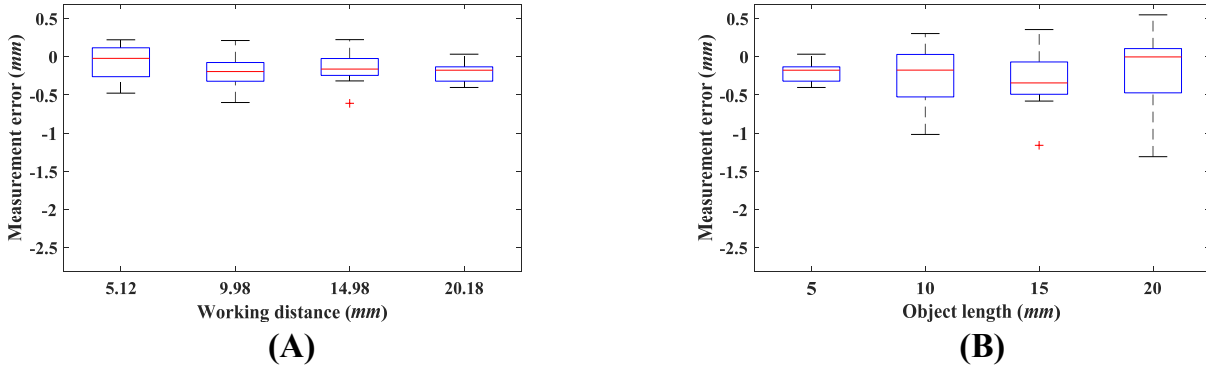


Figure 6.13. Performance of non-uniform model for general measurements: (A) effect of working distance, (B) effect of object length.

Comparing boxplots and average errors of both methods indicate that the uniform approach on average has three times more errors than the non-uniform method. These results demonstrate the advantage of the proposed non-uniform approach. Investigation of boxplots of figure 6.12 may indicate a general trend for errors of the uniform method. Specifically, the measurement error seems to increase with the working distance and length of the target object. In experiment 2 we saw a strong and positive correlation between uniform method error and these two parameters, which confirms this subjective observation. However, the objective analysis of ANOVA failed to detect a significant trend. Experiment 2 relied on the detection of circular shapes. This specific geometry enabled us to achieve sub-pixel resolution on measuring the length of target objects (*i.e.* radii of circles). However, experiment 4 was based on the detection of lines, which has the resolution of a pixel. Investigation of the performance of the non-uniform method also supports this. Specifically, the non-uniform method showed very weak correlations in experiment 2 (Table II). Therefore, we may expect to see a negligible trend for experiment 4, which subjective observation of figure 6.13 confirms.

## 6.5. Discussion

The phonatory mechanism of the larynx is the primary voice production system in humans. It can be modeled as a dynamic system that takes air stream as the input and produces an acoustic signal in the output. The parameters of this dynamic system (*e.g.* vocal fold length, glottal configuration, etc.) determine the relationship between its input and output. If we could measure and determine the input, the output, and the parameters of the system on calibrated scales, we would be able to express and model this dynamic system using mathematical equations. The method for measuring the input and output of this system, in particular for clinical voice assessment, has a long history.<sup>13</sup> The calibrated measurement of parameters of the phonatory system would help in achieving a more comprehensive physical model of voice production. This chapter presented a method that can measure spatial parameters of the phonatory mechanism on a calibrated scale (*i.e.* *mm*). It is expected that prospective horizontal measurements would improve our understanding of the function of normal and disordered phonatory mechanisms. Additionally, it could enable us to derive computational models tuned to each patient and hence make reliable predictions about the likely outcome of different treatment options. This computational approach would advance personalized medicine in the fields of laryngology and speech-language pathology. Last but not the least, calibrated horizontal measurements could allow us to make a direct evaluation of therapy efficacy (*e.g.* post-therapy reduction in the lesion size). The results of such prospective studies would advance evidence-based practice in the field of voice.

This chapter provided the method for horizontal calibration and measurements from a laser-projection transnasal fiberoptic HSV system, followed by a detailed analysis of its performance in different conditions and scenarios. Flexible endoscopy images have significant non-linear distortions, which leads to the dependence of the pixel length of an object on its spatial location.<sup>208</sup>



Chapter 5 established the radial symmetry of this distortion<sup>208</sup>; hence, the proposed calibration protocol was based on circular grids. The proposed non-uniform method has the potency of capturing and quantifying the effects of both working distance and spatial location simultaneously. To demonstrate the efficacy of the proposed method, its performance was contrasted with a uniform approach, which assumed the independence of the pixel size of an image from its spatial location. Such uniform model is the basis of most existing methods for horizontal measurements, including all parallel laser projection systems.<sup>122</sup>

The conducted experiments revealed several significant advantages for the non-uniform approach over its alternative uniform counterpart. Specifically, the analysis of figures 6.9 and 6.10 showed that the accuracy of radial measurements (experiment 2) using the non-uniform method was less dependent on the length of the target object and the working distance. For example, based on table 6.3 we see the average magnitude of error in the non-uniform case does not change significantly when working distance increases from 5 *mm* to 30 *mm*. However, the average magnitude of error in the uniform case shows an increase of 600%. The average $\pm$ std magnitude of error in uniform approach over the range of tested working distance was  $0.68\pm0.45$  *mm*. The average $\pm$ std magnitude of error in the non-uniform approach over a similar range of working distance was  $0.03\pm0.03$  *mm* which further highlights the advantage of the proposed non-uniform method.

Evaluation of both methods in general measurement scenario (experiment 4) showed trends similar to radial measurements. Specifically, figure 6.12 indicates that the accuracy of the uniform approach degrades with an increase in the length of the target object, whereas figure 6.13 does not show any trends for the non-uniform approach. When the length of the target object increases, it spans a wider spatial location in the FOV. Considering that non-linear distortion of flexible

endoscopy is spatially-dependent<sup>208</sup>, this may translate into a larger distortion of the final image. Therefore, we may expect to see a length-dependent error for the uniform approach. It is noteworthy that this dependence did not reach the significance level, which could be attributed to the small sample size and low spatial resolution of images. Average $\pm$ std magnitude of errors in the general measurement scenario resulted in  $0.27\pm0.24$  mm for non-uniform and  $0.86\pm0.6$  mm for uniform method, which shows an improvement of 318% for the non-uniform method.

## 6.6. Conclusion

This chapter was motivated by the importance of performing calibrated (*i.e.* mm) spatial measurements of the vocal folds and the surrounding laryngeal structures during phonation. Such measurements would improve our understanding of the normal and disordered phonatory mechanisms and enable us to derive more accurate computational models. It is expected that evidence-based practice and personalized medicine would benefit significantly from this line of research. However, the size of a target object in laryngeal images may depend on confounding factors, which prevents calibrated spatial measurements. This chapter investigated the effects of two confounding factors, namely the working distance and the spatial location of the target object. To that end, a set of circular grids were recorded at multiple working distances. These grids provided an efficient way of quantifying the effect of both factors. The information from these recordings was then used to train a statistical model that would take the spatial location and the working distance of the target object as the input, and estimate the calibrated length of the target object as the output. A laser projection fiberoptic endoscope was used to estimate the working distance from the positions of the laser points. The performance of the proposed method was investigated in different scenarios. The method was also compared with a uniform model approach, where the effect of spatial location is not considered. The overall measurement error from the

proposed method was  $-0.2 \pm 0.29$  *mm*, and the magnitude of error was  $0.27 \pm 0.24$  *mm*. These errors were more than three times lower than the uniform model approach.

## CHAPTER 7: VALIDITY AND ACCURACY OF HORIZONTAL AND VERTICAL MEASUREMENTS BASED ON DIRECT CALIBRATION

*Based on:*

Ghasemzadeh H., Deliyiski D. D., et al. External validity of calibrated vertical and horizontal measurements from a laser-projection fiberoptic transnasal endoscope, *in Preparation*.

Summary: Methods using laser-projection endoscopes allow for calibrated surface measurements. The design and evaluation of these methods are typically done in controlled settings, using benchtop recordings. However, many factors could be contributing to measurement errors from in-vivo images. This chapter investigates the effect of two such factors: imaging angle and surface topology. A laser-projection fiberoptic flexible endoscope was calibrated using benchtop recordings from flat surfaces (*i.e.* paper), perpendicular to the optical axis. Two experiments were conducted to evaluate its performance in situations modelling the in-vivo settings. (1) Images were acquired from tilted surfaces. (2) A target surface with known  $x$ -,  $y$ -,  $z$ -coordinates was 3D-printed, and its measurements accuracies were contrasted with the flat surface. The data analysis showed a significant effect of imaging angle on vertical measurement error. However, the effect of imaging angle on the magnitude of horizontal measurement error was not significant. Analysis of the effect of surface topology showed the reverse effects. The effect of surface type on vertical measurement error was not significant. But the magnitude of horizontal measurements errors from the 3D surface was significantly higher than the flat surface. The mean percent magnitude of horizontal measurement error increased from 5% (flat) to 10.6% (3D) at the working distance of 15 mm, which still represents satisfactory accuracy.

## 7.1. Introduction

Imaging techniques are an important part of the functional assessment of voice and diagnosis of voice disorders.<sup>167–171</sup> Previous studies have suggested that vibratory characteristics and kinematic measures from laryngeal images can be used for direct evaluation of treatment outcomes.<sup>35,254</sup> These applications would benefit significantly from the ability of performing calibrated spatial measurements from the acquired images. Chapters 4 and 6 of this dissertation presented the methods for calibrated vertical and horizontal measurements from a laser-projection fiberoptic HSV system. However, the methods were developed based on benchtop images, recorded in a very controlled setting. Specifically, images were acquired from flat surfaces perpendicular to the optical axis of the endoscope. Considering that the in-vivo environment is uncontrolled and with many variable factors, the performance of the proposed system may degrade significantly. The main aim of this chapter is to investigate how performances of the proposed vertical and horizontal measurements change as we move from the simple and controlled settings to more complex cases. To achieve this, the performance of the system is evaluated in two scenarios. First, vertical and horizontal measurement errors are evaluated on flat surfaces that are not perpendicular to the optical axis. This analysis will quantify the effect of imaging angle on the accuracy of horizontal and vertical measurements. Second, vertical and horizontal measurement errors are evaluated on a 3D surface. This analysis will quantify accuracy of horizontal and vertical measurements on non-flat surfaces.

In order to study the effect of variations in the imaging angle, first, we need to know the typical range of variations of this parameter. Reviewing the literature indicated that no rigorous study on the normative variations of the imaging angle during VSB or HSV has been done. However, a crude estimation could be made based on two different studies.<sup>204,255</sup> A single-subject

study with a flexible endoscope indicated that variations up to  $30^\circ$  in the imaging angle could be expected.<sup>204</sup> The second study was on differences in motions of a laryngoscope during endotracheal intubations between an expert, an intermediate-skilled practitioner, and a novice.<sup>255</sup> This study found  $10^\circ$  variations in the angle of the laryngoscope during the time that practitioners were trying to hold the view constant for placing the tube. Considering the very low sample size of the first study ( $n=1$ )<sup>204</sup>, and significant differences between the two laryngeal procedure (*i.e.* endotracheal intubation vs. laryngeal endoscopy) in the second study<sup>255</sup>, no clear conclusion can be made about the range of variability in the imaging angle. One possibility for this gap could be the lack of quantitative evaluation of the effect of imaging angle on objective measurements and subjective visual assessment of laryngeal images, which we tried to answer in chapter 5. It is noteworthy that, the topic of imaging angle in general, has found little attention. Hibi and colleagues were probably among the first people that investigated the effect of imaging angle on the endoscopic images.<sup>204</sup> They showed that distortions of endoscopic images significantly increase with an increase in the deviation of the imaging axis from the perpendicular angle.<sup>204</sup> Distortions as high as 20% were reported for a  $30^\circ$  deviation in the imaging angle. The other result is from a very recent work that used synthetic vocal folds to investigate the effect of different parameters of HSV recordings on the accuracy of the estimated subglottal air pressure and the cricothyroid activation from the glottal area waveform.<sup>205</sup> This work suggested that the imaging angle was the most influential factor in the estimation of parameters of the model. Based on this study, a  $10^\circ$  changes in the imaging angle could lead to a 10% error in the estimation of subglottal air pressure from the glottal area waveform.<sup>205</sup>

Considering the relevance and importance of horizontal and vertical measurements for the field of voice, it is expected for the quantitative results of this chapter to provide significant insights into the accuracy and reliability of the proposed measurement methods.

## 7.2. Aim and hypothesis

The main aims of this chapter are to investigate the effect of imaging angle on vertical and horizontal measurement errors from the laser-calibrated endoscope and to investigate the effect of non-flat surfaces on vertical and horizontal measurement errors from the laser-calibrated endoscope. The main research questions of this chapter are:

Q6a: How the imaging angle affects the performance of the vertical and horizontal measurements?

Q6b: How the topology of a 3D surface affects the vertical and horizontal measurements?

To answer these research questions four hypotheses were formed that are presented here.

The vertical distance of each individual laser point to the camera could be estimated using the method developed in chapter 4. In that regard, the method for vertical measurement may not have a direct dependency on the tilting angle of the target surface. However, changes in the imaging geometry would likely lead to changes in the shape of the laser points. Our initial visual observations have suggested that shapes of laser points change from circles to ellipses when the target surface is tilted. Considering that vertical distances were measured based on the position of the circular estimation of the laser points, it is very likely for the vertical measurement accuracy to be affected by the imaging angle too. Based on this rationale, it is hypothesized that,

H6a: The tilting angle of the target surface and the working distance will be good predictors of the vertical measurement error.

Chapter 5 investigated the effect of imaging angle on the distortion of a fiberoptic flexible endoscope. We showed that variations in the imaging angle could have a significant effect on non-calibrated horizontal measurements. Figure 7.1 shows a schematic of the imaging system when the target surface is tilted. Specifically, the points on one side of the surface would get closer to the camera (*e.g.* object B), whereas the points on the other side of the surface would get further away from the camera (*e.g.* object A).

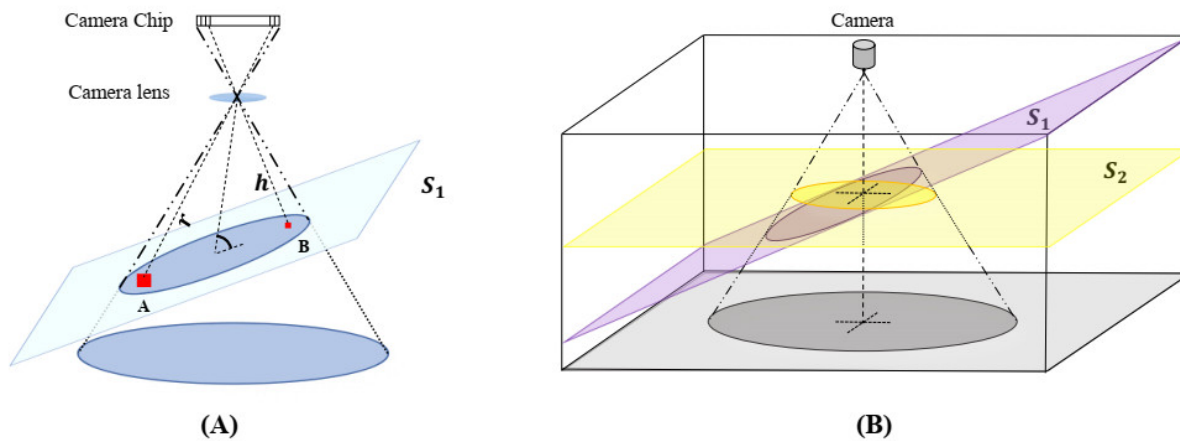


Figure 7.1. Imaging from a tilted surface: (A) effect of tilting the target surface on different objects within the FOV, (B) effect of tilting the target surface on the geometry of the FOV.

The uniform and non-uniform pixel-to-*mm* conversion scales were the basis of calibrated horizontal measurement from the laser-calibrated endoscope (chapter 6). Those models were developed based on a perpendicular imaging angle assumption and its corresponding geometry. However, that assumption does not reflect the geometry of the imaging system when the surface is tilted (figure 7.1(B)). Therefore, it is expected for this discrepancy to manifest as measurement errors. Combining the likely effect of imaging angle and the working distance on the measurement error, it is hypothesized that,

H6b: The tilting angle of the target surface and the working distance will be good predictors of the horizontal measurement error.



The topology of the target surface could also be a major contributor to measurement errors. Therefore, it is hypothesized that:

- H6c: The vertical measurement errors from a non-flat surface will be higher than those from a flat surface positioned at the same estimated average vertical distance.
- H6d: The horizontal measurement errors from a non-flat surface will be higher than a flat surface positioned at the same estimated average vertical distance.

### **7.3. Material and method**

To investigate the effect of imaging angle and the 3D shape of the target surface on horizontal and vertical measurement errors, different sets of benchtop recordings were collected. Considering the significant differences between the imaging angle and the 3D shape, the protocols for their data collection, and the methods for their data analysis were different and hence are described in two different sections.

#### **7.3.1. Material and method for the effect of the imaging angle**

##### **7.3.1.1. Data acquisition**

Different sets of benchtop recordings were used to pursue the aims of this chapter. The datasets were recorded using the same setup presented in section 1.4.1, figure 1.2. However, it had a major difference with the recordings of chapters 4 and 6. Recordings of those chapters were carried out using a setup with only one degree of freedom. That is, only the working distance was varied, but the tilting angle was fixed (perpendicular to the imaging axis). However, the setup for this chapter used two degrees of freedom.

First, we show that tilting the target surface and changing the imaging angle have comparable effects on the geometry of the imaging system. Referring to figure 7.2, we see two conditions. In the first condition, the target surface is fixed (surface  $S_1$ ) and the camera is rotated. In this case, the camera is perpendicular to  $S_1$  (position B), however, when the camera rotates by  $\theta^\circ$  (position A) it becomes perpendicular to a different surface ( $S_2$ ) that is, in fact, the rotation of  $S_1$  by the same  $\theta^\circ$ . Considering that in our setup (section 1.4.1, figure 1.2) the tilting angle of the target surface can be adjusted more conveniently and also more accurately, the target surface was tilted in this chapter.

