University of West Bohemia
Faculty of Applied Sciences

Glottis Detection and Evaluation in High-Speed Video Recording

Ing. Tomáš Ettler

DOCTORAL THESIS
Submitted in partial fulfillment of the requirements for a degree of Doctor of Philosophy in Computer Science

Supervised by Ing. Pavel Nový Ph.D.
Department of Computer Science and Engineering

Plzeň, 2022

Západočeská univerzita v Plzni
Fakulta aplikovaných věd

Detekce a vyhodnocení vysokorychlostního videozáznamu hlasivkové štěrbiny

Ing. Tomáš Ettler

DISERTAČNÍ PRÁCE
k získání akademického titulu doktor v oboru Informatika a výpočetní technika

Školitel: Ing. Pavel Nový Ph.D.
Katedra informatiky a výpočetní techniky

Plzeň, 2022

Detection and Evaluation of Glottis in Laryngeal High-Speed Videoendoscopy Recording

Ing. Tomáš Ettler

Abstract

This work summarizes the results of the study of vocal cords evaluation based on data extracted from recordings taken by a laryngoscopic system, specifically by Laryngeal High-Speed Videoendoscopy (LHSV). The main goal of this work is to process images contained in the recorded LHSV sequences, find and detect the vocal gap (glottis) using chosen image segmentation methods and evaluate the vocal cords' quality by analytical and statistical methods using a defined set of parameters.

The first part of this thesis focuses on the description of the nature and structure of the information that is obtained using the LHSV system. Therefore, the anatomy of the vocal cords and the physiology of voice creation are described concerning the information included in the image in the LHSV recording. Also, the basic types of vocal cords diseases are listed and the data gathering, structure, and problems affecting the quality of the LHSV recording are described. Furthermore, issues of image segmentation used on laryngoscopical image data taken from Laryngeal High-Speed Videoendoscopy are delineated together with a description of the developed method for glottis localization (finding ROI), segmentation, and parameter selection mainly based on geometry and glottis symmetry. The process is demonstrated in several case studies.

The important part of the work contains a description of new methods dealing with computed parameters and their relationships using correlation analysis. An approach based on expected and unexpected correlation relations resulting from the detailed analysis can provide a basic evaluation of the vocal cords' behavior. Other methods then provide a numeric evaluation of the glottis shape development based on statistical analysis and rating from the experts' examinations. The results are illustrated and explained.

Detekce a hodnocení videozáznamu pohybu hlasivek z vysokorychlostní kamery

Ing. Tomáš Ettler

Abstrakt

Tato práce shrnuje výsledky studia zabývajícího se hodnocením hlasivek na základě dat získaných ze záznamů pořízených laryngoskopickým systémem, konkrétně laryngeální vysokorychlostní videoendoskopií (Laryngeal High-Speed Videoendoscopy – LHSV) Hlavním cílem této práce je zpracovat obrazovou informaci, která je obsažena ve videosekvencích LHSV, najít a detekovat hlasivkovou štěrbinu (glottis) zvolenými metodami segmentace obrazu a vyhodnotit kvalitu hlasivek analytickými a statistickými metodami s využitím definovaného souboru parametrů.

První část této práce se zaměřuje na popis podstaty a struktury informace, která je získána pomocí systému LHSV. Proto je zde popsána anatomie hlasivek a fyziologie vzniku hlasu, to vše ve vztahu k informacím obsažených ve snímku v záznamu LHSV. Také jsou uvedeny základní typy onemocnění hlasivek a doplněn popis získávání dat, jejich struktura a poruchové jevy, které ovlivňují kvalitu záznamu LHSV. Dále je popsána problematika segmentace obrazu použitá na získaných obrazových datech z vyšetření pomoci LHSV a jsou shrnuté metody vyvinuté pro lokalizaci glottis, tzv. nalezení oblasti zájmu (Region of Interest – ROI), samotnou segmentaci a výběr parametrů založených především na geometrii a symetrii hlasivek. Proces je demonstrován na několika kazuistikách.

Důležitou částí práce je popis nových metod zabývajících se vypočítanými parametry a jejich vztahy pomocí korelační analýzy. Přístup založený na očekávaných a neočekávaných korelačních vztazích vyplývajících z podrobné analýzy může poskytnout základní hodnocení chování hlasivek. Další metody pak poskytují numerické hodnocení vývoje tvaru hlasivkové štěrbiny na základě statistické analýzy a expertního hodnocení. Výsledky jsou ilustrovány a vysvětleny.

Declaration / Prohlášení

This dissertation thesis was created at the end of the doctoral study at the Faculty of Applied Sciences, University of West Bohemia. I hereby declare that this thesis is my own original and sole work and all sources used in this work are listed in the bibliography.

Tato dizertační práce byla vytvořena na závěr doktorského studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni. Prohlašuji tímto, že tuto práci jsem vypracoval samostatně, s použitím odborné literatury a dostupných pramenů uvedených v seznamu literaury.

Ing. Tomáš Ettler

Acknowledgments

I would like to express my thanks and appreciation to my supervisor Ing. Pavel Nový, Ph.D. for his patience during leading my study, providing his expert knowledge in the image processing area, and his friendly approach during the long time I was progressing my work.

I am also grateful to the colleagues from the ENT department of University Hospital in Pilsen, especially Jiří Pešta (in memory), biomedical engineer, who was managing examination data and created a huge corpus containing results from many examinations, and Monika Vohlídková, ENT doctor, phoniatrician, and audiologist, for professional consultations, providing expert knowledge and cooperation in creating a data corpus.

Great thanks belong to my wife Zuzana for supporting me and for standing by my side when I needed it.

Contents

1 Introduction **1**
- 1.1 Thesis Outline . . . 1
- 1.2 Thesis Goals . . . 2

2 Anatomy and Examinations **4**
- 2.1 Larynx Anatomy . . . 4
- 2.2 Voice Formation . . . 5
- 2.3 Vocal Cords Diseases . . . 9
- 2.4 Examination Methods . . . 10
  - 2.4.1 Acoustic Methods . . . 10
  - 2.4.2 Aerodynamic Methods . . . 13
  - 2.4.3 Electrophysiological Methods . . . 14
  - 2.4.4 Optical Methods . . . 14
- 2.5 Used Methods in the ENT Department . . . 16
- 2.6 Commercial Systems and Used Methods . . . 18

3 Laryngoscopy High-Speed Video Data **20**
- 3.1 Used Camera . . . 20
- 3.2 Data Corpus Structure . . . 22
- 3.3 Input Data Quality . . . 24

4 Glottis Detection **26**
- 4.1 Overview of Published Methods . . . 26
- 4.2 Approach Used in this Work . . . 27
- 4.3 Region of Interest (ROI) . . . 29
  - 4.3.1 Methods . . . 29
  - 4.3.2 Thresholding Method . . . 29
  - 4.3.3 DFT Method . . . 31
  - 4.3.4 Results . . . 39
- 4.4 Detailed Glottis Segmentation within ROI . . . 41
  - 4.4.1 Original Thresholding Method . . . 42
  - 4.4.2 Cluster Analysis Segmentation Method . . . 43
  - 4.4.3 Results . . . 47

5 Glottis Symmetry **52**
- 5.1 Detection of the "Symmetry Axis" of Vocal Gap . . . 52
- 5.2 Floating Axis . . . 56

5.3 Evaluation of the Detection of the Symmetry Axis 56

6 Glottis Parameters **58**
6.1 Commonly Used Parameters 58
6.2 Additional Parameters 58
6.3 Center of Gravity of the Vocal Cords 59
6.3.1 Center of Gravity Parameters 59
6.3.2 Center of Gravity Trajectory 62
6.3.3 Selected Case Studies 64
6.3.4 Summary of Obtained Results 81

7 Parameter Analysis **83**
7.1 Correlation and Linear Approximation 83
7.2 Diagnostic Meaning of Correlation between Parameters 85
7.2.1 Searching for Useful Parameter Pairs 85
7.2.2 Correlation Structure Analysis 89
7.2.3 Results 94
7.2.4 Conclusion 94
7.3 Correlation Classification 95
7.3.1 Vocal Cords Rating 95
7.3.2 Method Description 96
7.3.3 Classification Using Medians 99
7.3.4 Classification Using Oriented Areas between Class EDF and Complement 108
7.3.5 Conclusion 114
7.3.6 Case Studies 115

8 Conclusion **129**
8.1 Achieved Goals 131
8.2 Author's Comment 133
8.3 Future Work 134

9 Resumé **135**

1 Introduction

Speaking is an elementary way of communication between people, irreplaceable in many activities. The essence of speaking is the voice, which is formed in the oral cavity, but it is created by the oscillating vocal cords in the larynx. Therefore, any disorder of voice production on the vocal cords or their dysfunction causes difficulties in people's lives and their interactions within the human community. Recognizing voice problems, early diagnosis and finding the cause of the voice disorder, and targeted treatment contribute to the elimination of these disorders, chronic consequences of the disease, and even averting possible fatal problems due to the occurrence of malignancies.

The diagnosis of the vocal cords is supported by a number of examination techniques and tools, which also include an analytical evaluation of the vocal cords' kinematics. One type of medical examination that provides data for analysis of vocal fold behavior is Laryngeal High-Speed Videoendoscopy (LHSV) which uses a high-speed camera to record the real movement of the vocal cords. This work describes the analysis and processing of this data with the aim of presenting methods and processes for obtaining the final evaluation of vocal cords according to defined criteria with the possibility of being used for early warning in case of any irregularity is detected.

Keywords: Vocal cords examination, Laryngeal High-Speed Videoendoscopy, Glottis segmentation, Vocal cords symmetry, ROI detection, Image cluster analysis, Correlation relationships, early warning medical tool

1.1 Thesis Outline

This work aims to contribute to the improvement of the quality of voice examination and diagnosis of vocal cord diseases. The methods presented are expected to help in the stage of incipient voice disorder, when the patient experiences unspecified difficulties, e.g. found by subjective Voice Handicap Index (VHI)[1], but standard investigative techniques do not provide a clear diagnosis. Similarly, the methods and procedures presented in this work could contribute to the early diagnosis of voice disorders in the case of inclusion of LHSV among screening examinations, when they may detect changes in the behavior of the vocal cords that other investigative methods are unable to distinguish.

A number of methods are used for voice examination, or to help diagnose diseases of the vocal cords, which are divided into different categories. We distinguish functional diagnostic methods and descriptive methods. Other classifications include acoustic, aerodynamic, electrophysiological, and optical examination methods, see e.g. [2], [3], [4].

The entire work is focused on functional diagnostics in the field of optical examination methods, specifically on the method of Laryngeal High-Speed Videoendoscopy (LHSV). It contains a comprehensive solution to the problem, including segmentation of the vocal gap (glottis) in each frame of the LHSV recording, estimation of the axis of glottis symmetry, and specification of a set of glottis parameters and their normalization for a comparable objective measurement of vocal cord behavior. The next part is data analysis, the use of several statistical methods to detect significant correlations, visualization of parameter values, and relationships between them. A multidimensional classification function that evaluates the quality of glottis behavior also contributes to the diagnosis. Visualization of parameters and their correlations and using classification functions can thus contribute to diagnosis in cases where mere observation of video sequences may not provide enough information to make a decision.

In this work, the solution to the mentioned tasks is exclusively based on the evaluation of LHSV image information. Acoustic recordings taken simultaneously with the video sequence, LHSV metadata, and the results of other types of examinations were used as supporting or supplementary information in the creation of an anonymized data corpus of LHSV records, from which this work is based. The data corpus of LHSV recordings was created thanks to long-term cooperation with the ENT department of the University Hospital in Pilsen, where the quality of the glottis closure was also evaluated by an ENT expert. For this work, a data corpus containing 692 video recordings (corpus no. 692) was used.

1.2 Thesis Goals

The goal of this work can be divided into three scientific objectives.

The first objective is to find a sequence of methods that lead to a successful segmentation of the glottis in each frame of the LHSV sequence. Especially in video sequences of poorer quality, when the image is out of focus, contains light artifacts, or the anatomical structures of the vocal cords are covered by moving fluid. In this field of Image processing and connection with glottis detection, many approaches and methods have already been published, summarized by e.g. [5] or [4], more in section 4.1. Despite this, an original sequence of LHSV preprocessing, Region of Interest (ROI) detection, and glottis segmentation methods were introduced to handle this task.

The second objective is to analyze the kinematics of the vocal cords, since mere observation of LHSV recordings or other examination results may not always be sufficient. Therefore, a set of parameters has to be defined that has a geometric basis and is derived from the shape of the glottis and may use the axis of symmetry. A statistical analysis of mutual relationships of the selected parameters is based on the assumption that most geometric parameters should have a strong relationship during standard vocal cords' behavior. From the diagnosis point of view, it is inter-

esting to note that the violation of the correlation relationship for some parameter pairs can be an indicator of a pathological state in the vocal cords. The development of individual parameter values, correlations, and their violations are visualized.

The third objective is the determination of classification functions for estimating the quality of the glottis behavior, which is based on LHSV recordings rated by an ENT expert and a set of selected glottis parameters with significant correlations. The correlation relationships of all combinations of parameters were analyzed and several parameter pairs were selected as important indicators of non-standard vocal cords' behavior. Another method was introduced for vocal cord classification based on statistical methods and robust analysis. The result of this method is a single number for easy glottis evaluation. These methods can help within vocal cords' examination to raise a warning in case of found irregularities or unexpected correlation values.

2 Anatomy and Examinations

2.1 Larynx Anatomy

The vocal cords (vocal folds, voice reeds) are located within the larynx at the top of the trachea (the cartilaginous tube that connects the larynx to the bronchi of the lungs). They are attached posteriorly to the arytenoid cartilages, and anteriorly to the thyroid cartilage. The length of the vocal fold of adults is about 8-16 mm. The opening between the vocal folds is called the glottis (vocal gap, rima glottidis). The larynx cut can be seen in figure 2.1.

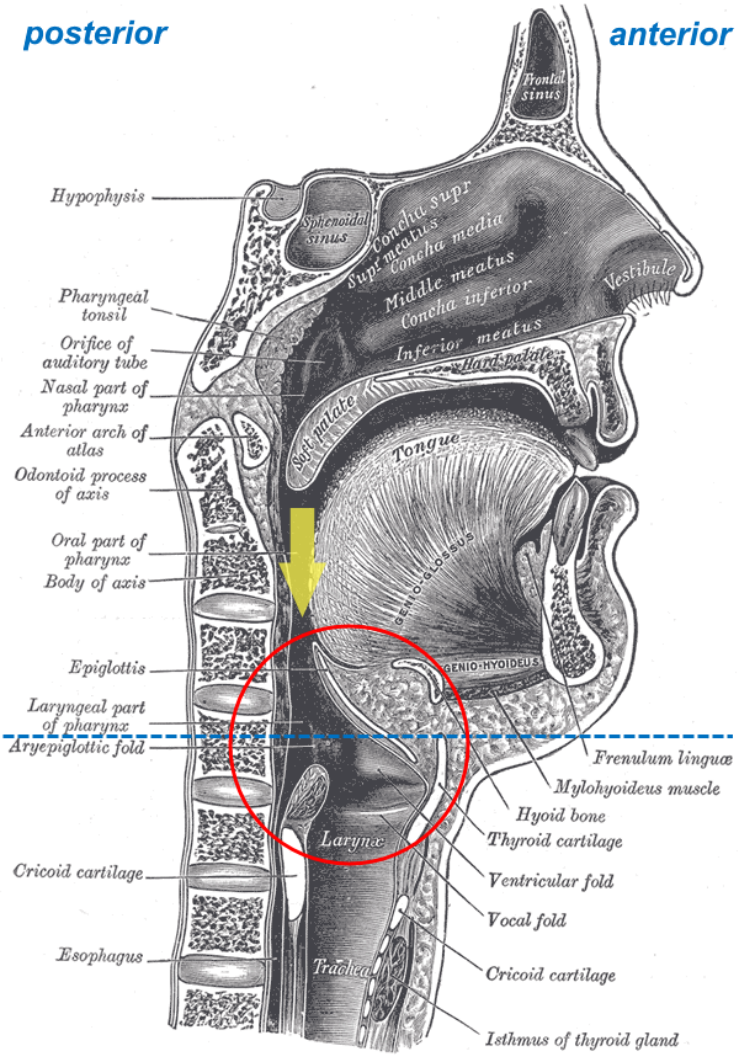


Figure 2.1: *Airways anatomy, sagittal section of the oral cavity, nose, and larynx. The section image is supplemented by a description of the anterior/posterior orientation, the transverse plane of the anatomical description of the vocal cords, and a schematic representation of the direction of the laryngoscopic view of the vocal cords (from [6], fig. 994 – Sagittal section of nose, mouth, pharynx, and larynx).*

The outer edges of vocal cords are attached to muscle in the larynx while their inner edges, or margins, are free forming the opening called the glottis. They are constructed from epithelium, but they have a few muscle fibers in them, namely the vocalis muscle which tightens the front part of the ligament near the thyroid cartilage. They are flat triangular bands and are pearly white. Above both sides of the glottis are the two vestibular folds or false vocal folds which have a small sac between them.

Glottis has the shape of an elongated isosceles triangle. The vocal cords can be seen from the oral cavity behind the laryngeal flap. This is used by laryngoscopic examination methods, see figure 2.3.

Because this work is about the processing of laryngoscopic images of the vocal cords, the following description is focused on anatomic structures which affect and condition the results of image processing from high-speed laryngoscopy, especially the glottis.

The frontal section of the larynx can be seen in figure 2.4, where the vocal folds are highlighted. In the upper part, there are false vocal cords (usually not too significant and don't affect the laryngoscopic view of the glottis) and vocal cords below.

In this work, the term "vocal cords" is used for the organ consisting of two "vocal folds". The "glottis" is then the gap between vocal folds.

2.2 Voice Formation

The main function of vocal cords is voice (vocal tone) creation. The phonation principle is creating discontinuous air flow by the periodic oscillation of vocal cords, the size of vocal cords affects the pitch of voice. The voice is completed by resonant cavities (supraglottic cavity, oral cavity, and nasal cavity). Different sizes and shapes lead to unique voice timbre [2], [3], [7].

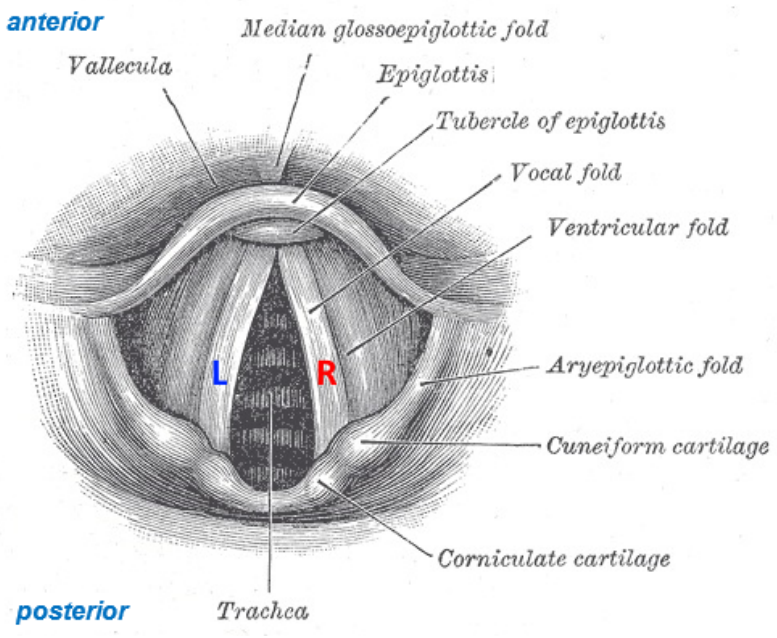


Figure 2.2: *Laryngoscopic view of the vocal cords, anatomical description in the transverse plane of the larynx, the picture is supplemented by a description of the anterior/posterior and left/right orientations (from [6], fig. 956 – Laryngoscopic view of the interior of the larynx).*

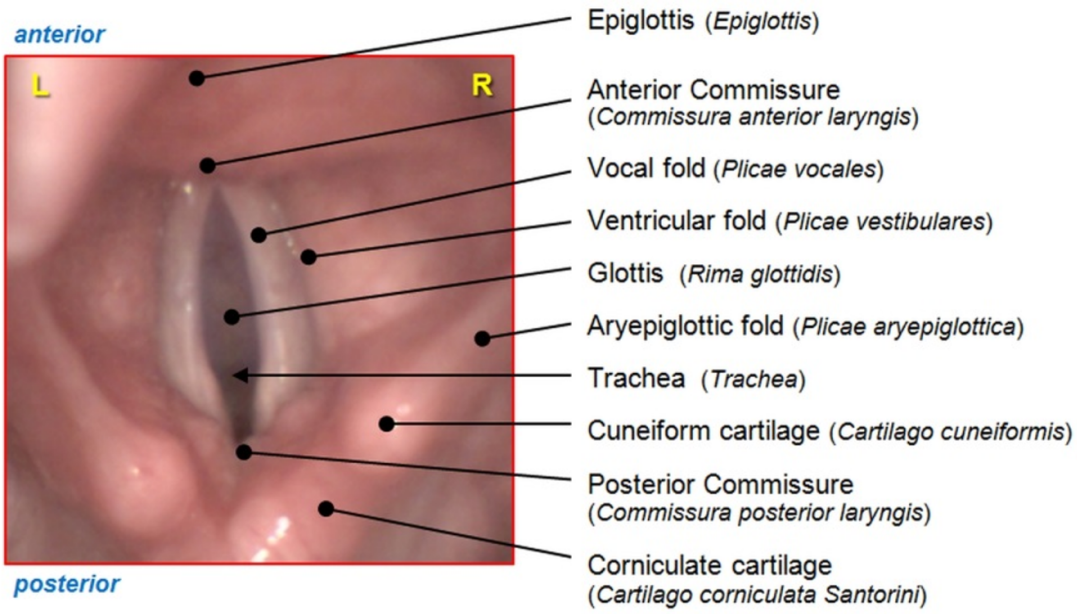


Figure 2.3: *Laryngoscopic image of the vocal cords.*
The image from LHSV is supplemented by a description of some anatomical structures that are important for the purposes of this work (an illustrative image is from the database of LHSV from the ENT department of the University Hospital in Pilsen).

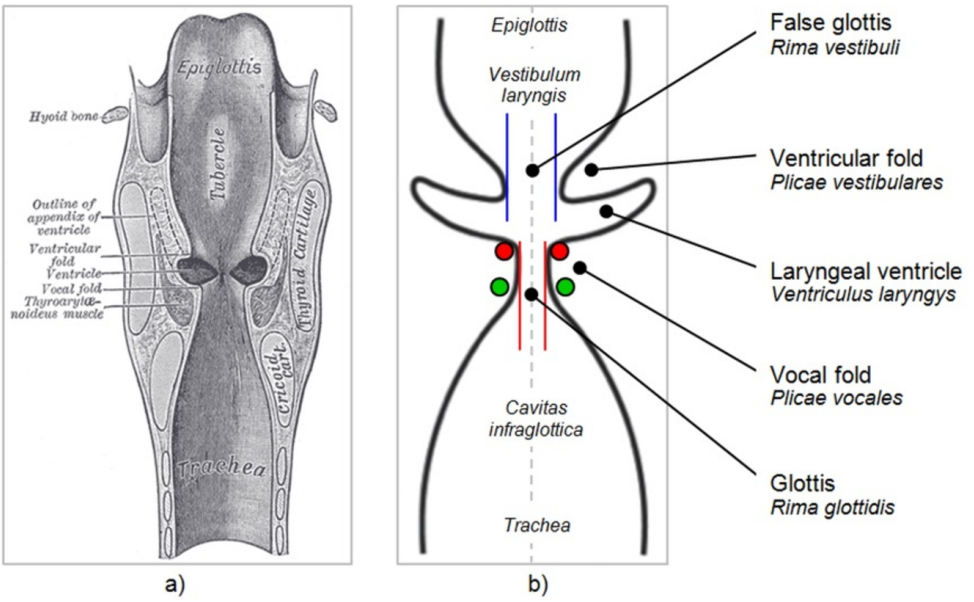


Figure 2.4: *Laryngeal anatomy, frontal section, laryngeal scheme.*
a) Anatomical description of the larynx in the frontal section (from [6], fig. 954 – Coronal section of the larynx and upper part of the trachea).
b) Schematic description of the larynx in the frontal section with marking of the upper and lower part of the vocal folds to show the so-called mucosal wave. Red mark – upper edge Plicae vocales; green mark – bottom edge Plicae vocales

There were two theories of voice creation spoken:

The myoelastic-aerodynamic theory was defined in 1958 by Dutch scientist Janwillem Van Den Berg (1920). According to this theory, the oscillation of vocal cords is created by the mass, tension, elasticity, and the situation created by exhaling from the lungs. Aerodynamic-aerostatic strength opens vocal cords, muscle, and ligament tensions return the folds to a closed position[1].

The other theory, the neurochronaxic one, was described by French physicist Raoul Husson in 1953. Vocal cords don't move passively but by nerve impulses. But opponents object that no nerve is able to transfer so many impulses needed e.g. during the singing when vocal cords oscillate more than 1000 times per second. Another argument was the moving of vocal cords in the case of one-sided paresis when nerves are damaged. Today, only myoelastic theory is accepted.

In the following figure, the movement during phonation is shown. It is divided into seven frames during the opening phase (abduction, fig. 2.5 frames 1–4) and the closing phase (adduction, fig. 2.5 frames 4–7) showing the state of vocal cords in schematic frontal cut from a laryngoscopic view in each movement phase. Figure 2.6 then shows movement on kymogram with comparison to progress of acoustic pressure creating sound (MIC) and electroglottograph signal (EGG).

[1]This physical rationale for myoelastic theory was given by Lieberman in 1968

Red and green marks on figures 2.4, 2.5, and 2.6 highlight the upper and lower parts of the vocal cords which create mucosal wave during the closing phase. Because of the laryngoscopic view direction (camera position), the mucosal wave is not clearly visible, only the vertical projection of the glottis can be detected. But the behavior of the mucosal wave was considered during the ENT expert rating, see section 7.3.1.

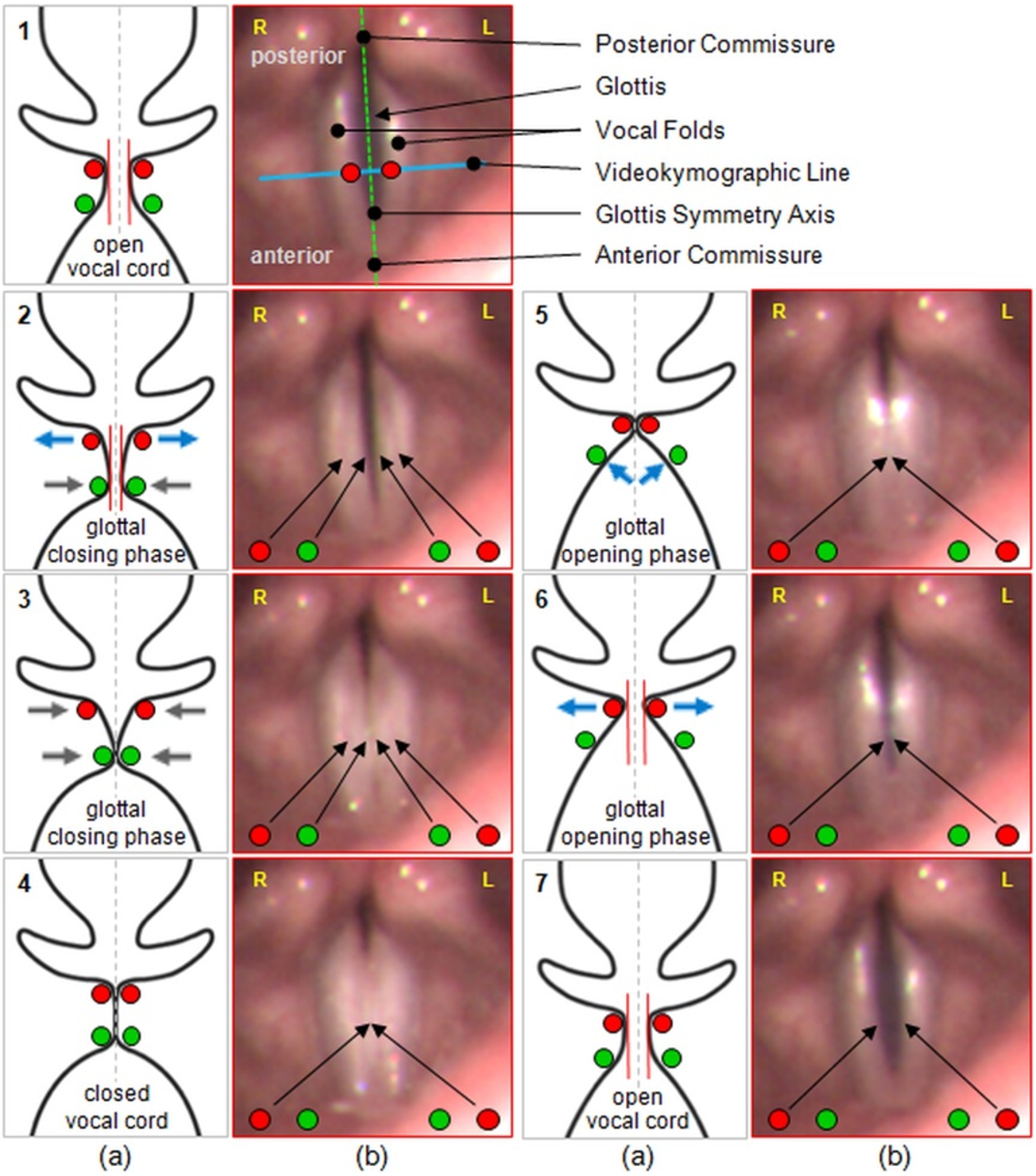


Figure 2.5: *Description and decomposition of vocal cord dynamics during the phonation phases opening (abduction, frames 1–4) and closing (adduction, frames 4–7). (a) scheme of the point position development of the vocal cords on the left and right side during the phonation of the vowel "i:" in the section of the frontal plane; (b) real images of glottis development in the transverse plane from LHSV recording.*

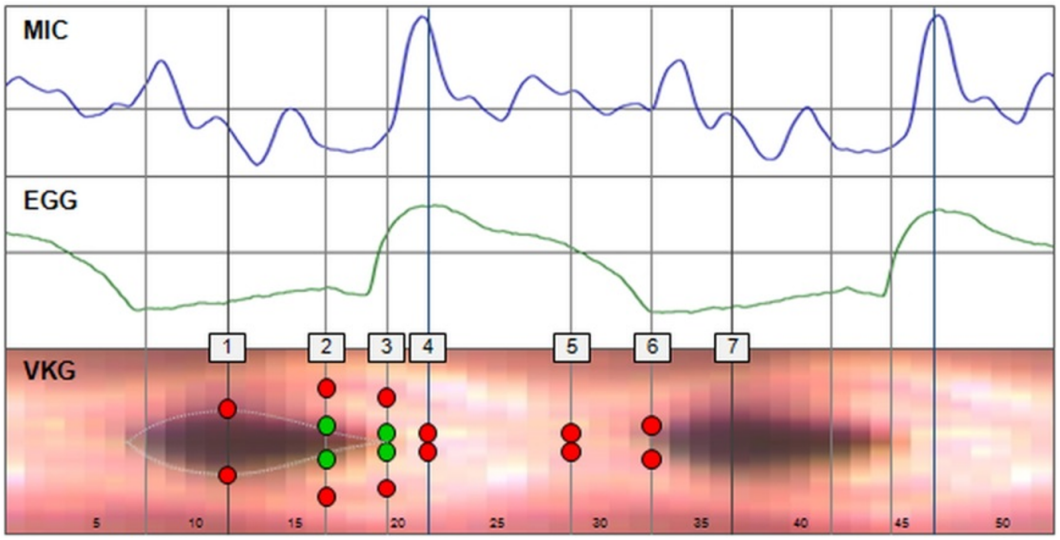


Figure 2.6: *Phase distribution of opening and closing of vocal cords in videokymographic section VKG (position shown in fig. 2.5 1b);*
MIC – recording of sound pressure by SPL microphone during vocal phonation "i:";
EGG – recording of electroglottographic signal;
VKG – videokymographic record in section, see fig. 2.5 1b)

2.3 Vocal Cords Diseases

Normal phonation can work only in specific circumstances where subglottal pressure, vocal cords tension and mass, length of the oscillating segment, and other parameters are in a certain range. During voice formation, healthy vocal cords periodically open and close in whole range with exception of loud shouting. Disruption of vocal cords function can lead to hoarseness, rasping or complete loss of voice.

The principle of vocal fold disease is the limitation of movement of one or both vocal folds by lowering elasticity or disability of nerve connection of vocal cord muscles [8].

Voice dysfunction can be divided into organic (caused by pathological anatomic changes of larynx structures) and functional (without any larynx problem)[3]. For visual observation, only organic issues are interesting like:

Polyp - mucous membrane filled with connective tissue.

Nodule - limited thickening of the tissue influencing flexibility, which prevents complete closure of the glottis.

Edema - submucosal ligament leakage occurs, the closure of the glottis is often only at the area of edema.

Cyst - encapsulated formation inside the vocal fold, an uneven edge is formed.

Aphonia - the vocal cords do not come close at all, no voice is formed.

Paresis - numbness occurs in various parts of the vocal cords.

There can be also combinations of mentioned and other more rare diagnoses.

All these diseases imply a change in the vocal cords' behavior and can be detected by optical methods, see examples in fig. 2.7. In the case of an advanced stage of the disease, the diagnosis is usually clear where no additional support is needed. But in the case of the first stage of the issues, no obvious result can be given from simple observation, then advanced methods can be used to provide objective measurement. To prevent the development of serious vocal cords diseases or larynx, several examination methods are described in the following section.