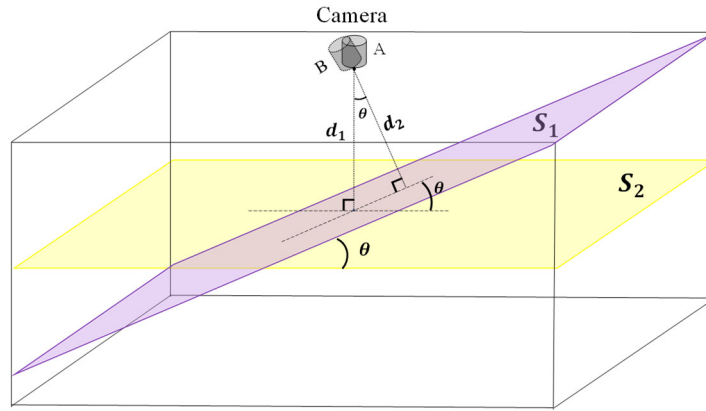


Figure 7.2. The effect of tilting the target surface vs. changing the imaging angle.

### 7.3.1.2. Database

Flexible endoscopes are equipped with a control handle that can bend its distal end. This feature enables the operator to change the FOV during the endoscopy. However, this feature means that the imaging angle of the system (at least) relies on the position of this handle. In other words, when the handle is at rest the optical axis may not be perpendicular to the target surface. To make reliable and accurate predictions about the likely effect of the imaging angle, this factor should be accounted for. In chapter 5 we showed that when the optical axis is perpendicular to the target

surface, the endoscope has radial symmetry. Furthermore, we saw that when the imaging angle deviates, the radial symmetry is disturbed significantly. Therefore, we may use the circular grid from chapter 6 to make the optical axis perpendicular. Figure 7.3(A and B) shows recordings from a circular grid at similar working distances, but opposite directions of tilting angle. Based on these images, we see when there is a tilting angle, the center of the circles moves away from the center of the image, and also the circles become ellipses. Additionally, as we predicted in chapter 5, the direction of the movement depends on the direction of the tilting angle. This observation could be utilized for achieving a perpendicular imaging angle.

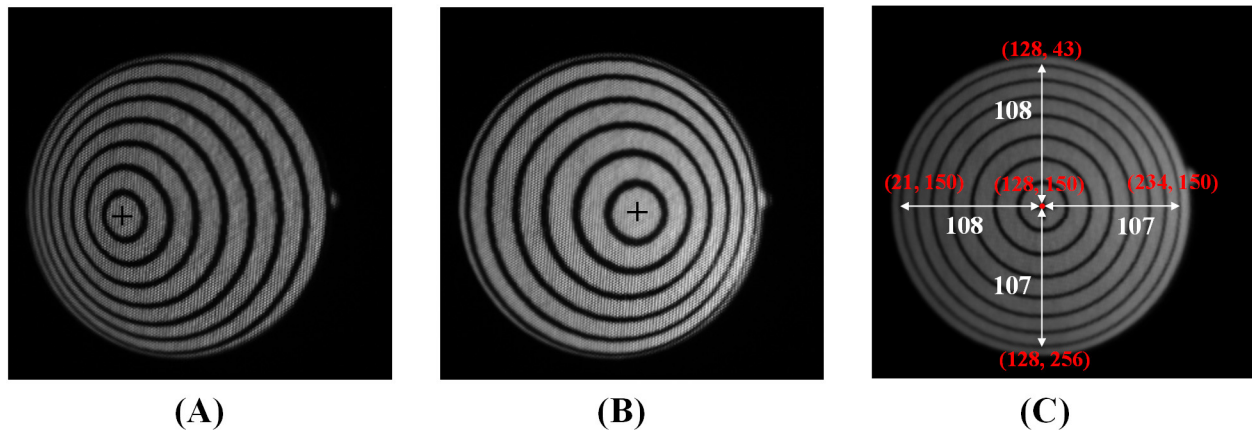


Figure 7.3. Recordings from a circular grid at the working distance of 8.66 mm: (A) the tilting angle of  $15^\circ$ , (B) the tilting angle of  $-15^\circ$ , (C) tilting angle of  $0^\circ$  after making the endoscopic tip perpendicular to the target surface.

The following procedure was followed to make the optical axis perpendicular. The target surface was leveled using a leveler. The coordinates of the center of the FOV was computed using the method described in section 4.3.3.1.2. A circular grid was attached to the metal sheet of the setup (figure 1.2), and then it was adjusted subjectively inside the FOV such that the largest visible circle had a uniform distance from the border of the FOV. The distal tip of the endoscope was passed through a mechanism that allowed its displacement in the left-right and front-back directions (figure 7.4). The distal tip was displaced until the center of the circular grid (the + mark

in figure 7.3) coincided with the center of the FOV. At this point, the optical axis of the endoscope would be perpendicular to the target surface. This position was fixed by tightening the fixtures in the displacement mechanism (figure 7.4). Another way for checking this would be to measure the radius of a certain circle in the four directions and make them as close as possible (*i.e.* the recorded image is a circle). Figure 7.3(C) shows an example of this.

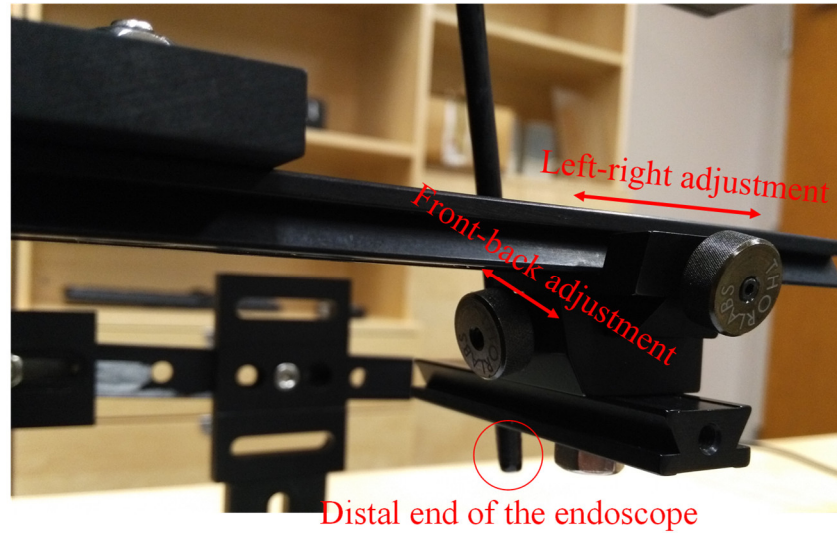


Figure 7.4. The setup that allowed precise adjustment of the distal tip of the endoscope.

#### 7.3.1.2.1. Database for vertical measurements

To test the effect of tilting angle and working distance on vertical measurement error, locations of the laser points at different working distances and imaging angles were recorded. The working distance was changed from 5 mm to 35 mm in 5-mm increments. The working distance was measured using a digital height gauge with an accuracy of 0.001" (approximately 0.03 mm). Additionally, five different tilting angles of 0° to 10° in 2.5°-increments were tried. The method for measurement of the tilting angle was described in section 5.4.2 and figure 5.3. In summary,  $7 \times 5 = 35$  different recording conditions were tested for this experiment. Figure 7.5 shows a schematic of different recording conditions. It is noteworthy that it is hard to adjust the setup for

achieving the exact target working distances and tilting angles; therefore, the actual values deviated from the target values. However, in the rest of this chapter, each condition will be referenced using its attempted values.

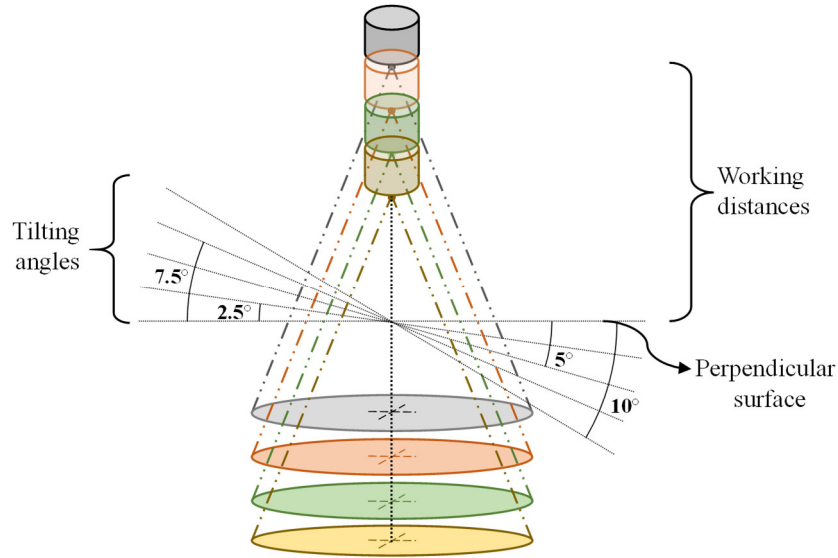


Figure 7.5. A diagram of the recording conditions. Different colors correspond to the FOV cone at different working distances. To simplify the visualization, the target surface is kept fixed and the camera is displaced. However, in the experiments it was the other way around.

We will see in the next section that estimation of vertical measurement error, depends on accurate measurement of the *mm* distance between two arbitrary points inside the FOV. Therefore, the following protocol was followed for the data collection. The setup was adjusted to a desired working distance and imaging angle. A white piece of paper was attached to the metal sheet, the laser source was turned on, the light source was turned off, and a recording was done. Then, a grid paper with a known *mm* spacing was attached parallel to the edges of the metal sheet, the laser source was turned off, the light source was turned on, and a recording was done. The reason for performing two separate recordings was as follows. The grid lines were printed in black, and they absorbed the green light of any laser points falling on them. This would introduce errors in the detection of the center of the laser points. However, having separate laser and grid recordings

would allow a more accurate segmentation outcome (*i.e.* centers of laser points from a laser recording, and equations of the grid lines from a grid recording). Then, we can combine segmented information and create a composite image for performing the analysis.

Our preliminary analysis indicated a likely effect for the rotation of endoscopic eyepiece inside the camera lens coupler. Such rotation can be quantified in terms of the fiducial angle ( $\alpha$  in figure 7.6). Therefore, the whole recording protocol was repeated for three different fiducial angles of  $32^\circ$ ,  $124^\circ$ , and  $309^\circ$ . In summary, the database for this experiment had  $7 \times 5 \times 2 \times 3 = 210$  different recordings.

#### **7.3.1.2.2. Database for horizontal measurements**

To test the effect of tilting angle and working distance on horizontal measurement error, a 5-*mm* line segment was recorded at different working distances and imaging angles. The working distance was changed from 5 *mm* to 35 *mm* in 5-*mm* increments. The working distance was measured using a digital height gauge with an accuracy of 0.001" (approximately 0.03 *mm*). Additionally, five different tilting angles of  $0^\circ$  to  $10^\circ$  in  $2.5^\circ$ -increments were tried. The method for measurement of the tilting angle is described in section 5.4.2 and figure 5.3. In summary,  $7 \times 5 = 35$  different recording conditions were tested for this experiment. Figure 7.5 shows a schematic of different recording conditions. It is noteworthy that it is hard to adjust the setup for achieving the exact target working distances and tilting angles; therefore, the actual values deviated from the target values. However, in the rest of this chapter, each condition will be referenced using its attempted values.

We showed in chapter 5 that fiberoptic flexible endoscopes have significant non-linear distortions. Additionally, in chapter 6 we showed that the spatial location of the target object is a confounding factor for calibrated horizontal measurements. Therefore, the 5-*mm* line segment was

positioned at eight different locations inside the FOV, per each recording condition. Considering, the systematic effect of spatial location on measurements (chapters 5 and 6), locations of the line segment inside the FOV was controlled. Specifically, the radius of the FOV and its center were estimated using the method described in section 4.3.3.1.2. Then the diameter of the FOV parallel to the  $x$ -axis was drawn (line AB in figure 7.6). Then, AB was divided into 5-equal spaced partitions (dashed gray line in figure 7.6). This process led to four lines inside the FOV that were parallel to the  $y$ -axis. The 5- $mm$  line segment was subjectively positioned on these four lines, such that the line AB was its perpendicular bisector. A similar process was repeated for the line AB parallel to the  $y$ -axis.

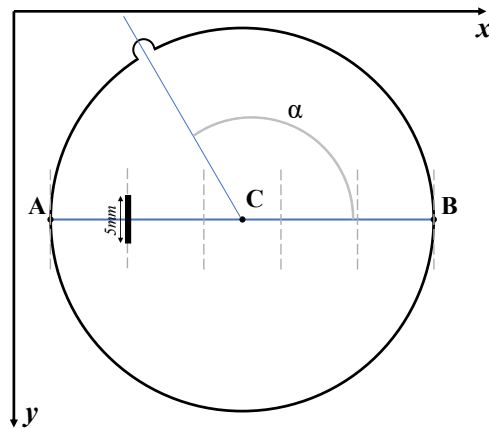


Figure 7.6. Placement of the 5- $mm$  line segment inside the FOV for horizontal measurements.

Similar to the previous section, the line segment was absorbing the green light of any laser points falling on it. This would introduce errors in the detection of the center of those laser points. Consequently, two different types of recordings were collected. A white piece of paper was attached to the metal sheet (figure 1.2), the laser source was turned on, the light source was turned off, and a recording was done. Then, the laser source was turned off, the light source was turned on, and then the 5- $mm$  line segment was placed at eight pre-determined spatial locations and it was

recorded. Consequently, there were 9 different recordings from each condition. The laser recording and the line segment recordings were segmented separately, then they were combined to create a set of composite images for the analysis.

Our preliminary analysis indicated a likely effect for the rotation of endoscopic eyepiece inside the camera lens coupler. Such rotation can be quantified in terms of the fiducial angle ( $\alpha$  in figure 7.6). Therefore, the whole recording protocol was repeated for three different fiducial angles of  $31.6^\circ$ ,  $124.1^\circ$ , and  $309.6^\circ$ . In summary, the database for this experiment had  $7 \times 5 \times 9 \times 3 = 945$  different recordings.

### 7.3.1.3. Analysis and measurements from a tilted surface

#### 7.3.1.3.1. Vertical measurements from a tilted surface

The estimation of vertical measurement error depends on the knowledge of the true vertical distance of each laser point. When the target surface is flat and perpendicular to the optical axis of the endoscope all laser points would have similar vertical distances. However, when the target surface is tilted, laser points would have dissimilar vertical distances. Figure 7.7 shows a schematic of the problem.

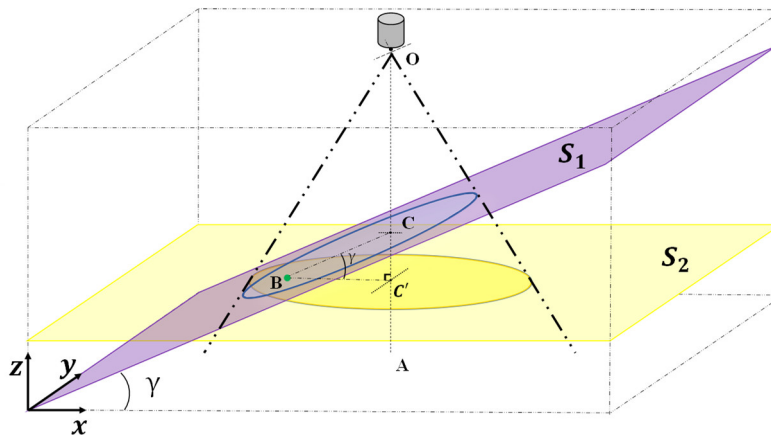


Figure 7.7. A schematic for estimation of the true vertical distance of the laser point B.



Let point O denotes the distal end of the endoscope and point B a laser point laying on the surface  $S_1$ , where  $S_1$  has a tilting angle of  $\gamma$  degree. We can pass the hypothetical surface  $S_2$  from point B, perpendicular to the optical axis of the endoscope (OA). The intersection of  $S_2$  with the optical axis (OA) is marked with  $C'$ . The *mm* length of  $OC'$  is defined as the vertical distance of the laser point B and is the aim of vertical measurement. The estimation of  $OC'$  can be done as follows. The length OC is the working distance and is known in *mm* from the recording condition. Assuming the availability of BC in *mm*, we can write,

$$CC' = BC \cdot \sin(\gamma) \quad (7-1)$$

$$OC' = OC + CC' = OC + BC \cdot \sin(\gamma) \quad (7-2)$$

It is noteworthy that in Equation 7-2 the length BC could be either positive (if point B has a larger vertical distance than point C), or negative (if point B has a smaller vertical distance than point C). Based on Equation 7-2, the *mm* length of BC is the only unknown factor that needs to be determined. We could use recordings from calibrated grids for measuring the *mm* distance between any two points inside the FOV. However, in chapter 5 we saw that the fiberoptic flexible endoscopes have significant non-linear distortion. Additionally, we showed that when the optical axis is not perpendicular to the target surface, point C would move away from the center of the FOV in the image (refer to figure 7.3 for an example). Consequently, determining the location of point C, which is the prerequisite of computing the *mm* length of BC, is not trivial and could be subject to significant error. To remedy this, a modified approach was taken. Let R denotes an arbitrary fixed laser point called the reference point. Now at each tilting angle  $\gamma$ , the true vertical difference ( $\Delta z$ ) between points R and B can be computed as,

$$\Delta z = BR \cdot \sin(\gamma) \quad (7-3)$$

where  $BR$  is the *mm* distance between points B and R in the direction of the tilt. Now, we could use recordings from calibrated grids for measuring  $BR$  in *mm*. Assuming a tilt in the  $x$ -direction, the process was as follows. Let  $N$  denotes the number of complete grids in the  $x$ -direction between points B and R ( $N=4$  in figure 7.8). The analytic equation of each grid line was determined during the segmentation process using the method described in section 5.4.3. Then, the pixel distance between the two  $y$ -direction lines enclosing the point R (length  $L_R$  in figure 7.8) was determined using the equation of the lines and in sub-pixel resolution. Additionally, the pixel distance between point R and the nearest  $y$ -direction line residing between points B and R (length  $l_R$  in figure 7.8) was determined using the equation of the line and in sub-pixel resolution. Values of  $L_B$  and  $l_B$  were computed, similarly, for point B. If  $\delta$  is the *mm* spacing of the grid lines, then  $BR$  can be computed in *mm* as,

$$BR = \left( N + \frac{l_R}{L_R} + \frac{l_B}{L_B} \right) \cdot \delta \quad (7-4)$$

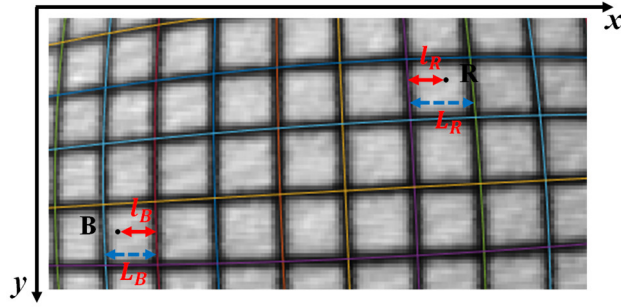


Figure 7.8. An example of computing the *mm* distance between two laser points B and R.

Now, the effects of working distance and tilting angle on vertical measurement error can be quantified using  $\Delta z$ . Let  $z_R$  and  $z_B$  denote the estimated vertical distance of points R and B using a vertical model, vertical measurement error ( $\mathcal{E}$ ) can be computed using Equation 7-5.

$$\mathcal{E} = \Delta z - (z_R - z_T) = BR \cdot \sin(\gamma) - z_R + z_T \quad (7-5)$$

### 7.3.1.3.2. Horizontal measurements from a tilted surface

Horizontal measurements from the 5-*mm* line segment recordings closely followed the method presented in chapter 6. To that end, a GUI was developed that showed recordings one at a time. The GUI used the mouse for selection of the proper laser points, in addition to the marking of the two ends of the line segment. We saw in chapter 6 that the horizontal measurement relies on the estimation of the working distance. Considering that the surface could be tilted, some of the laser points would be on a different vertical distance than the target 5-*mm* line segment. Therefore, only the laser points close to the line segment were marked. This will ensure that a correct vertical distance is estimated for the target object. The vertical distance of each selected laser point was estimated using the PCA-based vertical model (chapter 6, Equations 6-14 and 6-15) and their average values was used as the vertical distance of the target object. Calibrated *mm* length of the line segment was estimated by feeding the estimated working distance and locations of the two endpoints of the line segment into the uniform (Equations 6-5) and the non-uniform (Equations 6-3 and 6-4) horizontal models. Horizontal measurement error was computed as the difference between the true value (*i.e.* 5 *mm*) and the estimated value.

## 7.3.2. Material and method for the effect of the 3D surface

### 7.3.2.1. Data acquisition

To investigate possible effects of 3D surfaces on vertical and horizontal measurement errors, a set of benchtop recordings from a 3D shape was collected. To that end, a general 3D model was created in Matlab. The model had three peaks and three valleys. The maximum height difference between its peaks and its valleys was 15 *mm*, and the size (length and width) of the model was 50 *mm* × 50 *mm*. Figure 7.9(A) presents the created model. Investigation of the hypotheses of this

section requires an accurate registration of the acquired images to the model. Considering the significant non-linear distortion of fiberoptic flexible endoscopes, a matrix of  $20 \times 20$  fiducial markers was created (figure 7.9(B)). Each fiducial marker was a cuboid with a size of  $0.45 \text{ mm} \times 0.45 \text{ mm} \times 0.4 \text{ mm}$  (for length, width, and height). These fiducial markers were merged with the 3D model and a composite image was created. A Creality-Ender5 3D printer with a  $0.4 \text{ mm}$  nozzle size and Polylactic Acid filaments with  $1.75 \text{ mm}$  diameter was used to print the created composite model. The surface was printed layer by layer with a thickness of  $0.12 \text{ mm}$  in each layer. The temperature of the nozzle during the 3D printing was set to  $205^\circ \text{ Celsius}$ , and the temperature of the printing bed was set to  $55^\circ \text{ Celsius}$ . The precision of the print was  $\pm 0.12 \text{ mm}$ . Finally, to make the detection of the fiducial markers more accurate, all fiducial markers were painted in black. Figure 7.9(C) shows the printed 3D composite model, after painting its fiducial markers in black.

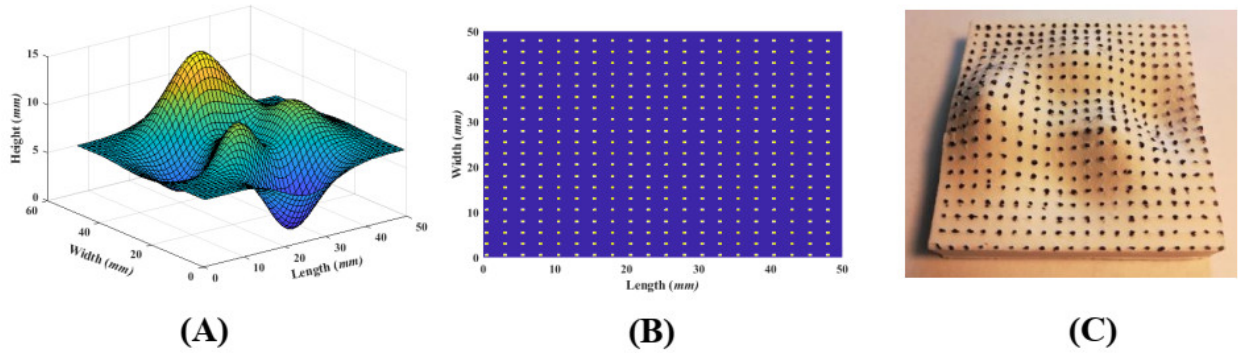


Figure 7.9. The data used for investigating the effect of 3D shape: (A) the 3D model, (B) fiducial markers, (C) the printed composite model.

Following the method described in section 7.3.1.2, the optical axis of the endoscope was made perpendicular to the target surface. Then, the printed 3D model was placed on the setup presented in section 1.4.1, figure 1.2 with one degree of freedom. Specifically, the tilting angle was kept fixed and at zero angle (*i.e.* perpendicular imaging angle) and only the working distance was varied. Five different subjective distances covering the range of close and far away, were used for

the recordings. Finally, the presence of the bright laser points was affecting the visual appearance of the fiducial markers. Therefore, at each working distance, two separate recordings were done. The external light source was turned on, and the laser source was turned off for the first recording. This data will be referred to as the model recordings for the rest of this chapter. The model recordings were used for the detection of the location of the fiducial markers in the recorded images. Then, the light source was turned off, and the laser source was turned on. This data will be referred to as the laser recordings for the rest of this chapter. The laser recordings were used for the detection of the location of the laser points in the recorded images. The laser recordings were analyzed with the PCA vertical measurement model (section 6.3.5). post-data collection. The average working distance of each data was measured and is reported in table 7.1.

Table 7.1. The estimated working distance from the 3D surface.

Working distance index	Estimated working distance	
	Mean ( <i>mm</i> )	std ( <i>mm</i> )
1	9.72	1.48
2	14.67	2.04
3	19.1	2.84
4	23.45	2.98
5	27.53	3.26

### 7.3.2.2. Analysis and measurements from a 3D surface

#### 7.3.2.2.1. Vertical measurements from a 3D surface

The locations of the laser points were detected using the method described in <sup>122</sup>. The fiducial markers were painted in black, this led to the absorption of the light from the laser points falling on them. For those points, the best circle representing the laser point was determined subjectively, and its center was used instead. A similar approach was followed for the detection of the fiducial

points, but instead of looking for the brightest point in the image (*i.e.* the laser points), we looked for the darkest points in the image (*i.e.* the fiducial markers). For the fiducial markers missing from the segmentation process, they were detected manually. The segmented information from a laser recording (*i.e.* the center of the laser points) was fused with the segmented information from its corresponding model recording (*i.e.* the center of the fiducial markers) and a composite image was created. Figure 7.10(A) shows an example.

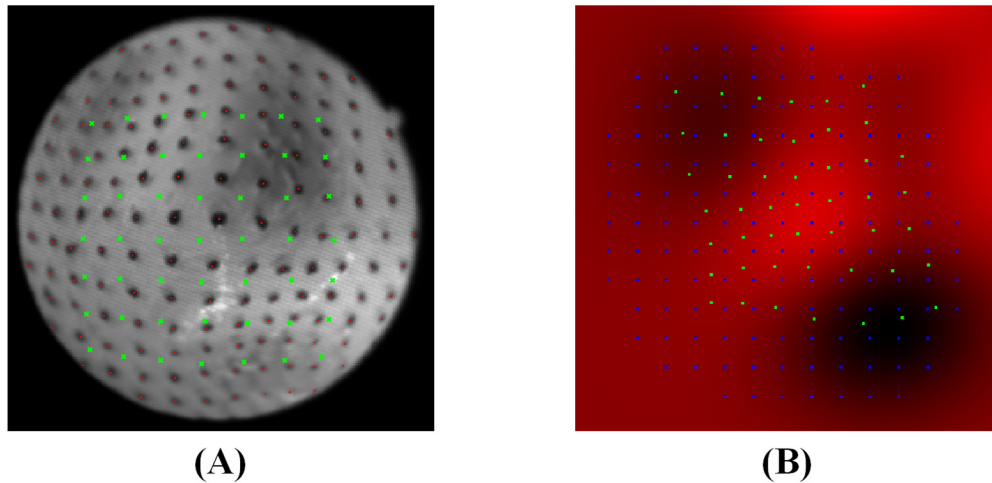


Figure 7.10. The outcome of the registration process: (A) a composite image before the registration. Centers of the fiducial markers are marked with a red dot. Centers of the laser points are marked with a green cross mark, (B) the registration outcome for the composite image.

Reliable estimation of vertical measurement errors depends on the knowledge of the ground truth (*i.e.* the real vertical distance). To achieve this, the four fiducial markers enclosing a laser points were determined. The indices of those fiducial markers in combination with the distances between them and the laser point were used to register the laser point on the 3D model (figure 7.9(A)). This process was repeated for all laser points with four enclosing fiducial markers. The points with 3 or less enclosing fiducial markers were omitted from the rest of the analysis. Figure 7.10(B) shows an example of the registration outcome. In figure 7.10(B) the height of the surface

is depicted in red, where a brighter color means a larger elevation at that point. Blue dots represent the center of the fiducial marker, and the green dots represent the center of the laser points.

Doing the registration step would give us the estimated true value of the height of a laser point on the model from the base of the model. Therefore, using the methodology presented in section 7.3.1.3.1 the laser point 25 (*i.e.* the laser point in the middle) was used as a reference, and differences in the height relative to laser point 25 were used to evaluate measurement error. Specifically, let  $z_R$  denotes the true height of the model at the center of laser point 25 after the registration. Also, let  $z_T$  denotes the true height of the model at the center of a target laser point after the registration. The true height difference between the two points ( $\Delta z$ ) would be equal to,

$$\Delta z = z_R - z_T \quad (7-6)$$

Additionally, let  $\tilde{z}_R$  and  $\tilde{z}_T$  denote the estimated vertical distances of the reference and the target points, respectively. Now, we could compute the difference in the vertical distance ( $\Delta \tilde{z}$ ) between the reference and the target laser points, using the PCA-based vertical measurement method. The vertical measurement error ( $\mathcal{E}$ ) can be computed as,

$$\mathcal{E} = \Delta z + \Delta \tilde{z} \quad (7-7)$$

where the plus is due to the fact that  $\Delta z$  is measured relative to the base of the model (*i.e.* a surface that is below the model), but  $\Delta \tilde{z}$  is measured relative to the endoscope (*i.e.* a surface that is above the model).

#### 7.3.2.2.2. Horizontal measurements from a 3D surface

In section 6.4.4 we saw that the non-uniform method offers significantly better measurement accuracy than the uniform method; therefore, the non-uniform method was used for this experiment. Horizontal measurement using the non-uniform method requires the  $x$ - $y$  coordinate of the two endpoints of the target object, in addition to its estimated working distance. This

information is readily available from the created composite images (figure 7.10 (A)). Specifically, the horizontal distance between each two adjacent fiducial markers is fixed and equal to 2.5 *mm*. Therefore, we could find a string of 3 adjacent fiducial markers that belong to the same row or column, and use *x-y* coordinates of the first and the last fiducial markers. The true horizontal distance for this selection would be equal to 5 *mm*. The composite images also include the location of the laser point. Therefore, the working distance of the selected string can easily be estimated using their nearby laser points and the PCA method.

For this experiment, all string of 3 adjacent fiducial markers that belong to the same row or column were detected. For each string, the number of laser points near to the string was computed, and if it was less than three, that string was omitted from the rest of the analysis. The vertical distances of all nearby laser points were estimated using the PCA-based vertical model (chapter 6, Equations 6-14 and 6-15) and their average was used as the working distance of the target object. Finally, the *x-y* coordinates of the first and the last fiducial markers from the string were used for performing the horizontal measurements.

## **7.4. Experiments and results**

Two experiments were conducted to answer the research questions of this chapter. Experiment 1 investigates the effect of the imaging angle on measurement accuracy. Experiment 2 presents measurement accuracies from a 3D surface. This section presents details of each experiment, followed by results and related discussions.

### **7.4.1. Experiment1: effect of the imaging angle**

The effects of imaging angle and working distance on measurement errors were investigated in this experiment.



#### 7.4.1.1. Experiment1a: effect of imaging angle on calibrated vertical measurements

This experiment was conducted to quantify the effects of imaging angle and working distance on vertical measurement errors from a flat surface. The following hypothesis was formed for this experiment.

H6a: The tilting angle of the target surface and the working distance will be good predictors of the vertical measurement error.

To test hypothesis H6a the dataset described in section 7.3.1.2.1 was used. Two different vertical models were presented in this dissertation. The first model was presented in chapter 4 and was published in the journal of voice. This model will be called the JOV model for the rest of this chapter. The second model was presented in chapter 6 and was based on the PCA analysis. This model will be referred to as the PCA model for the rest of this chapter. Vertical measurement error from each recording condition was evaluated using Equation 7-5.

The performance of the JOV model was evaluated after removing the top row laser points. Figure 7.11 shows boxplot of error for different working distances and imaging angles from the JOV model.

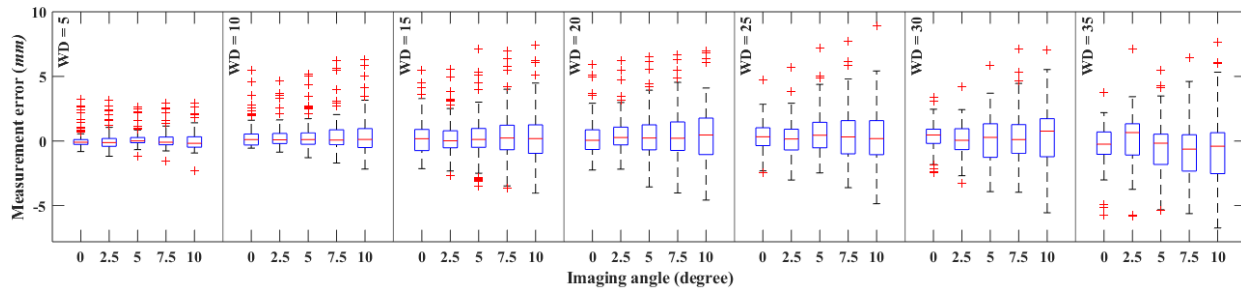


Figure 7.11. Boxplots of vertical measurement error using the JOV model at different working distances and imaging angles.

The performance of the PCA model was evaluated. Figure 7.12 shows boxplot of error for different working distances and imaging angles from the PCA model.

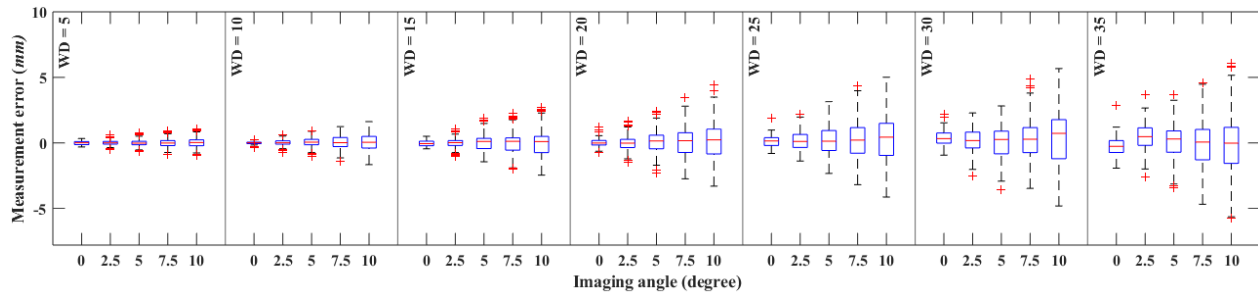


Figure 7.12. Boxplots of vertical measurement error using the PCA model at different working distances and imaging angles.

Investigating figures 7.11 and 7.12 indicates a higher magnitude of error in the JOV model. Additionally, boxplots of the PCA model show smaller variations in the error, which translates into a better agreement between measurements from different laser points. To test H6a a multiple linear regression analysis was used. Based on figures 7.11 and 7.12 measurement errors from both methods have outliers. Considering the sensitivity of regression analysis to the presence of outliers<sup>251</sup>, the robust multiple linear regression with iteratively reweighted least squares was used. Two different regression analyses were performed. In the first analysis, the imaging angle ( $\alpha$ ) and the working distance ( $wd$ ) were used as the predictor variables and measurement error was used as the outcome variable. This analysis determines whether the method tends to underestimate or overestimate the measurements. The second regression analysis was based on the same predictor variables but used the magnitude of measurement error as the outcome variable. This analysis determines the overall performance of the system.