2.4 Examination Methods

The first person who used the direct laryngoscopic method was singer M. Garsia in the second half of the 19th century [9]. He used a small mirror to watch vocal cords. Many doctors and scientists continued using this method and a laryngoscopy discipline was formed, that studies vocal cords disorders. There is also a discipline of phoniatry for the overall issue of voice creation, which includes physiology, pathophysiology, diagnostics, and treatment.

Various methods of vocal cord examinations were formed over time, they can be divided into four categories:

- Acoustic
- Aerodynamic
- Electrophysiologic
- Optical

2.4.1 Acoustic Methods

The first group of examination methods comprises the acoustic ones using sound recording and analysis. These are mainly the following methods:

Voice Range Profile

This method was introduced by P. H. Damsté (1970) under the name phonetography, but it was renamed to Voice Range Profile (VRP) examination because of naming

Figure 2.7: *Examples of LHSV images with specific diagnoses:*
(a) vocal nodule on the right side
(b) vocal polyp on the right side
(c) cyst vocal on the left side
(d) carcinoma on the left side
(e) papilloma
(f) granuloma
(g) Reinke's edema
(h) recurrent laryngeal n. paresis on the right side
(i) cordectomy on the left side

inaccuracy. This method is used for the evaluation of quantitative voice parameters leading to the creation of overall frequency voice range and dynamic range.

This method consists in recording the quietest and the loudest patient's voice in his complete frequency range. the result values are then carried into the diagram with frequency on the main axis x and intensity in dB on the axis y.

This method can be also used to evaluate a stress test when the voice before the stress and after it is compared. Long reading, theater performance, or singing can be considered as stressful situations. The tiredness usually leads to a narrowing of frequency or dynamic voice range.

Multi-Dimensional Voice Analysis

Multi-Dimensional Voice Analysis (MDVA, also called Multi-Dimensional Voice Program – MDVP) is a method based on watching time and signal acoustic parameters where objective values of laryngeal sound and base voice parameters are evaluated. The main parameters are:

- Soft phonation index (SPI) - is a ratio of harmonic frequency energy in the range of 70–1600 Hz to the energy in the range of 1600–4500 Hz.
- Jitter - parameter describing a degree of base frequency periodicity
- Shimmer - parameter of amplitude stability disturbances [10].

There are two types of values for these parameters:

- Absolute, which are calculated as the average of the deviations of the periods and amplitudes of subsequent cycles,
- Relative, which are related to the average values of all measured cycles [11].

One of the available commercial analyzers is the MDVA application from Pentax company (former Kay Elemetrics) computing up to 33 parameters [12]. This tool has become the standard for acoustic voice parameters evaluation because its result can be compared with the healthy population examination results.

Vocal Cords Closing Quality Evaluation

Vocal cords Closing Quality Evaluation (SCORE [13]) is based on Fourier analysis of one period of vocal cord oscillation. The main goal is to find a moment of pressure pulse response created by glottis closure. Normalized amplitudes computed from Fourier coefficients together with expert evaluation (from medical examination

results and experience from the field) determine the SCORE value on a simple scale. This SCORE value can be used for vocal cords function evaluation and it could lead to diagnosis determination of voice issue.

This method was developed in the ENT department of the University Hospital in Pilsen together with the Department of Computer Science and Engineering in the Faculty of Applied Sciences at the University of the West Bohemia.

2.4.2 Aerodynamic Methods

Aerodynamic methods are based on breath monitoring and measurement of air speed and volume passing through the glottis.

Pneumography

Pneumography was a previously used method that examined the movement of the chest and belly during breathing and phonation. This method was used for the first time by H. Gutzmann in the 1920s using a thin-walled rubber tube transferring pressure change to a kymograph[2]. This method was further improved and it was possible to observe far finer pressure changes.

The result consists of two curves – thoracic and abdominal (Gutzmalin-Oehmecke girdle pneumograph). After further processing, breath frequency, inhale-exhale ratio, curves synchronization, behavior during phonation, and other parameters were evaluated.

Pneumotachography

Pneumotachography replaces pneumography in the 1960s (Fleisch). It was based on air speed measurement and its volume passing through the glottis during phonation. Subglottic pressure (air pressure below vocal cords) measurement should be included in this method, but it was possible only in an invasive way, therefore it was abandoned. Another way was to introduce a catheter and measuring balloon to the area under the vocal cords, but it interfered with natural voice creation.

Another way of subglottal pressure measurement was using a tube placed on a tongue. This method was used by the Phonatory Aerodynamic System machine (PENTAX corp.) The pressure measurement is indirect during the phonation of syllable “pa:” because the air column is connected between subglottal and supraglottal space when phonating the consonant “p”. This way is also used in the ENT department of the University Hospital in Pilsen.

The basic parameters are the phonation volume, maximum phonation time, average

[2]Kymograph - motion recording device

air velocity, and phonation quotient (vital capacity of the lungs is needed to calculate that). However, this method can be used to assess the function of the larynx and the degree of damage, but not to diagnose the disease, as it cannot find the cause of the change in individual parameters.

2.4.3 Electrophysiological Methods

Another group of examination methods consists of the electrophysiological ones based on electrical values changes during phonation:

Elektroglottography

This method, which deals with the monitoring of changes in electrical resistance (conductivity) by high-frequency measurements during oscillations of the vocal cords, was introduced in 1957 by Fabre. The sensing electrodes are located above the wings of the thyroid cartilage and are defined by 6 points in the glottogram corresponding to the movement phases of the vocal cords and the shape of the vocal cord gap.

Elektromyography

Electromyography is a method that monitors muscle activity during phonation. It was first proposed by Seiffert in 1919 and it uses needle electrodes inserted into the muscles around the vocal cords. This method contributed to the elucidation of muscle activity during phonation and today it is mainly used to verify the disruption of the reversible nerve in polio.

The method measures the activity of individual muscles and the magnitude of the amplitudes of potentials.

2.4.4 Optical Methods

The last group is optical methods.

Already in the middle of the 19th century, laryngoscopy was used, where it is possible to observe the behavior of the vocal cords using a mirror. Due to the high frequency of oscillations, it was necessary to use additional devices to capture the motion. First, laryngostroboscopy was used (flashing light with a frequency slightly different, than the oscillation of the vocal cords), which made it possible to show their apparent movement. Due to the impossibility of capturing the real vocal cords' movement, it was not possible to find random phenomena within one oscillation.

Videokymography has already been able to record the real movement of the vocal cords in a straight line. The output is a kymogram, which is a graphical representa-

tion of oscillations, which is created by stacking individual image lines side by side [2].

Laryngoscopy

Laryngoscopy is a method of monitoring the larynx and vocal cords using a mirror (indirect laryngoscopy) or optical fibers (direct laryngoscopy). The mirror or laryngoscope is inserted into the mouth and vocal cords can be observed by the reflection. However, only the condition of the vocal cords can be monitored, their movement is not recognizable to the eye due to their rapid oscillation. This method was first used in 1854 by singer Garcia.

Laryngostroboscopy

Laryngostroboscopy allows the observation of vocal cord oscillations using a stroboscopic phenomenon[3] when the flashing light has a slightly different frequency from the oscillation frequency of the vocal cords, which is indistinguishable to the eye due to its high frequency. There is an optical slowing down of the vocal cords' oscillations and it is, therefore, possible to observe their behavior. Steady oscillation is required for laryngoscopic examination, otherwise, there is no effect of slow motion.

Videokymography

Videokymography is a method based on a mathematical-optical model in the cross-section of the vocal cords. It allows to examine them in any position from the front to the back commissure. It accurately graphically displays the frequency and amplitude of the oscillations of the vocal cords and also opening and closing behavior can be observed.

Recording requires equipment capable of capturing one line of the image at high speed (in the order of thousands of lines per second) and software for processing the scanned data and storing the resulting image. This method was first described and tested in [2].

The result of the examination is a picture composed of individual lines and shows the movement of the vocal folds at the monitored site.

[3]Stroboscopy is a phenomenon based on the inertia of perception of the human eye, where at a certain frequency of flashing light close to the frequency of a periodically fast moving object, the object's movement seems to move slowly.

Laryngoscopy High-Speed Videoendoscopy

Laryngoscopy High-Speed Videoendoscopy (LHSV) is a method for monitoring the vocal cords' behavior with a high-frequency camera that can record more than 1000 frames per second. Unlike videokymography, the whole image of the vocal cords is recorded. This method appears only with the development of image capture technology and replaces previous methods.

The main advantage over previous examinations is the possibility to observe whole vocal folds and their behavior during phonation. The shooting frequency is much higher than the vocal cords' oscillation so the real movement is recorded including any artifact during one period.

Nowadays, a combination of acoustic and optical methods is most often used for the examination and diagnosis of voice disorders, as they are the least demanding on the patient and can obtain a large number of relevant parameters. This work focuses on the processing and evaluation of images from high-speed camera recording.

2.5 Method Used in ENT Department of Faculty Hospital in Pilsen

At the ENT department of the University Hospital in Pilsen, subjective evaluation of the VHI (Voice Handicap Index [1]) is performed as standard using a questionnaire and examination by acoustic methods (voice field, multidimensional analysis, SCORE evaluation[13]), which are supplemented by examination by a high-speed camera.

In selected cases, other types of examinations are performed, such as electroglottography, STRESS tests [14], aerodynamic examinations of vocal cords [15], or detection of non-standard vibration of vocal cords (AOD) [16], [17].

The goal is to create a representative set of parameters over the data from different examinations, which are distributed over time at different stages of the disease or treatment (fig. 2.8). These data are then analyzed and the output is the vocal cords evaluation by the summary parameters. If the treatment of the disease requires microsurgery, the examinations are distributed over time according to the scheme in figure 2.9. However, the treatment of some diseases requires a long-term approach in the form of exercise, then data collection is individual.

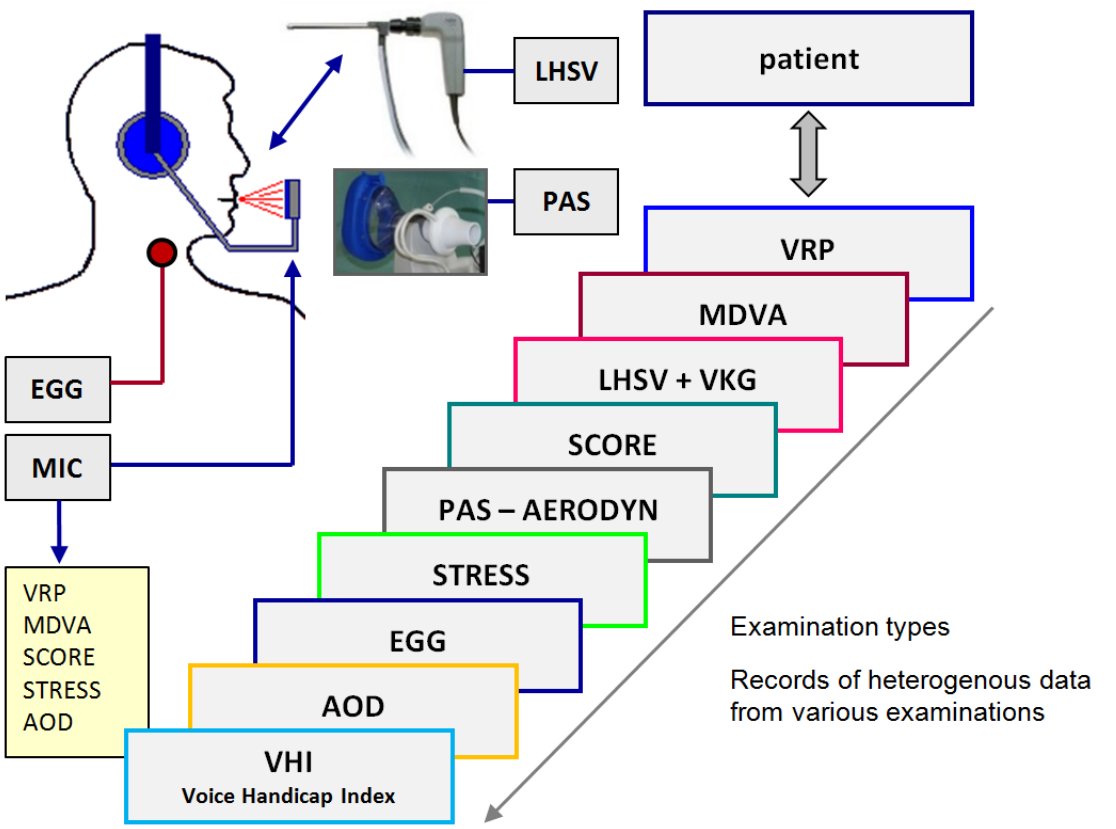


Figure 2.8: *Overview scheme of examination methods used at the ENT clinic of the University Hospital Pilsen.*
- VRP – Voice Range Profile
- MDVA – Multi-Dimensional Voice Analysis
- LHSV – Laryngeal High-Speed Videoendoscopy
- VKG – Videokymography
- SCORE – Quality of the Glottis Closure
- PAS – Phonatory Aerodynamic System
- STRESS – Stress Test of vocal cords
- EGG – Electroglottography System
- AOD – analysis of non-standard oscillation in voice sound recording
- VHI – Voice Handicap Index

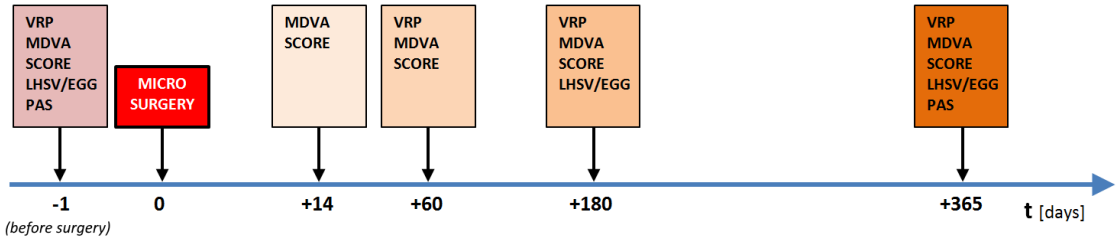


Figure 2.9: *Example of examination arrangement for microsurgery.*

2.6 Commercial Systems and Used Methods

This section summarizes the available LHSV systems and their features. These systems were developed from previously used methods due to technology improvement. Before LHSV was introduced, devices were using the stroboscopic effect, because it was not possible to construct a camera with the required sensitivity and frame rate. The videokymography[2], recording only one line, was introduced because it was difficult to fund expensive commercial systems. Nowadays, the technology allows the creation of small and efficient devices causing LHSV systems to be commonly used in hospitals. There is still development in progress in this area, but it is beyond this comparison.

At the beginning of our study, there were two main vendors of LHSV systems – Richard Wolf and KAY Pentax (former KAY Elemetrics). The parameters of their systems were corresponding to the year of introduction. Nowadays other companies provide LHSV systems with modern components, which are WEVOSYS and DiagNova. The summary of the properties is in table 2.1.

Table 2.1: LHSV systems and their properties.

LHSV systems (software)	high-speed camera parameters		
	light source	*frame rate [fps]*	*Resolution [px]*
HRES Endocam 5562 Richard Wolf GmbH	Xenon constant	2000 4000	256 x 256 256 x 256
KAY Pentax (KIPS – KAYPentax Image processing Software)	Xenon constant	2000 3000 4000	512 x 512 512 x 352 512 x 256
WEVOSYS – lingWAVES 4 (lingWAVES 4 High-Speed Videoendoscopy App)	BlueWhite, Luminance, LED	4000 max 8000	1440 x 1440 lower
DiagNova Technologies	Laser emitters NBI [405, 520, 638 nm]	4000 3200 2400	320 x 288 480 x 400 512 x 480

Together with the physical devices, proprietary software is provided to handle data output from the camera. The software is capable of displaying the image of vocal cords and playing recordings, also other functions are included. Besides technology and physical properties, the software also provides different functionality. Available functions related to this work (ROI detection, glottis segmentation, axis determination) are summarized in table 2.2.

Table 2.2: LHSV systems and available functions.

LHSV systems	basic methods of video sequence LHSV analysis		
	ROI setting	*glottis symmetry axis setting*	*glottis segmentation*
HRES Endocam 5562 Richard Wolf GmbH	manual	manual	manual, thresholding
KAY Pentax	manual	manual	edge detection, manual parameter setting
WEVOSYS – lingWAVES 4	manual	automatic	automatic
DiagNova Technologies	manual	manual	manual

Several features are briefly mentioned here for each LHSV system.

KAY Pentax system (Japan-USA) uses brightness change for glottis edge detection.

The WEVOSYS system has advanced functionality in glottis detection and axis assessment, which is the result of cooperation with the University of Erlangen.

The DiagNova system seems to have interactive ROI, glottis, and axis detection, which are not fully automatic. The advantage of this system is the possibility to use different light sources enabling the examiner to see through some tissues (Narrow Band Imaging – NBI). It is achieved by the different wavelengths of light. Also, Phonovibrogram (PVG[18]) is implemented in the DiagNova system.

Details about the Richard Wolf HRES camera are summarized in section 3.1.

This comparison was made for rigid endoscopy cameras, there are also flexible endoscopic devices with LHSV capabilities.

Information and pictures were taken from public resources:

- HRES Endocam 5562 Richard Wolf GmbH
  https://www.iis.fraunhofer.de/content/dam/iis/de/doc/il/bmt/mbv_broschuere_endocam_hres_5562_richard_wolf.pdf
- KAY Pentax –
  https://www.pentaxmedical.com/pentax/en/99/1/ENT-Speech/
- WEVOSYS – lingWAVES 4
  https://www.wevosys.com/products/lingwaves4/lingwaves4_high_speed_videoendoscopy.html
- DiagNova ALI Cam-HS1
  http://diagnova.eu/pages/offer/highspeed_camera.html

3 Laryngoscopy High-Speed Video Data

The Laryngoscopy High-Speed Videoendoscopy (LHSV) examination technique is one of the optical rigid laryngoscopic methods. It uses a special camera capable of capturing images at a frame rate much higher than the oscillation frequency of the vocal cords. The camera is inserted into the patient's mouth during a laryngoscopic examination and the patient's phonation of the vowel "i:". The biggest advantage of this method over stroboscopy or kymography is the possibility to capture the whole picture of vocal cords several times during one period of oscillation. From these images, we can observe the behavior during the opening and closing phases, compare it with other periods and see the exact location of the issue.

3.1 Used Camera

The camera used in the ENT department of the University Hospital in Pilsen is HRES ENDOCAM 5562 from Richard Wolf, shown in fig. 3.1. This camera became a source of our data. This camera was also used for capturing recordings in [18] or [19].

Figure 3.1: *Example of a high-speed video camera (HRES ENDOCAM 5562, Richard Wolf)*

The recording from the laryngoscopy high-speed video camera shows the whole image of the vocal cords (as seen from the mouth) and the real movement of the vocal cords (in figure 3.3). During the examination, a few seconds of the patient's phonation is recorded. A specified part of the recording containing hundreds of

frames is then taken to cover several periods during steady phonation.

The parameters of the used camera are in the following table 3.1:

Table 3.1: *Richard Wolf, HRES ENDOCAM 5562 camera parameters.*

Frame resolution	256 × 256 [px]
Frame rate	4000 [fps]
Color output	color (RGB)
Modes	kymogram, glottogram, hi-res mode
Maximum length of recording	4 seconds
Integrated microphone	yes
Light source	300 W Xenon light

When recording, the scene needs to be illuminated, so the camera has its light source. During the examination, the examiner must make sure that the vocal cords are correctly positioned, illuminated, and focused in the recording. Another problem may be the movement of the camera relative to the patient's vocal cords. Despite the very high scanning frequency, the shaking of the examiner's hand can cause a gradual change in the position of the glottis in the image within the sequence (more in section 3.3 about Input Data Quality).

The reason for the used vowel "i:" is to make the supraglottic space for the camera the most accessible [20]. In figure 3.2, there is a difference compared to other examinations, especially based on acoustic methods, where the vowel "a:" is used, but where the space above the vocal cords is not so accessible.

Figure 3.2: *The difference in the position of the larynx in magnetic resonance images (from [20]).*

An example of images from a high-speed camera is shown in figure 3.3. Due to the position of the camera, the right vocal cord is on the left side and vice versa. The

vocal cords' area is illuminated by the camera light source except for the glottis which is darker.

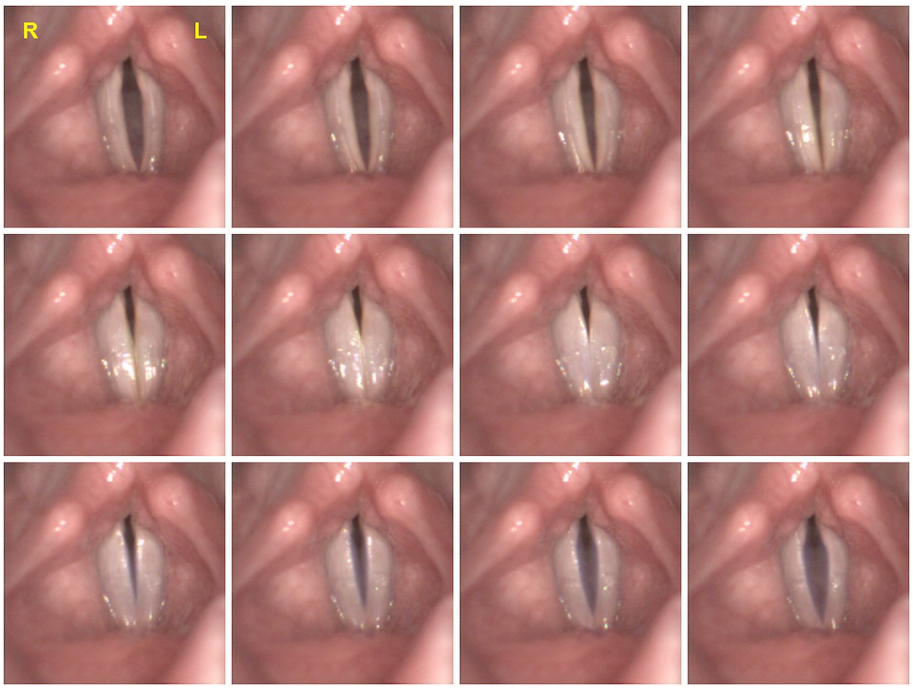


Figure 3.3: *Examples of frames from LHSV recording. Data are from the database of LHSV from the ENT department of the University Hospital in Pilsen.*

The position and size of the glottis in frames may be different in each recording because it depends on the camera position and anatomic differences of the patients. It means that pixel measurement needs to be adjusted individually and for comparison, normalization needs to be used.

3.2 Data Corpus Structure

The corpus of obtained data contains a diverse set of diagnoses and video recordings that are the source of a variety of anatomical structures and case studies in LHSV images. This diversity was useful for detailed testing of own detection methods regardless of diagnoses. There are also multiple records of individual patients in the corpus, usually for monitoring the progress of the disease or treatment, e.g. before and after microsurgery. From our point of view of testing detection methods, such video recordings are unique and independent.

There were two criteria to include video recording in the corpus. The first was that the video contains the vocal cords and the vocal cords were moving. The second criterium was to have data from other examinations and ENT expert evaluation of vocal cords' quality[1], see section 7.3.1. There was no condition about video quality or diagnosis to have real data to work with even though the quality of the recording

[1] All recordings were evaluated by biomedical engineer J. Pešta with ENT doctor M. Vohlídková

Table 3.2: *LHSV data corpus structure obtained from ENT clinic used for further analysis and processing.*

LHSV data corpus (no. 692)						
diagnosis	number of persons tested			video recordings		
	total sum	men	women	total sum	men	women
cyst	1	0	1	7	0	7
granuloma	1	1	0	2	2	0
vocal nodules	12	4	8	27	9	18
papilloma	2	2	0	10	10	0
recurrent lar. n. par.	61	19	42	220	63	157
Reinke's edema	7	1	6	17	1	16
vocal polyp	16	8	8	38	14	24
cordectomy	1	0	1	12	0	12
carcinoma	4	4	0	4	4	0
hemangioma	1	0	1	3	0	3
thyroidectomy	24	2	22	46	4	42
vocal fold leukoplakia	1	1	0	5	5	0
chronic laryngitis	1	1	0	3	3	0
healthy vocal cords	21	1	20	55	3	52
dg. was not available	101	51	50	243	127	116
total	254	95	159	692	245	447

is very low. It was possible to gather a corpus with 692 video recordings, which were used for the analysis and testing of our methods.

It should be mentioned that the corpus can contain recordings with various phonation intensities. The behavior of the vocal cords can be different in the case of loud phonation, where there is a higher probability of incomplete glottis closing. Also, the frequency F_0 is not known (the camera software tries to detect this value, but the detection is not always successful so this information was not used). The hospital database contains more information about every patient from other examinations described in section 2.5, but for vocal cords detection in this work, only image data from LHSV are used.

The structure of the data corpus can be seen in table 3.2, table 3.3 presents the age distribution of the patients examined and table 3.4 shows the number of available examinations by different methods like VRP, MDVA, and SCORE.

where the rating of movement behavior was assigned (used in section 7.3 and the axis was manually determined (used in section 5.3).

Table 3.3: *LHSV data corpus age distribution of the patients.*

	age distribution of patients				
	min	max	median	average	SD
men	21	88	53	51.05	16.22
women	13	80	51	47.03	16.51
total	13	88	51	48.45	16.51

Table 3.4: Available Data from VRP, MDVA, and SCORE examinations for the LHSV data corpus.

LHSV recordings + examinations VRP, MDVA, SCORE	number
LHSV total	692
LHSV and **VRP**	332
LHSV and **MDVA**	350
LHSV and **SCORE**	312
LHSV and **VRP** and **MDVA** and **SCORE**	273

3.3 Input Data Quality

The quality of the recording is a very important criterion that affects the success of glottis detection. Good quality images (according to the enumeration below) are segmented with high precision and do not show a high error rate.

The video recordings were included in the corpus regardless of the quality (as mentioned in section 3.2). The focus was to develop image processing methods that can recognize glottis for most of the videos.

Depending on the quality of the input data, the video sequences can be classified into 4 categories:

1. Good quality images where automatic detection of the vocal cord gap is usually successful.
2. Poor quality images, video can still be automatically processed with satisfactory results.
3. Poor quality images where the vocal cord gap can no longer be detected automatically and user input is required.
4. Unusable recording where the glottis is not visible or there is no movement.

However, for lower-quality images, the probability of error is higher, either in the selection of ROI (Region of Interest, more in section 4.3) or the glottis detection itself, especially at the edges.

The main quality issues are the following, more details can be found in [21]:

- Blurred image – incorrect focus
- Overexposure – too bright lighting
- Underexposure – too low lighting
- Overlapping – bad camera position
- Camera movement – a movement of the camera or patient
- Noise – incorrect settings, too low lighting
- Glottis out of the image – bad camera position
- Presence of fluids
- Glottis is too small – the incorrect camera position

4 Glottis Detection

The detection of vocal cords and glottis from LHSV video sequences has become a relatively frequent topic in universities and research centers around the world since the usage of these systems has expanded (see section 4.1). This is a special case of the application of image processing methods, which are also used in other areas of computer vision.

The main task of vocal cords detection is a segmentation of the glottis (the gap between vocal folds). There are many ways how to achieve this and there were already many methods published (mentioned in 4.1). Software from the camera manufacturers is usually using a manual or semi-automatic approach where a user needs to insert several inputs like more precise vocal cords location (delimit the Region of Interest – ROI), select several reference points or mark the axis of symmetry. In many publications, other methods were introduced to achieve glottis segmentation without user interaction, which can avoid errors of user input, increase speed, obtain objective results, and enable bulk processing of the recordings. Unfortunately, these automatic methods are usually tested on specific data corpus without detailed information about the input data quality. Results of such methods are usually presented on high-quality and specific video recordings. This work presents methods used for processing all data obtained from the ENT department where the quality may vary.

The segmentation can be done within a single frame from the LHSV recording, but usually, the advantage of the frames in the video sequence is used, e.g. to search the location where the movement is in progress or to find the best frame for the initial glottis detection. Such preprocessing of the video data is appropriate because segmentation methods are sensitive to disturbances in the picture like noise or irrelevant movement in surroundings. The noise reduction and other preprocessing methods were already described in [21].

4.1 Overview of Published Methods

This section briefly presents several published methods for image segmentation of LHSV. A detailed description is presented in [21].

Many methods are based on the analysis of the image histogram and its properties. This can be used to detect an object that is distinguished in the image by the value of brightness or at least one of the color components from the background. In the ideal case, the histogram is bimodal – containing two local maxima and a local minimum between them, which is zero or significantly lower than the maxima. The thresholding method deals with determining the boundary that separates the brightness of the points belonging to the glottis from the brightness of the points representing the background. These methods use the Otsu method [22], minimum error method [23], or heuristic methods.

Another approach for detecting the glottis in an image is to find edges and use them for segmentation. The principle of these methods is to find the boundaries of the object in the image by finding a steep brightness gradient or a sudden change in the value of color components in the image area. In the work [24] the Sobel operator is used to calculate the gradient as pre-processing. The work [25] uses the Sobel operator to specify the area of interest (ROI search). However, the article [26] mentions that these operations do not show good results for lower-quality images. There are also other methods based on edge detection, like the Canny edge detector algorithm or the usage of Mathematical morphology [27].

A different approach is to use Gabor filtering for image processing. It is a linear filter whose impulse response is defined as a Gaussian function modulated by the harmonic function [28]. This method is used in [24] together with the Wiener motion estimation method, which is used as a preprocessing to select the region of interest.

Another method is Active Contours. The principle of this method is based on minimizing the energy of the curve, which divides the image into individual areas (according to [29]). First, the initial position and the shape of the curve (needs not be continuous) and the energy are determined. Energy has 2 components, a component of internal energy (continuity and curvature of the curve) and a component of external energy, where the curve is favored if it leads to the image at the edge location. Gradually, the energy of the curve decreases iteratively until it reaches a minimum when ideally the curve copies the boundary of the object. The Active contours method is used to segment the vocal cord image in [30] (combination with thresholding) and [31] (3D active contours). Preprocessing can be used as the initial location of the curve – finding the area of interest a the approximate position and shape of the glottis.

An image topology can be also used for image segmentation. These methods use a layout of objects, colors, and shapes in the image. The Watershed segmentation method is used in [32], Region growing in [33].

4.2 Approach Used in this Work

Obtained data from the ENT department and the data structure are described in chapter 3. These data can be of different quality and be taken from patients with a variety of diagnoses or healthy volunteers. All such data need to be processed.

As mentioned before, the goal is to implement or develop an automatic method for glottis detection to offer independent support for examining doctors to evaluate vocal cords' behavior. To make the segmentation method as stable as possible, the ROI determination was selected as a preprocessing. This approach also helps to increase the efficiency of the whole image processing. During development, two methods of delimiting ROI were used (described in chapter 4.3).

Also for glottis segmentation itself, two methods were developed (described in sec-

tion 4.4). After the segmentation, a set of parameters was calculated for further evaluation and special focus was taken on data regarding vocal cords symmetry, which is the reason for searching for the symmetry axis (described in chapter 5).

In the further sections, two statistical methods were introduced to detect vocal cords irregularities (see 7.2) and to find a scoring function to provide a single value of vocal cords evaluation (see 7.3).

The overall schema of LHSV processing is presented in figure 4.1.

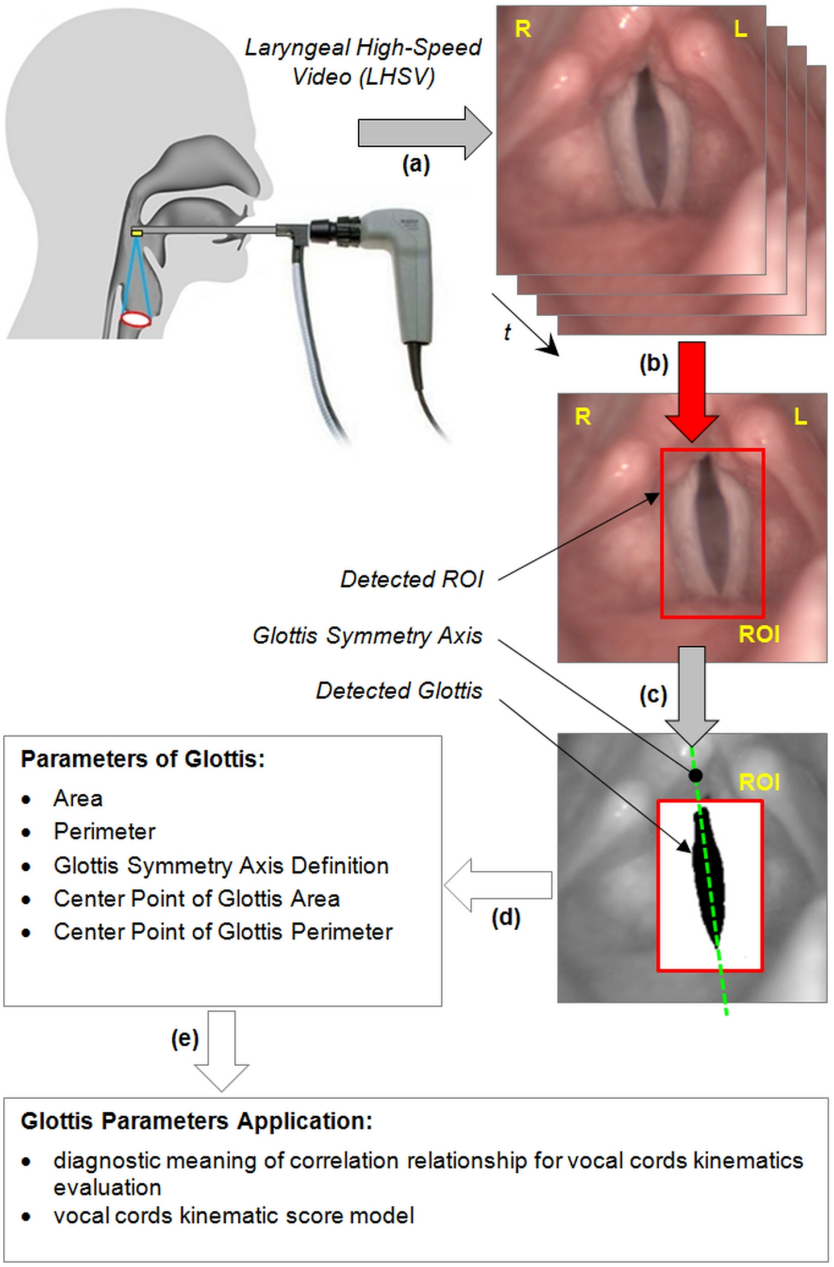


Figure 4.1: *Schema of the processing of LHSV recordings.*

4.3 Region of Interest (ROI)

The first step of image processing by the selected approach is determining the ROI. In almost all cases, it is not necessary to process the entire image, but only the part of it which contains the vocal cords. Restricting the area helps to speed up further processes and also to improve the results because the most of interfering elements causing false detection can be removed by restricting the area.