Table 7.2 shows the results of the regression analyses for the JOV model. Based on the results of table 7.2 we can make the following conclusions. There is a significant effect of working distance ( $p < 0.00001$ ) and imaging angle ( $p < 0.00001$ ) on the magnitude of the error. Also, the magnitude of measurement error was positively correlated with the working distance and the imaging angle. Additionally, the coefficient of imaging angle is 1.5 times larger than the working

distance. This indicates a higher sensitivity of error to the imaging angle. The overall model was able to account for 18.9% of variations in the magnitude of the error.

Table 7.2. Results of multiple linear regression for the JOV vertical measurement model. The symbols wd, a, and  $\varepsilon$  stands for the working distance, the imaging angle, and  $p < 0.00001$ .

Parameter	Error		Magnitude of error	
	Estimate	$p$	Estimate	$p$
Intercept	0.215	0.0002	0.174	$\varepsilon$
wd	-0.004	0.08	0.027	$\varepsilon$
a	-0.006	0.33	0.04	$\varepsilon$
R-squared	0.003		0.189	

The performance of the PCA model was evaluated using a similar approach. Table 7.3 shows the results. Based on the results of table 7.3 we can make the following conclusions. There is a significant effect of working distance ( $p < 0.00001$ ) and imaging angle ( $p < 0.00001$ ) on the magnitude of the error. Also, the magnitude of measurement error was positively correlated with the working distance and the imaging angle. Additionally, the coefficient of the imaging angle was 2 times larger than the working distance. This indicates a higher sensitivity of error to the imaging angle. The overall model was able to account for 34% of variations in the magnitude of the error.

Table 7.3. Results of multiple linear regression for the PCA vertical measurement model. The symbols wd, a, and  $\varepsilon$  stands for the working distance, the imaging angle, and  $p < 0.00001$ .

Parameter	Error		Magnitude of error	
	Estimate	$p$	Estimate	$p$
Intercept	-0.137	$\varepsilon$	-0.08	$\varepsilon$
wd	0.009	$\varepsilon$	0.019	$\varepsilon$
a	0.015	$\varepsilon$	0.038	$\varepsilon$
R-squared	0.03		0.34	

Comparing tables 7.2 and 7.3 we can quantify the advantages of the PCA model. The coefficients of working distance and imaging angles are smaller in the PCA model. This indicates

that measurements from the PCA model are more robust to variations in the working distance and tilting angle.

#### 7.4.1.2. Experiment1b: effect of imaging angle on calibrated horizontal measurements

This experiment was conducted to quantify the effects of imaging angle and working distance on horizontal measurement errors from a flat surface. The following hypothesis was formed for this experiment.

H6b: The tilting angle of the target surface and the working distance will be good predictors of the horizontal measurement error.

To test hypothesis H6b the dataset described in section 7.3.1.2.2 was used. Horizontal measurement error from each recording condition was computed as the difference between the true value and the estimated value from the vertical models. Two different horizontal models of uniform and non-uniform were presented in chapter 6. Both models were evaluated in this experiment.

The performance of the uniform model was evaluated. Figure 7.13 shows boxplots of error for different working distances and imaging angles from the uniform model.

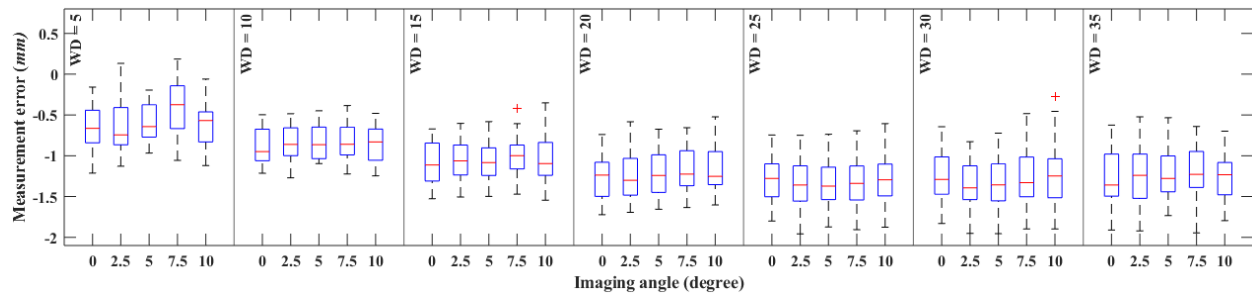


Figure 7.13. Boxplots of horizontal measurement error from the uniform model at different working distances and imaging angles.

The performance of the non-uniform model was evaluated. Figure 7.14 shows boxplots of error for different working distances and imaging angles from the non- uniform model.

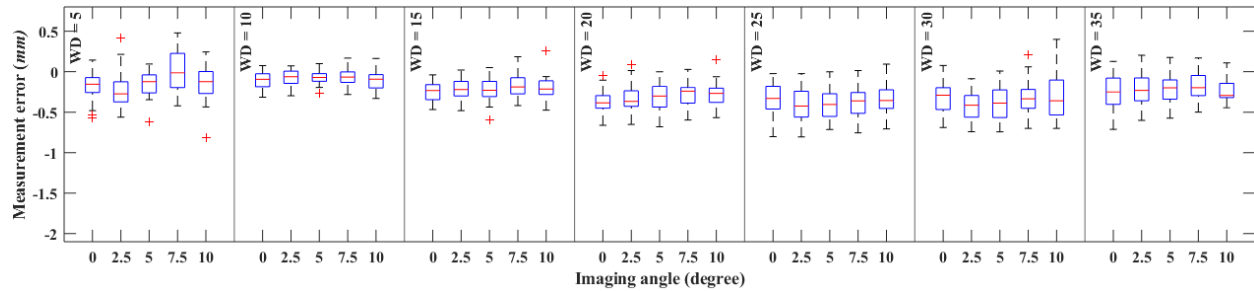


Figure 7.14. Boxplots of horizontal measurement error from the non-uniform model at different working distances and imaging angles.

Investigating figures 7.13 and 7.14 indicates a significantly higher magnitude of error in the uniform model. Additionally, comparing the two sets of boxplots reveals that the centers of non-uniform boxes are much closer to zero than their uniform counterparts. This indicates a random nature for measurement errors in the non-uniform model, compared to a systematic nature for measurement error in the uniform model. Putting it differently, we may achieve a very small error by averaging multiple measurements from the same object using the non-uniform model. To test H6b, a multiple linear regression analysis was used. To that end, two different regression analyses were performed. In the first analysis, the imaging angle ( $\alpha$ ) and the working distance ( $wd$ ) were used as the predictor variables and measurement error was used as the outcome variable. This analysis determines whether the method tends to underestimate or overestimate the measurements. The second regression analysis was based on the same predictor variables but instead the magnitude of measurement error was used as the outcome variable. This analysis determines the overall performance of the system.

Table 7.4 shows the results of regression analysis for the uniform model. Based on the results of table 7.4 we can make the following conclusions for the uniform model. There is a significant effect of working distance ( $p < 0.00001$ ). Also, the magnitude of the error and the working distance

were positively correlated. Finally, the imaging angle didn't reach the significant level ( $p=0.06$ ). The overall model was able to account for 34% of variations.

Table 7.4. Results of multiple linear regression for the uniform model for horizontal measurements. The symbols wd, a, and  $\varepsilon$  stands for the working distance, the imaging angle, and  $p<0.00001$ .

Parameter	Error		Magnitude of error	
	Estimate	$p$	Estimate	$p$
Intercept	-0.663	$\varepsilon$	0.667	$\varepsilon$
wd	-0.022	$\varepsilon$	0.022	$\varepsilon$
a	0.006	0.06	-0.006	0.06
R-squared	0.34		0.34	

The performance of the non-uniform model was evaluated using a similar approach. Table 7.5 shows the results of regression analysis for the non-uniform model. Based on the results of table 7.5 we can make the following conclusions for the non-uniform model. There is a significant effect of working distance ( $p<0.00001$ ), where the magnitude of the error and the working distance were positively correlated. Finally, the imaging angle did not reach the significant level ( $p=0.24$ ). The overall model was able to account for 8.9% of variations in the magnitude of the error.

Table 7.5. Results of multiple linear regression for the non-uniform model for horizontal measurements. The symbols wd, a, and  $\varepsilon$  stands for the working distance, the imaging angle, and  $p<0.00001$ .

Parameter	Error		Magnitude of error	
	Estimate	$p$	Estimate	$p$
Intercept	-0.135	$\varepsilon$	0.17	$\varepsilon$
wd	-0.007	$\varepsilon$	0.005	$\varepsilon$
a	0.005	0.008	-0.002	0.24
R-squared	0.11		0.089	

Comparing tables 7.4 and 7.5 we can quantify the advantages of the non-uniform model over its uniform counterpart. The coefficients for the working distance and the imaging angle were smaller in the non-uniform model. This indicates that measurements using the non-uniform model are more robust to these variations. For example, in the uniform model for every  $mm$  increase in

the working distance, the magnitude of error increases by 0.022 *mm*, which is 4.4 times higher than the non-uniform model. A similar argument can be made for imaging angle too, however, due to nonsignificant *p*-values, the effect of imaging angle should be interpreted with more caution.

#### **7.4.2. Experiment2: effect of a 3D surface**

Experiment 2 presents the performances of vertical and horizontal measurement methods on a 3D surface.

##### **7.4.2.1. Experiment2a: effect of a 3D surface on calibrated vertical measurements**

This experiment was conducted to quantify the effects of a 3D surface on vertical measurement errors. The following hypothesis was formed for this experiment.

H6c:      The vertical measurement errors from a non-flat surface will be higher than those from a flat surface positioned at the same estimated average vertical distance.

To test hypothesis H6c the dataset described in section 7.3.2.1 was used. In the previous section, we saw that the PCA model had superior performance compared to the JOV model; therefore, only the performance of the PCA model was investigated here.

First, the performance of the PCA model on a flat surface with a zero tilting angle should be computed. We had a total number of 124 data points from a flat surface at zero tilting angle. The data points were randomly divided into training (70%) and testing (30%) sets. The PCA model was trained using the training data. The trained model was applied to the testing samples, and the measurement error was computed. Figure 7.15 presents the fitted curves with their 95% confidence interval.

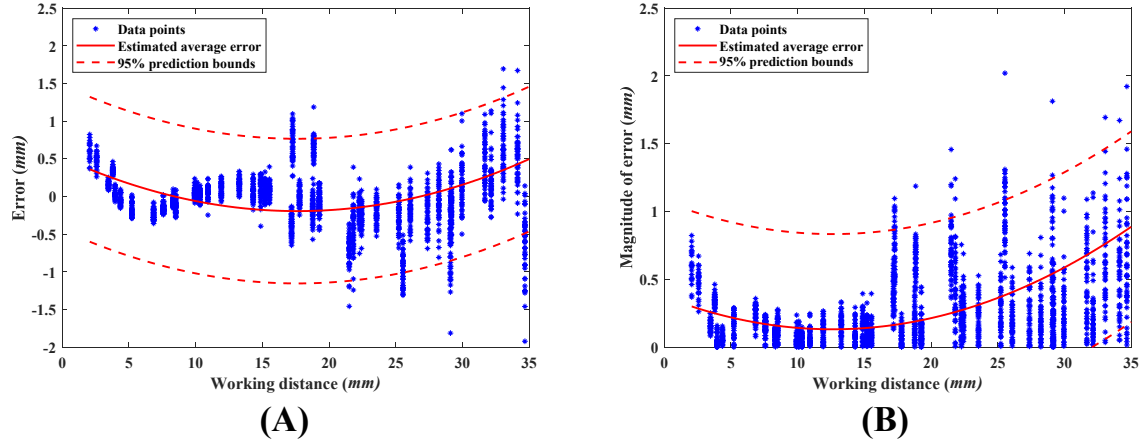


Figure 7.15. Performance of the PCA model on a flat surface: (A) vertical measurement errors, (B): magnitude of vertical measurement errors.

Then, vertical measurement errors from the 3D model were computed. Figure 7.16 present boxplots of this analysis. An interesting observation can be made from boxplot 7.16(A). The vertical measurement error has a random nature. That is, multiple measurements relative to a fixed reference can reduce the error significantly. This is evident from the fact that boxplots of vertical measurement error are relatively centered around zero. Considering that the endoscope would be utilized for studying the envelope of the vocal folds (and not the behavior of individual laser points), this characteristic is very beneficial.

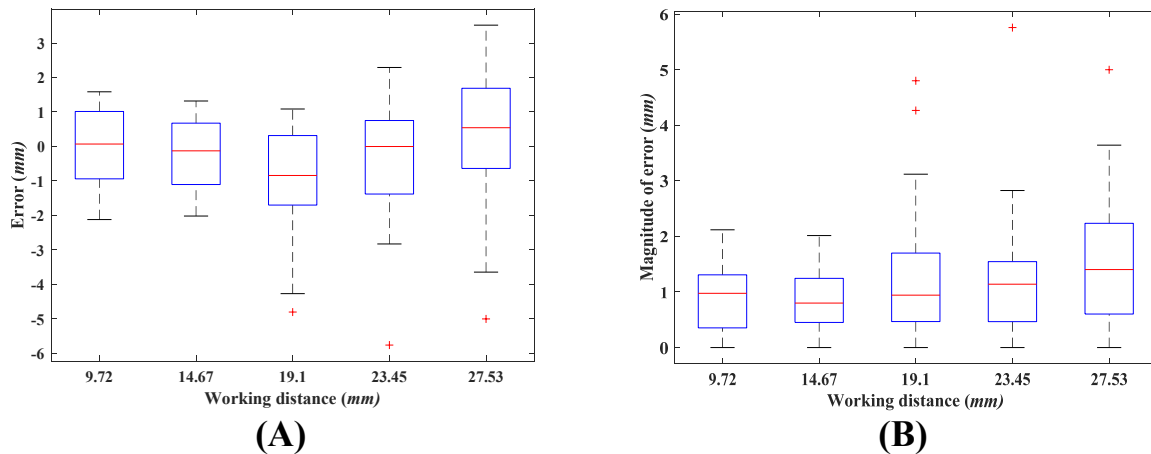


Figure 7.16. Performance of the vertical measurement errors on a 3D surface: (A) boxplot of error, (B) boxplot of the magnitude of error.



A two-way ANOVA could be used to test H6c. The dependent variable for this test was the vertical measurement errors, and the independent variables were the surface condition (flat vs. 3D), and the working distance group. Table 7.6 reflects the results of the analysis for measurement error and the magnitude of measurement error.

Table 7.6. Results of 2×5 ANOVA for vertical measurement errors.

Variable	Error ( <i>mm</i> )	Magnitude of error ( <i>mm</i> )
	<i>p</i>	<i>p</i>
Surface ( <i>S</i> )	0.59	<0.00001
Working distance ( <i>WD</i> )	<0.00001	<0.00001
<i>S</i> × <i>WD</i>	<0.00001	<0.00001

Based on the results of table 7.6 we see that the surface condition (flat vs. 3D) did not have a significant effect on the vertical measurement errors. However, the surface condition (flat vs. 3D) had a significant effect on the magnitude of vertical measurement errors. Running post-hoc analysis showed that the magnitude of vertical measurement error from the 3D surface was significantly higher at all working distances, except for the 23.45 *mm* group. Considering that the endoscope primarily would be utilized for studying the envelope of the vocal folds (and not the behavior of individual laser points), the non-significant difference of the error seems to be of higher practical value. Finally, the mean percent (magnitude of) error was defined as the mean of the ratio of (magnitude of) errors to the target value and is reported in table 7.7. As a final note, in all analyses we assumed that the 3D printing error and the registration error had a random nature.

Table 7.7. Mean percent error and mean percent magnitude of error for vertical measurement.

Estimated Working distance	Mean percent error %		Mean percent magnitude of error %	
	Flat	3D	Flat	3D
9.72	-0.2	0.2	2.5	9.2
14.67	0.8	-1.9	1.6	6.2
19.1	-0.1	-4.3	1.3	6.4
23.45	-3.7	-1.1	1	5
27.53	-0.7	0.7	0.9	5.4

#### 7.4.2.2. Experiment2b: effect of a 3D surface on calibrated horizontal measurements

This experiment was conducted to quantify the effects of a 3D surface on horizontal measurement errors. The following hypothesis was formed for this experiment.

H6d: The horizontal measurement errors from a non-flat surface will be higher than a flat surface positioned at the same estimated average vertical distance.

Horizontal measurement errors from the 3D model were computed. Figure 7.17 present boxplots of this analysis. Investigating boxplot of figure 7.17(A) suggests that at short working distances (less than 15 *mm*) the method is underestimating the measurements. However, at large working distances the method is overestimating the measurements. Investigation of figure 7.17(B) also shows that the magnitude of measurement is significantly higher at 9.72 *mm*. This could be because at shorter working distances the magnitude of registration error and/or printing error become more comparable with the measurement errors and therefore their contributions could become more significant.

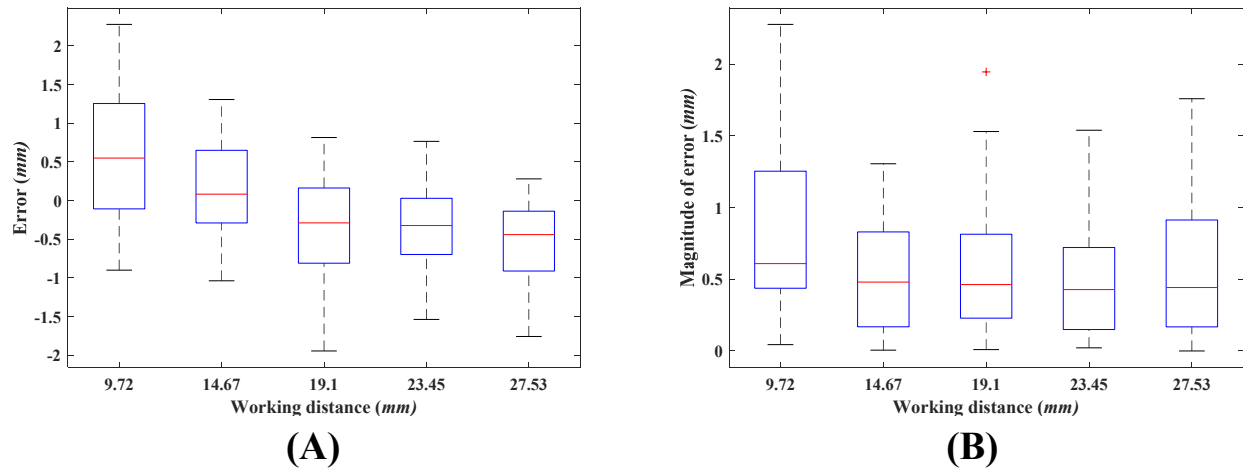


Figure 7.17. Performance of the horizontal measurement errors on a 3D surface: (A) boxplot of error, (B) boxplot of the magnitude of error.

A two-way ANOVA was used to test H6d. The dependent variable for this test was the horizontal measurement errors, and the independent variables were the surface condition (flat vs. 3D), and the working distance group. The data for the flat surface was the same as the experiment1b of this chapter, but only the data from the zero tilting angle were used. Table 7.8 reflects the results of the analysis for measurement error and the magnitude of measurement error. It is noteworthy that the values of working distance were a little different between different recordings, but we will report them using the same working distance group (for example the measurement from 9.72 mm will be referred to as 10 mm, etc.).

Table 7.8. Results of two-way ANOVA for horizontal measurement errors.

Variable	Error (mm)	Magnitude of error (mm)
	<i>p</i>	<i>p</i>
Surface ( <i>S</i> )	<0.00001	<0.00001
Working distance ( <i>WD</i> )	<0.00001	0.35
<i>S</i> × <i>WD</i>	0.0067	0.0008

Based on table 7.8 there was a significant effect of the surface (flat vs. 3D) on both horizontal measurement error, and the magnitude of horizontal measurement error. Post-hoc analysis was run

on the ANOVA model for the error. Only the measurement error from the shortest working distance ( $\sim 10\text{ mm}$ ) led to a significant difference between the two surface conditions. Interestingly, at this working distance the measurements from the flat surface were overestimated, but from the 3D surface were underestimated. Post-hoc analysis on the ANOVA model for the magnitude of errors showed a significant difference between measurement errors from the flat and the 3D surfaces at the working distance groups of  $10\text{ mm}$  and  $15\text{ mm}$ . The magnitude of error was significantly higher from the 3D surface. However, the test failed to detect a significant difference between the two conditions at the working distance of  $20\text{ mm}$ . Finally, the mean percent (magnitude of) error was defined as the average of the ratio of the (magnitude of) errors to the target value and is reported in table 7.9. The results of table 7.9 support the discussed finding. However, the more interesting finding is that the difference in the performance of horizontal measurements from the flat and the 3D surface decreases as the working distance increases. As a final note, in all analyses we assumed that the 3D printing error and the registration error had a random nature.

Table 7.9. Mean percent error and mean percent magnitude of error for horizontal measurement.

Working distance group	Mean percent error %		Mean percent magnitude of error %	
	Flat	3D	Flat	3D
10	-2.2	11.4	2.4	16.6
15	-5	2.5	5	10.6
20	-7.2	-7	7.2	11.7

## 7.5. Discussions

The laser-projection endoscope will be used for in-vivo data collection. However, the vertical and horizontal calibration methods presented in chapters 4 and 6 were developed based on benchtop recordings. There are significant differences between the two recording conditions.

Specifically, the benchtop recording presents the most controlled data acquisition scenario. For example, the surfaces were white, flat, there was a minimum light reflection, and the optical axis was perpendicular to the target surface. On the other hand, the in-vivo condition represents the least controllable data collection environment. For example, it is very likely for the in-vivo images to be acquired at a non-perpendicular imaging angle from the region of interest. Furthermore, the region of interest would definitely have a 3D topology and be non-flat. The in-vivo data will be collected in the presence of 300-Watt xenon light. This could add significant light reflections to the acquired images. In summary, the true performance of the method on actual in-vivo data could be significantly different from those estimated and reposted in chapters 4 and 6. Using calibrated intraoperative images would be one possible solution to remedy this.<sup>191</sup> We could use calibrated intraoperative images and determine if there are huge discrepancies between the performance of the two conditions. The main advantage of this approach would be its potency to mimic the true data collection condition. However, there are some limitations to this approach. For example, it can only be done in an operation room, which puts a practical restriction on its feasibility. Additionally, the number of possible factors is so high that if high measurement errors are found, it would be very difficult (if possible, at all) to determine the most contributing factors to the measurement errors. Obviously determining the most contributing factors would be necessary for devising better measurement approaches and/or instruments. Last not least, the ground truth is not known in intraoperative images. More precisely, using the intraoperative images as the ground truth assumes a perfect validity and reliability for its subsequent measurements, and attributes all measurement error to the method that is being tested. However, in reality, the estimated error would be a mixture of the two errors.

A different solution would be to simulate the most likely contributing factors in a controlled fashion. This approach has the potency of addressing the above-mentioned concerns; however, it depends on the selection of the most likely factors contributing to measurement errors, which requires enough knowledge to assist with the selection process. In chapter 5 we saw that imaging angle was the most contributing factor on uncalibrated measurement errors from a flat surface. Therefore, it is logical to select it as a contributing factor. Furthermore, figure 7.1 shows that variations in the imaging angle bring some parts of a flat surface closer to the endoscope while pushing the other parts further away. We could hypothesize that this factor may account for some of the observed increases in the measurement errors. In that case, imaging from a non-flat surface would be another instance of a non-uniform distance between the target surface and the endoscope. Therefore, the 3D structure of the target surface was selected as the second likely contributing factor.

The selected two most likely contributing factors (*i.e.* the imaging angle and 3D topology of the target surface) were changed in a systematic way to investigate their effect on horizontal and vertical measurement errors. Experiment 1 was conducted to quantify the effect of changes in the imaging angle. Our analysis showed that vertical measurement errors from the PCA-based method were 2 times more sensitive to variations in the imaging angle than the working distance. A similar analysis was conducted on horizontal measurement error. The analysis showed that horizontal measurement error from the non-uniform method was less sensitive to variation in the imaging angle than the working distance. Interestingly, the effect of imaging angle on the magnitude of error was non-significant. Comparing this outcome, with the high sensitivity of uncalibrated vertical measurement errors on imaging angle (chapter 5), highlights the efficacy of the proposed method for handling the effect of imaging angle.

Experiment 2 was conducted to see if there are significant differences in measurement accuracies between a flat and a non-flat surface. First, the effect of surface type (flat vs. 3D) was tested on vertical measurement errors. Interestingly, the surface type didn't have a significant effect on the vertical error. However, the magnitude of measurement from the 3D surface was significantly higher than a flat surface. The vertical measurement capability of the laser endoscope would primarily be used for vertical envelope estimation and not behavior of individual laser points. Therefore, the obtained non-significant result would be of higher practical value for this device. Additionally, the fact that the effect of surface type was non-significant for the error and was significant for error indicates a random nature for measurement errors from the 3D surface. More specifically, it indicates that the magnitude of overestimation and underestimation from the 3D surface is higher than a flat surface positioned at a comparable average working distance (hence the significant effect of the magnitude of error); however, the magnitudes of overestimation and underestimation are on the same level, and hence cancel each other out when averaged (hence the non-significant effect for error). Elaboration on the cause of this non-significant effect warrants some explanation. The calibrated endoscope projects a set of distinct laser points on the FOV, and each laser point would occupy a very small area of the whole image. Also, the vertical measurement method was designed such that measurement from each laser point was independent from other laser points. The combinations of these two characteristics mean that each laser point would only have access to information (including topology information) from a very limited area. If the target surface is smooth and without sudden changes in the vertical components, any small area can be approximated with a flat surface. Consequently, the area that each laser points have access to would be almost a flat surface, and we shouldn't see a significant difference in errors between the two conditions. Finally, a very peculiar trend can be seen in table 7.7 worth a

discussion. The mean percentage of the magnitude of error decreases with the working distance. When the working distance is shorter, a smaller portion of the 3D surface would be recorded, and in that regard, the vertical variation should be smaller. Additionally, in chapter 4 we saw that accuracy of vertical measurement was better at shorter working distances. So, we may expect to see smaller values for shorter working distances. Therefore, some other factors should be contributing to this observation. In all of the analyses we assumed that the 3D printing error and the registration error had a random nature. However, this could be an incorrect assumption. Specifically, at shorter working distances the magnification of the imaging system is higher, and therefore a smaller area of the target surface is recorded. This means that fewer fiducial markers would be present at images acquired from shorter working, and the number of fiducial markers would increase as the working distance increases. Considering that non-linear distortion of fiberoptic flexible endoscopes is location-dependent, and also the fact that registration accuracy relies on the fiducial markers; we could argue that at smaller working distances the registration error would be higher. This would translate into a less accurate estimation of the ground truth, which incorrectly may lead to a higher estimation of measurement errors. Finally, the statistical analysis indicated a significant effect of surface type on horizontal measurement error. However, post-hoc analysis did not find a consistent trend between different working distances. More specifically, the results suggested that at short working distances measurement errors from the 3D surface is higher than the flat surface, However, there was no significant difference between the two surface conditions at larger working distances. Table 7.9 shows that in fact as the working distance increase the difference in mean percent error and mean percent magnitude of error between the two surface condition decreases.



## 7.6. Conclusions

This work was motivated by the significant difference in the conditions that vertical and horizontal calibration methods were being developed and evaluated on, and the actual in-vivo imaging conditions. The in-vivo condition is uncontrolled with many different variable factors. These factors were not explicitly considered during the development of the algorithms. Therefore, measurement accuracies from in-vivo images could be very different from the estimated values in chapters 4 and 6. To address this concern, the two most likely parameters degrading the accuracy of developed horizontal and vertical measurement methods were investigated in this chapter. Those parameters included the imaging angle and the 3D topology of the surface. Doing the analysis showed that vertical measurement errors were two times more sensitive to variations in the imaging angle than the working distance. However, horizontal measurement errors were less sensitive to variation in the imaging angle, than the working distance. This highlights the robustness of the developed horizontal measurement method to variations in the imaging angle. Investigating the effect of surface type (flat vs. 3D) did not lead to significant differences in vertical measurement errors. Doing similar analysis on horizontal measurements indicated a significant effect of the surface type on horizontal measurement error. However, post-hoc analysis suggested that at short working distances measurement errors from the 3D surface were higher than the flat surface. But the two conditions were becoming more similar as the working distance was increased.

## CHAPTER 8: SUMMARY OF THE FINDINGS

Spatially calibrated measurements could offer significant advantages for voice science research and clinical applications. They could be used to derive criteria for more accurate and direct evaluation of intervention outcomes (*e.g.* post-intervention changes in the lesion size), and in that regards could advance the evidence-based practice in the field of laryngology and speech-language pathology. Spatially calibrated measurements could also be used to create comprehensive models that can link the input (*i.e.* airflow), the output (*i.e.* acoustic signal), and parameters of the phonatory system (*e.g.* calibrated glottal area waveform, vocal fold length, kinematic measures) together. It is expected for such computational models to explain the individual differences that we see in the intervention outcome of individual patients. This prospective line of research could advance precision and personalized medicine in the field of laryngology and speech-language pathology. Kinematic measures (*e.g.* the vocal fold velocity) are another possible outcome of calibrated measurements. Kinematic measures are closely related to biomechanics of the vocal fold vibration and provide wealth of information for modeling and patient-specific modeling applications. Additionally, vocal folds collision forces and vocal fold stiffness<sup>256</sup> are among important parameters of the phonatory system that indirectly may be estimated using the velocity measures. More accurate gradings of laryngeal diseases and studying the developmental aspects of the vocal folds are other topics on interest that could benefit from calibrated spatial measurements. Considering the significance of such prospective research, this dissertation was devoted to an in-depth treatment of spatial calibrated measurements from in-vivo high-speed videoendoscopy images.

Generally speaking, achieving the spatial calibration goals depend on the existence of some *auxiliary* information. This auxiliary information makes the conversion from the uncalibrated

lengths (*i.e.* pixel) to calibrated lengths (*i.e.* *mm*) possible. Depending on the source of the *auxiliary* information, two different categories of direct and indirect calibration approaches were identified and presented in this dissertation. The *auxiliary* information of direct method comes from the *same* image that we want to perform measurements from. While, in the indirect calibration approach the *auxiliary* information comes from a *different* image than the image that we want to perform measurements from. The definition of direct method stipulates the existence of some properly designed fiducial markers on the acquired images. Therefore, several important challenges should be addressed for direct methods. First, proper fiducial markers should be designed, such that calibrated measurements become possible, while the fiducial markers should not obstruct the clinical applications of the acquired images. Second, the fiducial markers should be delivered and projected on the field of view. Third, sophisticated calibration protocols and measurement techniques should be developed and implemented to achieve the measurement purposes. In summary, direct calibration could offer very reliable and accurate calibrated measurements but it requires specialized hardware and software capabilities, and because of that, it would only be accessible to very limited research labs. It is unlikely for such systems to become commercially available in the near future. Indirect calibration was the solution that was proposed in this dissertation to make calibrated measurements accessible to more research labs. Specifically, the indirect calibration uses the uncalibrated length (*i.e.* pixel length) of a common object for normalization of other spatial features of the image. Depending on the information available from the common object, either absolute *mm* measurement or percentage change of a target object can be computed. The downside of indirect calibration is the reliability of its subsequent measurement. Specifically, three main assumptions behind the validity of indirect calibration were presented in this dissertation. Often (if not always) direct evaluation of these three assumptions is not trivial

from in-vivo images, and hence the measurement errors from indirect calibration could not be estimated directly. However, two tests were proposed in this dissertation that could provide some levels of assurance regarding the validity of measurements. Figure 8.1 presents a diagram of the relationships among the different chapters of this dissertation.

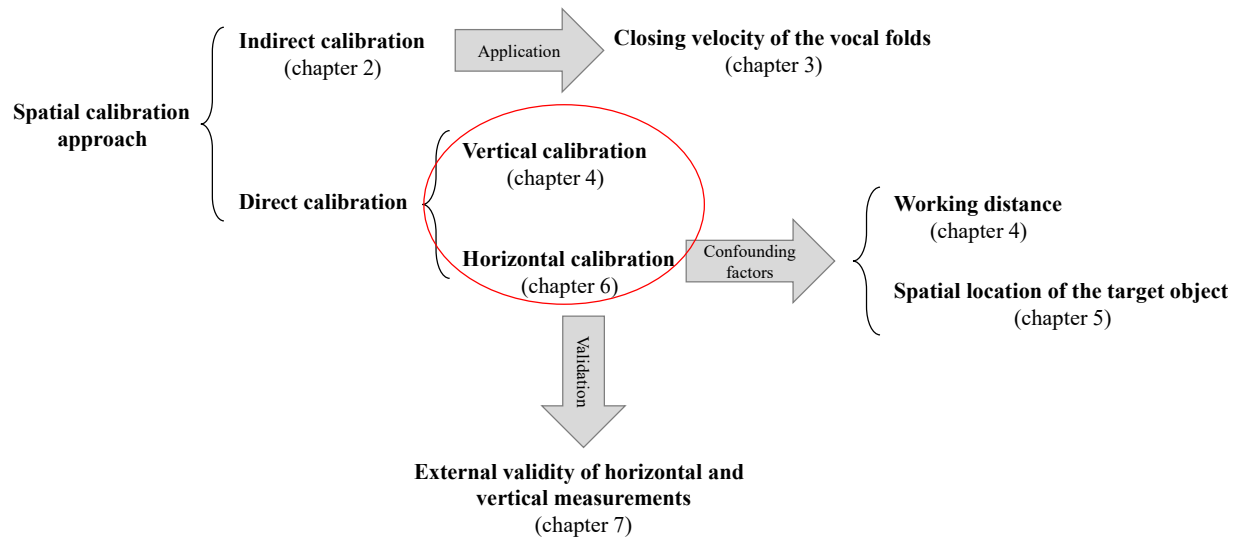


Figure 8.1. Graphical representation of the relationships among the chapters of this dissertation.

## 8.1. Specific contributions of each dissertation chapter

Chapter 2 was devoted to a formal treatment of indirect calibration method. The assumptions behind the validity of measurements were derived based on mathematical analysis of the pixel size. To make the problem tractable, it was assumed that the pixel size was only a function of the working distance (*e.g.* no non-linear image distortion) and the optical axis was perpendicular to the target surface. Under these conditions three main assumptions governing the validity of the indirect calibration were derived. First, the common attribute should be registered accurately in the target image. Second, the common attribute and the target object should be at the same vertical distance from the endoscope. Third, the calibrated length of the common attribute should not

change between different imaging sessions. Finally, these assumptions were tested and discussed in the context of laryngeal imaging and using a pre-existing HSV dataset.

Chapter 3 built on the results of chapter 2 and used the indirect calibration method for investigation of post-surgery changes in closing velocity of the vocal folds in patients with vocal fold mass lesions. HSV recordings from habitual pitch, habitual loudness of 16 subjects with VF mass lesions were collected pre-surgery and post-surgery. Spatially calibrated intraoperative images were acquired from each subject during the surgery. HSV data underwent temporal segmentation (to select the timestamps corresponding with different glottal phases), motion compensation (to remove the endoscopic motion artifacts), spatial segmentation (to detect the edges of the VF in sub-pixel resolution), and horizontal calibration processes. The pre-surgery HSV data were indirectly calibrated by registering the lesions from the intraoperative images to their corresponding HSV recording. The vocal fold width from each calibrated pre-surgery HSV data was selected, and then it was registered to its corresponding post-surgery HSV data. This step led to indirect calibration of the post-surgery HSV data. Three different experiments were conducted to investigate the (1) post-surgery changes in the closing velocity of the vocal folds, (2) differences in pre-surgery and post-surgery similarities between closing velocity of the two vocal folds, and (3) the association between post-surgery changes in the closing velocity of the vocal folds and the area of the lesion. Experiment 1 showed significant increases in the closing velocity of the vocal folds with the lesion, however, the increase for the contralateral side was limited more to the area in direct contact with the lesion. Experiment 2 showed that closing velocity of the two vocal folds become more similar after the surgery. Experiment 3 failed to detect a significant correlation between the post-surgery changes in the closing velocity of the vocal folds and the area of the lesion.