4.3.1 Methods

The determining of the ROI is part of many approaches to glottis detection thus several methods were already mentioned in section 4.1, following overview is focused exclusively on ROI detection.

There are many publications describing automatic ROI detection in LHSV videos. One of the methods is described in [33] and [34] which is based on a region-growing algorithm with a selected seed point. Another approach using morphological operations was presented in [35]. Publications [36], [37] mentioned ROI detection based on subtraction methods using more frames in the sequence, a similar approach was used in our method, described in section 4.3.2 and in [4]. Methods based on intensity variations of pixels are described in [33], [32] and [38], motion estimation is then used in [24] and [39]. Publication [40] describes the salient region method using topological structural information. Other methods like Deep Neural Networks were mentioned and the overall summary of methods used for ROI detection in LHSV recording or video frame was presented in [5].

It should be noted that according to the available information, the proprietary software provided with the LHSV systems usually allows only manual ROI determination.

4.3.2 Thresholding Method

The first developed method for ROI detection was described in [4]. It was a simple method based on subtracting the image with the most open and the most closed vocal cords. Pixels from the areas without any change have zero values after subtraction. Ideally, the only part of the resulting image with non-zero values is the vocal cord region. In most cases, however, it is necessary to threshold the resulting image and exclude random visual artifacts by finding the largest contiguous area (e.g. using Connected component labeling (CCL) [41]). The ROI thresholding algorithm is presented in figure 4.2:

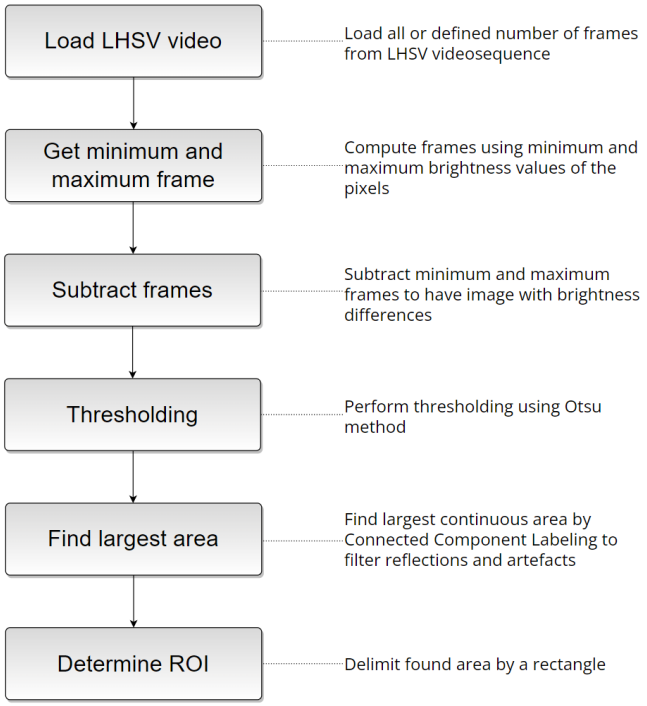
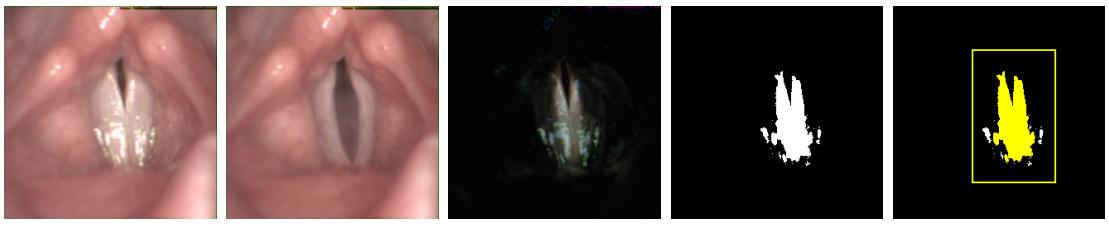


Figure 4.2: *Scheme of ROI thresholding method.*

The minimum frame in this method is a computed image consisting of minimum values of the pixels on the coordinates x and y from all frames or several periods. This method uses the fact that the glottis is usually darker than its surroundings and also deals with situations where the vocal folds are not moving synchronously. The same principle works for the maximum frame, which consists of the lightest pixels. The minimum frame is then subtracted from the maximum frame to get an image where the differences are visible in a lighter color. The resulting image is then thresholded and the largest continuous area found by CCL is then considered as a rough position of the glottis. This area with defined padding is delimited by a rectangle. Examples of the minimum and maximum frames, a frame after subtraction, thresholding, and finding the glottis area can be seen in figure 4.3.



(a) Maximum (b) Minimum (c) Subtracted (d) Thresholded (e) CCL

Figure 4.3: *Maximum and minimum frame computed from LHSV, resulting image of subtraction, thresholded image, and found largest continuous area with determined ROI.*

The results of this method are sufficient for most cases. However, in some images,

for example with little movement of the vocal cords or other disturbing elements like camera movement or present fluids, which can be seen in figure 4.4, this method reacts incorrectly, for example by marking the entire image area or a completely different position as ROI.

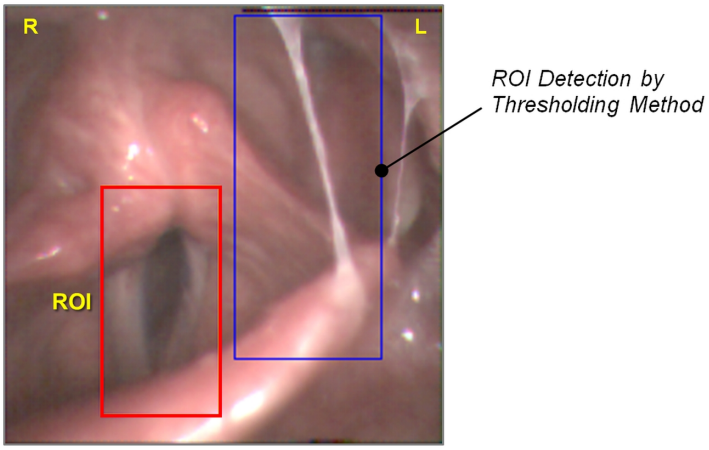


Figure 4.4: *A case study with fluid where the thresholding method of ROI detection failed. See the behavior later in fig. 4.7. Displayed picture is computed minimum image.*

4.3.3 DFT Method

Because a well-defined ROI is a prerequisite for subsequent successful glottis detection, it was decided to test a different approach and use the method using frequency analysis of the brightness (or color components) change of the pixels to evaluate the oscillation of the anatomical structure corresponding to these points. This method is using laryngotopography, see [42], [43], [44], [45], [47]. Based on the defined decision rules, the tested pixel is then included in the ROI or not. The methods created in this way are denominated as "DFT ROI detection methods".

Principles Used for DFT ROI Detection Methods

According to the physiology of voice formation and the characteristics of vocal cords behavior, an approach for ROI detection, which is based on the oscillation of anatomical structures in LHSV frames, can be used. The movement of the vocal cords is repeated regularly and the periodicity contained in the movement of vocal folds corresponds to the fundamental vocal cord frequency F_0 (or near to F_0). This periodicity is subsequently reflected in the brightness change of individual pixels in the LHSV frames (value of brightness Y or the values of the color components R, G, B), see [42],[43],[44].

To detect the periodicity in the pixels of LHSV images, the discrete Fourier transform (DFT) is used, defined for the finite number of samples of the equidistantly sampled input signal, and the method of spectral analysis of the signal using the

DFT amplitude spectrum. This input signal is the brightness or color component value at the pixel (x,y) of LHSV images. The higher rate of oscillation was recognized in the red color component, all further operations use R values. This method was described more in detail in [48].

The illustrative result after using DFT, based on the analysis of the oscillation of anatomical structures, is presented in fig. 4.5, which shows the overall pixel distribution with the detected frequency F_0, see fig. 4.5(a), and the distribution of DFT-amplitude spectrum values at these pixels, see fig. 4.5(b).

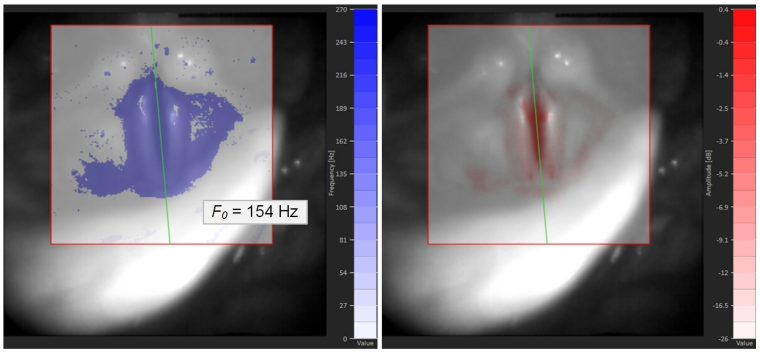


Figure 4.5: *Distribution of calculated values in LHSV: (a) detected frequency F_0, (b) DFT-amplitude spectrum.*

Following figure 4.6 shows the pixel brightness change and DFT amplitude spectrum values of two points. One point is in the glottis area and another one is in the surroundings. There is visible oscillation and the frequency can be easily detected by DFT.

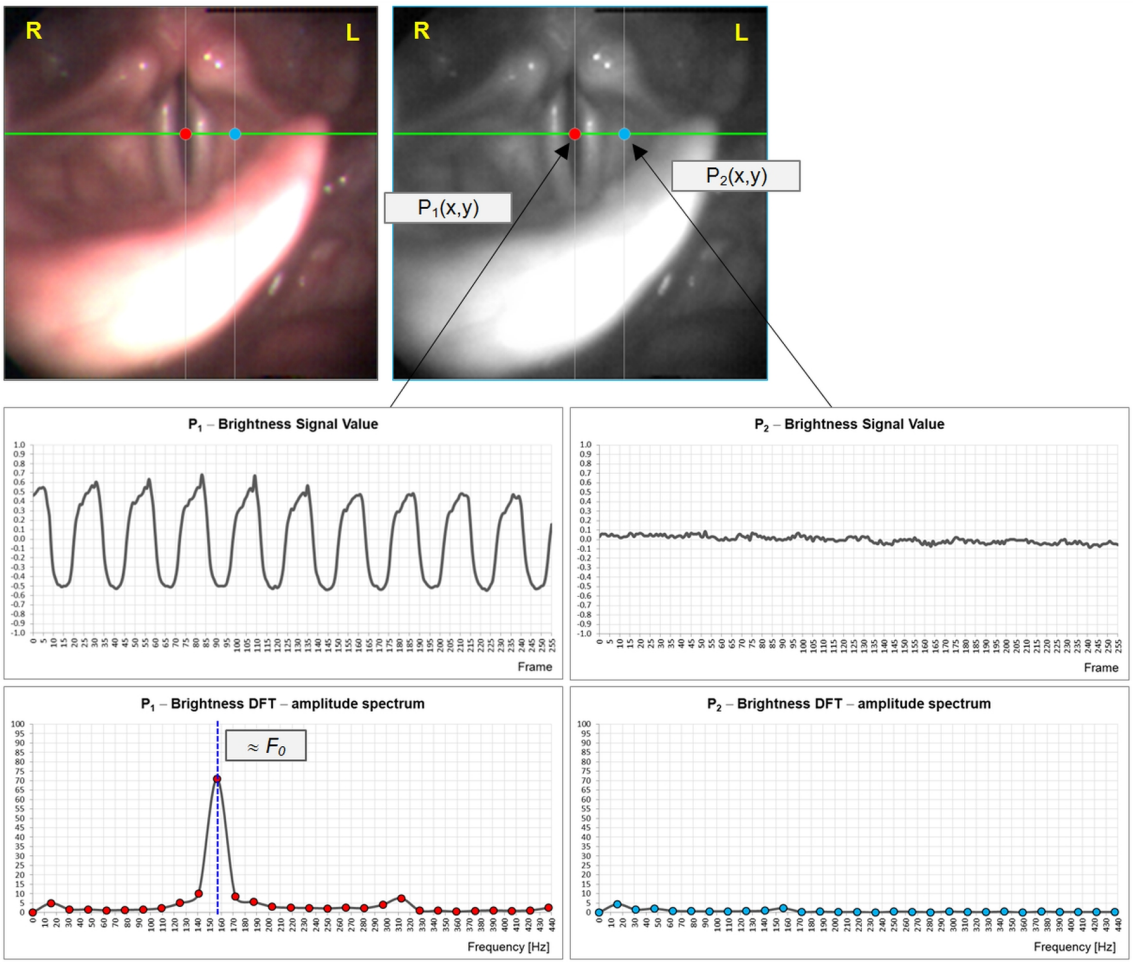


Figure 4.6: *Pixel brightness change and DFT-amplitude spectrum value of two points in an LHSV recording of healthy vocal cords.*

Figure 4.7 shows the situation with the image containing fluid, which affects the ROI recognition by the thresholding method in fig 4.4. Again, two points were selected to show the pixel brightness change and DFT amplitude spectrum values. One point is in the glottis area (on the edge of the vocal fold) and the second one is in the area of fluid. It can be seen that brightness values oscillate in both cases, but the frequency is different.

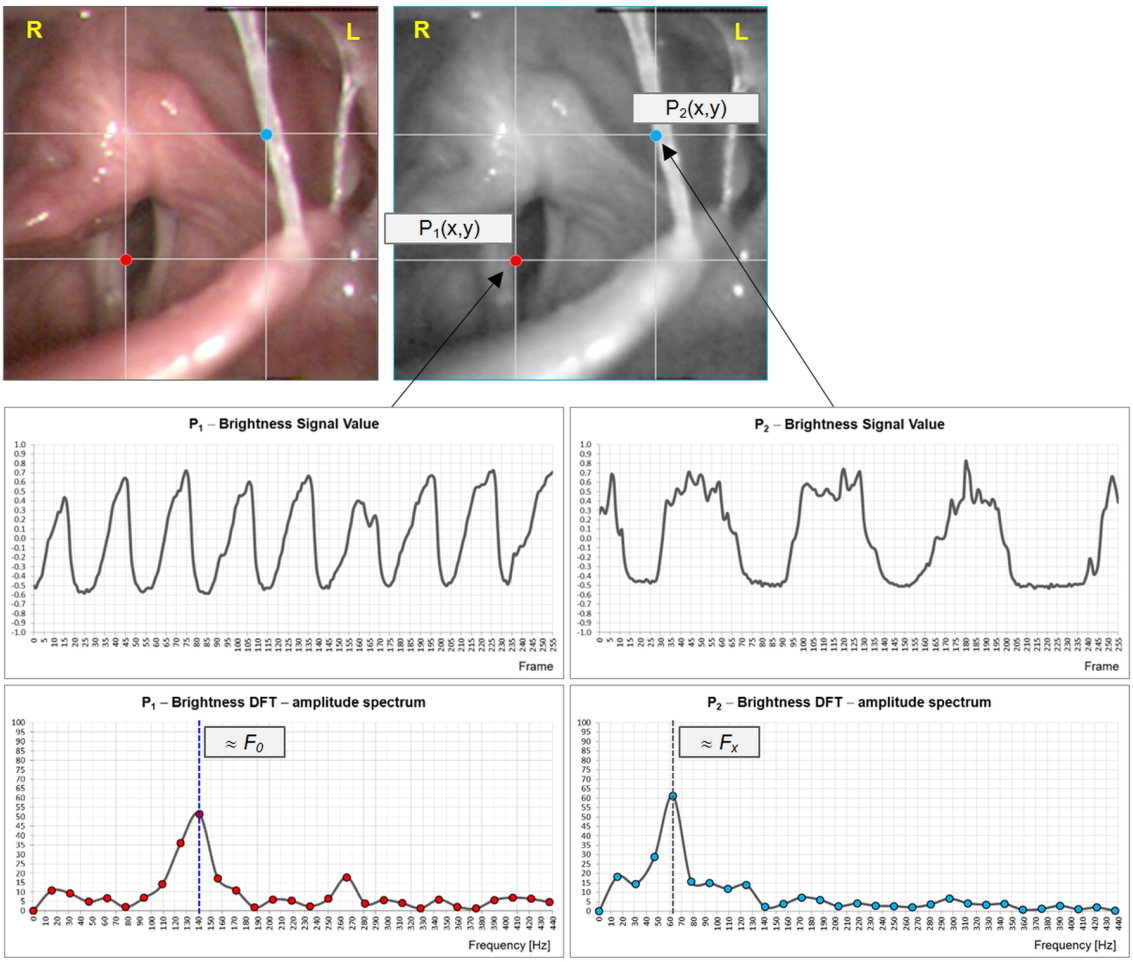


Figure 4.7: *Pixel brightness changes and DFT-amplitude spectrum values for two points in an LHSV recording containing fluid. The first point is in the glottis area, the second point is in the area with fluid. There is an apparent difference in the most significant frequency in the glottis area (F_0) and the fluid area (F_x).*

Using these findings, the following methods were introduced.

General ROI DFT Variant

As input data for this method, values of pixels (R component was used, but also brightness value can be used) $pv(x, y, i)$ are taken. The x and y are the coordinates of the pixel in the frame and i is the frame index (see example in fig. 4.8).

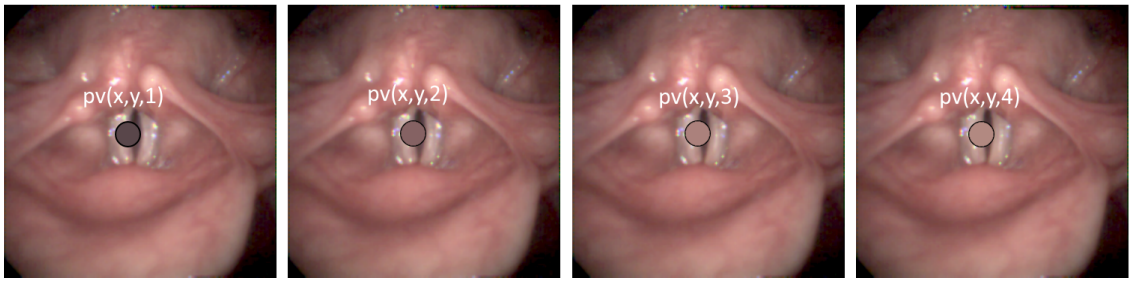


Figure 4.8: *Example of points pv(x,y,i) in LHSV sequence with color change.*

Every array of values $pv(x, y, k)$ where $k = 0$ to $N - 1$ is multiplied by a rectangular time window (other windows can be used like Kaiser or Hamming window) and normalized first to optimize the results according to eq. (4.1)[45]. N is the number of frames in the recording.

$$pv^{w}_{norm}(x, y, i) = \frac{pv(x, y, i)}{\frac{1}{N}\sum_{i=0}^{N-1} pv(x, y, i)} \tag{4.1}$$

Then the Fourier spectrum $PV(k)_{(x,y)}$ and DFT amplitude spectrum $|PV(k)|_{(x,y)}$ is calculated for $\forall k = 0, 1, 2, ..., N - 1$:

$$PV(k)_{(x,y)} = \sum_{i=0}^{N-1} pv^{w}_{norm}\,(x, y, i) \cdot e^{-j\frac{2\pi}{N}ik} \tag{4.2}$$

$$|PV(k)|_{(x,y)} = \sqrt{\left(Re[PV(k)_{(x,y)}]\right)^2 + \left(Im[PV(k)_{(x,y)}]\right)^2} \tag{4.3}$$

For each pixel (x, y) of the video sequence, the maximum value of the DFT-amplitude spectrum was further determined together with the index corresponding to this maximum $k_{max}(x, y)$. This index corresponds to the frequency of change (periodicity) of the pixel value at the point (x, y). For $\forall(x, y)$, an array of maximum amplitudes $|PV(k_{max})|_{(x,y)}$ and their corresponding indices $k_{max}(x, y)$ were obtained.

Because of working with the pixel value change, which corresponds to the anatomical structure of the vocal cords moving with the fundamental vocal cords' frequency F_0, the frequency spectrum can be limited to the frequencies f_{LOW} and f_{HIGH} for further analysis and processing. These frequencies are set according to the lowest and highest possible voice frequencies during the LHSV examination. Values out of this range are considered as a noise or a not relevant movement in the vocal cords image.

The f_{LOW} and f_{HIGH} values were determined by statistics from other examinations or camera microphone on the same corpus no. 692. The distribution of frequency F_0 of many LHSV examinations is shown in figure 4.9. As can be seen, almost all values are between 80 and 430 Hz.

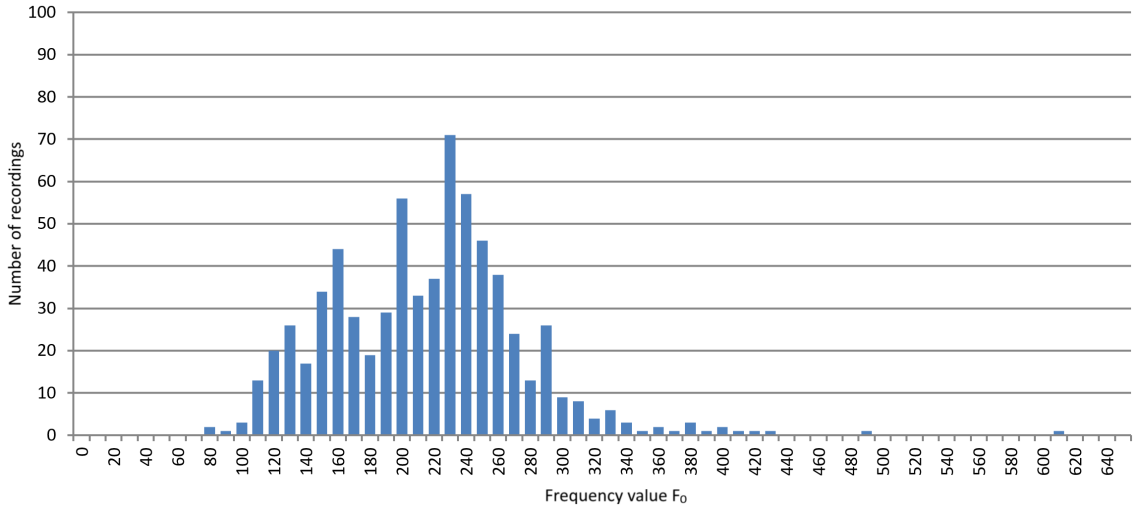


Figure 4.9: *Distribution of frequencies F_0 from other examinations to determine f_{LOW} and f_{HIGH}.*

Because of the DFT principle, where the frequencies are represented by an integer index k_i, minimum and maximum frequencies are set according to the nearest k. According to the LHSV camera frame rate and the number of frames, the precision (frequency bin) is 15.625 Hz. So the resulting limits are following:

$$k_{LOW} = 5 \implies f_{LOW} = 78.13\ Hz \tag{4.4}$$

$$k_{HIGH} = 28 \implies f_{HIGH} = 437.50\ Hz \tag{4.5}$$

Because the frequency F_0 of vocal cords in a currently analyzed LHSV video is unknown, it needs to be determined e.g. from the highest amplitudes of DFT results. The pixels with amplitude higher than $T_{|PV|}$ (threshold for value $|PV(k_{max})|_{(x,y)}$) are selected and a histogram of corresponding $k_{max}(x,y)$ is computed. The most occurred index $k_{max}(x,y)$ within the k_{LOW} and k_{HIGH} is selected and labeled as k_{ref}. The corresponding frequency is an estimate of F_0 of the vocal cord's oscillation. $T_{|PV|}$ value was set as a value of a pixel with the 256th (square root of the pixel number) highest value, i.e. the top 256 pixel values were used to find k_{ref}.

The amplitude values of the k_{ref} of all pixels are then used to determine the areas which oscillate near F_0. The pixels which meet the condition (4.6) are then considered to be a first ROI estimate $ROI^{(0)}$. The value of threshold T_{ROI} was set to 0.5 after extensive testing during this study.

$$|PV(k_{ref})|_{(x,y)} \geq T_{ROI} \cdot \max_{\forall(x,y)} \left(|PV(k_{ref})|_{(x,y)} \right) \tag{4.6}$$

In some cases, the isolated pixels or small groups of pixels were found within the first ROI estimate $ROI^{(0)}$ due to noise or reflections. To avoid such cases, these pixels were filtered out using a morphological transformation $OPENING(n)$ (Erosion

followed by Dilation using the same structuring element 3×3 for both operations, [27], [46]). The depth n of the opening was set to 2 to filter isolated small groups of pixels. The result is the ROI estimate $ROI^{(1)}$.

Geometrized ROI DFT Variant

Because it is not necessary to have ROI delimited with one-pixel precision, it is possible to geometrize this task. The input frame can be split by a grid, each resulting square is for further processing represented by a single pixel, see example in figure 4.10. The value of the representative pixel can be set as an average value of all pixels within the square. Also, different methods could be used to define representative pixels (like minimum, maximum, random...), but the average value led to the best results during testing. The squares created by applying the grid were called Elementary squares (ESq). Several sizes of the Elementary squares were tested (4×4, 8×8, 16×16) and the results of 8×8 were returning the best results. Because of averaging pixels, this approach also eliminates the isolated pixels in the result and further filtering is not needed, unlike the general variant.

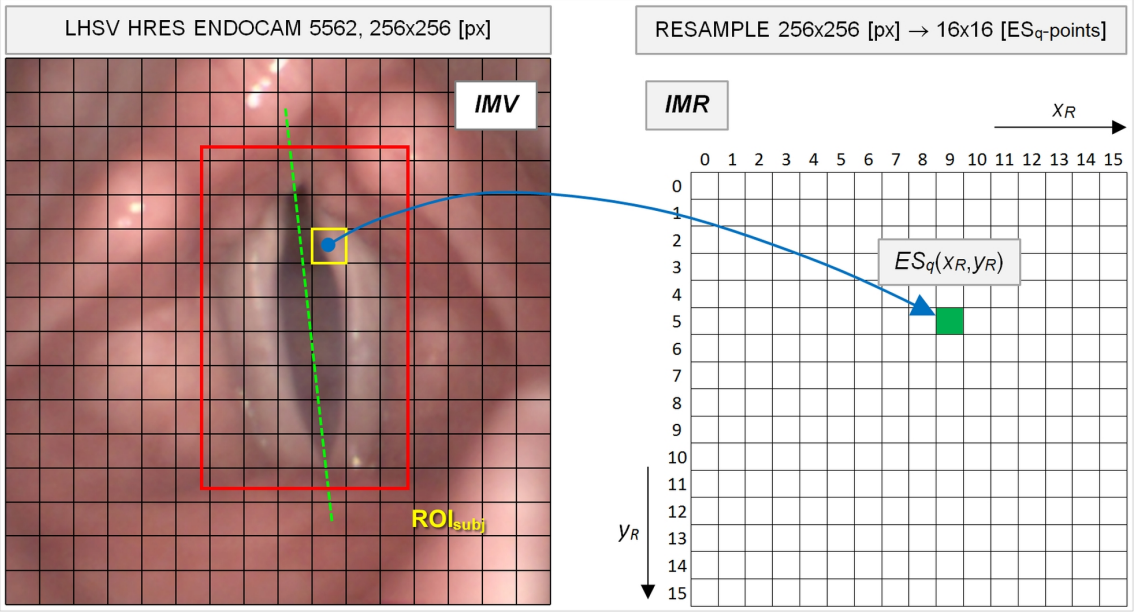


Figure 4.10: *Example of resampling of LHSV using a grid* 16×16. *Each elementary square is then represented by a single pixel.*

A new smaller image will be created from the created pixels representing elementary squares, which can be processed like in the general method. Image is smaller, so the computing is faster and further post-processing is not necessary. The resulting set of pixels is then applied back to elementary squares and all pixels within these squares are then part of the ROI estimate $ROI^{(1)}$.

Final ROI Estimation

In general, ROI estimate ($ROI^{(1)}$) can have any shape, it is just a set of pixels located in the area covering the vocal cords. But for easier delimiting, the circumscribing rectangle is considered to be an ROI. This rectangle is usually enlarged in all directions to avoid too close position from the edges of the vocal cords. This is then considered to be a final ROI. The complete scheme is shown in figure 4.11.

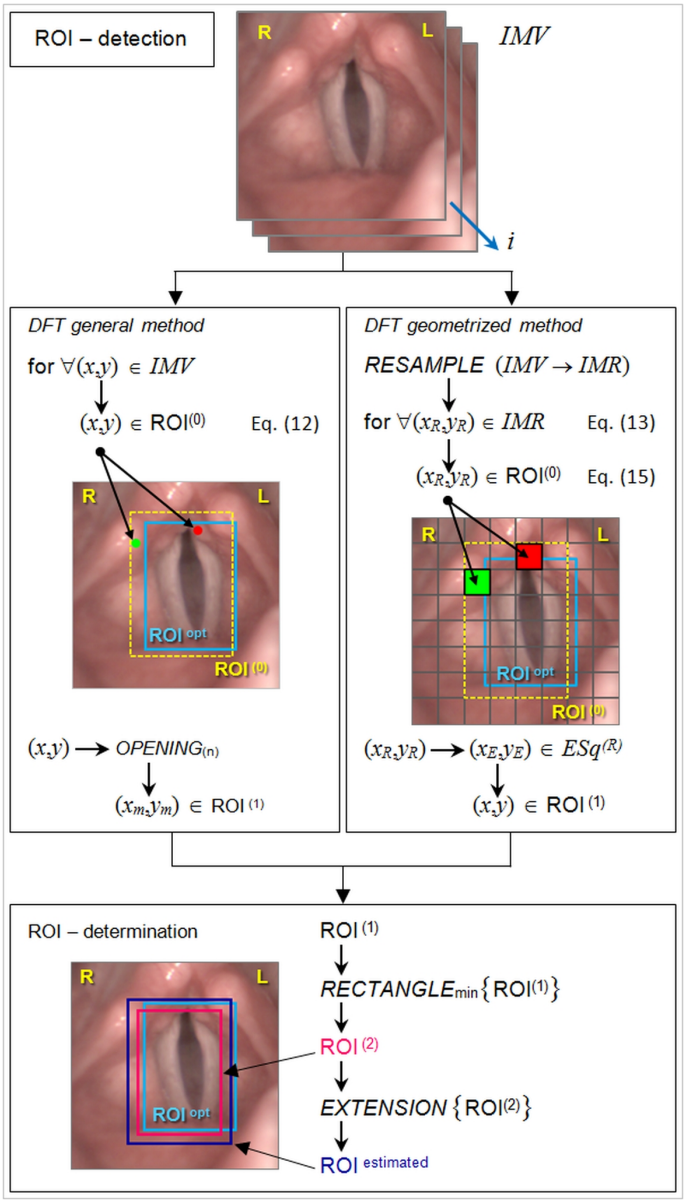


Figure 4.11: *Principle diagram of the procedure for obtaining the resulting ROI estimate by DFT General and DFT Geometrized methods.*

4.3.4 Results

In the article [48], the results were presented on the data corpus containing 412 LHSV recordings, see table 4.1. This corpus was further extended to the current 692 recordings, the testing results are presented in table 4.2.

Table 4.1: *Results of ROI methods from LHSV data corpus 412 from article [48].*

ROI Detection	Thresholding method		DFT General method		DFT geometrized (8×8) method	
CORRECT	285	69.17%	347	84.22%	368	89.32%
FAILURE	116	28.16%	32	7.77%	35	8.50%
FAILURE TIGHT	11	2.67%	33	8.01%	9	2.18%
TOTAL	412	LHSV data corpus No. 412				

Table 4.2: *Results of ROI methods from LHSV data corpus no. 692.*

ROI Detection	Thresholding method		DFT General method		DFT geometrized (8×8) method	
CORRECT	491	70.95%	593	85.69%	624	90.17%
FAILURE	183	26.45%	49	7.08%	54	7.80%
FAILURE TIGHT	18	2.60%	50	7.23%	14	2.02%
TOTAL	692	LHSV data corpus no. 692				

The ROI was correctly determined in more than 90% of LHSV recordings which is much better comparing it with the 70% success rate of the thresholding method. The FAILURE TIGHT result category contains cases, where the resulting ROI was not correct, because of a small part of the glottis was missing, but the result would probably not affect the further processing of glottis segmentation. But the ROI itself was not correct so it was considered to be a failure.

In the following examples, the thresholding and DFT methods were compared. Figure 4.12 shows a situation where the randomly moving mucus caused significant differences in the maximum and minimum frames leading to the wrong ROI determination by the thresholding method (a similar situation was shown in fig. 4.7, the mucus oscillation frequency is smaller than f_{LOW} and is filtered when using DFT methods, which found the ROI correctly).