Chapter 4 presented the methodology for direct vertical calibration of HSV images using a laser-projection fiberoptic transnasal endoscope. The access to calibrated vertical measurement could provide significant and clinically valuable information regarding the vertical movements of the vocal folds in normal and disordered populations. Furthermore, the vertical calibration is the prerequisite for horizontal calibrated measurements from the laser-projection endoscope. The  $x$ -,  $y$ -coordinates of the laser points is the primary factor that encodes the vertical distance. However, investigating the position of the laser points showed that, besides the vertical distance, they also depended on the parameters of the lens coupler, including the field of view (FOV) position within the image frame and the rotation angle of the endoscope. An automatic calibration method was developed to compensate for the effect of these parameters. Statistical image processing and pattern recognition were used to detect the FOV, the center of FOV, and the fiducial marker. This step normalized the HSV frames to a standard coordinate system and removed the dependence of the laser-point positions on the parameters of the lens coupler. Then, using a statistical learning technique, a calibration protocol was developed to model the trajectories of all laser points as the working distance was varied. Finally, a set of experiments was conducted to measure the accuracy and validity of every step of the procedure. The system was able to measure vertical distance with mean percent error in the range of 1.7% to 4.7%, depending on the working distance.

Accurate calibrated horizontal measurements require the determination of its confounding factors, and then accounting for them. Working distance is the most trivial confounding factor for horizontal measurements, and the method for its estimation was presented in chapter 4. Chapter 5 investigated the possibility of a second confounding factor for calibrated horizontal measurements, namely the spatial location of the target object. To that end, the effect of the fiberoptic flexible endoscope distortions on calibrated horizontal measurements were studied and quantified. It was

shown that two sources of nonlinear distortions could deviate captured images from the reality. The first distortion stemmed from the wide-angle lens used in flexible endoscopes. It was shown that endoscopic images have a significantly higher spatial resolution in the center of the FOV than its periphery. The difference between the two could lead to as high as 26.4% error in calibrated horizontal measurements. The second distortion stemmed from variations in the imaging angle. It was shown that the disparity between spatial resolution in the center and periphery of endoscopic images increases as the imaging angle deviates from the perpendicular position. Furthermore, it was shown that when the imaging angle varies, the symmetry of the distortion was also affected significantly. Our analyses showed that the combined distortions could lead to calibrated horizontal measurement errors as high as 65.7%.

Chapter 6 built on the results and outcomes of chapters 4 and 5 and presented the methodology for accurate horizontal measurements from a laser-projection fiberoptic transnasal endoscope. To that end, a set of circular grids were recorded at multiple working distances. A statistical model was trained that would map from pixel length of the object, the working distance, and the spatial location of the target object into its *mm* length. This non-uniform model was contrasted with a second model that was not compensating for the effect of spatial location of the target object. This property led to a model with similar pixel sizes for all part of the image, and hence it was named the uniform model. The uniform model is the basis of existing methods for calibrated horizontal measurements, and it is significant in that regard. A detailed analysis of the performance of both models was presented. The analyses showed that the accuracy of the uniform method depended significantly on the working distance and also the length of the target object. However, the non-uniform model was quite robust to those variations. The estimated average magnitude of error from the non-uniform method was 0.27 *mm*, which was three times less than the uniform model.

Chapters 4 and 6 presented the methods for calibrated vertical and horizontal measurements from a laser-projection fiberoptic transnasal endoscope. The design and evaluation of those methods was done in controlled settings and using benchtop recordings. However, many factors could be contributing to measurement errors from in-vivo images.

Chapter 7 investigated the effect of two factors that were more likely to contribute significantly to increased measurement errors from in-vivo images. These factors were the imaging angle and the surface topology. To that end, the calibrated vertical and horizontal measurement models trained in chapters 4 and 6 were used. Two experiments were conducted to evaluate their performances in situations modelling the in-vivo settings. The first experiment was based on images acquired from tilted surfaces. The second experiment was based on a target surface with known  $x$ -,  $y$ -,  $z$ -coordinates that was 3D-printed. The measurement accuracies from the tilted surface and the 3D-printed surface were contrasted with the accuracy from the flat surface. The data analysis showed a significant effect of imaging angle on vertical measurement error. However, the effect of imaging angle on the magnitude of horizontal measurement error was not significant. Analysis of the effect of surface topology showed the reverse effects. The effect of surface type on vertical measurement error was not significant. But the magnitude of horizontal measurements errors from the 3D surface was significantly higher than the flat surface. The mean percent magnitude of horizontal measurement error increased from 5% (flat) to 10.6% (3D) at the working distance of 15 *mm*, which still represents satisfactory accuracy.

## **8.2. Directions for further investigations**

This dissertation can be expanded in several directions for future works. Chapter 2 presented the concept of indirect calibration. It was shown that the vocal fold width was a robust feature for calibration. However, this conclusion was based on a small sample size and may not be very



generalizable. Conducting a study with a bigger sample size and more phonatory behaviors (*e.g.* resting state, combinations of different loudness levels and pitches) could lead to an attribute that governs the consistency of the common attribute assumption to the maximum extent. Additionally, devising a test that can validate the vertical distance assumption was left as an open problem for future research. Chapter 3 may be expanded in several directions. Specifically, the dependent variable of chapter 3 was the magnitude of maximum closing velocity at different scanning lines. However, the developed method could be used to investigate the phase-differences between different scanning lines of the vocal folds. It is quite possible for this variable to explain some of phenomena that the magnitude of velocity cannot. Relating the post-surgery changes in the closing velocity and output of the system (*e.g.* acoustic changes) would be another line for future research. Considering that calibrated images are often not available, devising a non-calibrated proxy for closing velocity could remove a significant obstacle in application of kinematic measures for other studies. Chapter 5 presented the effects of non-linear distortion and imaging angle on horizontal measurements from a fiberoptic flexible endoscope. However, rigid endoscope and distal-chip flexible endoscopes are more widely used in clinical practice. Investigation and quantification of the effects of non-linear distortion and imaging angle from rigid and distal-chip flexible endoscopes could be of significant value for clinical practices. Additionally, chapter 5 showed that variations in the imaging angle is a significant confounding factor for horizontal measurement. However, the method for estimation and compensation of the imaging angle is still lacking. Our initial experimentations with the laser-projections endoscope showed promising results that could lead to an innovative application for the laser-projection endoscope and requires further investigations. Finally, chapters 4 and 6 presented the method for vertical and horizontal calibrated

measurements; however, applications of these methods were not part of this dissertation. The applications of these methods would be a whole avenue for future research.

## REFERENCES

## REFERENCES

1. Connor NP, Cohen SB, Theis SM, Thibeault SL, Heatley DG, Bless DM. Attitudes of children with dysphonia. *J Voice*. 2008;22(2):197-209.
2. Lass NJ, Ruscello DM, Bradshaw KH, Blankenship BL. Adolescents' perceptions of normal and voice-disordered children. *J Commun Disord*. 1991;24(4):267-274.
3. Branski RC, Cukier-Blaj S, Pusic A, et al. Measuring quality of life in dysphonic patients: a systematic review of content development in patient-reported outcomes measures. *J voice*. 2010;24(2):193-198.
4. Merati AL, Keppel K, Braun NM, Blumin JH, Kerschner JE. Pediatric voice-related quality of life: findings in healthy children and in common laryngeal disorders. *Ann Otol Rhinol Laryngol*. 2008;117(4):259-262.
5. Murry T, Rosen CA. Outcome measurements and quality of life in voice disorders. *Otolaryngol Clin North Am*. 2000;33(4):905-916.
6. Scott S, Robinson K, Wilson JA, Mackenzie K. Patient-reported problems associated with dysphonia. *Clin Otolaryngol Allied Sci*. 1997;22(1):37-40.
7. Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J voice*. 1999;13(4):557-569.
8. Allen MS, Pettit JM, Sherblom JC. Management of vocal nodules: a regional survey of otolaryngologists and speech-language pathologists. *J Speech, Lang Hear Res*. 1991;34(2):229-235.
9. Ramig LO, Verdolini K. Treatment efficacy: voice disorders. *J Speech, Lang Hear Res*. 1998;41(1):S101--S116.
10. Roy N, Merrill RM, Gray SD, Smith EM. Voice disorders in the general population: prevalence, risk factors, and occupational impact. *Laryngoscope*. 2005;115(11):1988-1995.
11. Cutiva LCC, Vogel I, Burdorf A. Voice disorders in teachers and their associations with work-related factors: a systematic review. *J Commun Disord*. 2013;46(2):143-155.
12. Titze IR. *Principles of Voice Production*. Prentice-Hall, Englewood Cliffs, NJ; 1994.
13. Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. Cengage Learning; 2000.
14. Rothenberg M. Acoustic interaction between the glottal source and the vocal tract. *Vocal fold Physiol*. 1981;1:305-323.

15. Rothenberg M. Source-tract acoustic interaction in breathy voice. In: *Proceedings of the International Conference on Physiology and Biophysics of the Voice, Iowa City, IA.* ; 1983:465-481.
16. Huffman MK. Measures of phonation type in Hmong. *J Acoust Soc Am.* 1987;81(2):495-504.
17. Fischer-Jørgensen E. *Phonetic Analysis of Breathly (Murmured) Vowels in Gujarati.*; 1970.
18. Södersten M, Lindestad P-Å. Glottal closure and perceived breathiness during phonation in normally speaking subjects. *J Speech, Lang Hear Res.* 1990;33(3):601-611.
19. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am.* 1990;87(2):820-857.
20. Alku P, Vilkman E. A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. *Folia Phoniatr Logop.* 1996;48(5):240-254.
21. Shue Y-L, Chen G, Alwan A. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. In: *Eleventh Annual Conference of the International Speech Communication Association.* ; 2010.
22. Bergan CC, Titze IR, Story B. The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice. *J Voice.* 2004;18(3):305-317.
23. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am.* 1988;84(2):511-529. doi:10.1121/1.396829
24. Titze IR. Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *J Speech, Lang Hear Res.* 2006;49:439-447.
25. Schuberth S, Hoppe U, Döllinger M, Lohscheller J, Eysholdt U. High-precision measurement of the vocal fold length and vibratory amplitudes. *Laryngoscope.* 2002;112(6):1043-1049.
26. Holmberg EB, Doyle P, Perkell JS, Hammarberg B, Hillman RE. Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy. *J Voice.* 2003;17(3):269-282. doi:10.1067/S0892-1997(03)00076-6
27. Iwahashi T, Ogawa M, Hosokawa K, Kato C, Inohara H. A detailed motion analysis of the angular velocity between the vocal folds during throat clearing using high-speed digital imaging. *J Voice.* 2016;30(6):770.e1-770.e8.
28. Dromey C, Stathopoulos ET, Sapienza CM. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *J Voice.* 1992;6(1):44-54.

29. Titze IR, Sundberg J. Vocal intensity in speakers and singers. *J Acoust Soc Am*. 1992;91(5):2936-2946. doi:10.1121/1.402929
30. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech-Language Pathol*. 2013;22(2):212-226.
31. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. *J Speech, Lang Hear Res*. 1993;36(1):21-40.
32. De MSB, de Heyning Van PH, Wuyts FL, Lambrechts L. The perceptual evaluation of voice disorders. *Acta Otorhinolaryngol Belg*. 1996;50(4):283-291.
33. Kent RD. Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *Am J Speech-Language Pathol*. 1996;5(3):7-23.
34. Oates J. Auditory-perceptual evaluation of disordered voice quality. *Folia Phoniatr Logop*. 2009;61(1):49-56.
35. Behrman A. Common practices of voice therapists in the evaluation of patients. *J Voice*. 2005;19(3):454-469.
36. Murugappan S, Boyce S, Khosla S, Kelchner L, Gutmark E. Acoustic characteristics of phonation in “wet voice” conditions. *J Acoust Soc Am*. 2010;127(4):2578-2589.
37. Warms T, Richards J. “Wet voice” as a predictor of penetration and aspiration in oropharyngeal dysphagia. *Dysphagia*. 2000;15(2):84-88.
38. Arvedson JC. Feeding children with cerebral palsy and swallowing difficulties. *Eur J Clin Nutr*. 2013;67(S2):S9.
39. Baker BM, Fraser AM, Baker CD. Long-term postoperative dysphagia in oral/pharyngeal surgery patients: subjects’ perceptions vs. videofluoroscopic observations. *Dysphagia*. 1991;6(1):11-16.
40. Kempster GB, Gerratt BR, Abbott KV, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech-Language Pathol*. 2009;18(2):124-132.
41. Takahashi H. rating using the grbas scale. *Japan Soc Logop phoniatr*. 1995.
42. Honjo I, Isshiki N. Laryngoscopic and voice characteristics of aged persons. *Arch Otolaryngol*. 1980;106(3):149-150.
43. Gobl C, Chasaide AN. The role of voice quality in communicating emotion, mood and attitude. *Speech Commun*. 2003;40(1-2):189-212.
44. Laver JDM. Voice quality and indexical information. *Br J Disord Commun*. 1968;3(1):43-

- 54.
45. Yuasa IP. Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *Am Speech*. 2010;85(3):315-337.
  46. Gobl C, Chasaide AN. Acoustic characteristics of voice quality. *Speech Commun*. 1992;11(4-5):481-490.
  47. Kreiman J, Gerratt BR. Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am*. 2000;108(4):1867-1876.
  48. Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am*. 1998;104(3):1598-1608.
  49. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. *J Speech, Lang Hear Res*. 1993;36(1):14-20. doi:10.1044/jshr.3601.14
  50. Chan KMK, Yiu EML. The effect of anchors and training on the reliability of perceptual voice evaluation. *J Speech, Lang Hear Res*. 2002.
  51. Eddins DA, Anand S, Camacho A, Shrivastav R. Modeling of breathy voice quality using pitch-strength estimates. *J Voice*. 2016;30(6):774--e1.
  52. Kopf LM, Skowronski MD, Anand S, Eddins DA, Shrivastav R. The Perception of Breathiness in the Voices of Pediatric Speakers. *J Voice*. 2017.
  53. Lieberman P. Perturbations in vocal pitch. *J Acoust Soc Am*. 1961;33(5):597-603.
  54. Lieberman P. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *J Acoust Soc Am*. 1963;35(3):344-353.
  55. Koike Y. Application of Some Acoustic Measures for the Evaluation of Laryngeal Dysfunction. 1973.
  56. Koike Y, Takahashi H, Calcaterra TC. Acoustic measures for detecting laryngeal pathology. *Acta Otolaryngol*. 1977;84(1-6):105-117.
  57. Wendahl RW. Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. *Folia Phoniatr Logop*. 1966;18(2):98-108.
  58. Qi Y, Weinberg B, Bi N, Hess WJ. Minimizing the effect of period determination on the computation of amplitude perturbation in voice. *J Acoust Soc Am*. 1995;97(4):2525-2532. doi:10.1121/1.411972
  59. Klingholz F. The measurement of the signal-to-noise ratio (SNR) in continuous speech. *Speech Commun*. 1987;6(1):15-26.
  60. Yumoto E, Gould WJ, Baer T. Harmonics-to-noise ratio as an index of the degree of

- hoarseness. *J Acoust Soc Am*. 1982;71(6):1544-1550.
61. Qi Y, Hillman RE. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *J Acoust Soc Am*. 1997;102(1):537-543. doi:10.1121/1.419726
  62. Ebihara S, Ogawa S. Normalized noise energy as an acoustic measure to evaluate pathologic voice. *J Acoust Soc Am*. 1986;80(5):1329-1334. doi:10.1121/1.394384
  63. Michaelis D, Gramss T, Strube HW. Glottal-to-Noise Excitation Ratio - A New Measure for Describing Pathological Voices. *Acustica*. 1997;83(4):700-706.
  64. Ghasemzadeh H, Arjmandi MK. Toward Optimum Quantification of Pathology-induced Noises: An Investigation of Information Missed by Human Auditory System. *IEEE/ACM Trans Audio, Speech, Lang Process*. 2020;28:519-528.
  65. Klich RJ. Relationships of vowel characteristics to listener ratings of breathiness. *J Speech, Lang Hear Res*. 1982;25(4):574-580.
  66. Stevens KN. Physics of laryngeal behavior and larynx modes. *Phonetica*. 1977;34(4):264-279.
  67. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech, Lang Hear Res*. 1994;37(4):769-778.
  68. Godino-Llorente JJ, Gómez-Vilda P. Automatic Detection of Voice Impairments by Means of Short-Term Cepstral Parameters and Neural Network Based Detectors. *IEEE Trans Biomed Eng*. 2004;51(2):380-384. doi:10.1109/TBME.2003.820386
  69. Arjmandi MK, Pooyan M. An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomed Signal Process Control*. 2012;7(1):3-19. doi:10.1016/j.bspc.2011.03.010
  70. Ghasemzadeh H, Searl J. Modeling Dynamics of Connected Speech in Time and Frequency Domains with Application to ALS. *11th Int Conf Voice Physiol Biomech*. 2018;(August).
  71. Vaziri G, Almasganj F, Behroozmand R. Pathological assessment of patients' speech signals using nonlinear dynamical analysis. *Comput Biol Med*. 2010;40(1):54-63.
  72. Jiang JJ, Zhang Y. Nonlinear dynamic analysis of speech from pathological subjects. *Electron Lett*. 2002;38(6):294-295. doi:10.1049/e1
  73. Ghasemzadeh H, Tajik Khass M, Khalil Arjmandi M, Pooyan M. Detection of vocal disorders based on phase space parameters and Lyapunov spectrum. *Biomed Signal Process Control*. 2015;22:135-145. doi:10.1016/j.bspc.2015.07.002
  74. Kalman RE. On the general theory of control systems. In: *Proceedings First International Conference on Automatic Control, Moscow, USSR*. ; 1960.



75. Kalman RE. Mathematical description of linear dynamical systems. *J Soc Ind Appl Math Ser A Control*. 1963;1(2):152-192.
76. Lindblom B, Sundberg J. Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement. *J Acoust Soc Am*. 2005;48(1A):120-120. doi:10.1121/1.1974958
77. Stevens KN, House AS. Development of a Quantitative Description of Vowel Articulation. *J Acoust Soc Am*. 2005;27(3):484-493. doi:10.1121/1.1907943
78. Yunusova Y, Rosenthal JS, Rudy K, Baljko M, Daskalogiannakis J. Positional targets for lingual consonants defined using electromagnetic articulography. *J Acoust Soc Am*. 2012;132(2):1027-1038. doi:10.1121/1.4733542
79. Stevens KN. On the quantal nature of speech. *J Phonetics*. 1989;17:3-45.
80. Stevens KN, Keyser SJ. Quantal theory , enhancement and overlap. *J Phon*. 2010;38(1):10-19. doi:10.1016/j.wocn.2008.10.004
81. Honda K, Takano S, Takemoto H. Effects of side cavities and tongue stabilization: Possible extensions of the quantal theory. *J Phon*. 2010;38(1):33-43.
82. Fujimura O. Remarks on quantitative description of the lingual articulation. *Front speech Commun Res*. 1978:17-24.
83. Perkell JS, Matthies ML, Tiede M, et al. The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *J speech, Lang Hear Res*. 2004.
84. Gick B, Stavness I, Chiu C, Fels S. Categorical variation in lip posture is determined by quantal biomechanical-articulatory relations. *Can Acoust*. 2011;39(3):178-179.
85. Moisik SR, Gick B. The Quantal Larynx: The Stable Regions of Laryngeal Biomechanics and Implications for Speech Production. *J Speech, Lang Hear Res*. 2017;60(3):540-560. doi:10.1044/2016\_jslhr-s-16-0019
86. Moisik S, Gick B. The quantal larynx revisited. *J Acoust Soc Am*. 2013;133(5):3522-3522. doi:10.1121/1.4806322
87. Perkell JS. Movement goals and feedback and feedforward control mechanisms in speech production. *J Neurolinguistics*. 2012;25(5):382-407.
88. Williamson G. *Human Communication: A Linguistic Introduction*. Speechmark; 2001.
89. Deliyski DD, Powell MEG, Zacharias SRC, Gerlach TT, De Alarcon A. Experimental investigation on minimum frame rate requirements of high-speed videoendoscopy for clinical voice assessment. *Biomed Signal Process Control*. 2015;17:21-28. doi:10.1016/j.bspc.2014.11.007

90. Zacharias SRC, Deliyski DD, Gerlach TT. Utility of laryngeal high-speed videoendoscopy in clinical voice assessment. *J Voice*. 2018;32(2):216-220.
91. Bonilha HS, Deliyski DD. Mucosal wave: A normophonic study across visualization techniques. *J Voice*. 2008;22(1):23-33.
92. Olthoff A, Woywod C, Kruse E. Stroboscopy versus high-speed glottography: a comparative study. *Laryngoscope*. 2007;117(6):1123-1126.
93. Powell ME, Deliyski DD, Zeitels SM, et al. Efficacy of Videostroboscopy and High-Speed Videoendoscopy to Obtain Functional Outcomes From Perioperative Ratings in Patients With Vocal Fold Mass Lesions. *J Voice (in Press)*. 2019. doi:10.1016/j.jvoice.2019.03.012
94. Bonilha HS, Deliyski DD, Whiteside JP, Gerlach TT. Vocal Fold Phase Asymmetries in Patients With Voice Disorders: A Study Across Visualization Techniques. *Am J Speech-Language Pathol*. 2012;21(1):3-15. doi:10.1044/1058-0360(2011/09-0086)
95. Rosen CA. Stroboscopy as a research instrument: development of a perceptual evaluation tool. *Laryngoscope*. 2005;115(3):423-428.
96. Bonilha HS, O'Shields M, Gerlach TT, Deliyski DD. Arytenoid adduction asymmetries in persons with and without voice disorders. *Logop Phoniatr Vocology*. 2009;34(3):128-134. doi:10.1080/14015430903150210
97. Braunschweig T, Flaschka J, Schelhorn-Neise P, Döllinger M. High-speed video analysis of the phonation onset, with an application to the diagnosis of functional dysphonias. *Med Eng Phys*. 2008;30(1):59-66.
98. Mehta DD, Deliyski DD, Quatieri TF, Hillman RE. Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings. *J Speech, Lang Hear Res*. 2011;54(1):47-54.
99. Verikas A, Gelzinis A, Bacauskiene M, Uloza V. Integrating global and local analysis of color, texture and geometrical information for categorizing laryngeal images. *Int J Pattern Recognit Artif Intell*. 2006;20(08):1187-1205.
100. Orlikoff RF, Deliyski DD, Baken RJ, Watson BC. Validation of a Glottographic Measure of Vocal Attack. *J Voice*. 2009;23(2):164-168. doi:10.1016/j.jvoice.2007.08.004
101. Lohscheller J, Švec JG, Döllinger M. Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects. *Logop Phoniatr Vocology*. 2013;38(4):182-192.
102. Patel RR, Dubrovskiy D, Döllinger M. Measurement of glottal cycle characteristics between children and adults: physiological variations. *J Voice*. 2014;28(4):476-486.
103. Patel R, Donohue KD, Unnikrishnan H, Kryscio RJ. Kinematic measurements of the vocal-fold displacement waveform in typical children and adult populations: quantification of

- high-speed endoscopic videos. *J Speech, Lang Hear Res.* 2015;58(2):227-240.
104. Hillman RE, Mehta DD. The science of stroboscopic imaging. In: Kendall KA, Leonard RJ, eds. *Laryngeal Evaluation: Indirect Laryngoscopy to High-Speed Digital Imaging*. Thieme New York, NY; 2010:101-109.
  105. Deliyski D. Laryngeal high-speed videoendoscopy. In: Kendall K, Leonard R, eds. *Laryngeal Evaluation: Indirect Laryngoscopy to High-Speed Digital Imaging*. Thieme Medical, New York, NY; 2010:245-270.
  106. Sprecher A, Olszewski A, Jiang JJ, Zhang Y. Updating signal typing in voice: addition of type 4 signals. *J Acoust Soc Am.* 2010;127(6):3710-3716.
  107. Mehta DD, Deliyski DD, Hillman RE. Why Laryngeal Stroboscopy Really Works: Clarifying Misconceptions Surrounding Talbot's Law and the Persistence of Vision. *J Speech, Lang Hear Res.* 2010;53(5):1263-1267. doi:[https://doi.org/10.1044/1092-4388\(2010/09-0241\)](https://doi.org/10.1044/1092-4388(2010/09-0241))
  108. Mehta DD, Hillman RE. Current role of stroboscopy in laryngeal imaging. *Curr Opin Otolaryngol Head Neck Surg.* 2012;20(6):429.
  109. Deliyski DD, Hillman RE. State of the art laryngeal imaging: research and clinical implications. *Curr Opin Otolaryngol Head Neck Surg.* 2010;18(3):147.
  110. Patel RR, Eadie T, Paul D, et al. Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *Am J Speech-Language Pathol.* 2018;27(3):887-905. doi:10.1044/2018\_ajslp-17-0009
  111. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE. Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. *Folia Phoniatr Logop.* 2008;60(1):33-44.
  112. Švec JG, Schutte HK. Videokymography: high-speed line scanning of vocal fold vibration. *J Voice.* 1996;10(2):201-205.
  113. Švec JG, Šram F, Schutte HK. Videokymography in Voice Disorders: What to Look For? *Ann Otol Rhinol Laryngol.* 2007;116(3):172-180.
  114. Golla ME, Deliyski DD, Orlikoff RF, Moukalled HJ. Objective comparison of the electroglottogram to synchronous high-speed images of vocal-fold contact during vibration. *Model Anal Vocal Emiss Biomed Appl - 6th Int Work MAVEBA 2009.* 2009;9:1-4.
  115. Mehta DD, Zañartu M, Quatieri TF, Deliyski DD, Hillman RE. Investigating acoustic correlates of human vocal fold vibratory phase asymmetry through modeling and laryngeal high-speed videoendoscopy. *J Acoust Soc Am.* 2011;130(6):3999-4009. doi:10.1121/1.3658441

116. Naghibolhosseini M, Deliyski DD, Zacharias SRC, de Alarcon A, Orlikoff RF. Temporal Segmentation for Laryngeal High-Speed Videoendoscopy in Connected Speech. *J Voice*. 2018;32(2):256--e1.
117. Mehta DD, Deliyski DD, Zeitels SM, Quatieri TF, Hillman RE. Voice production mechanisms following phonosurgical treatment of early glottic cancer. *Ann Otol Rhinol Laryngol*. 2010;119(1):1-9.
118. Schwarz R, Döllinger M, Wurzbacher T, Eysholdt U, Lohscheller J. Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *J Acoust Soc Am*. 2008;123(5):2717-2732.
119. Yan Y, Damrose E, Bless D. Functional analysis of voice using simultaneous high-speed imaging and acoustic recordings. *J Voice*. 2007;21(5):604-616.
120. Skalski A, Zielinski T, Deliyski D. Analysis of vocal folds movement in high speed videoendoscopy based on level set segmentation and image registration. In: *2008 International Conference on Signals and Electronic Systems*. ; 2008:223-226.
121. Yan Y, Chen X, Bless D. Automatic tracing of vocal-fold motion from high-speed digital images. *IEEE Trans Biomed Eng*. 2006;53(7):1394-1400.
122. Ghasemzadeh H, Deliyski DD, Ford DS, Kobler JB, Hillman RE, Mehta DD. Method for Vertical Calibration of Laser-Projection Transnasal Fiberoptic High-Speed Videoendoscopy. *J Voice*. 2020;34(6):847-861. PMID: 31151853; PMCID: PMC6883161.
123. Moukalled HJ, Deliyski DD, Schwarz RR, Wang S. Segmentation of Laryngeal High-Speed Videoendoscopy in Temporal Domain Using Paired Active Contours. *Sixth Int Work Model Anal Vocal Emiss Biomed Appl MAVEBA*. 2009;9(d):137-140.
124. Karakozoglou S-Z, Henrich N, D Alessandro C, Stylianou Y. Automatic glottal segmentation using local-based active contours and application to glottovibrography. *Speech Commun*. 2012;54(5):641-654.
125. Deliyski DD. Endoscope motion compensation for laryngeal high-speed videoendoscopy. *J Voice*. 2005;19(3):485-496. doi:10.1016/j.jvoice.2004.07.006
126. Sulica L. Laryngoscopy, stroboscopy and other tools for the evaluation of voice disorders. *Off Proced Laryngol An Issue Otolaryngol Clin*. 2012;46(1):21.
127. Milstein CF, Charbel S, Hicks DM, Abelson TI, Richter JE, Vaezi MF. Prevalence of laryngeal irritation signs associated with reflux in asymptomatic volunteers: impact of endoscopic technique (rigid vs. flexible laryngoscope). *Laryngoscope*. 2005;115(12):2256-2261.
128. Yanagisawa E, Yanagisawa K. Stroboscopic videolaryngoscopy: A comparison of fiberoptic and telescopic documentation. *Ann Otol Rhinol Laryngol*. 1993;102(4):255-265.

129. Eller R, Ginsburg M, Lurie D, Heman-Ackah Y, Lyons K, Sataloff R. Flexible laryngoscopy: a comparison of fiber optic and distal chip technologies part 2: laryngopharyngeal reflux. *J Voice*. 2009;23(3):389-395.
130. Chandran S, Hanna J, Lurie D, Sataloff RT. Differences between flexible and rigid endoscopy in assessing the posterior glottic chink. *J Voice*. 2011;25(5):591-595.
131. Ng ML, Bailey RL. Acoustic changes related to laryngeal examination with a rigid telescope. *Folia Phoniatr Logop*. 2006;58(5):353-362.
132. Kobler JB, Zeitels SM, Hillman RE, Kuo J. Assessment of vocal function using simultaneous aerodynamic and calibrated videostroboscopic measures. *Ann Otol Rhinol Laryngol*. 1998;107(6):477-485.
133. Mehta DD, Deliyski DD, Zeitels SM, Zañartu M, Hillman RE. Integration of transnasal fiberoptic high-speed videoendoscopy with time-synchronized recordings of vocal function. *ePhonoscope*. 2015:105-114.
134. Zañartu M, Mehta DD, Ho JC, Wodicka GR, Hillman RE. Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: a case study. *J Acoust Soc Am*. 2011;129(1):326-339.
135. Rosen CA, Murry T. Diagnostic laryngeal endoscopy. *Otolaryngol Clin North Am*. 2000;33(4):751-757.
136. Gray SD, Smith ME, Schneider H. Voice disorders in children. *Pediatr Clin North Am*. 1996;43(6):1357-1384.
137. Chait DH, Lotz WK. Successful pediatric examinations using nasoendoscopy. *Laryngoscope*. 1991;101(9):1016-1018.
138. Clark BS, Gao WZ, Bertelsen C, et al. Flexible versus rigid laryngoscopy: A randomized crossover study comparing patient experience. *Laryngoscope*. 2020.
139. Rothenberg M. Source-tract acoustic interaction and voice quality. In: *Transcripts of the 12th Symposium Care of Professional Voice, Part I*. New York, NY: The Voice Foundation. ; 1983:25-31.
140. Ben-David BM, Icht M. Voice Changes in Real Speaking Situations during a Day, with and Without Vocal Loading: Assessing Call Center Operators. *J Voice*. 2016;30(2):247e1-247e11. doi:10.1016/j.jvoice.2015.04.002
141. Laukkanen AM, Ilomäki I, Leppänen K, Vilkman E. Acoustic Measures and Self-reports of Vocal Fatigue by Female Teachers. *J Voice*. 2008;22(3):283-289. doi:10.1016/j.jvoice.2006.10.001
142. Laukkanen AM, Kankare E. Vocal loading-related changes in male teachers' voices investigated before and after a working day. *Folia Phoniatr Logop*. 2006;58(4):229-239.

doi:10.1159/000093180

143. Laukkanen A-M, Jarvinen K, Artkoski M, et al. Changes in Voice and Subjective Sensations during a 45-min Vocal Loading Test in Female Subjects with Vocal Training. *Folia Phoniatr e Logop.* 2004.
144. Jonsdottir V, Laukkanen A-M, Siiki I. Changes in Teachers ' Speech during a Working Day with and without Electric Sound Amplification. *Folia Phoniatr e Logop.* 2003;601:282-287. doi:10.1159/000066149
145. Lehto L, Laaksonen L, Vilkmann E, Alku P. Occupational voice complaints and objective acoustic measurements - Do they correlate? *Logop Phoniatr Vocology.* 2006;31(4):147-152. doi:10.1080/14015430600654654
146. Wolfe VI, Long J, Youngblood HC, Williford H, Olson MS. Vocal parameters of aerobic instructors with and without voice problems. *J Voice.* 2002;16(1):52-60. doi:10.1016/S0892-1997(02)00072-3
147. Yang A, Stingl M, Berry DA, et al. Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *J Acoust Soc Am.* 2011;130(2):948-964.
148. Yang A, Lohscheller J, Berry DA, et al. Biomechanical modeling of the three-dimensional aspects of human vocal fold dynamics. *J Acoust Soc Am.* 2010;127(2):1014-1031.
149. Šidlof P, Švec JG, Horáček J, Veselý J, Klepáček I, Havlík R. Geometry of human vocal folds and glottal channel for mathematical and biomechanical modeling of voice production. *J Biomech.* 2008;41(5):985-995.
150. Titze IR, Alipour F. *The Myoelastic Aerodynamic Theory of Phonation.* National Center for Voice and Speech; 2006.
151. Thomson SL, Mongeau L, Frankel SH. Aerodynamic transfer of energy to the vocal folds. *J Acoust Soc Am.* 2005;118(3):1689-1700.
152. Fulcher LP, Scherer RC. Phonation threshold pressure: Comparison of calculations and measurements taken with physical models of the vocal fold mucosa. *J Acoust Soc Am.* 2011;130(3):1597-1605.
153. Patel RR, Donohue KD, Lau D, Unnikrishnan H. In vivo measurement of pediatric vocal fold motion using structured light laser projection. *J Voice.* 2013;27(4):463-472.
154. Verdolini Abbott K, K., Hersan, R., Hammer, D., & Potter Reed J. *Adventures in Voice: A whole new way of doing things for kids.* 2015.
155. Selby JC, Gilbert HR, Lerman JW. Perceptual and acoustic evaluation of individuals with laryngopharyngeal reflux pre-and post-treatment. *J Voice.* 2003;17(4):557-570.