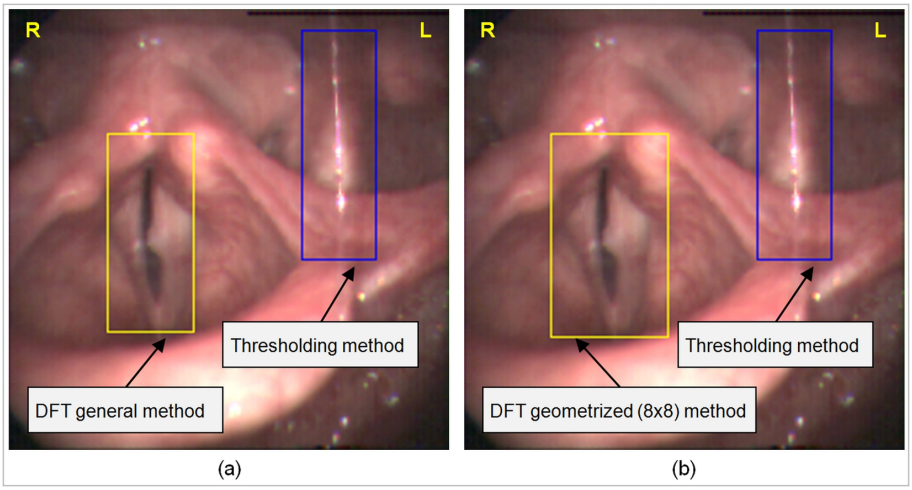


Figure 4.12: *Comparison of DFT methods and thresholding method. In this case, the thresholding method incorrectly determined the ROI outside of the glottis area because of randomly moving mucus. The DFT methods found the ROI correctly.*

figures 4.13 and 4.14 show the cases where the DFT general method failed to find ROI correctly and part of the glottis was missing in the result (example of FAILURE TIGHT result). The geometrized variant then found the glottis correctly.

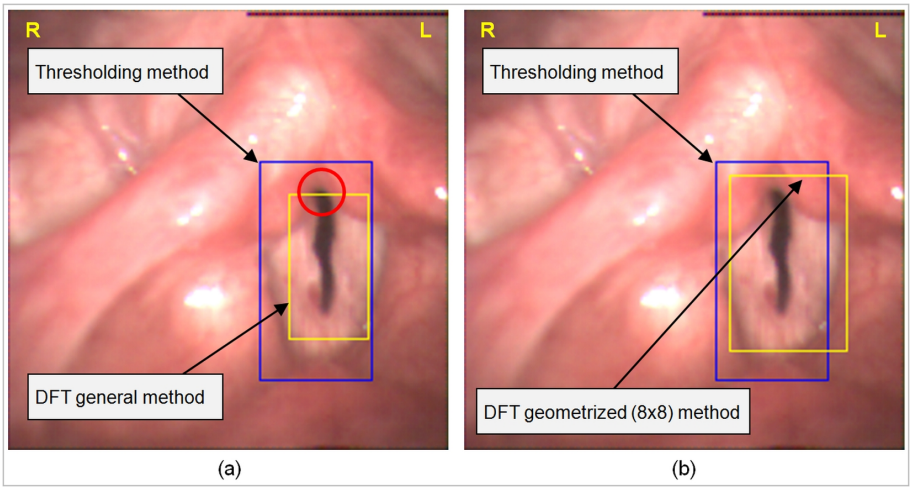


Figure 4.13: *Comparison of DFT methods and thresholding method. In this case, the thresholding method found the ROI correctly, but DFT general method missed part of the glottis. The geometrized method then determined the ROI correctly.*

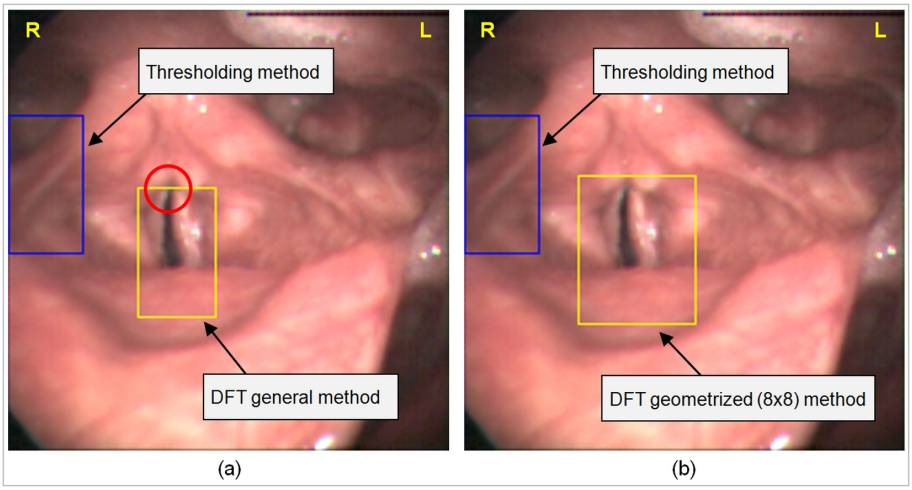


Figure 4.14: *Comparison of DFT methods and thresholding method. In this case, the thresholding method determined ROI completely out of glottis. DFT general method missed part of the glottis, the geometrized variant then delimited the ROI correctly.*

The last example in figure 4.15 presents the situation, where both DFT methods failed to determine ROI correctly. Again, a part of the glottis was missing. On the contrary, the thresholding method determined the ROI correctly in this case. This is one of the rare occurrences where the latter methods were less successful than the previously used method.

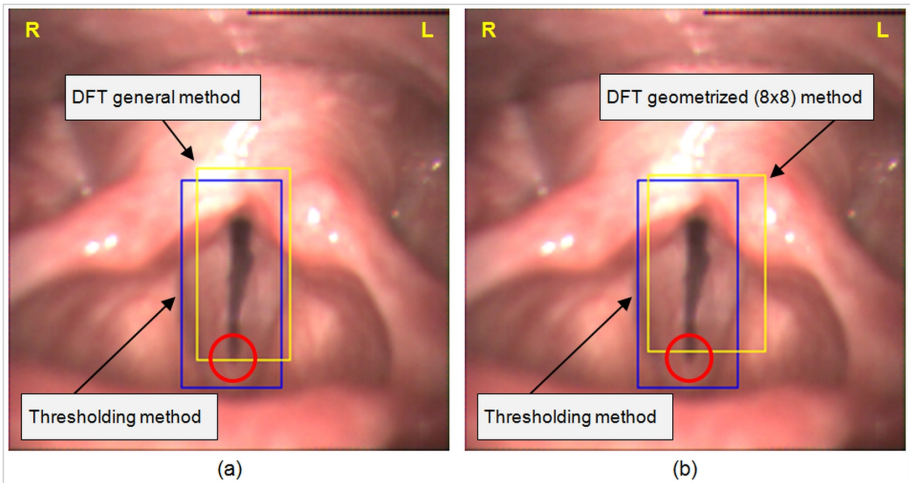


Figure 4.15: *Comparison of DFT methods and thresholding method. In this case, the thresholding method correctly determined the ROI, but the DFT methods missed the bottom part of the glottis.*

4.4 Detailed Glottis Segmentation within ROI

Two methods were used to detect the glottis. Within the work [4], the first of them was developed, based on searching for the correct threshold using "Max-Min-

Thresholding" within previously determined ROI.

The second method also requires determined ROI and the glottis detection itself is based on cluster analysis, in which the pixels were divided into classes and one of these classes corresponds to the glottis. This method was mentioned in [21] and shows interesting results where the detection success rate is higher than using the first method.

4.4.1 Original Thresholding Method

The thresholding method was described in [4]. It is based on searching for a single threshold to separate darker glottis from lighter surroundings. In the ideal case, the histogram is bimodal and the threshold can be found. This method uses the Minimum error thresholding[49] approach on the already found ROI. The threshold is initially found for the minimum image, which contains maximally open vocal cords (described in section 4.3.2), and then the same threshold is then used for all frames in the LHSV sequence. The result is a black and white image for every frame showing the shape of the glottis in all phases of vocal cords movement, which can be further processed for computing parameters.

The steps of this method are described in figure 4.16.

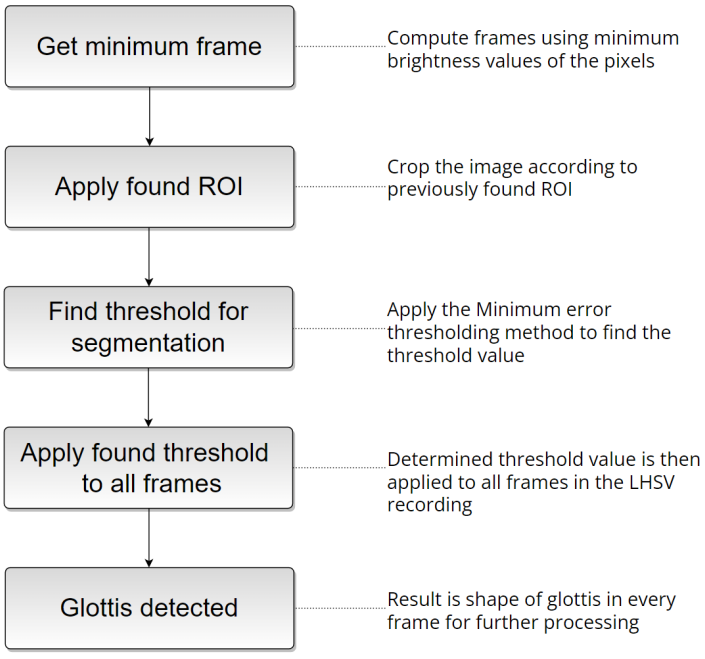


Figure 4.16: *Scheme of thresholding method for glottis detection.*

The glottis is usually successfully found in a case of a bimodal histogram, where the glottis area is darker than the whole surroundings. But it is quite often that the histogram of the minimum image is not bimodal and there are also places in surroundings with lower brightness than the glottis. This leads to the result including

not only glottis but also other parts of the image. In that case, a manual correction was necessary to obtain a correct result. Figure 4.17 shows the segmentation result, where the glottis was detected successfully, figure 4.18 shows an incorrect result, where the part of the glottis was not segmented completely.

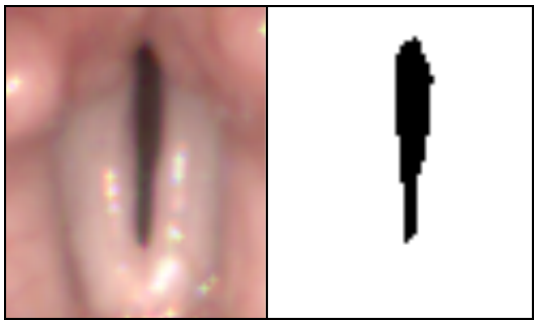


Figure 4.17: *Result of the thresholding method with the successfully detected glottis.*

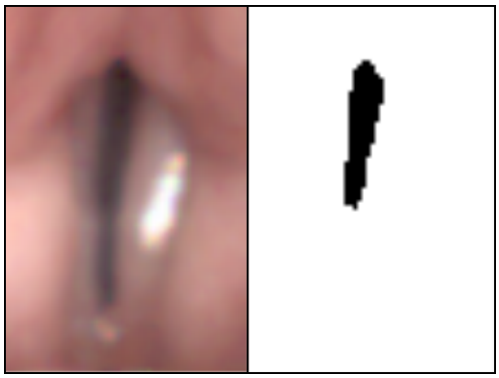


Figure 4.18: *Result of the thresholding method where the glottis was not fully detected.*

4.4.2 Cluster Analysis Segmentation Method

To improve glottis detection in LHSV recording where the thresholding method is not able to make successful segmentation, another method was introduced based on cluster analysis, which provided promising results from the beginning. This method was presented in [50].

In most cases, the glottis is darker than the surroundings, most notably in the red component of the RGB color model. For this approach, individual pixels can be considered as objects with parameters derived from image data. The objects representing the pixels can be then divided into classes with similar properties using cluster analysis, which is described e.g. in [51].

In general, cluster analysis represents the procedure of grouping objects into more or less homogeneous groups based on their similarity. The most used method of cluster analysis is the K-means method, see [51]. According to individual parameters,

objects are classified into classes ($CLASS_j$) with the smallest possible parameter differences within the class or with the largest possible parameter difference among classes. Individual classified objects are placed in m-dimensional space, where m is the number of monitored parameters. The coordinates of the m-dimensional space can be considered as indices to the parameter values, and each parameter can have a different weight. In this space, the distances between objects are then calculated and clusters of objects with similar properties are searched. The method requires an initial setup of so-called centers ($Center_j$), which can be some selected objects or newly created abstract ones for quick and accurate division of objects into classes.

The first step of the method is the initial division of the objects xx_i ($i = 1$ to n, where n is the number of objects) into classes $CLASS_j$ according to the selected criterion. The smallest distance between the object and individual initial centers $Center_j$ is used for every object, see equation (4.7), where $j = 1$ to k, k is the number of classes $CLASS_j$, i is the object index.

$$CLASS_j(xx_i) = \arg \min_{j=1...k} ||xx_i - Center_j|| \tag{4.7}$$

After dividing all objects into classes, in the second step, the new centers $Center_j$ of individual classes are determined by averaging all objects xx_i, which belong to class $CLASS_j$, see (4.8), where n_j is the number of xx_i objects in the $CLASS_j$ class (pic. 4.19).

$$Center_j = \frac{1}{n_j} \sum_{i \in CLASS_j} xx_i \tag{4.8}$$

This method can be repeated with new centers until a stable state when the groups of points no longer change, or the classification can be terminated after a specified number of cycles.

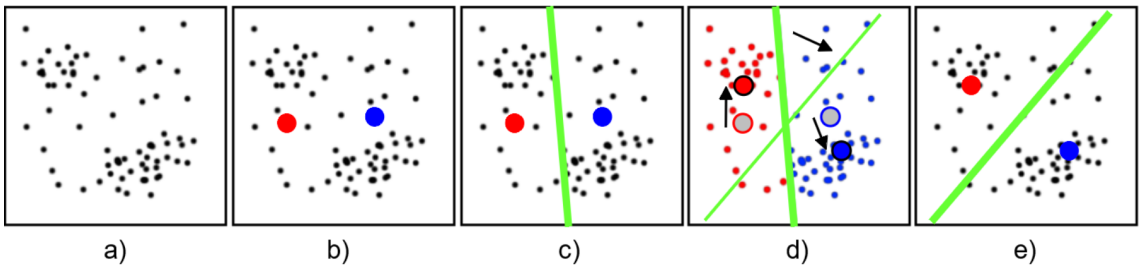


Figure 4.19: *Visualization of the principle of cluster analysis in individual steps:*
a) Objects in space xx_n in the frame;
b) Selected starting center points – $Center_j$, $j = 1, 2$;
c) Initial classification of xx_j to two classes $CLASS_j$;
d) Calculation of new center points $Center_j$ and a new classification of xx_j to classes $CLASS_j$;
e) Positions of new CPs $Center_j$ and classes $CLASS_j$ after next cycle.

If the pixels are treated as objects, several parameters can be selected for classification, including setting the weights of individual parameters. Optimal weight setting is a very complex task of the heuristic type, in this method, the values were determined by testing various settings. The description and final setting of parameter weights are given in table 4.3.

Table 4.3: *Description of individual parameters with settings of weights.*

parameter	value range	weight
R – red component value	0–255	1,0
G – green component value	0–255	0,1
B – blue component value	0–255	0,1
X – coordinate x in the frame	0–255	1,0
Y – coordinate y in the frame	0–255	0,1
RmB – difference between R and B	-255–255	1,0
C – distance from frame center in pixels	0–255	1,0

The red component value (R) has a high weight because the value change is most significant in this color component. Because of the vertical orientation of the vocal cords in the image, the coordinate x value (X) has also a high value of weight together with the distance from the center C due to the glottis position. RmB helps with distinguishing more red (surroundings) and black (glottis) areas. Other parameters have a lower influence on the detection thus the weight values are lower.

Before applying the cluster analysis method, it is also necessary to determine the number of classes, resp. the number of initial centers. Two classes should be sufficient to detect points belonging to the vocal cord region. However, in low-contrast images or ones where the glottis area contains a greater scale of brightness, it is advantageous to use 5 classes. This also helps to find correct segments when using x and y as parameters.

The result of segmentation when 5 classes were used and final weight coefficients can be seen in figure 4.20. The calculated centers at the frame with the most open vocal cord are then used for all frames in the video sequence.

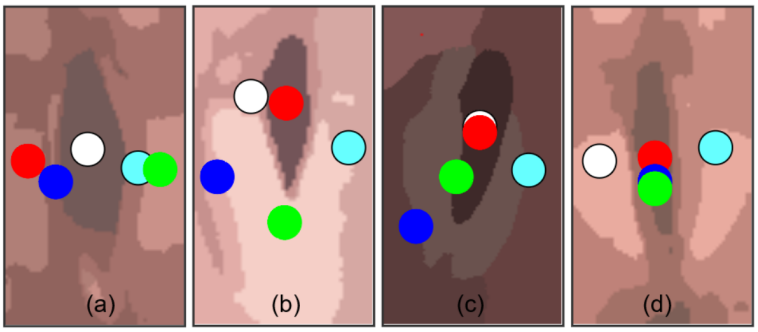


Figure 4.20: *Result of classification after the test: number of classes $j = 5$, coeff $_1$*

The initial selection of centers affects the speed and outcome of the classification. For the fastest classification, it is advisable to identify centers that have the most different parameters. In space, therefore, they should be the most distant. If the centers are selected at random, the result may be different when using a cluster repeatedly analysis. The method may not always find the global optimum.

Cluster Analysis with Adaptive Parameter Weight

In the cluster analysis method (see section 4.4.2), the parameters have weights set statically for all recordings. But because every LHSV recording is different, a modification of the method has been proposed based on computing adaptive weights for each video recording calculated from the input image data. This method was part of [50].

Several initial values for computing adaptive parameters were set and calculated from input LHSV images. These values are following:

- X_{max} – ROI width of the input image.
- Y_{max} – ROI height of the input image.
- R_{max} – Maximum R value in the selected ROI.
- R_{min} – Minimum R value in the selected ROI.
- R_{sum} – Sum of R values of all points in the ROI.

In the case of ROI, where the height is greater than the width, the change in the X value has a bigger influence on the glottis segmentation. On the contrary, in the wider ROI region, the significance of the X parameter decreases. Making the weight of the parameter X dynamic according to the ROI width should prevent poor detection at the right/left edges of the vocal cords due to the large distance from the center. The same applies to the R parameter where images with higher contrast need a higher weight of the R parameter than low contract images to avoid the wrong classification of the darkest or lightest points in the glottis area.

To find relationships between the initial values and the weights, experimental testing was performed on sample videos and by iterative approach, the configuration of the relationships was tuned. The condition for calculating these relationships was the simplicity of calculation. In this way, the following resulting relationships for weights of each parameter from table 4.3 were formed:

$$W_R = 2 \cdot min\left(\frac{110}{R_{max} - R_{min}}, 5\right) \tag{4.9}$$

$$W_{RmB} = \frac{W_R}{2} \tag{4.10}$$

$$W_X = \frac{120}{X_{max}} \tag{4.11}$$

$$W_C = \frac{W_R}{2} + W_X \tag{4.12}$$

$$W_{LRDiff} = \frac{R_{max} - R_{min}}{200} \tag{4.13}$$

The weights for parameters G, B, and Y were set to zero for low effect in the result to speed up the segmentation process.

4.4.3 Results

The cluster analysis method significantly increases the success of accurate glottis detection compared to the Max-Min Thresholding method. The initial tests of the cluster analysis method were performed on a smaller set of 130 LHSV (a subset of corpus no. 692) with manually delimited ROI used as a training set for parameter weight configuration. During the testing, the glottis was correctly detected by cluster analysis in 97 cases, using the thresholding method in 42 cases (see table 4.4). Of these, the result of the thresholding method was subjectively better only in four cases in comparison to the cluster analysis method.

Table 4.4: *Detection success rate of cluster analysis method and thresholding method during initial testing[21].*

	recordings	percentage
Selected video recordings	130	100 %
Successful detections by Cluster analysis	97	74.6 %
Successful detections by Thresholding method	42	32.3 %
Successful detections by at least one method	101	77.7 %

Images were considered as successfully detected when the segmented area corresponded to the entire area of the glottis and did not contain any other larger areas from the surroundings. In the majority of cases where the image was evaluated as a failure, the detected area corresponded to the glottis in the original image but contained also other parts of the surroundings, or a small part of the glottis area was missing in lower contrast images. Examples of complete or partial failure to find the correct threshold value for the Max-Min-Thresholding method compared to cluster analysis, are shown in figure 4.21.

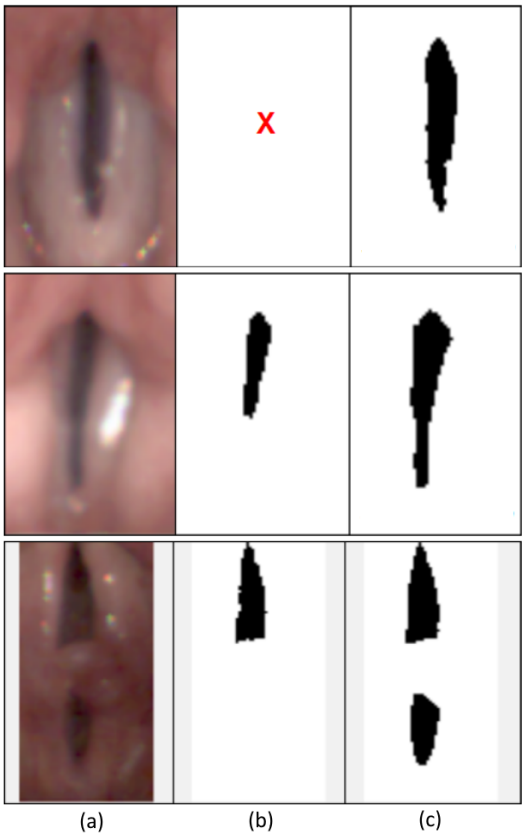


Figure 4.21: *Examples of complete and partial failure of glottis detection by the Max-Min-Thresholding method in comparison with detection by cluster analysis. (a) ROI with the glottis, (b) segmentation using thresholding method, (c) segmentation using cluster analysis.*

The overall results of introduced segmentation methods on final data corpus no. 692 are summarized in table 4.5. During this testing, where the ROI needed to be determined first automatically, the DFT ROI 8×8 method was selected as it returns the best results in section 4.3.4. In case ROI was not detected successfully, the segmentation method was not processed.

Table 4.5: *Overall results of segmentation methods.*

	number of recordings	
Total recordings in Corpus no. 692	692	100% of LHSV recordings
Unsuccessfully detected ROI (DFT 8×8)	68	9.8% of LHSV recordings
Successfully detected ROI (DFT 8×8)	624	90.2% of LHSV recordings
Glottis segmented by Thresholding method	126	20.2% from successful ROI
Glottis segmented by Cluster analysis method	377	60.4% from successful ROI
Glottis segmented by at least one method	384	61.5% from successful ROI

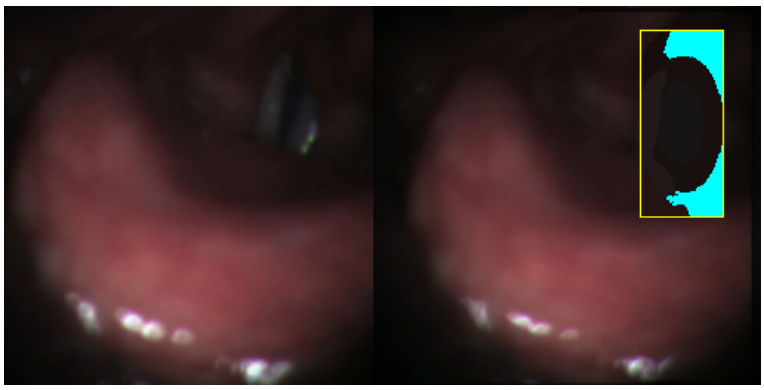
According to the results, the glottis was detected in 377 cases of 692 LHSV recordings by the cluster analysis segmentation method. The success rate was lower compared

to initial testing with a value of about 60%, caused by a bigger amount of LHSV recordings with worse quality in the final corpus. Used corpus no. 692 was not filtered in any way (see section 3.2) and contains also recordings where the segmentation is difficult even manually for an ENT expert. The cluster analysis method is much more successful than the thresholding method, which is unsuitable for recordings of poor quality. Only in seven cases, the thresholding method returned better results than the cluster analysis method.

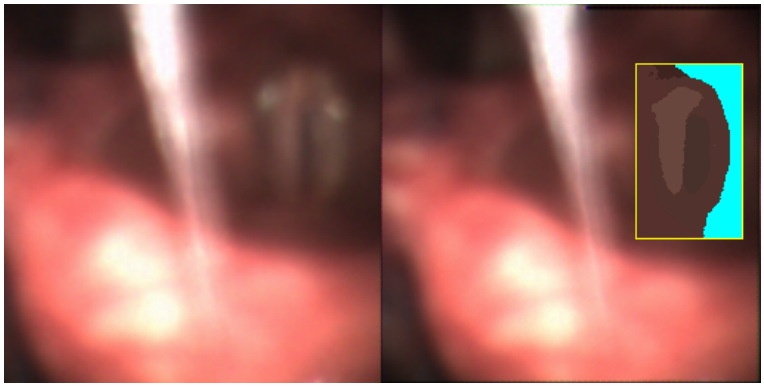
All 315 cases unsuccessfully segmented by the cluster analysis method, were analyzed and the reason for incorrect detection was estimated, see table 4.6. The biggest problem was the occurrence of a very blurred image or insufficient illumination. In 84 cases, there was no obvious reason found related to the image quality, but the method with automatically adjusted adaptive parameters failed. Examples of unsuccessful detection can be seen in pic. 4.22. It should be noted that many recordings with the following issues were segmented successfully among 377 previously mentioned cases.

Table 4.6: *Reasons for unsuccessful glottis detection.*

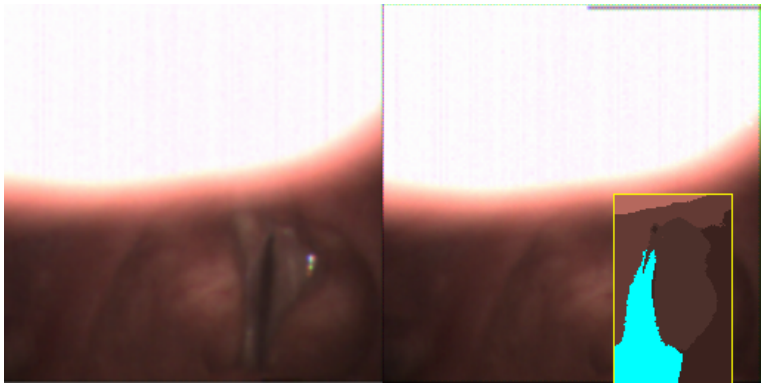
issue	number of occurrences
Insufficient illumination	58
Blurred image	76
Strong reflection	34
Not whole glottis visible	11
Low contrast or not clear edges	52
Insufficient parameter configuration	84



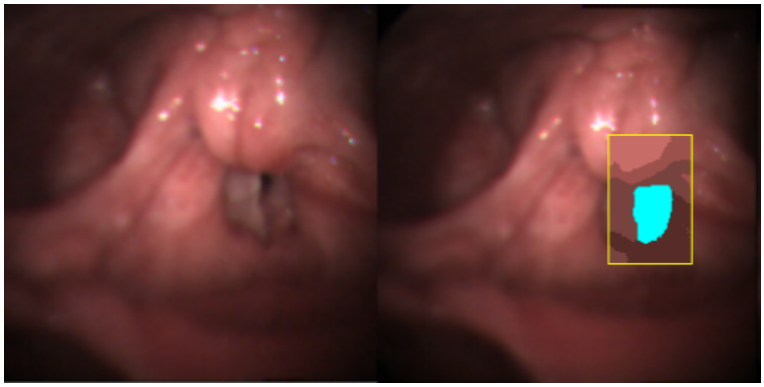
(a) Insufficient illumination



(b) Blurred image



(c) Strong reflection



(d) Not whole glottis visible

Figure 4.22: *Examples of unsuccessful glottis detection*

For further analysis and processing, the results of recordings with incorrectly determined glottis were adjusted manually to get the correct parameters. It means the K-means method was used for segmentation but parameters were modified manually for each LHSV recording or the location of the glottis was selected to get the correct segmentation result and further parameters.

5 Glottis Symmetry

In commercial systems, automatic axis detection is usually not performed or the axis position is determined manually by two points that correspond to the position of the anterior and posterior commissures. Automating this process is complicated due to the different shapes of the larynx in different patients, and the position of the camera varies from examination to examination. In addition, part of the vocal cords can be covered by one of the cartilages, the glottis may be at the edge of the recording area, another anatomical structure covers the vocal cords, or the commissures are blurred. The axis detection method must therefore accept these differences to a reasonable extent.

The area of the glottis is, after its detection, defined by black pixels in the black and white image, see pic. 5.1.

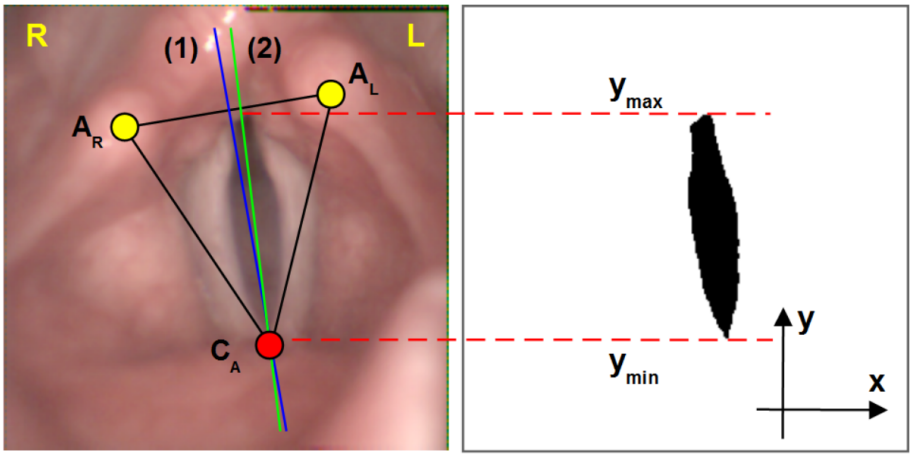


Figure 5.1: *The original image of the vocal cords captured by an LHSV camera. The vocal cord is in the phase of maximum opening during the phonation of the vocal "i:":*
(1): Anatomical vocal cords axis defined by ENT expert;
(2): Computed symmetry axis;
A_L, A_R*: left and right c. arytenoidea;*
C_A*: anterior commissura.*

5.1 Detection of the "Symmetry Axis" of Vocal Gap

Based on the monitoring of the vocal cords and glottis behavior with respect to the position of the anterior and posterior commissures and the anatomical axis of the vocal cords, the following initial assumptions were made and a robust method for estimating the parameters of the major vocal cord axis was developed.

Assumptions:

a) the axis corresponds to the slot when it is closed;

b) when the vocal cords are not closed, the axis passes through the center of the glottis in the most closed state;

c) the location of the vocal cords does not change in the sequence of pictures, there is position conformity of the frames.

Detection of the axis in the closed slit phase (in assumption a), is very difficult. However, its position can be determined according to the area of the slit just before closing and after opening it. These pixels are present in the glottis area for most frames. However, the vocal cords may not open evenly along their entire length and their movement may begin at one of the edges. Therefore, it is necessary to monitor the behavior of the vocal cords until they are fully open, see fig. 5.2

(a) Symmetrical vocal cords

(b) Asymmetrical vocal cords

Figure 5.2: *Glottis shape development of symmetrical (a) and asymmetrical (b) vocal cords; opening-closing-opening phase.*

Here, it is possible to use the assumption c that the images are parallel and their position does not change in the sequence. By projecting all the frames of the sequence and summing the identical pixels, a matrix is obtained that is important for the frequency of occurrences of the vocal cord at each point in the frame. To visualize the information obtained in this way, divide the value of all points by the number of frames in the sequence. This creates a point filtering effect by averaging according to the same pixels and creates a grayscale image in which the darkest areas are those where the vocal cord is open for the longest time. At these points,

the probable location of the symmetry axis of the vocal cords is assumed according to this method. (pic. 5.3)

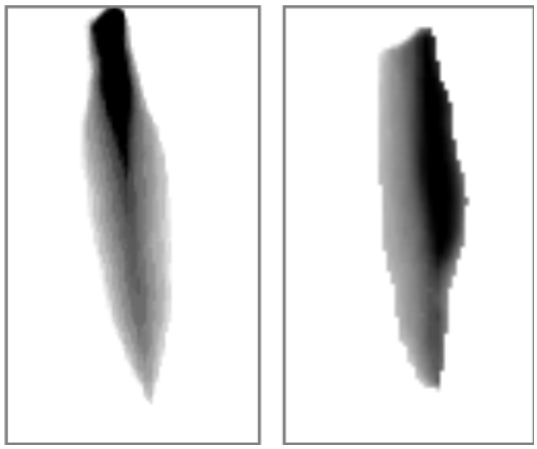


Figure 5.3: *Grayscale images obtained by summing the individual images in a defined sequence and subsequent point filtering in the case of symmetrical and asymmetrical vocal cords behavior.*

Thanks to the method of scanning with a high-speed camera, a vertical orientation of the axis can be assumed. Therefore, the exact position can be specified by finding the darkest point in each line of the calculated image. When the vocal cords are not completely closing, several equally dark pixels appear in one line. In such cases, according to assumption b, the middle of the darkest points was denoted, see pic. 5.4.

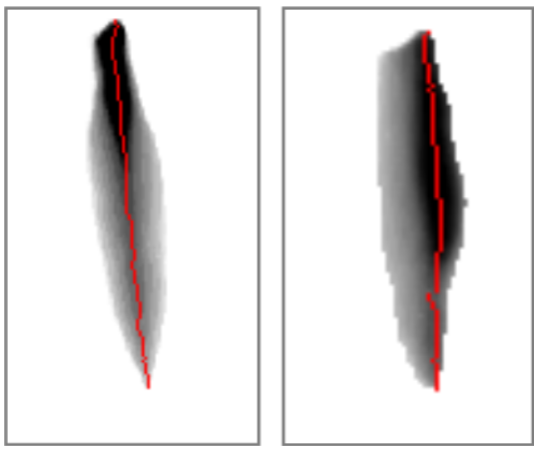


Figure 5.4: *Visualized most frequent points of the detected glottis area in every line for symmetrical and asymmetrical vocal cord behavior.*

At the lower and upper edges of the vocal cords, there may be the aforementioned problems of covering part of the glottis or trimming it. Therefore, inaccuracies occur at the edges of the slit (pic. 5.5). For this reason, it is advisable to use only points in the middle of the vocal cords to determine the axis.

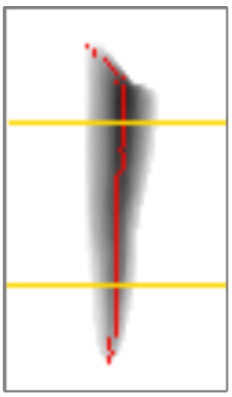


Figure 5.5: *Demonstration of the problem of determining the axis when posterior commissure covers the glottis.*

The obtained points determine the estimate of the vocal cord axis points according to the assumptions set out above. In most cases, they already form a line or correspond to a line in close proximity. In some cases, due to damage to the vocal cords, the glottis is divided into several parts and inaccuracies may occur. Therefore, to specify the points of the line that represents the estimate of the vocal cord axis, the linear regression method is used. The obtained line is constrained only to the area of the glottis and it is an estimate of the axis of symmetry of the vocal cords (pic. 5.6).

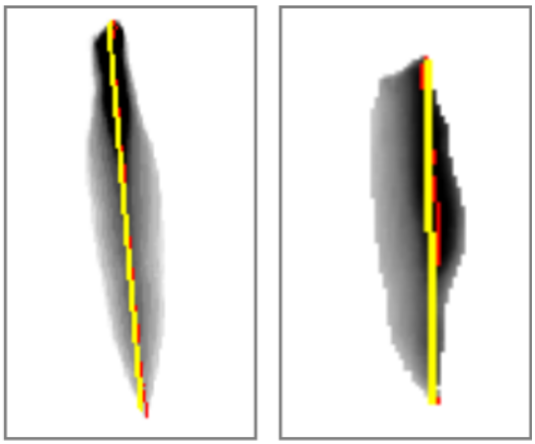


Figure 5.6: *Estimation of points of the glottis axis obtained by applying the linear regression method.*

From the points at the edge of the glottis where the axis intersects the border of vocal cords, the formula of the axis can be created. The precision of the result may be affected by rasterization. p_1x and p_1y are the coordinates of the first point, p_2x and p_2y are the coordinates of the second point.

$$ax + by + c = 0 \tag{5.1}$$

$$a = p_2y - p_1y$$
$$b = p_1x - p_2x$$
$$c = -a \cdot p_1x - b \cdot p_1y = p_2x \cdot p_1y - p_2y \cdot p_1x$$

5.2 Floating Axis

During the recording of the vocal cords with a high-speed camera, it is not uncommon that the vocal cords' position changes in the sequence of images due to the relative movement of the larynx and the camera. In this case, assumption c) of the position conformity of the individual frames is violated. The method is therefore further modified to determine the estimate of the so-called floating axis. The change in the position of the vocal cords is not significant due to the inertial mass of the camera and the high scanning speed in individual consecutive images. At a rate of 4000 frames per second and a basic vocal cord frequency of F_o is 200 Hz, there are 20 frames per vocal cord opening/closing period, ie a time of one period T_{period} is 5 ms. If we consider a sequence longer than 100 frames, the frame sequence time will be $T_{sequence} = 25$ ms and a shift of several pixels may occur. In such cases, it is not appropriate to sum all the frames in the sequence, but only a limited number of frames before and after the frame where the axis should be detected. The number of frames must be greater than the number corresponding to one oscillation of the vocal cords so that a maximum closure state occurs in the selection.

5.3 Evaluation of the Detection of the Symmetry Axis

The result of the application of this method is a straight line, which is an estimate of the axis of the vocal cords and at the same time the axis of the assumed symmetry. According to this axis, it is possible to compare the shape symmetry and observe the differences in properties of both vocal folds.

To evaluate the quality (accuracy) of the vocal cord axis estimation, the AxisConformity (AC) parameter was used. This parameter describes the conformity of the found axis and the axis determined by the ENT expert. The AC can have values from zero (no conformity) to one (the lines are equal). The details about the method for the evaluation are described in [21].

$$AC = \frac{k}{k + |d_1| + |d_2| + p \cdot (d_1 - d_2)^2} \qquad AC \in (0, 1\rangle \tag{5.2}$$

$$d_1 = \frac{x_1 - x_{orl1}}{h} \qquad d_2 = \frac{x_2 - x_{orl2}}{h} \tag{5.3}$$

where k = tolerance coefficient $(k = 0{,}2)$,
p = coefficient angle penalization $(p = 3)$,
d_i = distance of points,
h = ROI height,
x_i = x-coordinate of the point of the estimated axis by the method,
x_{orli} = x-coordinate of the point of the anatomical axis according to the ENT expert.

233 selected LHSV recordings (a random subset of corpus no. 692) were used as a testing set, where the axis was determined by the ENT expert. The results of AC values are presented in the histogram in fig. 5.7.

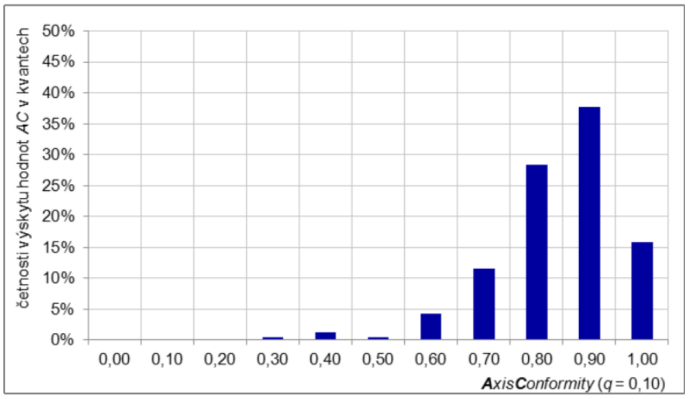


Figure 5.7: *Frequency graph of occurrence of AxisConformity values.*

The correct detection of the axis can be considered for values $AC > 0.65$. In this case, 93% of detections were correct with axis rotation error lower than 5.7° or position estimation error less than 5px. This is shown in table 5.1 together with a comparison when $AC > 0.75$. A detailed result explanation is in [21].

Table 5.1: *Histogram of AxisConformity values in individual subintervals of AC for a set of tested video sequences.*

quantum	number	%	AC > 0.65	AC > 0.75
0.00	0	0.00		
0.10	0	0.00		
0.20	0	0.00		
0.30	1	0.43		
0.40	3	1.29		
0.50	1	0.43		
0.60	10	4.29		
0.70	27	11,59	93.56 %	
0.80	66	28.33		81.97 %
0.90	88	37.76		
1.00	37	15.88		
total	233	100		

6 Glottis Parameters

In this part of the work, several parameters based on the geometry of the vocal cords are introduced that are commonly used in medical systems of high-speed laryngoscopy, parameters that are published in a number of scientific articles, and proposed some new parameters that are tested in cooperation with the ENT department of University Hospital in Pilsen, see [4], [52].

6.1 Commonly Used Parameters

The first group of parameters consists of the ones which are part of the most high-speed laryngoscopy systems based on basic measurement. This group includes the glottis area size, the length of its perimeter, the glottis height, and width, see the overview in table 6.1.

Table 6.1: *Overview of basic parameters for single frame and their symbols, see [4].*

parameter	symbol
Glottis area size	A
Ratio of max. and min. area size in a period	A_r
Border length of glottis	P
Glottis height (symmetry axis length)	D_{axis}
Glottis width (normal line abscissa length)	D_{norm}

6.2 Additional Parameters

For the processing and evaluation of high-speed laryngoscopy images in [4] and [52], the list of parameters is supplemented by additional ones in table 6.2. This set of parameters is tested for cross-correlation and correlation with other parameters of some examinations, see chapter 2.5.

Table 6.2: *Overview of additional parameters for single frame and their symbols, see [4], [52].*

parameter	symbol
Left part of glottis area size (from symmetry axis)	A_{left}
Right part of glottis area size (from symmetry axis)	A_{right}
Left and right area size difference	A_{diff}
Left part of glottis border length (from symmetry axis)	P_{left}
Right part of glottis border length (from symmetry axis)	P_{right}
Left and right border length difference	P_{diff}
Segmentation of the border	S
Segmentation of the left part border	S_{left}
Segmentation of the right part border	S_{right}
Oblongness	O
Border approximation ellipsis height	H_e
Border approximation ellipsis width	W_e
Oblongness of ellipsis	O_e

Most of these parameters are obtained by direct measuring from the detected glottis, resp. vocal cords, in a single frame or in a sequence of frames, some parameters are derived by calculation from the values of already computed parameters.

6.3 Center of Gravity of the Vocal Cords

The parameters of the center of gravity (also called center point) of the vocal cords were added to the set of additional parameters, which help explicitly detect symmetrical/asymmetrical behavior of the vocal folds. These parameters also form a set of other parameters for evaluating the symmetry and asymmetry of the vocal cords. The glottis area and the border are used for the calculation of center points.

These parameters were introduced in the [53].

6.3.1 Center of Gravity Parameters

The parameters of the center of gravity C_k can be understood as the distance of the center of gravity of the glottis area S_k in the image k from the main axis (glottis symmetry axis) of the vocal cords D_x and its normal D_y. By their nature, these parameters are both static and dynamic. The importance of monitoring the dynamics of the center of gravity development is evident from the schematic representation in fig. 6.1 (this is a schematic description of the evolution of the center of gravity position regardless of the type of center of gravity and the method of position calculation).

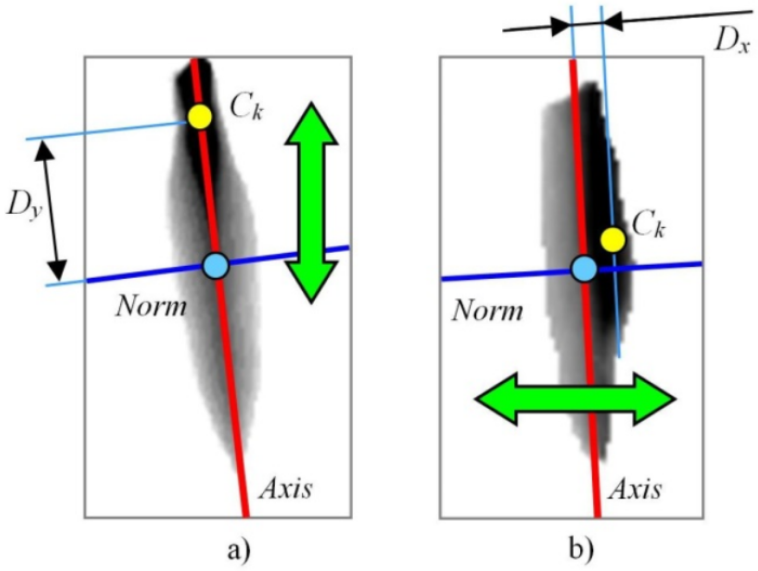


Figure 6.1: *Scheme and description of center point movement during one period of vocal cords oscillation;*
a) Symmetrical vocal cords, change in distance D_y (distance of the center point from normal) is dominant;
b) Non-symmetrical vocal cords, change in distance D_x (distance of the center point to symmetry axis) becomes more significant.

Two types of glottis S_k centers of gravity are used:

- The center of gravity of the area CA_k, where the following calculation formulas (6.1) and (6.2) apply:

$$CA_k = CA_k\left(x_c^{(S)}, y_c^{(S)}\right) \tag{6.1}$$

$$x_c^{(S)} = \frac{1}{A}\sum_{x_S \in S_k}\sum_{y_S \in S_k} x_S, \qquad y_c^{(S)} = \frac{1}{A}\sum_{x_S \in S_k}\sum_{y_S \in S_k} y_S \tag{6.2}$$

  x_s and y_s are the values of coordinates of individual pixels, which form the detected area of the glottis S_k detected by the methods (see chapter 4.4), A is the size of the glottis area S_k in pixels.

- The center of gravity of the (inner) borderline CH_k of the detected glottis S_k; where calculation formulas (6.3) and (6.4) apply:

$$CH_k = CH_k\left(x_c^{(H)}, y_c^{(H)}\right) \tag{6.3}$$

$$x_c^{(H)} = \frac{1}{P}\sum_{x_H \in H_k}\sum_{y_H \in H_k} x_H, \qquad y_c^{(H)} = \frac{1}{P}\sum_{x_H \in H_k}\sum_{y_H \in H_k} y_H \tag{6.4}$$

  In the formula (6.4), x_H and y_H are the values of coordinates of individual pixels, which form the inner boundary H_k of the glottis area S_k, P is the length of the boundary H_k in pixels.

Based on the position of both types of center of gravity, other parameters were defined to contribute to the description of the symmetry of the vocal cords and the dynamics of the vocal folds, such as distances of the center of gravity from the axis (D_x) and the normal line (D_y). To calculate these parameters, the relations (6.5) and (6.6) apply, where a, b and c are the line coefficients of the axis resp. normal, see section 5.1, x_c, and y_c are the coordinates of the center of gravity.

$$D_x = \frac{ax_c + by_c + c}{\sqrt{a^2 + b^2}} \tag{6.5}$$

$$D_y = \frac{a_{norm}x_c + b_{norm}y_c + c_{norm}}{\sqrt{a_{norm}^2 + b_{norm}^2}} \tag{6.6}$$

An overall overview of the parameters derived from the position of the center of gravity is in table 6.3.

Table 6.3: *Overview of center point parameters for single frame and their symbols, see [52].*

parameter	symbol
Area center point position	CA_k
Distance of the Area CP from the symmetry axis	D_xS
Distance of the Area CP from the normal	D_yS
Border center point position	CH_k
Distance of the Border CP from the symmetry axis	D_xH
Distance of the Border CP from the normal	D_yH
Difference between Area and Border CPs in normal direction	$D_x(diff)$
Difference between Area and Border CPs in symmetry axis direction	$D_y(diff)$

Based on the acquired knowledge about the development of the area and the border (perimeter) of the glottis, it is also possible to derive the symmetry parameters of the vocal folds calculated according to the development of the area and the detected inner boundary of the vocal folds during phonation. Because the symmetry of the vocal cords seems to be an important factor, a visual representation and a description of some case studies are introduced as an example. In individual cases, the development of the parameters of the area and the border of the vocal gap during the video sequence is compared with the development of the position of the center of gravity of the area and the boundary line of the glottis. The selection of cases in section 6.3.3 was made from the total number of 692 processed video recordings from the LHSV examination at the ENT department of the University Hospital in Pilsen as typical representatives of cases that occur in the set of records.

6.3.2 Center of Gravity Trajectory

Also, the movement and trajectory of the center of gravity were briefly analyzed with the focus to see different progress in the opening and closing phases. This approach is similar to monitoring the distances of the center of gravity from the axis and the normal line, but instead of measuring distances from the axis, the movement of these points is recorded. The LHSV recording was split into single periods of vocal cords oscillation and for each period the position of the center of gravity was computed for each frame. Then the area A_{CA} and perimeter P_{CA} of the closed curve created by the center of gravity position in each frame were computed together with the ratio of area size to the perimeter. The simple polygon was expected in this study, thus the area can be computed according to eq. (6.7) and the perimeter according to eq. (6.8), where x_i and y_i are coordinates of the points, n is the number of points, and $x_n = x_0$ and $y_n = y_0$.

$$A_{CA} = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \tag{6.7}$$

$$P_{CA} = \sum_{i=0}^{n-1} \left(\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \right) \tag{6.8}$$

The following figures contain approximated shapes of the glottis together with the computed centers of gravity of the area CA_k. Frames within one period are drawn into the same place. The area, perimeter, and their ratio are displayed below each period. Figure 6.2 shows the behavior of healthy and symmetrical vocal cords, where the area is quite small and the resulting ratio is near zero. Figure 6.3 contains vocal cords with asymmetrical movement and it corresponds with the bigger area value and the ratio.

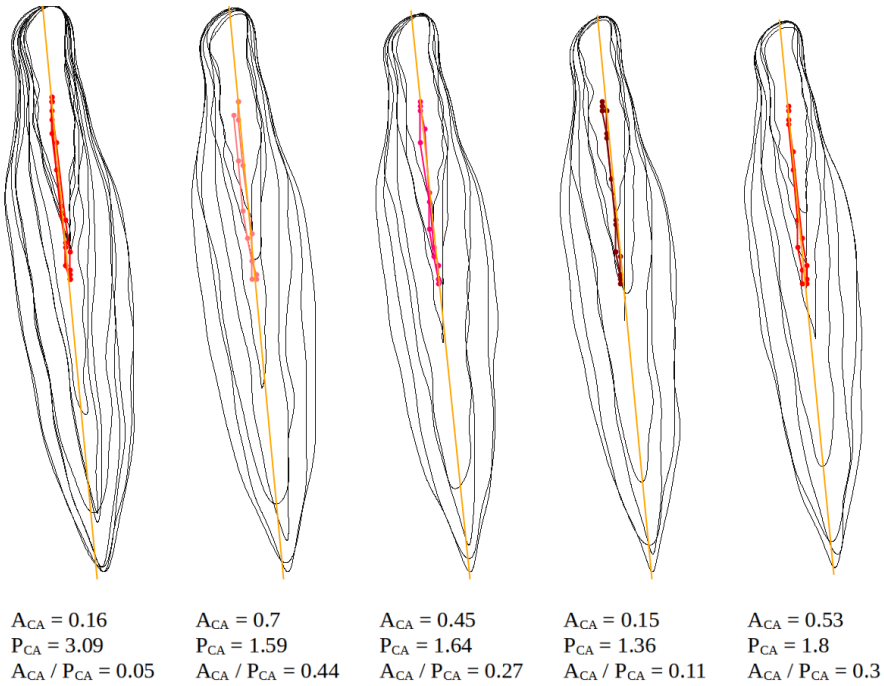


Figure 6.2: *Symmetrical healthy vocal cords with the trajectory of the center of gravity. The area is very small and the ratio is near zero.*

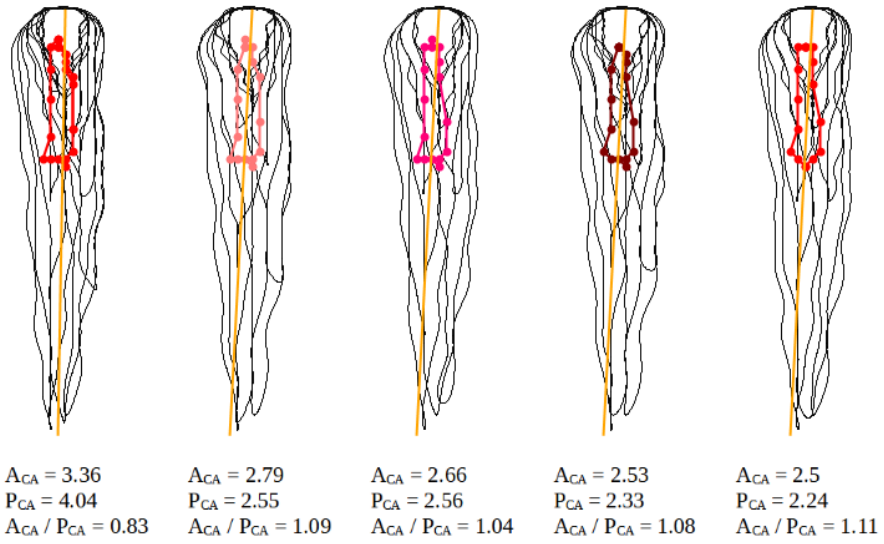


Figure 6.3: *Vocal cords with different behavior between the closing and opening phases with the trajectory of the center of gravity. The area is relatively big and the ratio is around one.*

A computed ratio could be also interesting for the evaluation of the vocal cords' behavior. This study was just an idea of another direction, it was not further used.

6.3.3 Selected Case Studies

This section contains several case studies with descriptions and parameter development to demonstrate the usefulness of monitoring center point behavior. Mentioned cases are then also used in section 7.3.6. The data are computed from 400 frames of LHSV sequences, but for better visualization, some graphs show only 100 samples. The left and the right side are related to the anatomy, because of the way of capturing images, the right side is shown on the left.

Case 1 – Healthy symmetrical vocal cords

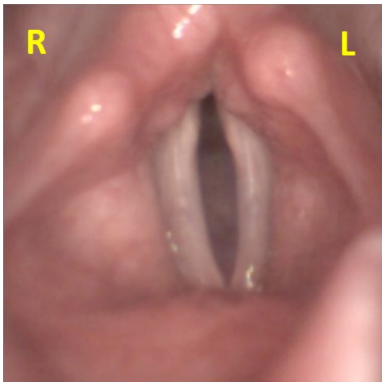
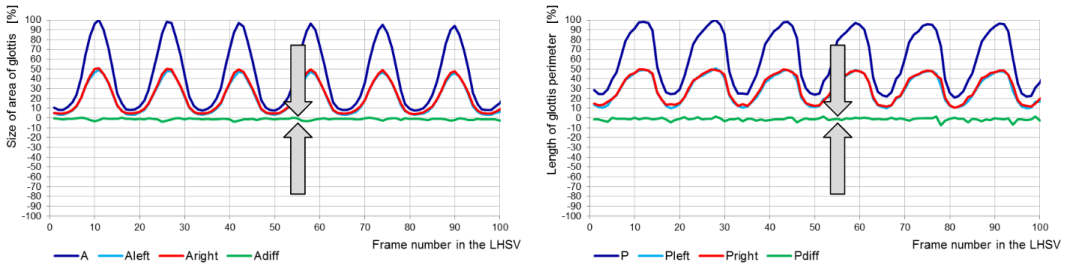


Figure 6.4: *LHSV image of healthy symmetrical vocal cords.*

- Healthy symmetrical vocal cords (pic. 6.4).
- Female, age: 20 years.
- Phonation "i:", $SPL_{min} = 81$ dB, $SPL_{max} = 92$ dB, $F_0 = 252$ Hz (from sound recording).
- Strong phonation, healthy vocal cords without movement limitation.
- Analysis of the area A and perimeter length P shows symmetry, see also no changes of parameters A_{diff} and P_{diff} (fig. 6.5). According to the minimum of A and P, the vocal cords don't completely close because of the strong phonation.
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the behavior of the symmetrical vocal cords, i.e. the expected movement of the center points in the direction of the symmetry axis (D_yS and D_yH) and no movement of the center points in the direction of the normal (D_xS and D_xH), see fig. 6.6 (a) and (b).
- The relative positions of the center point of the area CA_k and the border CH_k of the glottis during phonation are almost identical (fig. 6.6 (c) and (d)).

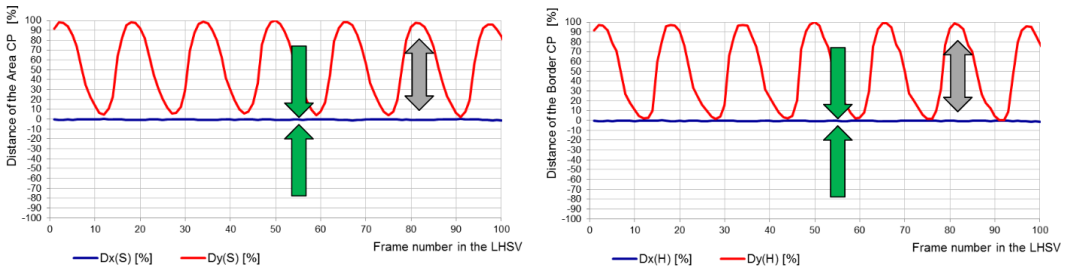
Conclusion:
Analysis of the development of the area size A, the border length P, the position of

the center point of the area CA_k, and the border CH_k of the glottis during phonation confirms **the symmetry of vocal cords** in this case.

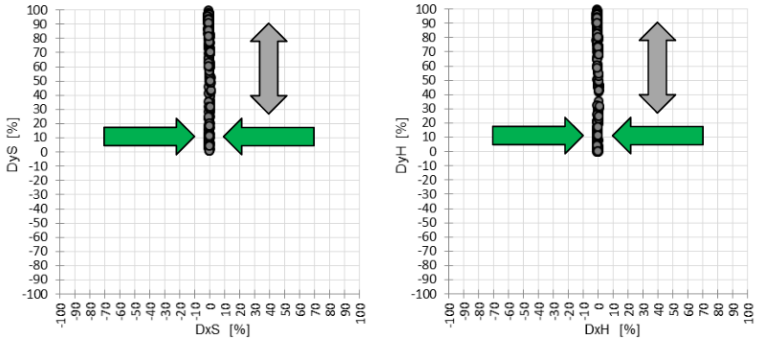


(a) (b)

Figure 6.5: *(a) Glottis area size (A) and (b) border length (P) development. Differences A_{diff} and P_{diff} are close to zero.*



(a) (b)



(c) (d)

Figure 6.6: *Development of the area center point distances D_xS and D_yS (a) and border center point distances D_xH and D_yH (b), with standardized trajectory summary (c), (d). The movement in the x-axis is minimal for both center points.*

Case 2 – Non-symmetrical vocal cords with carcinoma

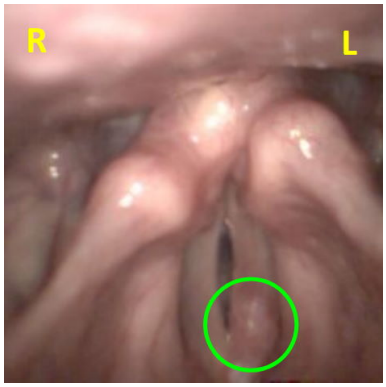
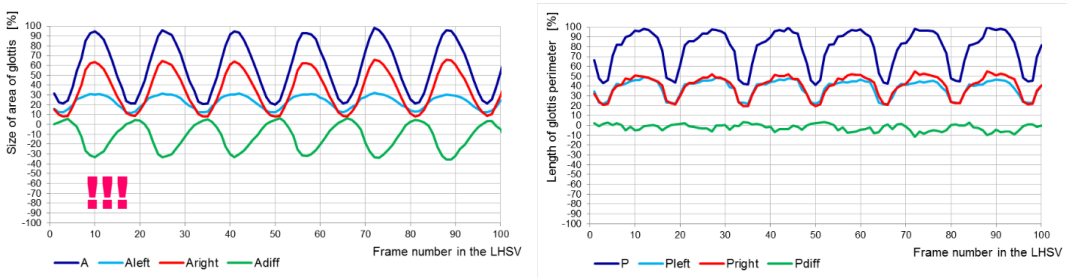


Figure 6.7: *LHSV image of the non-symmetrical vocal cords with carcinoma.*

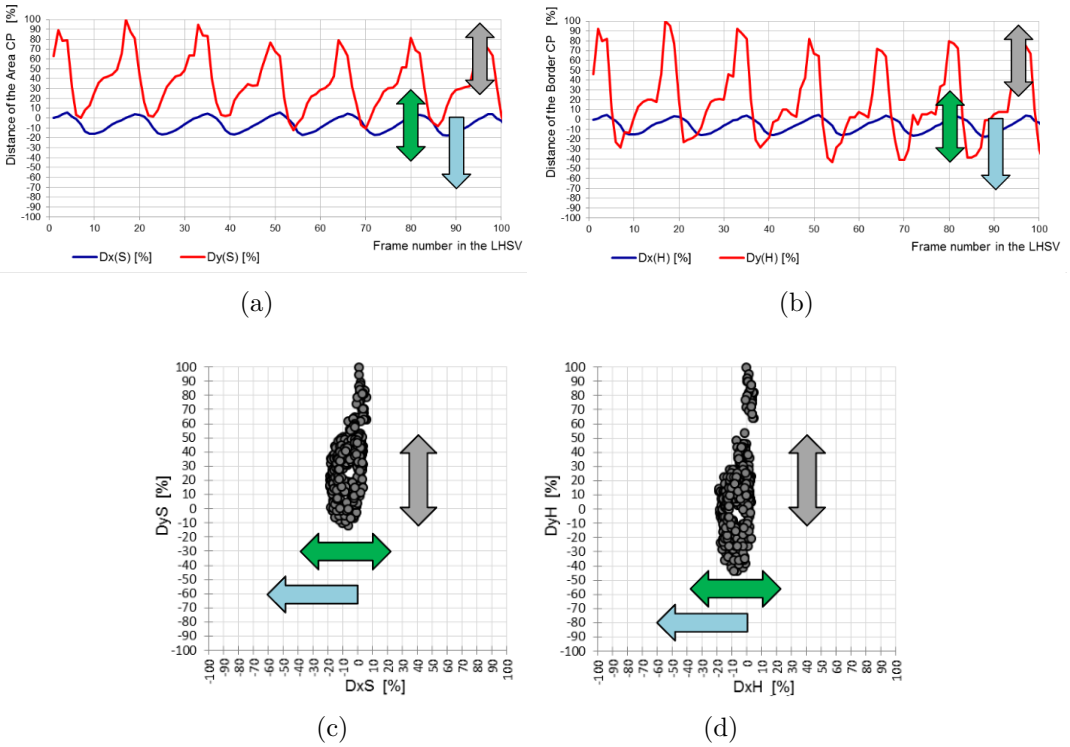
- Non-symmetrical vocal cords, diagnosed carcinoma, movement limitation on the left side (pic. 6.7).
- Male, age: 82 years.
- Phonation "i:", $SPL_{min} = 77$ dB, $SPL_{max} = 81$ dB, $F_0 = 298$ Hz (from sound recording).
- Analysis of the area size A development shows asymmetry, see also parameter A_{diff}. Parameter P has no significant irregularities (fig. 6.8).
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the behavior of the non-symmetrical vocal cords. Movement of the center points in the direction of the symmetry axis (D_yS and D_yH) is supplemented with significant movement of the center point in the direction of the normal (D_xS and D_xH), see fig. 6.9 (a) and (b). The center points are deflected to the right side because of the carcinoma object, which limits the movement of the left vocal fold.
- The difference in mutual position of CA_k and CH_k center points is increased (fig. 6.9 (c) and (d)).

Conclusion:
Analysis of the development of area A shows asymmetry caused by carcinoma object, which limits the movement. Border length does not show any significant issue, Center point movements confirm **the asymmetry of vocal cords** in this case.



(a) (b)

Figure 6.8: *(a) Glottis area size (A) and (b) Inner border length (P) development. Difference A_{diff} shows asymmetry.*



(a) (b)

(c) (d)

Figure 6.9: *Development of the (a) area center point distances (D_xS and D_yS) and (b) border center point distances (D_xH and D_yH), with standardized trajectory summary (c), (d). The movement in the x-axis is significant for both center points.*

Case 3 – Non-symmetrical vocal cords with nodule

This case contains two LHSV recordings from different times. The first one was taken before microsurgery and the second one two months after the microsurgery, see timeline in fig. 6.10.

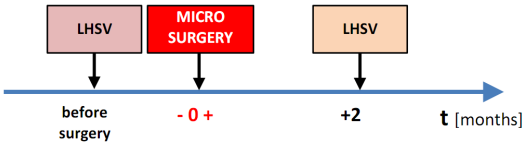


Figure 6.10: *Timeline of the LHSV examinations and the microsurgery.*

Case 3a:

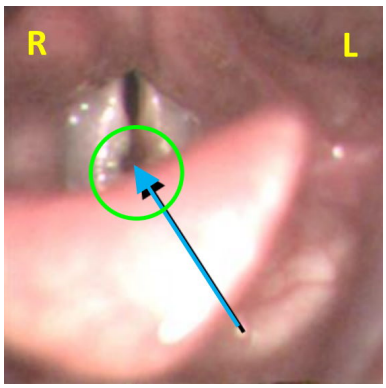
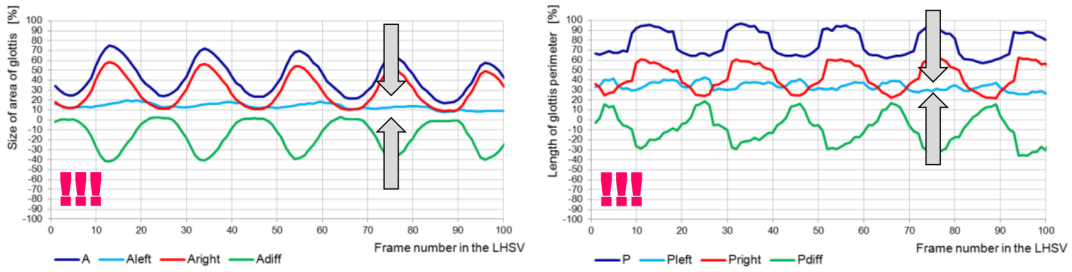


Figure 6.11: *LHSV image of the non-symmetrical vocal cords with a nodule.*

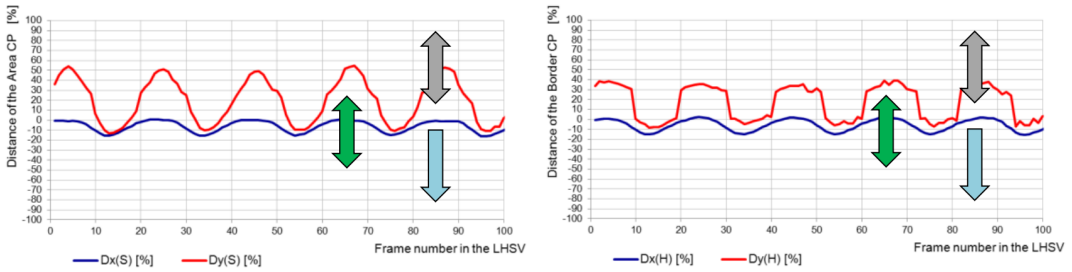
- Non-symmetrical vocal cords, nodule on the left side (pic. 6.11), condition before microsurgery.
- Female, age: 48 years.
- Phonation "i:", $SPL_{min} = 78$ dB, $SPL_{max} = 87$ dB, $F_0 = 192$ Hz (from sound recording).
- Analysis of area A and perimeter P developments shows asymmetry, caused by movement limitation of the left vocal fold, see also parameters A_{diff} and $Pdiff$ (fig. 6.12).
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the asymmetry, The movement of the center points in the axis direction (D_yS a D_yH) is complemented with significant movement in the normal direction (D_xS and D_xH), see fig. 6.13 (a) and (b). The positions are deflected to the right side from the symmetry axis.