156. Schindler A, Mozzanica F, Maruzzi P, Atac M, De Cristofaro V, Ottaviani F. Multidimensional assessment of vocal changes in benign vocal fold lesions after voice therapy. *Auris Nasus Larynx*. 2013;40(3):291-297.
157. Rydell R, Schalén L, Fex S, Elnér Å. Voice evaluation before and after laser excision vs. radiotherapy of T1A glottic carcinoma. *Acta Otolaryngol*. 1995;115(4):560-565.
158. Chen SH, Hsiao T-Y, Hsiao L-C, Chung Y-M, Chiang S-C. Outcome of resonant voice therapy for female teachers with voice disorders: Perceptual, physiological, acoustic, aerodynamic, and functional measurements. *J Voice*. 2007;21(4):415-425.
159. Fex B, Fex S, Shiromoto O, Hirano M. Acoustic analysis of functional dysphonia: Before and after voice therapy (accent method). *J Voice*. 1994;8(2):163-167. doi:10.1016/S0892-1997(05)80308-X
160. Tezcaner CZ, Ozgursoy SK, Sati I, Dursun G. Changes after voice therapy in objective and subjective voice measurements of pediatric patients with vocal nodules. *Eur Arch Oto-Rhino-Laryngology*. 2009;266(12):1923-1927.
161. Roy N, Bless DM, Heisey D, Ford CN. Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short-and long-term treatment outcomes. *J Voice*. 1997;11(3):321-331.
162. Gillespie AI, Dastolfo C, Magid N, Gartner-Schmidt J. Acoustic analysis of four common voice diagnoses: moving toward disorder-specific assessment. *J Voice*. 2014;28(5):582-588.
163. Holmberg EB, Hillman RE, Perkell JS, Guiod PC, Goldman SL. Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice. *J Speech, Lang Hear Res*. 1995;38(6):1212-1223. doi:10.1044/jshr.3806.1212
164. Döllinger M, Gómez P, Patel RR, Alexiou C, Bohr C, Schützenberger A. Biomechanical simulation of vocal fold dynamics in adults based on laryngeal high-speed videoendoscopy. *PLoS One*. 2017;12(11):e0187486.
165. Bohr C, Kraeck A, Eysholdt U, Ziethe A, Döllinger M. Quantitative analysis of organic vocal fold pathologies in females by high-speed endoscopy. *Laryngoscope*. 2013;123(7):1686-1693.
166. Stevens KN. *Acoustic Phonetics*. Vol 30. MIT press; 2000.
167. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur Arch Oto-rhino-laryngology*. 2001;258(2):77-82.
168. Rosen CA, Gartner-Schmidt J, Hathaway B, et al. A nomenclature paradigm for benign midmembranous vocal fold lesions. *Laryngoscope*. 2012;122(6):1335-1341.

doi:10.1002/lary.22421

169. Naunheim MR, Carroll TL. Benign vocal fold lesions: Update on nomenclature, cause, diagnosis, and treatment. *Curr Opin Otolaryngol Head Neck Surg*. 2017;25(6):453-458. doi:10.1097/MOO.0000000000000408
170. Spiegel JR, Sataloff RT, Hawkshaw MJ. Stroboscovideolaryngoscopy: results and clinical value. *Ann Otol Rhinol Laryngol*. 1991;100(9):725-727.
171. Woo P, Colton R, Casper J, Brewer D. Diagnostic value of stroboscopic examination in hoarse patients. *J voice*. 1991;5(3):231-238.
172. Titze IR. The physics of small-amplitude oscillation of the vocal folds. *J Acoust Soc Am*. 1988;83(4):1536-1552.
173. Titze IR, Talkin DT. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *J Acoust Soc Am*. 1979;66(1):60-74.
174. Titze IR, Jiang JJ, Hsiao T-Y. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. *Ann Otol Rhinol Laryngol*. 1993;102(1):58-63.
175. Boutin H, Smith J, Wolfe J. Laryngeal flow due to longitudinal sweeping motion of the vocal folds and its contribution to auto-oscillation. *J Acoust Soc Am*. 2015;138(1):146-149.
176. Hirano M. Phonosurgery: basic and clinical investigations. *Otol*. 1975;21:239-242.
177. Krausert CR, Olszewski AE, Taylor LN, McMurray JS, Dailey SH, Jiang JJ. Mucosal wave measurement and visualization techniques. *J Voice*. 2011;25(4):395-405.
178. Titze IR. Phonation threshold pressure: A missing link in glottal aerodynamics. *J Acoust Soc Am*. 1992;91(5):2926-2935.
179. Verdolini-Marston K, Titze IR, Druker DG. Changes in phonation threshold pressure with induced conditions of hydration. *J voice*. 1990;4(2):142-151.
180. Chan RW, Titze IR. Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics. *J Acoust Soc Am*. 2006;119(4):2351-2362.
181. Imaging H. 28 Laryngeal High-Speed Videoendoscopy. *Laryngeal Eval*. 2014. doi:10.1055/b-0034-81468
182. Eller R, Ginsburg M, Lurie D, Heman-Ackah Y, Lyons K, Sataloff R. Flexible laryngoscopy: a comparison of fiber optic and distal chip technologies. Part 1: vocal fold masses. *J Voice*. 2008;22(6):746-750.
183. Yamauchi A, Yokonishi H, Imagawa H, et al. Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging. *J Voice*. 2016;30(2):205-214.
184. Powell ME, Deliyski DD, Hillman RE, Zeitels SM, Burns JA, Mehta DD. Comparison of



- Videostroboscopy to Stroboscopy Derived From High-Speed Videoendoscopy for Evaluating Patients With Vocal Fold Mass Lesions. 2016;25(Andrade 2009):2011-2013. doi:10.1044/2016
185. Gardner GM, Parnes SM. Status of the mucosal wave post vocal cord injection versus thyroplasty. *J Voice*. 1991;5(1):64-73.
  186. Rihkanen H, Reijonen P, Lehtikainen-Söderlund S, Lauri E-R. Videostroboscopic assessment of unilateral vocal fold paralysis after augmentation with autologous fascia. *Eur Arch Oto-Rhino-Laryngology Head Neck*. 2004;261(4):177-183.
  187. Hsiung M-W, Kang B-H, Su W-F, Pai LU, Lin Y-H. Combination of fascia transplantation and fat injection into the vocal fold for sulcus vocalis: long-term results. *Ann Otol Rhinol Laryngol*. 2004;113(5):359-366.
  188. González-Herranz R, García EH, Granda-Rosales M, Eisenberg-Plaza G, Woodeson JM, Plaza G. Improved mucosal wave in unilateral autologous temporal fascia graft in sulcus vocalis type 2 and vocal scars. *J Voice*. 2019;33(6):915-922.
  189. Tsuji DH, de Almeida ER, Sennes LU, Butugan O, Pinho SMR. Comparison between thyroplasty type I and Arytenoid rotation: a study of vocal fold vibration using excised human larynges. *J Voice*. 2003;17(4):596-604.
  190. Schade G, Leuwer R, Kraas M, Rassow B, Hess MM. Laryngeal morphometry with a new laser “clip on” device. *Lasers Surg Med Off J Am Soc Laser Med Surg*. 2004;34(5):363-367.
  191. Kobler JB, Rosen DI, Burns JA, et al. Comparison of a flexible laryngoscope with calibrated sizing function to intraoperative measurements. *Ann Otol Rhinol Laryngol*. 2006;115(10):733-740.
  192. Herzon GD, Zeale DL. New laser ruler instrument for making measurements through an endoscope. *Otolaryngol Neck Surg*. 1997;116(6):689-692.
  193. Hertegård S. Measurement of human vocal fold vibrations with laser triangulation. *Opt Eng*. 2002;40(9):2041. doi:10.1117/1.1396324
  194. Luegmair G, Kniesburges S, Zimmermann M, Sutor A, Eysholdt U, Dollinger M. Optical reconstruction of high-speed surface dynamics in an uncontrollable environment. *IEEE Trans Med Imaging*. 2010;29(12):1979-1991.
  195. Deliyski DD, Shishkov M, Mehta DD, Ghasemzadeh H, Bouma B, Zañartu M, de Alarcon A, Hillman RE. Laser-Calibrated System for Transnasal Fiberoptic Laryngeal High-Speed Videoendoscopy. *J Voice*. 2019 Aug 2:S0892-1997(19)30278-4. Epub ahead of print. doi: 10.1016/j.jvoice.2019.07.013. PMID: 31383516; PMCID: PMC6995434.
  196. Awan SN, Roy N, Jetté ME, Meltzner GS, Hillman RE. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual

- judgements from the CAPE-V. *Clin Linguist Phon.* 2010;24(9):742-758.
197. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J Speech, Lang Hear Res.* 1996;39(2):311-321.
  198. Eadie TL, Doyle PC. Classification of dysphonic voice: acoustic and auditory-perceptual measures. *J Voice.* 2005;19(1):1-14.
  199. Peterson EA, Roy N, Awan SN, Merrill RM, Banks R, Tanner K. Toward validation of the cepstral spectral index of dysphonia (CSID) as an objective treatment outcomes measure. *J Voice.* 2013;27(4):401-410.
  200. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J voice.* 2010;24(5):540-555.
  201. Godino-Llorente JJ, Gomez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Trans Biomed Eng.* 2006;53(10):1943-1953. doi:10.1109/TBME.2006.871883
  202. Noordzij JP, Woo P. Glottal area waveform analysis of benign vocal fold lesions before and after surgery. *Ann Otol Rhinol Laryngol.* 2000;109(5 Pt 1):441-446. doi:10.1177/000348940010900501
  203. Patel RR, Unnikrishnan H, Donohue KD. Effects of vocal fold nodules on glottal cycle measurements derived from high-speed videoendoscopy in children. *PLoS One.* 2016;11(4):e0154586.
  204. Hibi SR, Bless DM, Hirano M, Yoshida T. Distortions of videofiberscopy imaging: reconsideration and correction. *J Voice.* 1988;2(2):168-175.
  205. Deng JJ, Hadwin PJ, Peterson SD. The effect of high-speed videoendoscopy configuration on reduced-order model parameter estimates by Bayesian inference. *J Acoust Soc Am.* 2019;146(2):1492-1502.
  206. Speyer R, Wieneke GH, Kersing W, Dejonckere PH. Accuracy of measurements on digital videostroboscopic images of the vocal folds. *Ann Otol Rhinol Laryngol.* 2005;114(6):443-450.
  207. Alzamendi GA, Manriquez R, Hadwin PJ, et al. Bayesian estimation of vocal function measures using laryngeal high-speed videoendoscopy and glottal airflow estimates: An in vivo case study. *J Acoust Soc Am.* 2020;147(5):EL434--EL439.
  208. Ghasemzadeh H, Deliyski DD. Non-Linear Image Distortions in Flexible Fiberoptic Endoscopes and their Effects on Calibrated Horizontal Measurements Using High-Speed Videoendoscopy. *J Voice.* 2020 Sep 18:S0892-1997(20)30331-3. Epub ahead of print. doi: 10.1016/j.jvoice.2020.08.029. PMID: 32958427.

209. Ghasemzadeh H, Deliyski D, Hillman RE, Mehta DD. Method for Horizontal Calibration of Laser-Projection Transnasal Fiberoptic High-Speed Videoendoscopy. *In Preparation*.
210. Johns MM. Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. *Curr Opin Otolaryngol Head Neck Surg*. 2003;11(6):456-461. doi:10.1097/00020840-200312000-00009
211. Oppenheim Alan V, Willsky Alan S, Hamid Nawab S. Signals and systems. *ISBN-10, Pearson Press USA*. 1996.
212. Hsiao T-Y, Wang C-L, Chen C-N, Hsieh F-J, Shau Y-W. Noninvasive assessment of laryngeal phonation function using color Doppler ultrasound imaging. *Ultrasound Med Biol*. 2001;27(8):1035-1040.
213. DeJonckere PH, Lebacq J. Vocal Fold Collision Speed in vivo: The Effect of Loudness. *J Voice*. 2020. doi:10.1016/j.jvoice.2020.08.025
214. Subbotina M V. Evaluation the velocity of vocal fold movements in adults by duplex Doppler scanning. *Vestn Otorinolaringol*. 2019;84(5):38-43.
215. Colton RH, Woo P, Brewer DW, Griffin B, Casper J. Stroboscopic signs associated with benign lesions of the vocal folds. *J Voice*. 1995;9(3):312-325. doi:10.1016/S0892-1997(05)80240-1
216. Wallis L, Jackson-Menaldi C, Holland W, Giraldo A. Vocal fold nodule vs. vocal fold polyp: Answer from surgical pathologist and voice pathologist point of view. *J Voice*. 2004;18(1):125-129. doi:10.1016/j.jvoice.2003.07.003
217. Benninger MS. Microdissection or Microspot CO 2 Laser for Limited Vocal Fold Benign Lesions: A Prospective Randomized Trial. *Laryngoscope*. 2000;110(S92):1-1. doi:10.1097/00005537-200002001-00001
218. Altman KW. Vocal Fold Masses. *Otolaryngol Clin North Am*. 2007;40(5):1091-1108. doi:10.1016/j.otc.2007.05.011
219. Dejonckere PH, Kob M. Pathogenesis of vocal fold nodules: New insights from a modelling approach. *Folia Phoniatr Logop*. 2009;61(3):171-179. doi:10.1159/000219952
220. De Vries MP, Schutte HK, Veldman AEP, Verkerke GJ. Glottal flow through a two-mass model: comparison of Navier--Stokes solutions with simplified models. *J Acoust Soc Am*. 2002;111(4):1847-1853.
221. Benninger MS, Alessi D, Archer S, et al. Vocal fold scarring: current concepts and management. *Otolaryngol - Head Neck Surg*. 1996;115(5):474-482.
222. Rousseau B, Hirano S, Scheidt TD, et al. Characterization of vocal fold scarring in a canine model. *Laryngoscope*. 2003;113(4):620-627.

223. Cavallo SA, Baken RJ. Prephonatory laryngeal and chest wall dynamics. *J Speech, Lang Hear Res.* 1985;28(1):79-87.
224. Shiba TL, Chhetri DK. Dynamics of phonatory posturing at phonation onset. *Laryngoscope.* 2016;126(8):1837-1843.
225. Chhetri DK, Neubauer J, Berry DA. Neuromuscular control of fundamental frequency and glottal posture at phonation onset. *J Acoust Soc Am.* 2012;131(2):1401-1412.
226. Faaborg-Andersen K. Electromyography of laryngeal muscles in humans. technics and results. *Aktuel Probl Phoniatr Logop.* 1965;12:1.
227. Deliyski D, Petrushev P. Methods for objective assessment of high-speed videoendoscopy. *Proc Adv Quant Laryngol.* 2003:1-16.
228. Titze IR. Mechanical stress in phonation. *J Voice.* 1994;8(2):99-105.
229. Sapienza C, Ruddy BH. *Voice Disorders.* Plural Publishing; 2016.
230. Hunter EJ, Titze IR, Alipour F. A three-dimensional model of vocal fold abduction/adduction. *J Acoust Soc Am.* 2004;115(4):1747-1759.
231. Manneberg G, Hertegard S, Liljencrantz J. Measurement of human vocal fold vibrations with laser triangulation. *Opt Eng.* 2001;40(9):2041-2045.
232. Larsson H, Hertegård S. Calibration of high-speed imaging by laser triangulation. *Logop Phoniatr Vocology.* 2004;29(4):154-161.
233. George NA, de Mul FFM, Qiu Q, Rakhorst G, Schutte HK. New laryngoscope for quantitative high-speed imaging of human vocal folds vibration in the horizontal and vertical direction. *J Biomed Opt.* 2008;13(6):64024. doi:10.1117/1.3041164
234. Wurzbacher T, Voigt I, Schwarz R, et al. Calibration of laryngeal endoscopic high-speed image sequences by an automated detection of parallel laser line projections. *Med Image Anal.* 2008;12(3):300-317.
235. Semmler M, Kniesburges S, Birk V, Ziethe A, Patel R, Döllinger M. 3D reconstruction of human laryngeal dynamics based on endoscopic high-speed recordings. *IEEE Trans Med Imaging.* 2016;35(7):1615-1624.
236. Luegmair G, Mehta DD, Kobler JB, Döllinger M. Three-Dimensional Optical Reconstruction of Vocal Fold Kinematics Using High-Speed Video With a Laser Projection System. *IEEE Trans Med Imaging.* 2015;34(12):2572-2582.
237. Ji Z, Leu M-C. Design of optical triangulation devices. *Opt Laser Technol.* 1989;21(5):339-341.
238. Smith WJ, Smith WJ. *Modern Optical Engineering.* (3rd, ed.). McGraw-hill New York;

- 2000.
239. Bayer BE. Color imaging array. 1976.
  240. Atherton TJ, Kerbyson DJ. Size invariant circle detection. *Image Vis Comput.* 1999;17(11):795-803.
  241. Yuen HK, Princen J, Illingworth J, Kittler J. Comparative study of Hough transform methods for circle finding. *Image Vis Comput.* 1990;8(1):71-77.
  242. Duda RO, Hart PE. *Use of the Hough Transformation to Detect Lines and Curves in Pictures.*; 1971.
  243. Ballard DH. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognit.* 1981;13(2):111-122.
  244. Dougherty ER, Lotufo RA. *Hands-on Morphological Image Processing.* Vol 59. SPIE press; 2003.
  245. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med.* 2010;363(4):301-304.
  246. Neal ML, Kerckhoffs R. Current progress in patient-specific modeling. *Brief Bioinform.* 2009;11(1):111-126.
  247. Kendall KA, Leonard RJ, eds. *Laryngeal Evaluation: Indirect Laryngoscopy to High-Speed Digital Imaging.* Thieme; 2011.
  248. Dailey SH, Kobler JB, Hillman RE, et al. Endoscopic measurement of vocal fold movement during adduction and abduction. *Laryngoscope.* 2005;115(1):178-183.
  249. Bonilha HS, Deliyski DD, Gerlach TT. Phase asymmetries in normophonic speakers: visual judgments and objective findings. *Am J Speech-Language Pathol.* 2008;17(4):367-376.
  250. Fannin TE, Grosvenor T. *Clinical Optics.* Butterworth-Heinemann; 2013.
  251. Field A, Miles J, Field Z. *Discovering Statistics Using R.* Sage publications; 2012.
  252. Wilcox RR. *Introduction to Robust Estimation and Hypothesis Testing.* Academic press; 2011.
  253. Patel RR, Donohue KD, Johnson WC, Archer SM. Laser projection imaging for measurement of pediatric voice. *Laryngoscope.* 2011;121(11):2411-2417.
  254. Bonilha HS, Focht KL, Martin-Harris B. Rater methodology for stroboscopy: a systematic review. *J Voice.* 2015;29(1):101-108.
  255. Carlson JN, Das S, la Torre F, Callaway CW, Phrampus PE, Hodgins J. Motion capture measures variability in laryngoscopic movement during endotracheal intubation: a

- preliminary report. *Simul Healthc J Soc Simul Healthc*. 2012;7(4):255.
256. Stepp CE, Hillman RE, Heaton JT. A virtual trajectory model predicts differences in vocal fold kinematics in individuals with vocal hyperfunction. *J Acoust Soc Am*. 2010;127(5):3166-3176.