Conclusion:
The analysis of the development of area A and the border length P shows the asymmetry which is caused by the nodule on the left. Development of the position of area center point CA_k and the border center point CH_k of the glottis during phonation confirms **the asymmetry of the vocal cords**.

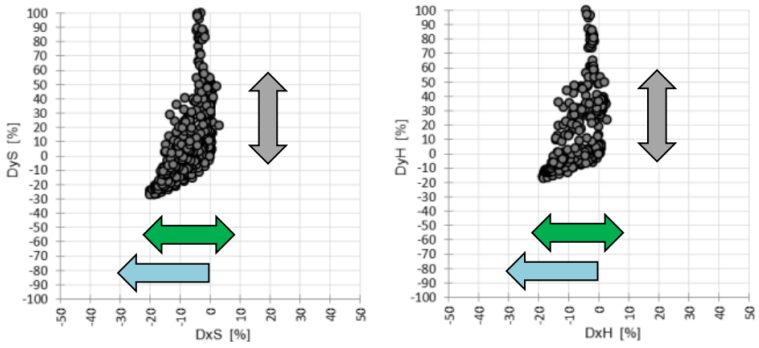


(a) (b)

Figure 6.12: *(a) Glottis area size A and (b) border length P development. Minimal movement of the A_{left} and P_{left} together with significant differences in the development of A_{diff} and P_{diff} show asymmetry.*



(a) (b)



(c) (d)

Figure 6.13: *Development of the (a) area center point distances (D_xS and D_yS) and (b) border center point distances (D_xH and D_yH), deflected to the right side. Movement on the axis direction is standard. Graphs (c) and (d) show a standardized trajectory summary.*

Case 3b:

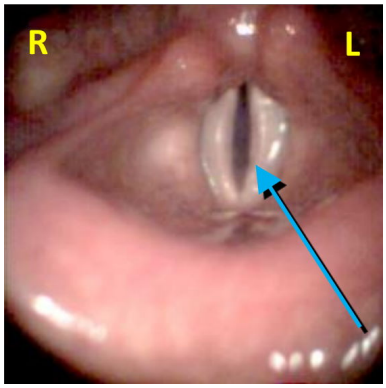
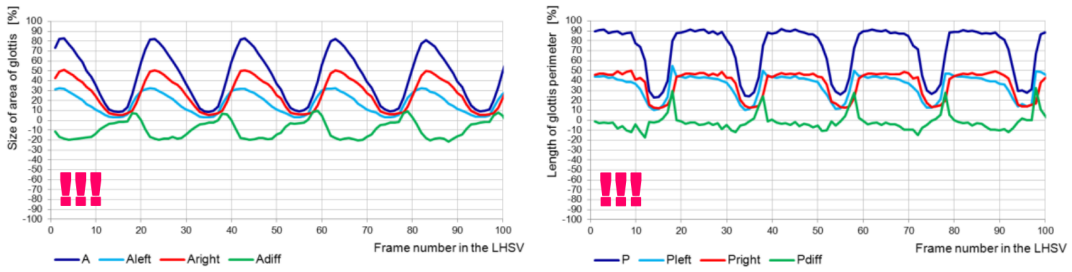


Figure 6.14: *LHSV image of the vocal cords after microsurgery (original nodule position marked by the arrow).*

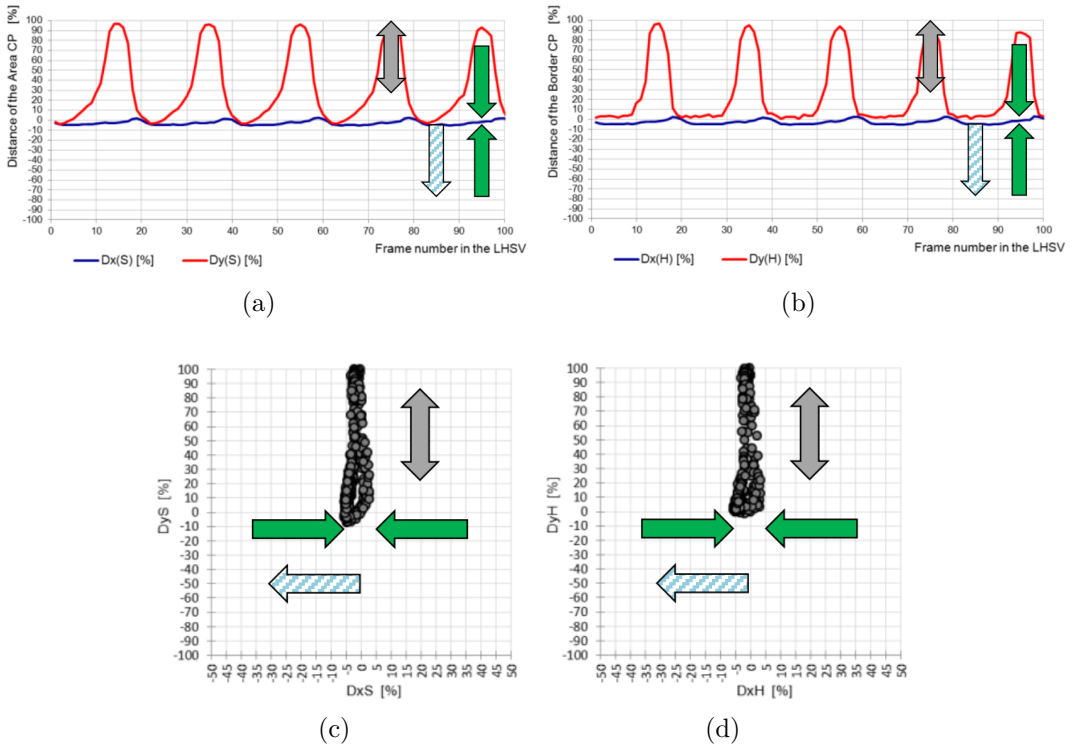
- Non-symmetrical vocal cords, nodule on the left side (pic. 6.14), condition of two months after microsurgery.
- Female, age: 48 years.
- Phonation "i:", $SPL_{min} = 82$ dB, $SPL_{max} = 84$ dB, $F_0 = 191$ Hz (from sound recording).
- Analysis of the area A and the perimeter P developments still shows asymmetry, see also parameters A_{diff} and $Pdiff$, but the movement of the left vocal fold has increased (fig. 6.15).
- The development of the area and perimeter center point positions (CA_k and CH_k) shows improvement in the vocal cord movement, which is almost symmetrical along the axis, there is only slight deflection to the right (fig. 6.16).
- The difference in the mutual position between the CA_k and CH_k is minimal (fig. 6.16).

Conclusion:
The analysis of the development of area A and the border length P indicates a slight asymmetry, which is caused by the state after surgery. However, the development of the position of the area center point CA_k and the border center point CH_k during phonation no longer confirms this **asymmetry of the vocal cords**.



(a) (b)

Figure 6.15: *(a) Glottis area size A and (b) border length P development. The movement of A_{diff} and P_{diff} is less significant. Area size A_{left} and border length P_{left} shows limited movement caused by the state after surgery, causing asymmetry.*



(a) (b)

(c) (d)

Figure 6.16: *Development of the (a) area center point distances (D_xS and D_yS) and (b) border center point distances (D_xH and D_yH). The movement in the normal direction is minimal and almost symmetrical, slight deflection to the right is still observable. Graphs (c) and (d) show a standardized trajectory summary.*

Case 4 – Non-symmetrical vocal cords with polyps

This case contains two LHSV recordings from different times. The first one was taken before microsurgery and the second one three months after the microsurgery, see timeline in fig. 6.17.

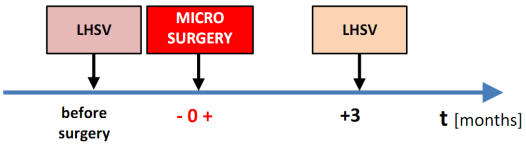


Figure 6.17: *Timeline of the LHSV examinations and the microsurgery.*

Case 4a:

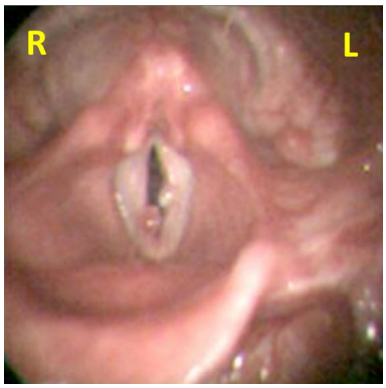
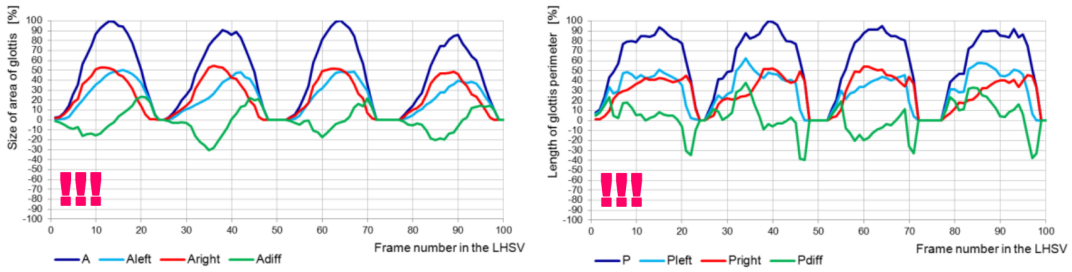


Figure 6.18: *LHSV image of the non-symmetrical vocal cords with polyps before microsurgery.*

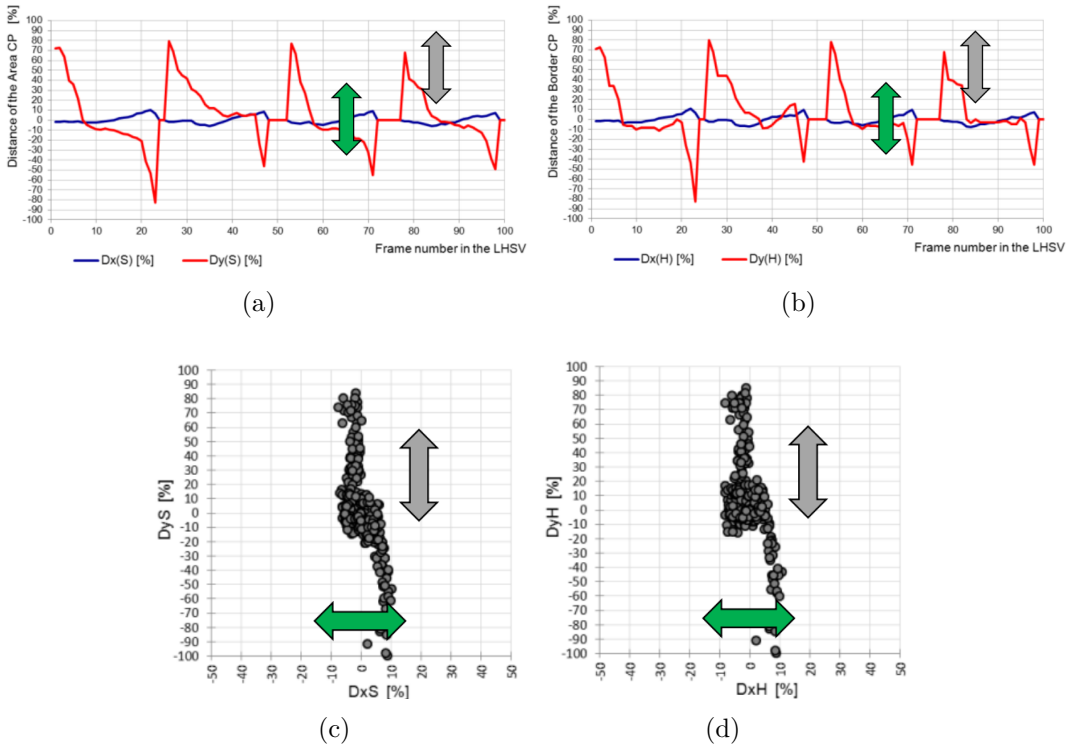
- Non-symmetrical vocal cords, polyps on both sides (pic. 6.18), condition before microsurgery.
- Female, age: 37 years.
- Phonation "i:", $SPL_{min} = 68$ dB, $SPL_{max} = 76$ dB, $F_0 = 152$ Hz (from sound recording).
- Analysis of the area A and the perimeter P developments shows asymmetry, see also parameters A_{diff} and P_{diff} (fig. 6.19). The asymmetry is significant even vocal folds move periodically, but not synchronously because of differently positioned polyps.
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms asymmetry. The movement of the center points in the axis direction (D_yS a D_yH) is complemented by movement in the normal direction (D_xS and D_xH), see fig. 6.20 (a) and (b).
- Because of the overlapping of polyps, in some cases, the glottis is divided into two parts causing bigger changes in center point positions.

Conclusion:
The analysis of the development of area A and the border length P shows asymmetry, which is caused by the polyps on the left and right sides. The development of the position of the area center point CA_k and the border center point CH_k of the glottis during phonation confirms this **asymmetry of the vocal cords**.



(a) (b)

Figure 6.19: *(a) Glottis area size A and (b) inner border length P development. Significant changes can be seen in signals of A_{diff} and P_{diff} regardless the movement is periodical. Different positions of polyps are causing asymmetrical movement.*



(a) (b)

(c) (d)

Figure 6.20: *Due to the asymmetry of the vocal cords, there is a more pronounced movement in the case of both centers of gravity CA_k and CH_k in the direction of the normal. The bigger changes in positions are caused by the occasional splitting of the glottis onto two parts.*

Case 4b:

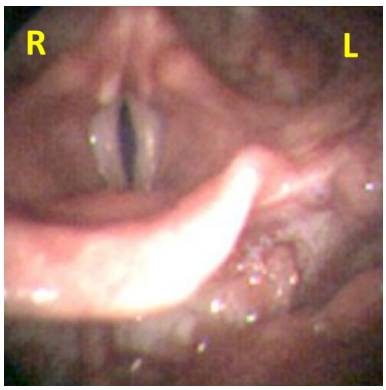
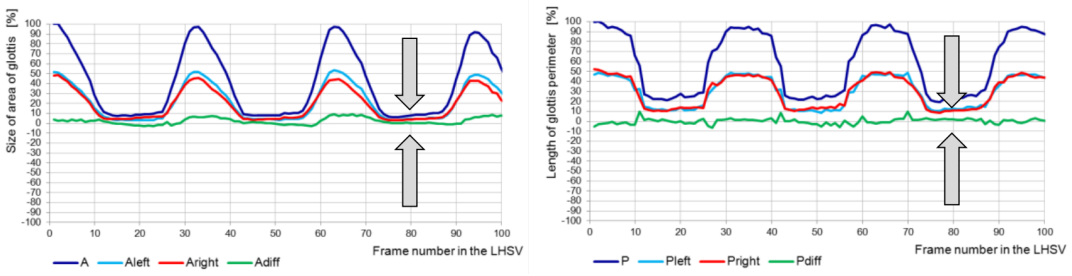


Figure 6.21: *Vocal cords with surgically removed polyps.*

- Non-symmetrical vocal cords, polyps were on both sides (pic. 6.18), condition after microsurgery.
- Female, age: 37 years.
- Phonation "i:", $SPL_{min} = 62$ dB, $SPL_{max} = 69$ dB, $F_0 = 129$ Hz (from sound recording).
- Analysis of the area A and the perimeter P developments show symmetrical vocal cords, see also very small changes of parameters A_{diff} and P_{diff} (fig. 6.19).
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the symmetry. The movement of the center points in the axis direction (D_yS and D_yH) is complemented by almost no movement in the normal direction (D_xS and D_xH), see fig. 6.20 (a) and (b).

Conclusion:
The analysis of the development of area A and the border length P together with the center points CA_k and CH_k positions show **symmetry of the vocal cords**.



(a) (b)

Figure 6.22: *(a) Glottis area size A and (b) border length P development. The differences between A_{diff} and P_{diff} are minimal.*

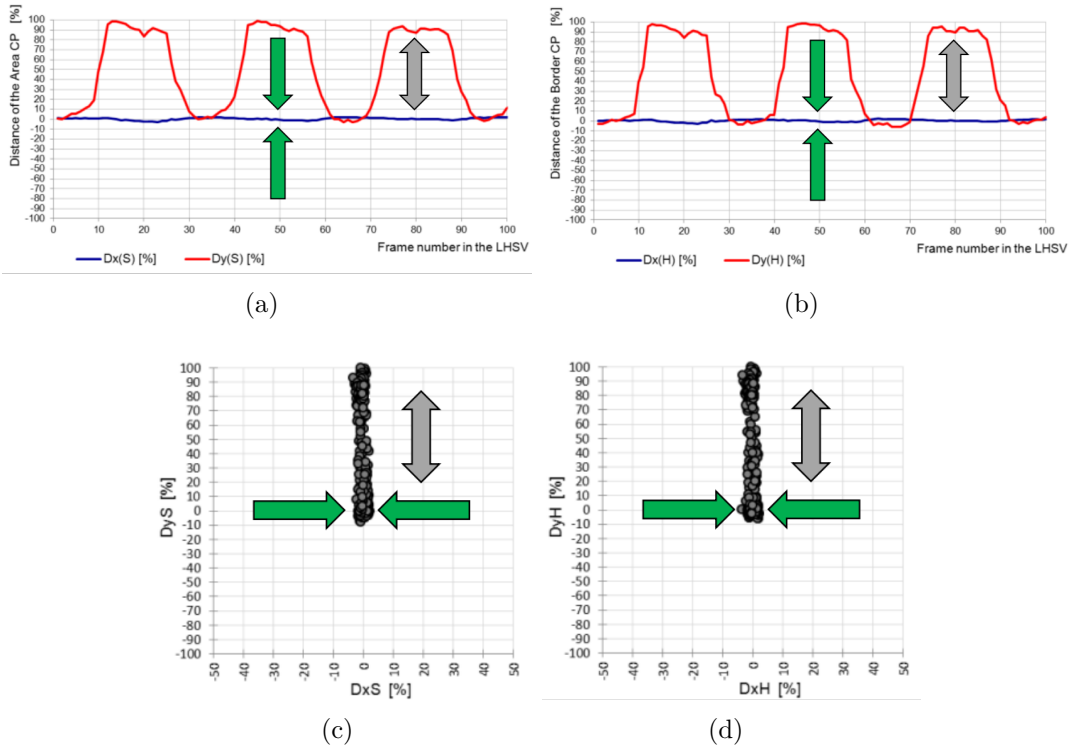


Figure 6.23: *Development of the (a) area center point distances (D_xS and D_yS) and (b) border center point distances (D_xH and D_yH). A significant movement of the center points along the axis of the vocal cords with a shift towards the posterior commissure and minimal movement in the direction of the normal is evident. The analysis of the behavior of the center points CA_k and CH_k confirms bf the symmetry of the vocal cords after microsurgery.*

Case 5 – Non-symmetrical vocal cords with cyst

This case contains three LHSV recordings at different times. The first one was taken before microsurgery, the second one three months after the microsurgery, and the third one three years after the microsurgery, see timeline in fig. 6.24.

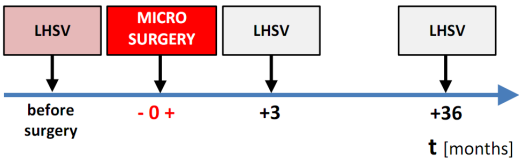


Figure 6.24: *Timeline of the LHSV examination and the microsurgery.*

Case 5a:

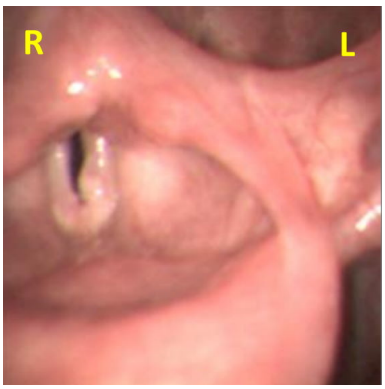
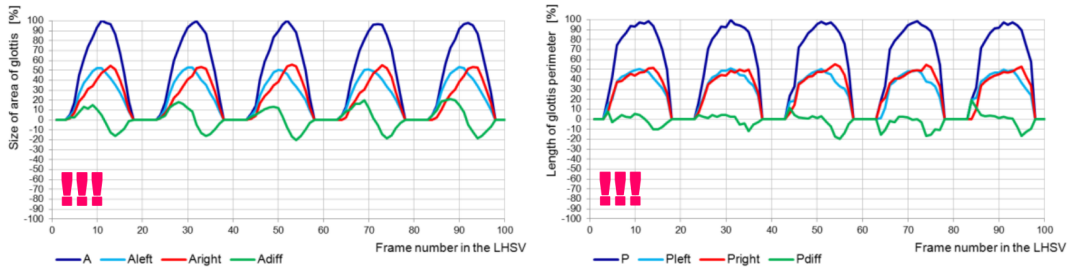


Figure 6.25: *LHSV image of the non-symmetrical vocal cords with a cyst before microsurgery.*

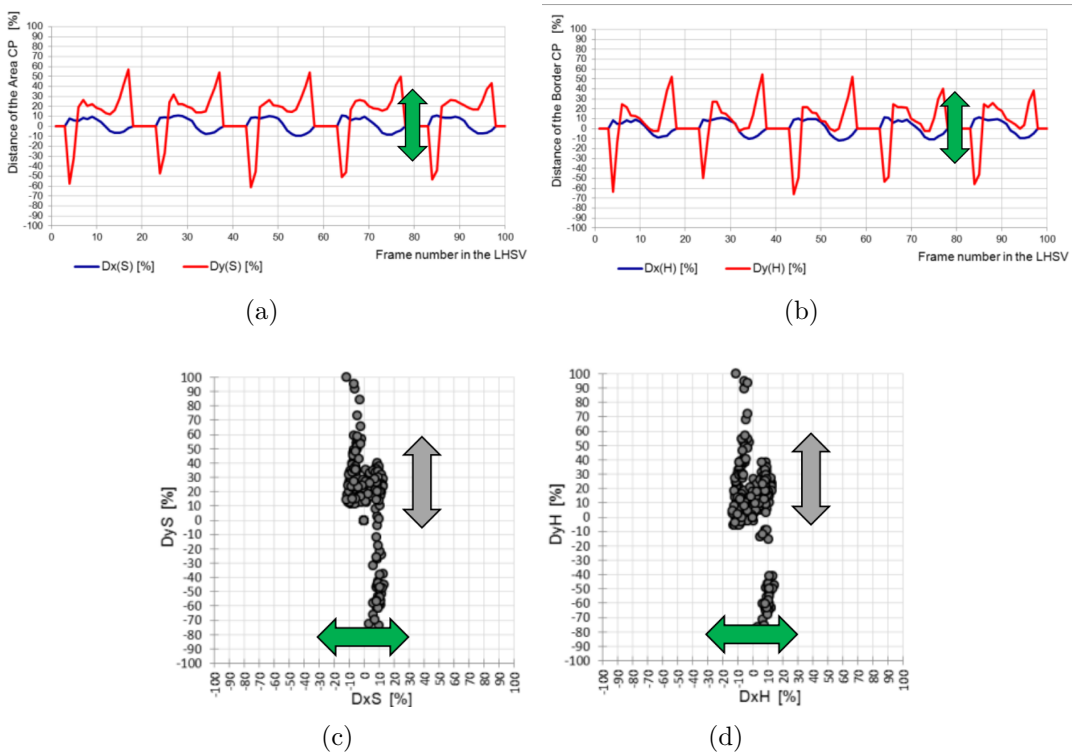
- Non-symmetrical vocal cords, cyst on the left side (pic. 6.25), condition before microsurgery.
- Female, age: 57 years.
- Phonation "i:", $SPL_{min} = 81$ dB, $SPL_{max} = 87$ dB, $F_0 = 201$ Hz (from sound recording).
- Analysis of area A and perimeter P development shows asymmetry, see also parameters A_{diff} and $Pdiff$ (fig. 6.26). The asymmetry can be observed especially at area size development.
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the asymmetry. The movement of the center points in the axis direction (D_yS a D_yH) is complemented with significant movement in the normal direction (D_xS and D_xH), see fig. 6.27 (a) and (b).

Conclusion:
The analysis of the development of area A and the border length P shows asymmetry, which is caused by the cyst on the left side. The development of the position of area center point CA_k and the border center point CH_k of the glottis during phonation confirms this **asymmetry of the vocal cords**.



(a) (b)

Figure 6.26: *(a) Glottis area size A and (b) border length P development. Significant differences in A_{diff} confirm asymmetrical vocal cords. The change of P_{diff} is not conclusive.*



(a) (b)

(c) (d)

Figure 6.27: *Due to the asymmetry of the vocal cords, there is a more pronounced movement in the case of both centers of gravity CA_k and CH_k in the direction of the normal. This fact confirms* **the asymmetry of the vocal cords**.

Case 5b:

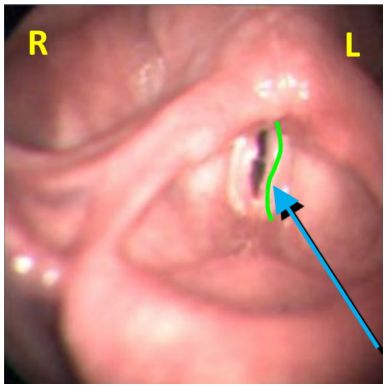
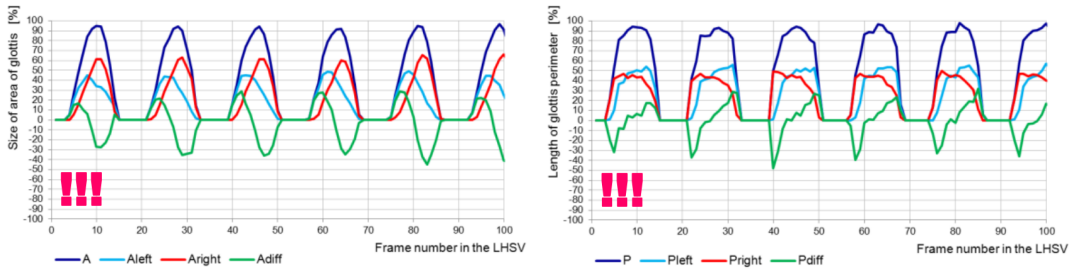


Figure 6.28: *An example of the initially asymmetrical vocal cord with a cyst on the left side, 3 months after microsurgery (the original location of the cyst is indicated by an arrow, the left vocal fold shows undulation).*

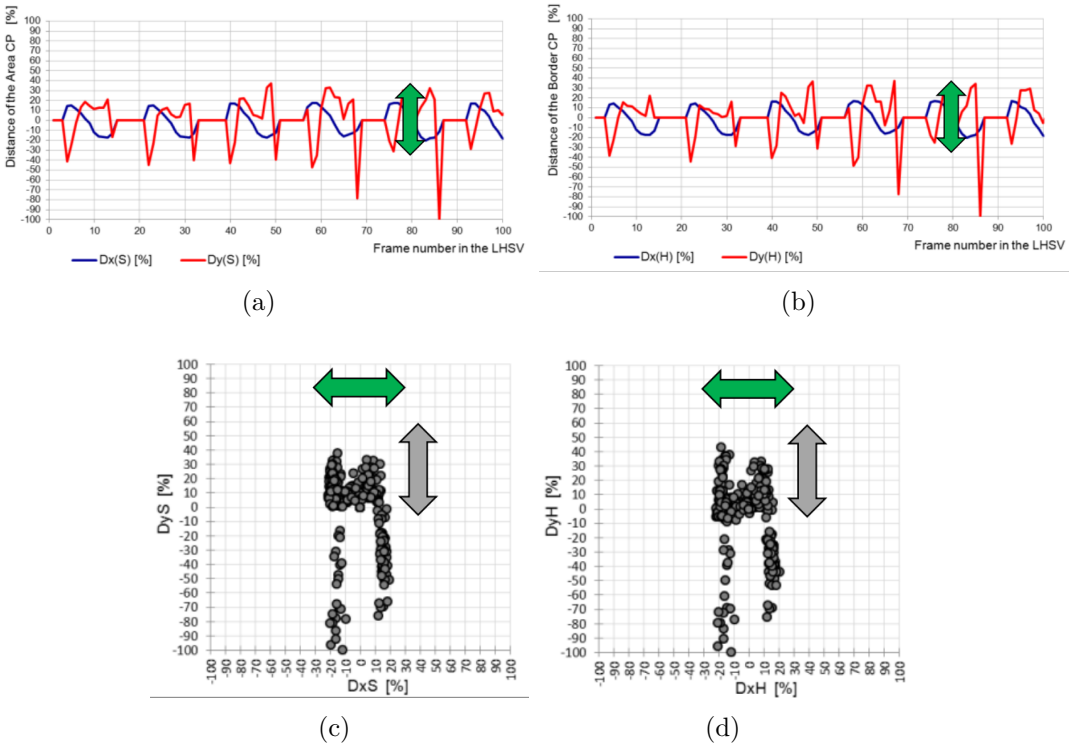
- Non-symmetrical vocal cords, cyst on the left side (pic. 6.28), condition of three months after microsurgery.
- Female, age: 57 years.
- Phonation "i:", $SPL_{min} = 82$ dB, $SPL_{max} = 89$ dB, $F_0 = 222$ Hz (from sound recording).
- Analysis of the development of the area size A and the border length P indicates asymmetry, see parameters A_{diff} and $Pdiff$. Compared to the condition before the surgery, this deterioration can be seen in the LHSV sequence, when the left vocal fold is undulating (fig. 6.29).
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the behavior of the non-symmetrical vocal cords. Movement of the center points in the direction of the symmetry axis (D_yS and D_yH) is supplemented with significant movement of the center points in the direction of the normal (D_xS and D_xH), see fig. 6.30 (a) and (b).

Conclusion:
The analysis of the development of area A and the border length P shows asymmetry, which is caused by the state after microsurgery. The development of the position of area center point CA_k and border center point CH_k of the glottis during phonation confirms this **asymmetry of the vocal cords**.



(a) (b)

Figure 6.29: *(a) Glottis area size A and (b) border length P development. The changes of A_{diff} and P_{diff} are more significant, it indicates asymmetry.*



(a) (b)

(c) (d)

Figure 6.30: *Development of the (a) area center point distances (D_xS and D_yS) and (b) border center point distances (D_xH and D_yH). Due to the asymmetry of the vocal cords, there is a more pronounced movement in the case of both center points CA_k and CH_k in the direction of the normal. This fact confirms* **the asymmetry of the vocal cords**.

Case 5c:

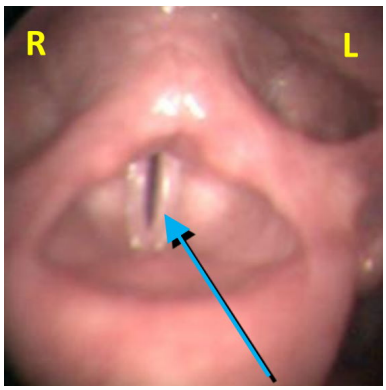
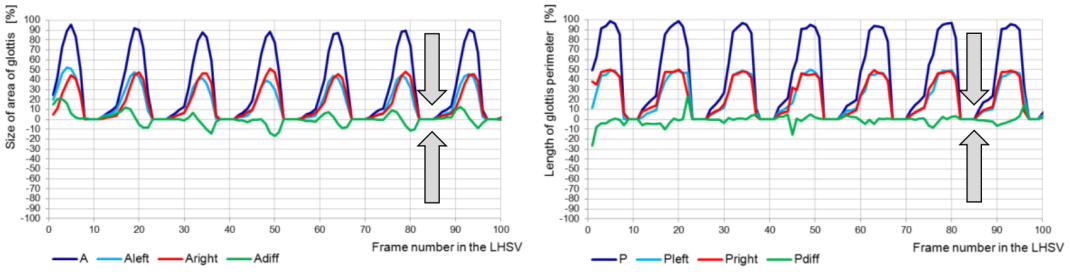


Figure 6.31: *An example of the initially asymmetrical vocal cord with a cyst on the left side, 36 months after microsurgery (the original location of the cyst is indicated by an arrow).*

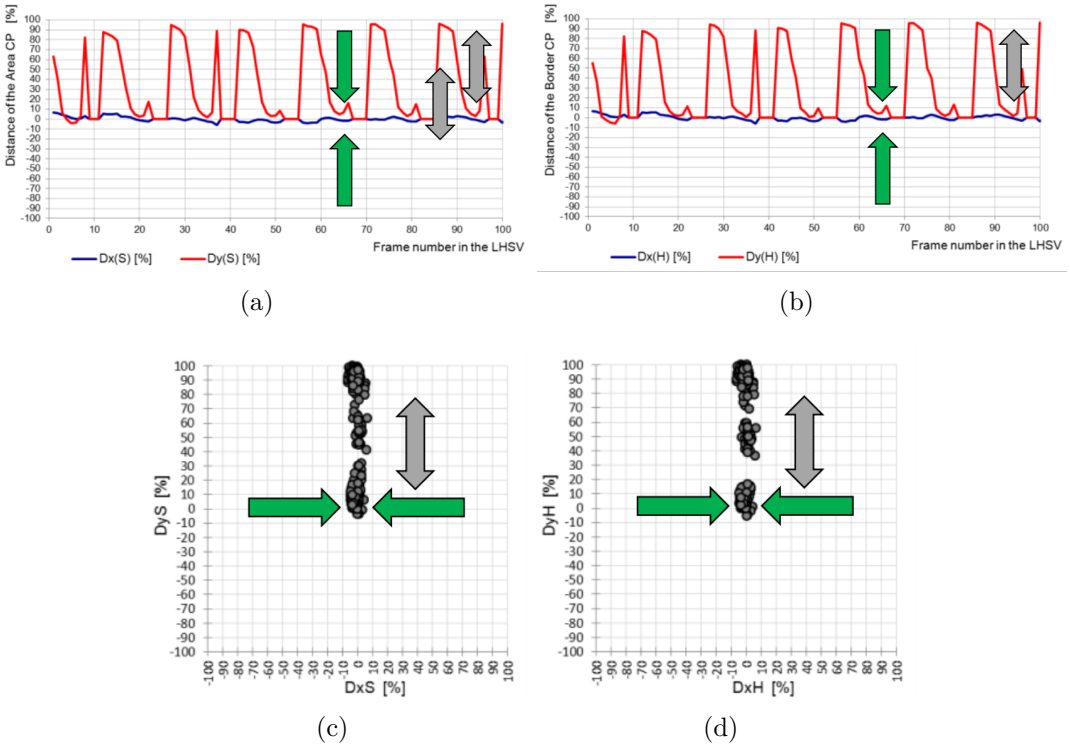
- Originally Non-symmetrical vocal cords with a cyst on the left side (pic. 6.31), condition of three years after microsurgery.
- Female, age: 60 years.
- Phonation "i:", $SPL_{min} = 79$ dB, $SPL_{max} = 79$ dB, $F_0 = 249$ Hz (from sound recording).
- Analysis of the development of the area size A and the border length P indicates symmetry, see also parameters A_{diff} and $Pdiff$. Compared to the state before the surgery and the state 3 months after the procedure, there is a noticeable improvement in vocal cords' behavior (fig. 6.32).
- The development of the area and perimeter center point positions in the normal direction (D_xS and D_xH) confirms the behavior of the symmetrical vocal cords. Movement of the center points in the direction of the symmetry axis (D_yS and D_yH) is supplemented with minimal movement of the center points in the direction of the normal (D_xS and D_xH), see fig. 6.33 (a) and (b).

Conclusion:
The analysis of the development of area A and the border length P shows symmetry, the vocal cords look recovered after the cyst removal. The development of the position of area center point CA_k and border center point CH_k of the glottis during phonation confirms this **symmetry of the vocal cords**.



(a) (b)

Figure 6.32: *(a) Glottis area size A and (b) border length P development. The changes of A_{diff} and P_{diff} are minimal, it indicates symmetry.*



(a) (b)

(c) (d)

Figure 6.33: *Development of the (a) area center point distances (D_xS and D_yS) and (b) border center point distances (D_xH and D_yH). Both center points have minimal movement in the direction of the normal. This fact confirms* **the symmetry of the vocal cords**. *The zero-value gaps are caused by fully closed vocal cords where center point positions cannot be determined.*

6.3.4 Summary of Obtained Results

According to the results so far, the significance of the newly introduced parameters of the area center point CA_k and border center point CH_k of the vocal cords can be summarized and compared with the parameters of the area size A and border length

P. The results of the comparison of individual type cases are given in table 6.4. Both center points showed almost identical motion. For symmetry, the movement is interesting, especially in the direction of x.

Table 6.4: *Overview of some typical cases found after the analysis of case studies*

vocal cords	area, border	CP movement	description
symmetrical	symmetry	$D_x \to min$	parameters correspond – confirmed symmetry
symmetrical	asymmetry	$D_x \to min$	CP development shows symmetry, but it is not confirmed by A and P development
asymmetrical	asymmetry	$D_x \to max$	CP – normal distance is significant – confirmed asymmetry
asymmetrical	asymmetry	$D_x \to$ high values	one-sided deflection from symmetry axis – confirmed asymmetry

The reliability and accuracy of the Area size, Border length, and Center point position parameters depend on the quality of the LHSV video recording, i.e. on the size of the vocal cords and the recording angle.

7 Parameters Analysis Using Correlation Relationships

The recording of the LHSV examination allows monitoring of the real movement of the vocal cords in the form of a video sequence. Since the goal is to evaluate the kinematics of the vocal cords and mere observation may not always be sufficient, it is necessary to find parameters and evaluation techniques that can characterize changes in glottis shape over time during one or more periods and evaluate the quality of the glottis closure and opening.

Almost all parameters computed from the glottis are based on the geometry, vocal cords' axis, and symmetry. An interesting approach seems to be based on the assumption that most of these geometric parameters of the glottis should have a strong "correlation relationship" (see below) during the phonation of healthy vocal cords. From the point of view of diagnostics, the fact that the "correlation relationship" in the specific recording is unexpectedly low for some parameters can be important. Violation of this relationship may be an indicator of a pathological condition on the vocal cords.

This chapter deals with the measurement of the "correlation relationship", searching for a set of parameter pairs that can have diagnostic significance in terms of correlations, and introduces evaluation methods. They can identify LHSV recordings with unexpected behavior and evaluate the glottis quality. The result should be to raise a warning to the doctor in case any potential issue or irregularity is detected. The glottis evaluation in section 7.3 also uses the rating of the glottis behavior by the ENT expert where the opening and closing process, the shape, results from other examinations, and overall state of vocal cords is considered (as mentioned in section 7.3.1). This led to an approach to make automatic classification of the glottis into 5 disjoint classes based on 14 selected geometric parameters.

After obtaining the data from the LHSV video, data analysis is needed to evaluate the vocal cords' behavior and to find any irregularities. Two methods using statistics and correlation between parameters were introduced.

7.1 Correlation and Linear Approximation

In general, the correlation is the relationship between two sets of paired values, measured in the value range $\langle -1; 1 \rangle$. This sample correlation coefficient is represented by formula (7.1).

$$r = \frac{\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2} \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (y_j - \bar{y})^2}} \tag{7.1}$$

In statistics, the definition is valid for the random and independent data, which is not met in the case of working with the set of parameters from LHSV, but we can work with the linear approximation which has the same meaning and the same properties. The line $y_i = ax_i + b$, where x_i, y_i are the data values for $i = 1, ..., n$, can be used for linear approximation using the ordinary least square methods according to the following formulas[54]:

$$a = \frac{\sum_{j=1}^{n} (x_j - \bar{x})\, y_j}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \tag{7.2}$$

$$b = \bar{y} - a\bar{x}; \qquad \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j, \tag{7.3}$$

Because of the $\sum_{j=1}^{n}(x_j - \bar{x})\bar{y} = \bar{y}\sum_{i=1}^{n}(x_j - \bar{x}) = 0$, then the following applies:

$$\begin{aligned} a = \frac{\sum_{j=1}^{n} (x_j - \bar{x})\, y_j}{\sum_{j=1}^{n} (x_j - \bar{x})^2} = \frac{\sum_{j=1}^{n} (x_j - \bar{x})\, (y_j - \bar{y})}{\sum_{j=1}^{n} (x_j - \bar{x})^2} = \\ = \frac{\sum_{j=1}^{n} (x_j - \bar{x})\, (y_j - \bar{y})}{\sqrt{\sum_{j=1}^{n} (x_j - \bar{x})^2}\sqrt{\sum_{j=1}^{n} (y_j - \bar{y})^2}} \frac{\sqrt{\sum_{j=1}^{n} (y_j - \bar{y})^2}}{\sqrt{\sum_{j=1}^{n} (x_j - \bar{x})^2}} = \\ = \frac{\frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})\, (y_j - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2}\sqrt{\frac{1}{n-1}\sum_{j=1}^{n} (y_j - \bar{y})^2}} \frac{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n} (y_j - \bar{y})^2}}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2}} \end{aligned} \tag{7.4}$$

The first part of the resulting formula 7.4 is equivalent to the sample correlation coefficient in (7.1) and it is valid regardless of x_j and $y_j; j = 1, ..., n$ are identically independent distributed (iid) samples or not.

Analogously in the same terms, the sample standard deviations formulas are numerically equivalent to:

$$s_x = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n} (y_j - \bar{y})^2}; \qquad s_y = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2} \tag{7.5}$$

According to these facts, the following formulas are numerically equal to regression line coefficients:

$$a = r\frac{s_y}{s_x}, \qquad b = \bar{y} - r\frac{s_y}{s_x}\bar{x} \tag{7.6}$$

A squared error of the linear approximation can be computed as follows:

$$\begin{aligned} Q(r) = \sum_{i=1}^{n} (y_i - ax_i - b)^2 = \sum_{i=1}^{n} \left(y_i - r\frac{s_y}{s_x}x_i - \left(\bar{y} - r\frac{s_y}{s_x}\bar{x}\right)\right)^2 = \\ = \sum_{i=1}^{n} \left((y_i - \bar{y}) - r\frac{s_y}{s_x}(x_i - \bar{x})\right)^2 = \\ = \sum_{i=1}^{n} \left[(y_i - \bar{y})^2 + r^2\frac{s_y^2}{s_x^2}(x_i - \bar{x})^2 - 2r\frac{s_y}{s_x}(x_i - \bar{x})(y_i - \bar{y})\right] = \\ = (n-1)s_y^2 + (n-1)r^2\frac{s_y^2}{s_x^2}s_x^2 - 2r\frac{s_y}{s_x}s_x s_y (n-1)r = \\ = (n-1)s_y^2(1 - r^2) \quad (7.7) \end{aligned}$$

Thus for $r = \pm 1$ is $Q(r) = 0$ and for $r = 0$ is $Q(r) = (n-1)s_y^2$. The squared error of linear approximation of the data by the ordinary least square method has analogous properties to the squared error of linear regression. Hence for $r \to \pm 1$ the squared error drops to zero and for $r \to 0$ the squared error increases to $(n-1)s_y^2$.

Therefore, the formula (7.1) is a measure of the "linear relationship strength" between x and y, similar to the standard correlation coefficient (or its estimate). Therefore, it makes sense to use the correlation name for the r (even if it is not in the strict sense of the word). The same applies to other probabilistic and statistical terms in this text.

7.2 Diagnostic Meaning of Correlation between Parameters

Most parameters from LHSV recordings are geometric or easily derivable from other geometric parameters. Therefore for healthy glottis, a strong relationship for some parameter pairs can be expected.

To find parameter pairs where the correlation value could be useful for the glottis evaluation, the following method was introduced, the idea and introduction can be found in [55]. This method deals with the mutual correlations of all 91 possible parameter pairs and the main task is to find specific parameter pairs where the correlation value is expected to be high. An example of a such relationship could be between area size A and perimeter length P, where the correlation value would depend on the smooth or wrinkled edge.

7.2.1 Searching for Useful Parameter Pairs

For the study of correlation relationships, several parameters mentioned in section 6 were selected and used for further analysis. The group of 14 selected geometric

or geometric derivable parameters is listed in table 7.1. Combining all these parameters, 91 pairs were created. For every recording, each pair is represented by a single value computed as a correlation coefficient between the development of the parameter values.

Table 7.1: *List of parameters used for correlation analysis.*

parameter	symbol
Glottis area size	A
Border length of glottis	P
Left part of glottis area size (from symmetry axis)	A_{left}
Right part of glottis area size (from symmetry axis)	A_{right}
Left and right area size difference	A_{diff}
Left part of glottis border length (from symmetry axis)	P_{left}
Right part of glottis border length (from symmetry axis)	P_{right}
Left and right border length difference	P_{diff}
Distance of the Area CP from the symmetry axis	D_xS
Distance of the Area CP from the normal	D_yS
Distance of the Border CP from the symmetry axis	D_xH
Distance of the Border CP from the normal	D_yH
Difference between Area and Border CPs in normal direction	D_xDiff
Difference between Area and Border CPs in symmetry axis direction	D_yDiff

Because of the geometric basis, a strong relationship is expected, mostly linear or well-linearizable. For example, the relation is expected for the parameter pairs $A - A_{left}$ and $A - A_{right}$ at healthy and symmetrical vocal cords. It means that the parameter A can be computed from A_{right} or A_{left} or vice versa. Figure 7.1 shows healthy vocal cords where the correlation value for the parameter pair $A - A_{left}$ is 0.999, and the correlation for pair $A - A_{right}$ is 0.999 too. The advantage of independence in scale is used here as the A and A_{left} or A_{right} have different value ranges.

When such expected relation is broken, the warning can be raised and it can have diagnostical meaning. Fig. 7.2 shows vocal cords with a post-traumatic paresis of the reversible nerve on the left side. Because of the movement limitation, the correlation value of the parameter pair $A - A_{left}$ is only 0.294 while the $A - A_{right}$ is still high, specifically 0.994.

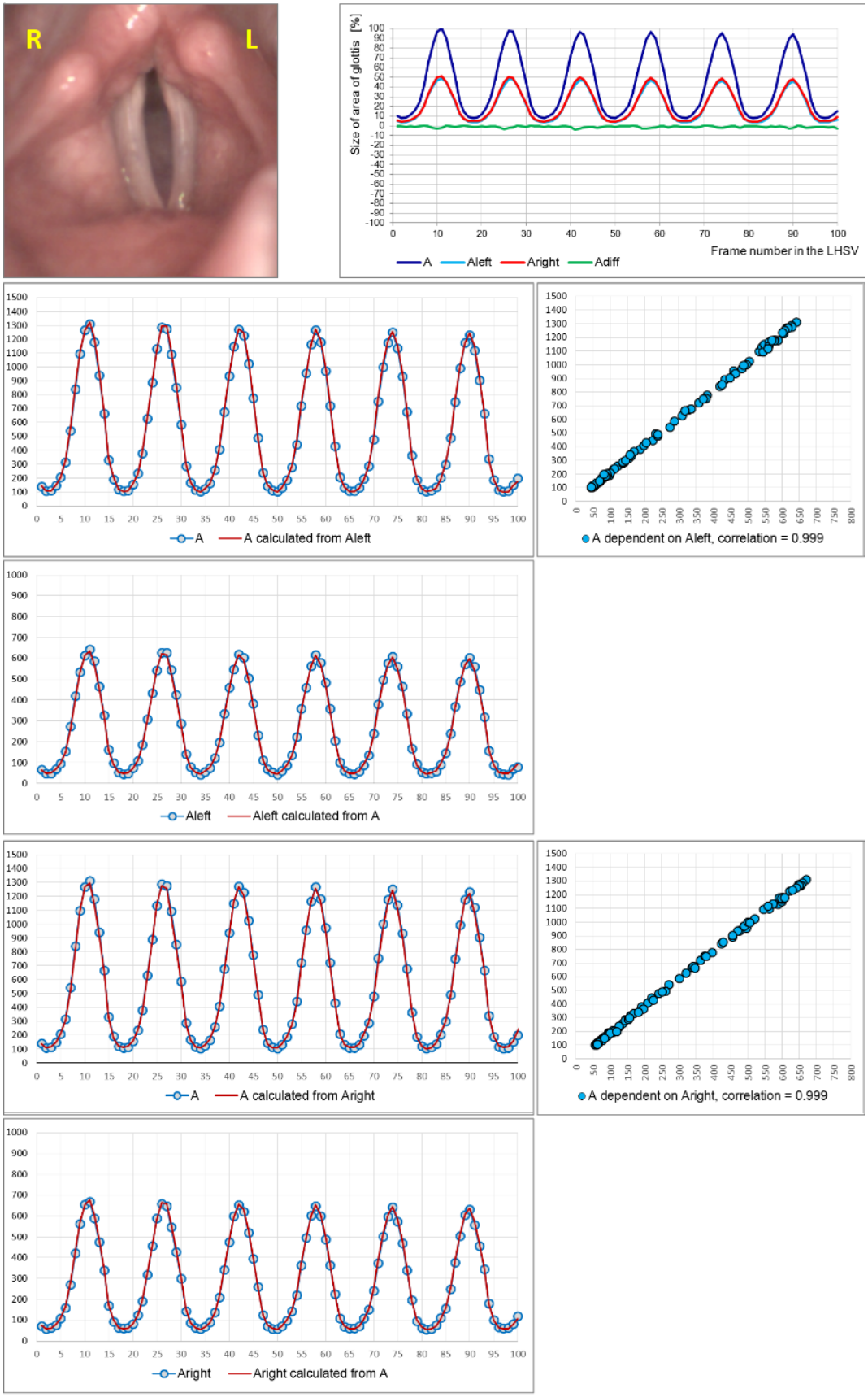


Figure 7.1: *Demonstration of correlation relationship for healthy vocal cords, the correlation value for the parameter pair* $A - A_{left} = 0.999$ *and for* $A - A_{right} = 0.999$.

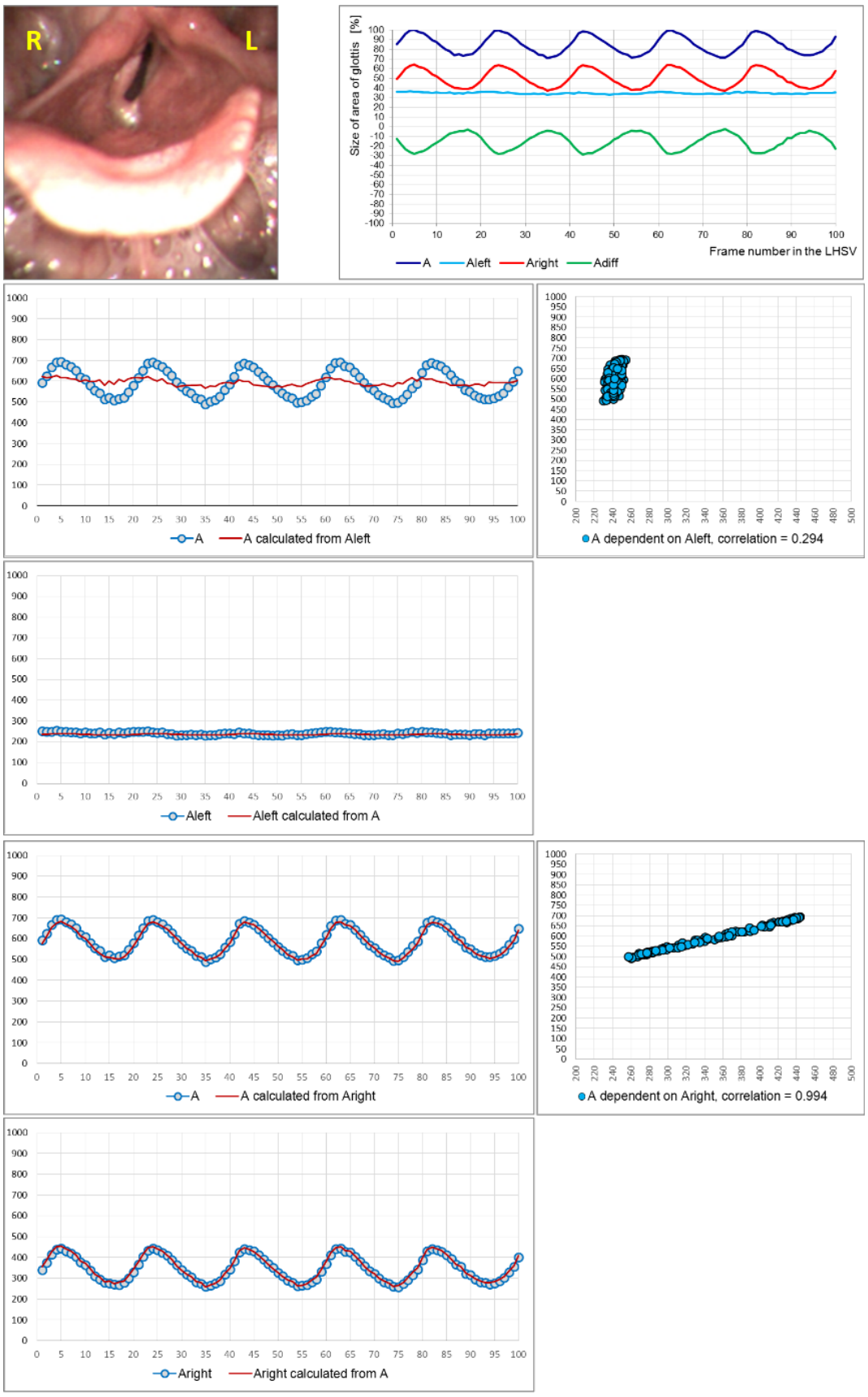


Figure 7.2: *Demonstration of correlation relationship for vocal cords with paresis on the left side, the correlation value for the parameter pair* $A - A_{left} = 0.294$ *and for* $A - A_{right} = 0.994$. *The low correlation value indicates a problem.*

The correlation structure was analyzed by several methods to determine which correlations correspond to the normal (standard) state and which to the pathological state.

7.2.2 Correlation Structure Analysis

In this section, the simplest version of the analysis is presented, applicable for making decisions about correlation relationships that can be understood as significant from a diagnostic point of view, and which is illustrative and clear.

The correlation relationships were divided into three groups:

- SIGNIFICANT – the correlation value is high (near 1 or -1) for the most of recordings;
- EXCEPTIONAL – the correlation value is low (near 0) for the most of recordings;
- All other values

The visualization of the SIGNIFICANT and EXCEPTIONAL values is in figures 7.3 and 7.4.

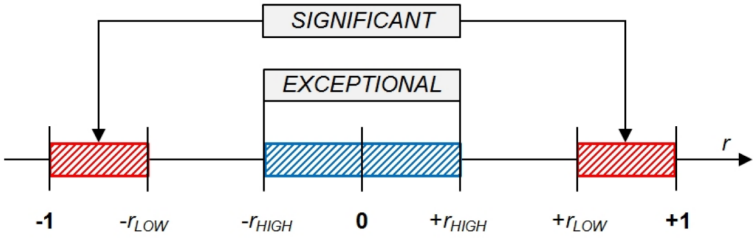


Figure 7.3: *Significant and exceptional correlation value based on limits* r_{LOW} *and* r_{HIGH}

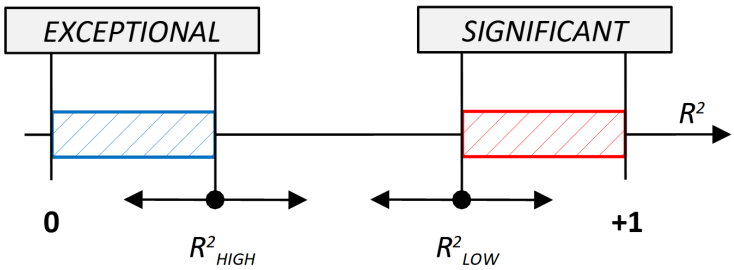


Figure 7.4: *Significant and exceptional correlation value based on determination coefficient* R^2

To tentatively determine what value is significant or exceptional, the limits were introduced according to the data analysis:

- Lower limit $r_{LOW} \in \langle -1; 1\rangle$ to indicate SIGNIFICANT correlations.

$$(r_{LOW})^2 = R^2_{LOW}; \qquad R^2_{LOW} \in \langle 0; 1\rangle \tag{7.8}$$

  R^2_{LOW} is the determination coefficient. The meaning of this limit is if we estimate one of the pair parameters through the other, we reduce its variability at least by the value R^2_{LOW} [%].

- Upper limit $r_{HIGH} \in \langle -1; 1\rangle$ to indicate EXCEPTIONAL correlations.

$$(r_{HIGH})^2 = R^2_{HIGH}; \qquad R^2_{HIGH} \in \langle 0; 1\rangle \tag{7.9}$$

  R^2_{HIGH} is the determination coefficient. The meaning of this limit is if we estimate one of the pair parameters through the other, we reduce its variability at most by the value R^2_{HIGH} [%].

To determine which parameter pairs can be important for vocal cords evaluation, we need to analyze the number of SIGNIFICANT or EXCEPTIONAL correlation values. The frequency analysis of the correlation values of the LHSV recordings was done and the following thresholds were introduced:

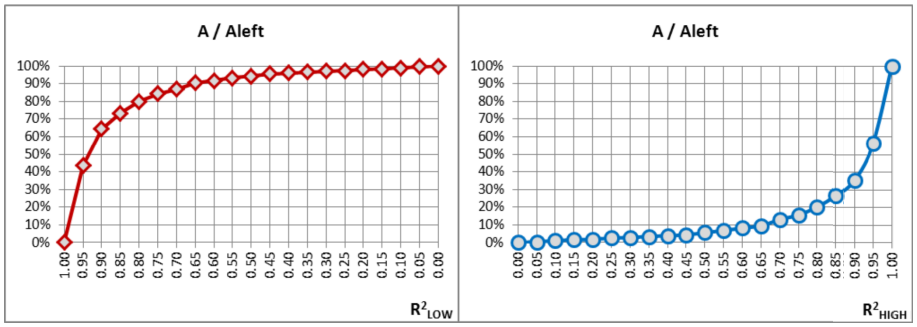
- Lower frequency limit Nr_{LOW} for marking SIGNIFICANT correlations, which means at least Nr_{LOW}% of records has SIGNIFICANT correlation value for the selected parameter pair.
- Upper frequency limit Nr_{HIGH} for marking EXCEPTIONAL correlations, which means at most Nr_{HIGH}% of records has EXCEPTIONAL correlation value for the selected parameter pair.

The parameter pairs, we are searching for, need to fulfill the criteria when at least Nr_{LOW} of records have correlation value in $\langle -1; -r_{LOW})$ or $(r_{LOW}; 1\rangle$ and at the same time at most Nr_{HIGH} of records have correlation values in $(-r_{HIGH}; r_{HIGH})$.

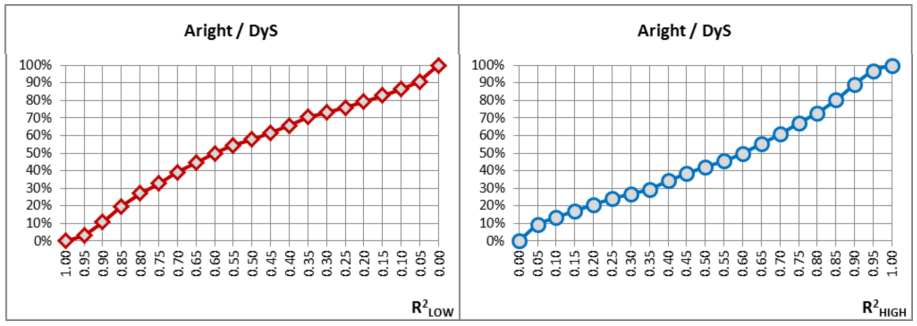
The analysis of the SIGNIFICANT and EXCEPTIONAL correlation values was done over corpus no. 692 (see chapter 3). First, the dependence of the frequency of occurrence of SIGNIFICANT and EXCEPTIONAL correlations on the value of the coefficient of determination $R^2_{LOW} \in \langle 0; 1\rangle$ was tested, see fig. 7.5.

For 91 possible pairs, for which the mutual correlation was calculated, we detected three basic types with the specific behavior of dependence between the number of occurrences and the value of determination coefficient R^2_{LOW} or R^2_{HIGH}. These types are important for the selection of the resulting parameter pairs and marking them as the EXCEPTIONAL relationships, suitable for diagnostic purposes. The

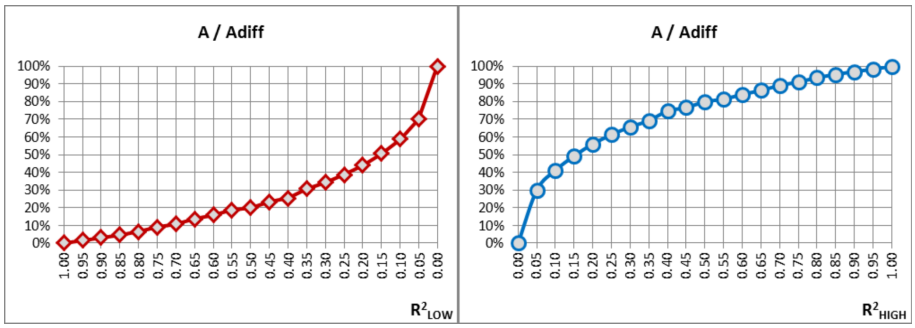
graphs of the detected types of dependency are shown in figure 7.5. Given the nature of the problem and the significance of the analyzed parameter pairs, we selected parameter pairs with the course (a), supplemented by parameter pairs that correspond to almost linear courses (b). In total, there are 34 parameter pairs out of a possible 91 for further consideration.



(a) The course where many records contain high values of correlation. 19 parameter pairs have this type of course, this example is for parameter pair $A - A_{left}$.



(b) Almost linear course. 15 parameter pairs have this type of course, this example is for parameter pair $A_{right} - D_yS$.



(c) The course where just a few records contain high values of correlation. 57 parameter pairs have this type of course, this example is for parameter pair $A - A_{diff}$.

Figure 7.5: *The course types of dependency between the number of records (in [%]) and determination coefficient across all 91 parameter pairs. Frequency distribution of parameter pairs occurrence in the population according to determination coefficient R^2 is useful for determining SIGNIFICANT and EXCEPTIONAL correlation relationships.*

In a further analysis, frequency limits were tested for the frequency of SIGNIFICANT and EXCEPTIONAL correlations in a population sample of 692 LHSV records. From the previous analysis of the dependence between the frequency and determination coefficient, and the course type, the following ranges for the four limits were determined:

- For SIGNIFICANT correlation values

$$R^2_{LOW} \in \langle 0.55; 0.95 \rangle ; \qquad Nr_{LOW} \in \langle 60\%; 95\% \rangle \tag{7.10}$$

- For EXCEPTIONAL correlation values

$$R^2_{HIGH} \in \langle 0.05; 0.30 \rangle ; \qquad Nr_{HIGH} \in \langle 5\%; 20\% \rangle \tag{7.11}$$

Table 7.2 shows the number of parameter pairs with the SIGNIFICANT relationship which meet the criteria of R^2_{LOW} threshold in rows and frequency of occurrences Nr_{LOW} in columns. We are searching for specific settings of the R^2_{LOW} and Nr_{LOW} to get just several parameter pairs.

In the same sense, table 7.3 shows the number of parameter pairs with the EXCEPTIONAL relationship which meet the criteria or R^2_{HIGH} threshold in rows and the frequency of occurrences Nr_{HIGH} in columns. We are searching for specific settings of the R^2_{HIGH} and Nr_{HIGH} to get just several parameter pairs.

According to this analysis, the values for the four limits were set as follows:

- Value 0.894 for r_{LOW} was chosen, it corresponds to the determination coefficient $R^2_{LOW} = 0.800$.

$$r_{LOW} = \sqrt{0.800} = 0.894 \tag{7.12}$$

- Value 0.447 for r_{HIGH} was chosen, it corresponds to the determination coefficient $R^2_{HIGH} = 0.200$.

$$r_{HIGH} = \sqrt{0.200} = 0.447 \tag{7.13}$$

- The threshold value 66.7% was chosen for Nr_{LOW}. It means at least 2/3 of records indicate a SIGNIFICANT correlation value at specific parameter pair.

$$Nr_{LOW} = 66.7\% \tag{7.14}$$

- The threshold value 10.0% was chosen for Nr_{HIGH}. It means at most 10% of records indicate an EXCEPTIONAL correlation at specific parameter pair.

$$Nr_{HIGH} = 10\% \tag{7.15}$$

Table 7.2: Development of the number of parameter pairs depending on the frequency range [%] of the occurrence of the pair in the population sample and determination coefficient R^2 for the cases of SIGNIFICANT correlation relationships.

R^2_{LOW}	Frequency of occurrence of SIGNIFICANT correlations Nr_{LOW} [%]																				
	100%	95%	90%	85%	80%	75%	70%	**65%**	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%	0%
1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91
0.95	0	0	0	0	0	1	1	1	2	2	2	4	6	6	6	6	9	10	14	16	91
0.90	0	0	1	1	2	2	2	2	6	6	6	6	6	9	11	13	15	20	27	29	91
0.85	0	0	1	2	2	3	6	6	7	7	7	9	11	16	21	26	26	29	30	34	91
0.80	0	1	2	2	5	7	7	**7**	7	10	14	20	21	25	26	27	29	31	31	37	91
0.75	0	1	2	5	7	7	7	11	14	18	20	21	26	26	29	30	31	31	33	43	91
0.70	0	2	4	7	7	11	13	16	17	21	21	26	27	29	30	31	31	32	36	46	91
0.65	0	3	6	7	10	14	17	17	20	21	25	27	29	31	31	31	32	34	41	56	91
0.60	0	3	7	9	13	15	17	17	21	24	28	29	31	31	31	32	33	37	44	72	91
0.55	0	4	7	13	15	17	17	21	23	27	29	31	31	31	32	33	36	41	56	77	91
0.50	0	5	8	14	16	17	19	21	26	29	31	31	32	32	34	34	39	49	68	82	91
0.45	0	6	12	15	17	17	20	25	29	31	32	32	33	34	34	36	44	60	77	90	91
0.40	0	8	13	16	17	19	21	28	31	32	32	33	34	34	36	41	54	71	79	90	91
0.35	0	8	15	17	18	20	28	31	32	32	34	34	34	35	40	50	68	78	84	91	91
0.30	0	10	15	17	19	22	31	32	34	34	34	34	35	39	48	66	75	81	90	91	91
0.25	0	13	15	17	20	29	32	34	34	34	34	35	40	48	64	75	80	86	91	91	91
0.20	0	13	17	19	25	33	34	34	34	34	37	40	50	68	76	82	86	90	91	91	91
0.15	1	15	17	22	32	34	34	34	34	38	41	60	73	79	83	87	90	91	91	91	91
0.10	3	15	20	31	34	34	34	36	41	56	70	79	82	87	90	91	91	91	91	91	91
0.05	4	17	30	34	34	37	49	68	78	84	86	90	91	91	91	91	91	91	91	91	91
0.00	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91

Table 7.3: Development of the number of parameter pairs depending on the frequency range [%] of the occurrence of the pair in the population sample and determination coefficient R^2 for the EXCEPTIONAL correlation relationships.

R^2_{LOW}	Frequency of occurrence of EXCEPTIONAL correlations Nr_{HIGH} [%]																				
	0%	5%	**10%**	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
0.00	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91	91
0.05	4	17	30	34	34	37	49	68	78	84	86	90	91	91	91	91	91	91	91	91	91
0.10	3	15	20	31	34	34	34	36	41	56	70	79	82	87	90	91	91	91	91	91	91
0.15	1	15	17	22	32	34	34	34	34	38	41	60	73	79	83	87	90	91	91	91	91
0.20	0	13	**17**	19	25	33	34	34	34	34	37	40	50	68	76	82	86	90	91	91	91
0.25	0	13	15	17	20	29	32	34	34	34	34	35	40	48	64	75	80	86	91	91	91
0.30	0	10	15	17	19	22	31	32	34	34	34	34	35	39	48	66	75	81	90	91	91
0.35	0	8	15	17	18	20	28	31	32	32	34	34	34	35	40	50	68	78	84	91	91
0.40	0	8	13	16	17	19	21	28	31	32	32	33	34	34	36	41	54	71	79	90	91
0.45	0	6	12	15	17	17	20	25	29	31	32	32	33	34	34	36	44	60	77	90	91
0.50	0	5	8	14	16	17	19	21	26	29	31	31	32	32	34	34	39	49	68	82	91
0.55	0	4	7	13	15	17	17	21	23	27	29	31	31	31	32	33	36	41	56	77	91
0.60	0	3	7	9	13	15	17	17	21	24	28	29	31	31	31	32	33	37	44	72	91
0.65	0	3	6	7	10	14	17	17	20	21	25	27	29	31	31	31	32	34	41	56	91
0.70	0	2	4	7	7	11	13	16	17	21	21	26	27	29	30	31	31	32	36	46	91
0.75	0	1	2	5	7	7	7	11	14	18	20	21	26	26	29	30	31	31	33	43	91
0.80	0	1	2	2	5	7	7	7	7	10	14	20	21	25	26	27	29	31	31	37	91
0.85	0	0	1	2	2	3	6	6	7	7	7	9	11	16	21	26	26	29	30	34	91
0.90	0	0	1	1	2	2	2	2	6	6	6	6	6	9	11	13	15	20	27	29	91
0.95	0	0	0	0	0	1	1	1	2	2	2	4	6	6	6	6	9	10	14	16	91
1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91

7.2.3 Results

The relationships between all 91 possible pairs of geometric parameters from 692 LHSV recordings were analyzed, each recording contained at least 400 frames for computing the correlation values. Following 7 parameters, mentioned in the table 7.4, fulfill all conditions introduced and configured in the previous analysis.

Table 7.4: *List of parameter pairs that meet the set criteria.*

parameter pair	number of SIGNIFICANT	number of EXCEPTIONAL
$D_xS - DxH$	96.5%	0.1%
$D_yS - D_yH$	92.0%	0.4%
$A - A_{right}$	82.1%	1.0%
$P - P_{left}$	81.4%	0.3%
$A - P$	81.1%	0.7%
$A - A_{left}$	79.9%	1.9%
$P - P_{right}$	77.6%	2.6%

During the LHSV video evaluation, the correlation values of these parameter pairs are computed and compared with the pairs above. In case of any "broken" correlation relationship, the warning can be raised. The number of "broken" correlation relationships could be an indicator of the seriousness of the issue.

7.2.4 Conclusion

The presented method identified seven parameter pairs where the mutual correlation relation is strong for most of the analyzed LHSV recordings and is expected to be high in any recording with healthy vocal cords. In case of the unexpectedly low value of one or more correlation values between these parameters, the warning should be raised leading to possible more attention. This method can be used as one of the tools during the vocal cords examination.

The results are computed from 692 records from the data corpus regardless of the video quality or diagnosis. If a particular recording shows small correlation values of one or more parameter pairs, it can be expected that the behavior is different from "standard" behavior and it is worth examining the vocal cords more in detail.

This method was a result of the first analysis of correlation values between parameters before any other data (like ratings from the ENT expert) were used. There could be many ways how to improve this method, like searching for the parameter pairs of significant and exceptional correlation values from healthy vocal cords only to improve detection of a potential issue, or using different statistical methods (like Wilk's limits).

All the parameter pairs from the results are related to the vocal cords' symmetry. It supports the idea of watching the behavior of the left and right sides separately.

The ideas of this method were first presented in [55] together with several case studies. More case studies are presented in section 7.3.6.

7.3 Correlation Classification

Analysis of computed parameters from the detected glottis is the essential task for the evaluation of the vocal cords' kinematics. The goal mentioned in this section is to obtain a single number for a specific LHSV recording that is easily understandable. It is probably too ambitious to determine a diagnosis, but it should be possible to create a simple rating of the vocal cords' behavior to get a warning about a potential issue.

To accomplish this objective, a new method based on correlations and statistics was introduced. This method is called Segment comparison using correlation. The input data contains computed values for 14 parameters and ratings from the ENT expert. 396 LHSV recordings (part of corpus no. 412, some recordings were removed due to duplicates with corpus no. 692) were used as a training set, the rest of corpus no. 692 was then analyzed with the setup to test the method. The data sets contain a single value for each of 91 parameter pairs representing their correlation for every LHSV recording in the set.

7.3.1 Vocal Cords Rating

Every LHSV recording from the data corpus mentioned in section 3 was evaluated by an ENT expert and was classified into one of 5 classes like school grades. The evaluation was based on the subjective observation of the recording, symmetry specifically, development of the shape during the opening and closing phases, and movement of mucosal wave (pic. 2.5). Also results from other examinations were considered if they were available, like MDVA, VRP, and SCORE. The audio recordings, diagnoses, and many years of experience in this field were also taken into account. The classes are following:

1. Healthy vocal cords without any detected issues in the video and audio-based examinations.
2. Vocal cords in good shape but one of the video or audio-based examinations indicates a potential issue.
3. Vocal cords with detected minor pathological findings.
4. Vocal cords with severe pathological findings, asymmetry, or limited movement.
5. Vocal cords in bad condition, very limited movement detected or containing significant pathological finding, matching with audio analysis.

This rating was added to metadata and used for parameter setup in the training set. It was also used for comparison with the classification results of this method.

The vocal cords' quality rating considers the following factors:

- The symmetry of the left and right vocal folds in all phases of opening and closing of vocal cords.
- Undulations, observable on the tissue of the vocal cords in individual phases, see fig. 2.4 and 2.5, should be the same on both vocal folds.
- Closing of the vocal cords in the adduction phase. The glottis should close completely or remain partially open towards the posterior commissure depending on the intensity of the phonation and the frequency.
- Presence of unexpected objects on the vocal folds, which may primarily not affect the symmetry during movement, but changes the flexibility and weight of the vocal fold. It can also affect the above-mentioned undulation of the vocal fold tissue.
- The regularity and periodicity of oscillations corresponding to F_0, which is often disturbed in post-traumatic states, e.g. in diagnosed reversible nerve paresis (innervation disorder).

7.3.2 Method Description

The first step of this method is to find pairs of parameters (measured by their correlation as explained in section 7.1) that characterize the membership to the classes "in the best way". Every LHSV recording will be classified based on the difference between a correlation of such parameter pairs and the values which were found during the processing of the data from the training set. Because the classification is done separately for each class whether an LHSV recording belongs to it or not, the recording can be classified into multiple classes and the final result is then rounded average value of the class numbers.

According to the fact, that computed values for every of m parameters are synchronously sampled in time $(t_1, t_2, ..., t_n)$, $t_1 < t_2 < ... < t_n$, the Sample correlation coefficient $r(X, Y)$ (eq. (7.16)) could be chosen as a comparison method [1].

$$r(X,Y) = \frac{\sum_{i=1}^{n}(X(t_i) - \bar{X})(Y(t_i) - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X(t_i) - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y(t_i) - \bar{Y})^2}} \tag{7.16}$$

Reasons for this choice and several properties or $r(X, Y)$:

[1] Because $(X(t_i), Y(t_i))$ for $i = 1, ..., n$ does not have to be a sample in general, the $r(X, Y)$ is not correlation coefficient estimate, but equivalent data processing with the same properties as described in the section 7.1.

- $r(X, Y)$ is symmetrical, i.e. $r(X, Y) = r(Y, X)$.
- $r(X, Y)$ is independent on translation in both operands, i.e. $r(X, Y) = r(X + a, Y + b)$ for $\forall a, b \in R_1$.
- $r(X, Y)$ is independent on scale in both operands, i.e. $r(X, Y) = r(uX, vY)$ for $\forall u, v \in R_1; u, v > 0$.
- $-1 \leq r(X, Y) \leq +1$, which is based on Cauchy-Schwarz inequality.
- if $|r(X, Y)| = +1$, there are numbers $a, b \in R_1$ where $Y(t) = aX(t) + b$ for $t \in \{t_1, t_2, ..., t_n\}$[56]. $a > 0$ for $r(X, Y) = +1$, $a < 0$ for $r(X, Y) = -1$.
- $1 - r^2(X, Y)$ well characterizes quality (mean square error) where X approximates linear approximation from Y and vice versa by the ordinary least squares method.

Because we are searching for the similarity of the values, it is a good approach to modify the resulting correlation coefficients the way, which would differentiate the close and far values more significantly. This transformation still has to keep the following properties which are used in further processing:

- $r(X, Y) = r(X', Y') \implies t[r(X, Y)] = t[r(X', Y')]$
- $r(X, Y) > r(X', Y') \implies t[r(X, Y)] > t[r(X', Y')]$
- $r(X, Y) = 0 \implies t[r(X, Y)] = 0$
- $r(X, Y) = s \ \wedge \ r(X', Y') = -s \implies t[r(X, Y)] = -t[r(X', Y')]$

Among many transformations which meet the criteria, the Fisher transformation[2][54] was used, which is defined as eq. (7.17).

$$t\left[r(X, Y)\right] = 0.5 \ln\left(\frac{1 + r(X, Y)}{1 - r(X, Y)}\right); \qquad r(X, Y) \neq \pm 1 \tag{7.17}$$

The graph of the Fisher transformation can be seen in fig. 7.6. The example of classified values before and after Fisher transformation can be seen in figures 7.7 and 7.8, where their differentiation of the values near 1 is significant, also the medians of the values are more distinguished. All further calculations containing correlation values of parameter pairs use the values after Fisher transformation.

[2]Fisher transformation is usually used for confidence interval estimates of sample correlation coefficient for normally distributed variables.

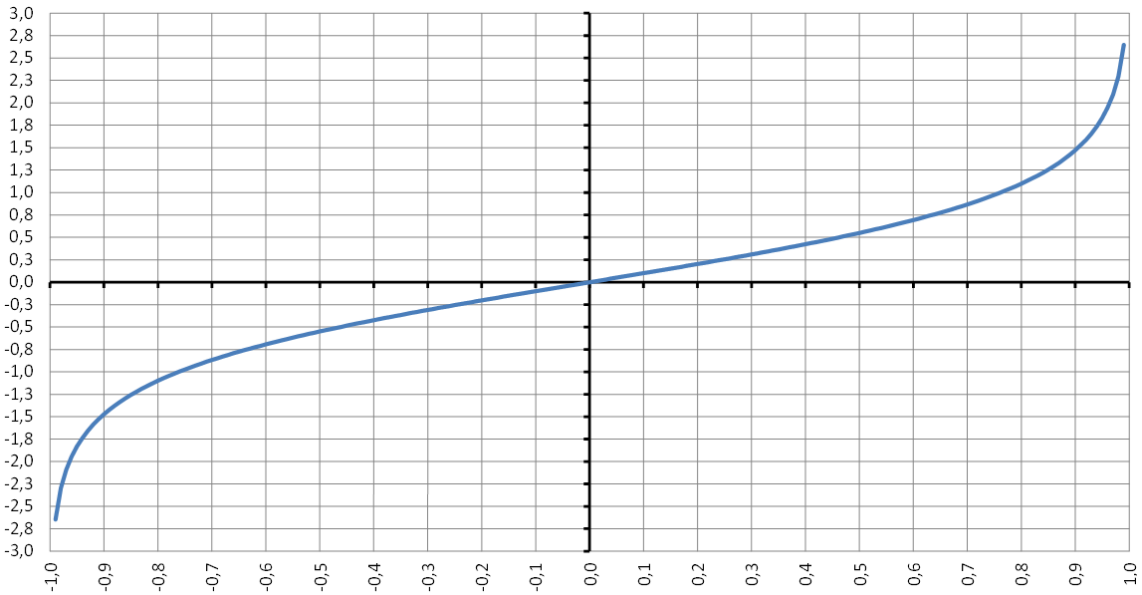


Figure 7.6: *Values of the Fisher transformation.*

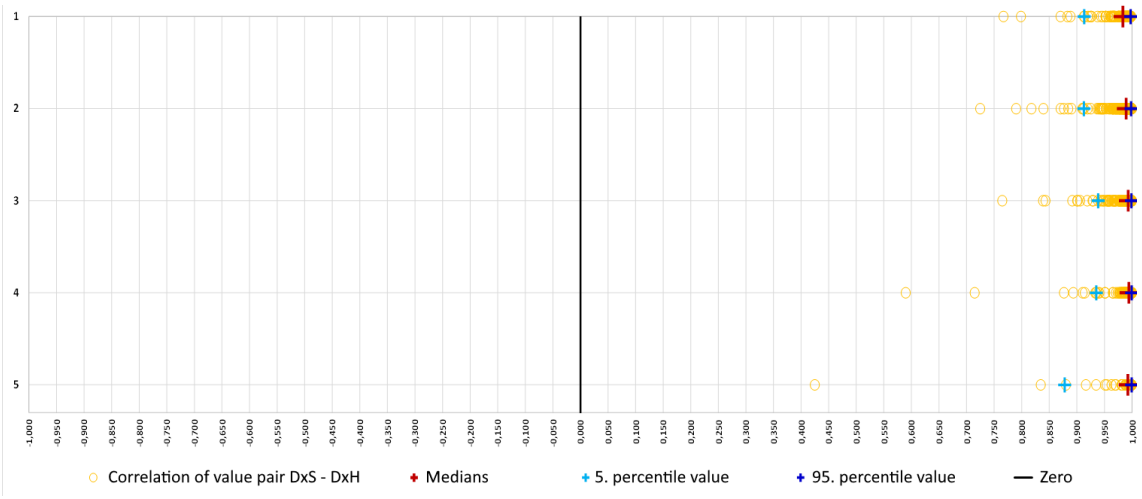


Figure 7.7: *Original values of the value pair $D_xS - D_xH$ before Fisher transformation.*

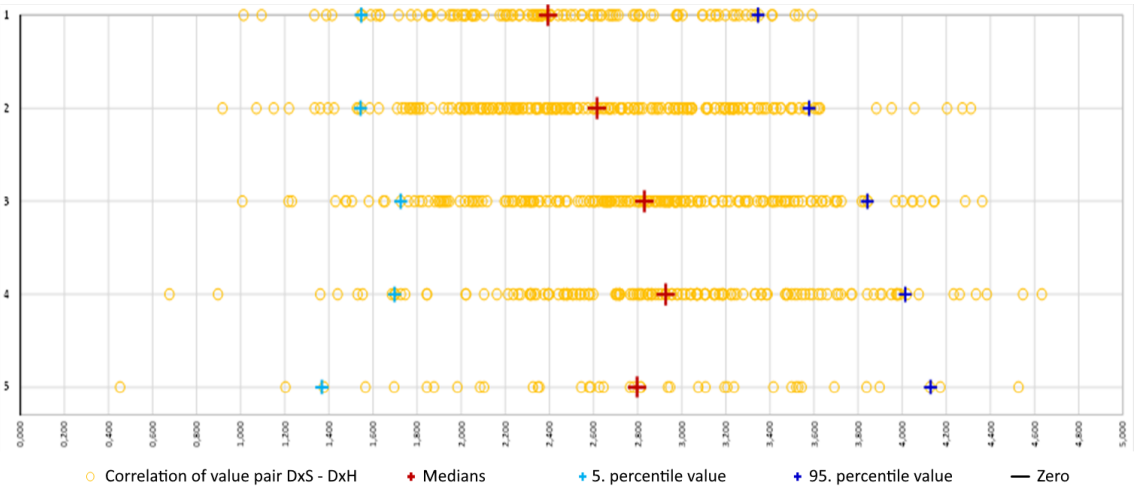


Figure 7.8: *Values from fig. 7.7 after the Fisher transformation. The distribution of the values is better observable, which is represented by the position of the 5th and 95th percentile marks.*

7.3.3 Classification Using Medians

The classification itself is done by splitting data into groups of the values belonging to the specified classes and their complements which comprises values from other classes. To have the classification as precise as possible, for each class, we searched for the parameter pair with the greatest distance between the median of the correlation values belonging to a specified class and the median of the correlation values NOT belonging to a specified class (complement median). All correlation values are after the Fisher transformation. The median is defined as the following.

Lets have values $x_1, x_2, ..., x_p$ where p is number of values, then sorted sequence of these values is $x_{(1)}, x_{(2)}, ..., x_{(p)}$, where $x_{(i)} \leq x_{(i+1)}; i = 1, ..., p-1$. The $x_{(1)}$ is the lowest value in this sequence and the $x(p)$ is the greatest. The median is then defined as in eq. (7.18) for odd p, in eq. (7.19) for even p.

$$med(x_1, x_2, ..., x_p) = x_{\left(\frac{p+1}{2}\right)}; p \in \{1, 3, 5, ...\}\,; \tag{7.18}$$

$$med(x_1, x_2, ..., x_p) = \frac{x_{\left(\frac{p}{2}\right)} + x_{\left(\frac{p}{2}+1\right)}}{2}; p \in \{2, 4, 6, ...\} \tag{7.19}$$

The evaluation of the ENT expert was used and the individual recordings from the training set were classified into one of the 5 classes. Within each class and for each parameter pair, a median of correlation values belonging to the class was calculated, and the median of the values belonging to any of the other classes (class complement) was calculated. Then we picked the parameter pair for each class where the distance of the medians was greatest. Values and numbers can be seen in table 7.5.

Table 7.5: *Classes with the parameter pair where the median distance was greatest.*

Class	Parameter pair	Median distances	Class median	Complement median
1	$D_yH - P_{right}$	0.848	-1.817	-0.970
2	$P - P_{right}$	0.607	2.424	1.818
3	$A - A_{right}$	0.320	1.868	2.188
4	$D_yS - P$	1.035	-0.553	-1.589
5	$P - P_{right}$	1.217	0.888	2.104

As can be seen in table 7.5, the parameter pair $P - P_{right}$ is there twice, for classes 2 and 5. The training sets of LHSV are different for each class and it is understandable that one significant pair can determine more classes. According to the median values, the higher correlation value of this parameter pair is rather classified into class 2 (healthy), and the lower value into class 5 (bad).

The interesting question can be why these parameter pairs are different from chapter 7.2. The reason is the different approach in this method, where we are not searching for the highest correlation values in the whole set, but we are searching for the biggest difference between the set of videos rated as a specific class and the set of videos not belonging to this class. If the correlation value is high for all classes, then the significance for classification is low.

Figure 7.9 shows the distribution of the values of parameter pair $D_yH - P_{right}$ after Fisher transformation for all classes on the y-axis. The cross symbol is the median of the specific class and the triangle mark is the complement median. The yellow circles are the single correlation values belonging to the class, the gray dots below are the values of the complement. This example shows the greatest distance for class 1 in the first row.

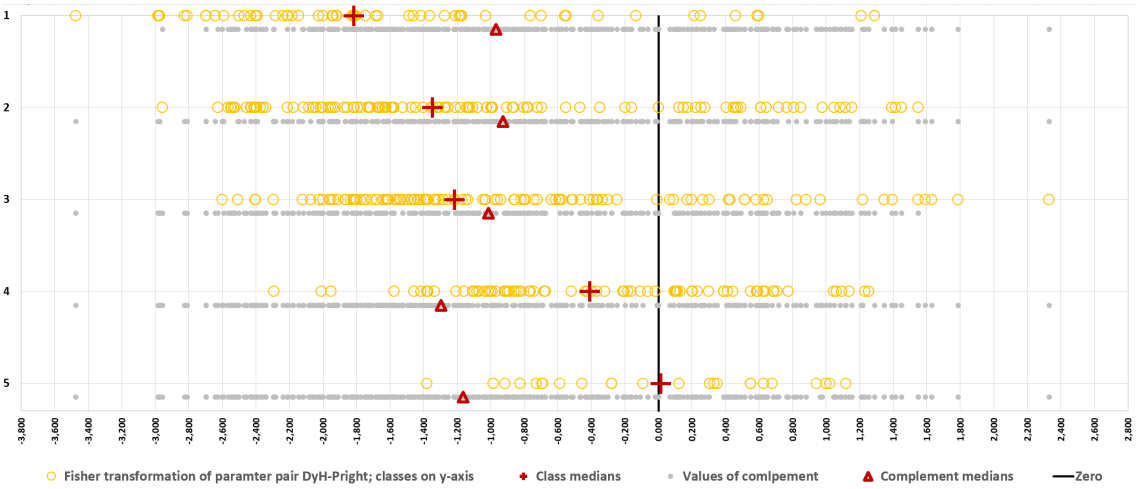


Figure 7.9: *Distribution of the correlation values for parameter pair $D_yH - P_{right}$ for the classes within the training set of 396 LHSV recordings. The biggest distance between the class median and complement median can be seen within class 1.*

Similarly, figures 7.10, 7.11, and 7.12 show the parameter pairs and the distribution of data together with the medians for all classes.

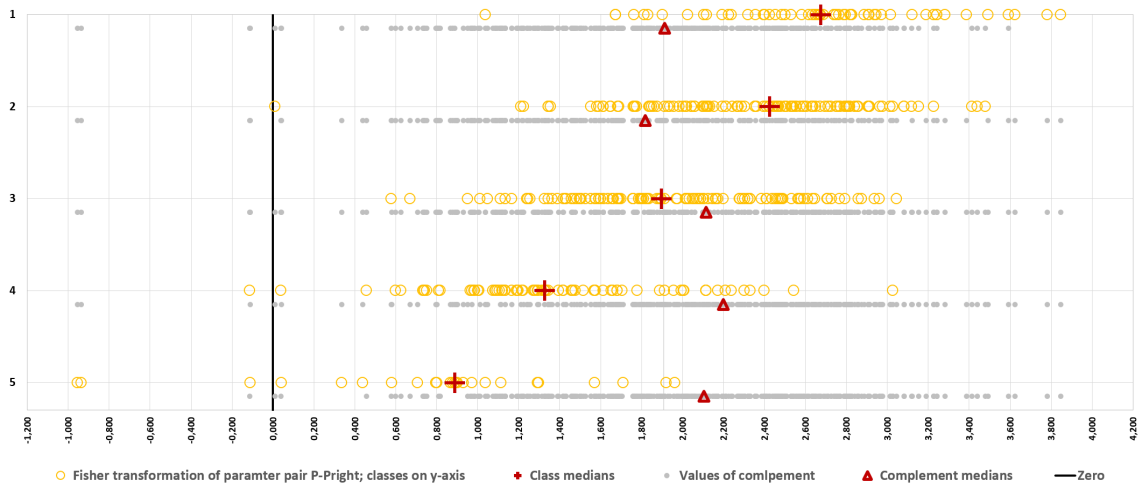


Figure 7.10: *Distribution of the correlation values for parameter pair* $P - P_{right}$*, where the biggest distance between the class median and complement median can be seen within classes 2 and 5.*

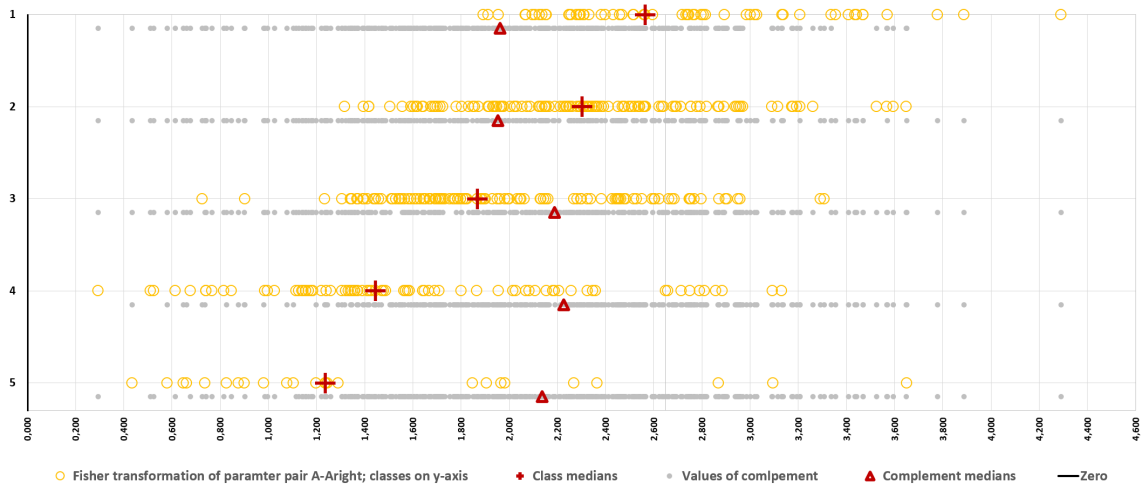


Figure 7.11: *Distribution of the correlation values for parameter pair* $A - A_{right}$*, where the biggest distance between the class median and complement median can be seen within class 3.*

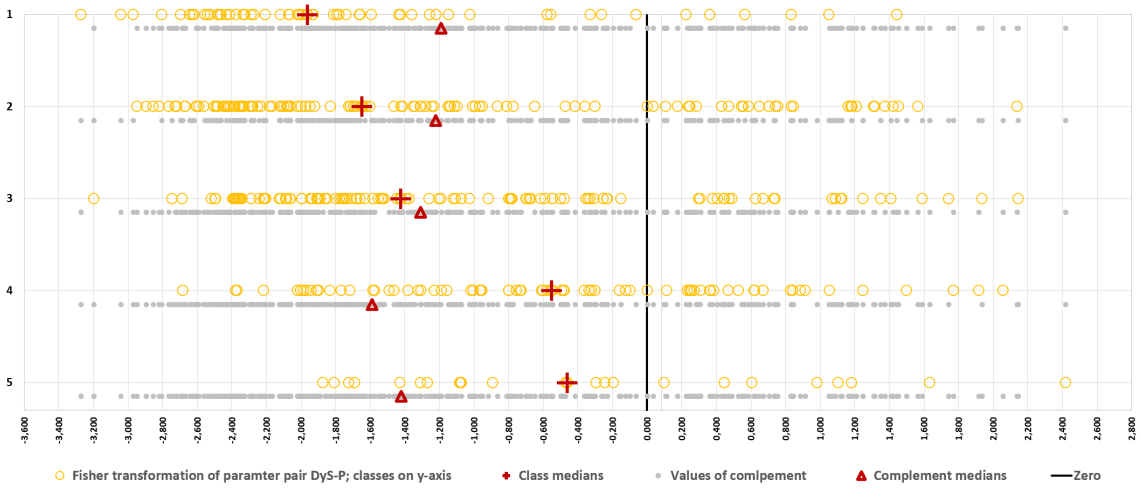


Figure 7.12: *Distribution of the correlation values for parameter pair D_yS-P, where the biggest distance between the class median and complement median can be seen within class 4.*

The classification of any LHSV recording works as follows. The correlation values of parameter pairs of classes 1 to 5 are compared with the corresponding computed class medians and complement medians. In the case of a lower distance to the class median, the recording is classified as a member of the class, if the distance to the complement median is lower, the recording is not classified into the class. The following situations may happen:

1. The recording is classified into just one class. Then the class is final.
2. The recording is classified into more classes. In this case, the final class is a rounded average value of the class numbers.
3. The recording is not classified into any class. In this case, the record cannot be classified automatically and is marked as unclassified.

Validation

As a validation of this process, the training recordings were classified using found medians. To compare it with the original rating from the ENT expert, the differences between the original and the found classes were computed, the results can be seen in table 7.6. There were 11 recordings not classified[3].

[3]From the 11 not classified recordings, one was rated by ENT expert as class 1, 3 recordings were in class 2, 3 were rated as class 3, 2 as class 4 and 2 as class 5.

Table 7.6: *Classification of training data set (396 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage	cumulative percentage
0	157	39.7%	39.7%
1	182	46.0%	85.6%
2	42	10.6%	96.2%
3	4	1.0%	97.2%
4	0	0%	97.2%
not classified	11	2.8%	100%

Results

The LHSV recordings from corpus no. 692 were used as a testing sample for the method evaluation. Table 7.7 shows the results of the data which were not included in the training set. The results are very similar to the validation results. Hence we can consider found configuration of the parameter pairs as correct.

Table 7.7: *Classification of testing data set (296 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage	cumulative percentage
0	107	36.2%	36.2%
1	144	48.7%	84.8%
2	35	11.8%	96.6%
3	7	2.4%	99.0%
4	0	0%	99.0%
not classified	3	1.0%	100%

Table 7.8 shows the complete results of corpus no. 692, training and testing set together. As can be seen in the column with cumulative percentage, the difference between rated vocal cords by the ENT expert and the classification from the method is in 85.3% of cases equal to or less than 1, which is a very good result for such heterogeneous data.

Table 7.8: *Classification of complete data corpus no. 692 (692 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage	cumulative percentage
0	264	38.2%	38.2%
1	326	47.1%	85.3%
2	77	11.1%	96.4%
3	17	1.6%	98.0%
4	0	0%	98.0%
not classified	14	2.0%	100%

The previous tables show the differences without differentiation whether the classification result is greater or lesser than the rating from the ENT expert. This is split in table 7.9 where the positive values mean that the method was more strict (gives a worse rating) than the ENT expert. For a better idea, the distributions of the results are shown in fig. 7.13.

Table 7.9: *Classification of complete data corpus no. 692 (692 recordings). The first column contains rating differences, which is a difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage
-4	0	0.0%
-3	8	1.2%
-2	20	2.9%
-1	119	17.2%
0	264	38.2%
1	207	29.9%
2	57	8.2%
3	3	0.4%
4	0	0.0%
not classified	14	2.0%

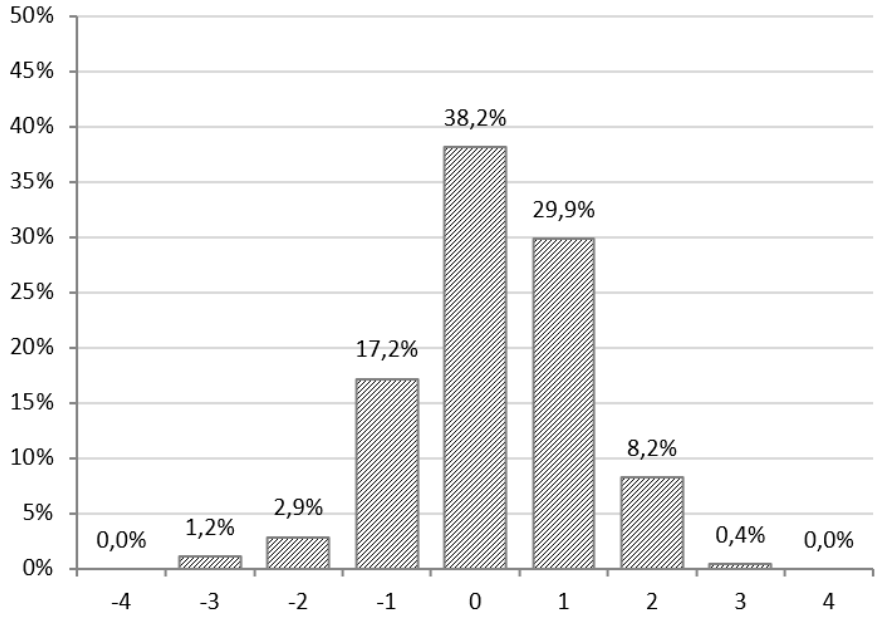


Figure 7.13: *Detailed distribution of the results of the correlation classification. Displayed data are the differences between the classification results and the rating of the ENT expert.*

Discussion

According to the class characteristic and variability of source data, the difference ± 1 is acceptable. Thus about 85% of recordings are well classified, which is a very good result.

The training set gives the best results which are expected as the method setup was done on these data. But the result of testing data gives very similar numbers, which shows consistency of the data and well-configured parameters.

From the distribution in figure 7.13, we can see that there are more values with a positive difference between ENT expert and the method results than negative. It is due to numerical rounding, where the higher class was determined in the case of classification into two adjacent classes. It is better to be more strict and evaluate the recording with a worse class because there is less probability that the potential issue was not detected.

Error Estimation

The distances from both medians are used for classification, but they are not important for the error estimation of specific classification. The error can be estimated according to the number of occurrences in the basic set. The following situations may happen:

1. False positive – It is decided to belong to the complement of the given class but the classified record actually belongs to the given class.

2. False negative – It is decided to belong to the given class but the classified record does not belong to it.

Empirical distribution functions (EDF) for given data pairs are a good starting point for such assessments (probability estimates of individual types of errors). The empirical distribution function is the number of observations less than or equal to the value on the x-axis to the number of all observations from a given group (here the values from a parameter pair correlation of a given class or its complement)[4].

In the best case, the EDF of the corresponding correlations after Fisher transformation for a given parameter pair and class is "on the right side" or "on the left side" (see fig. 7.14) of the EDF Fisher transformed correlations for that parameter pair for the complement. Of course, there are pairs of parameters in which the mutual position of both EDFs cannot be simply determined like in the previously mentioned situations, because they intersect (i.e. they are stochastically incomparable groups in the probability case sense), see fig 7.15.

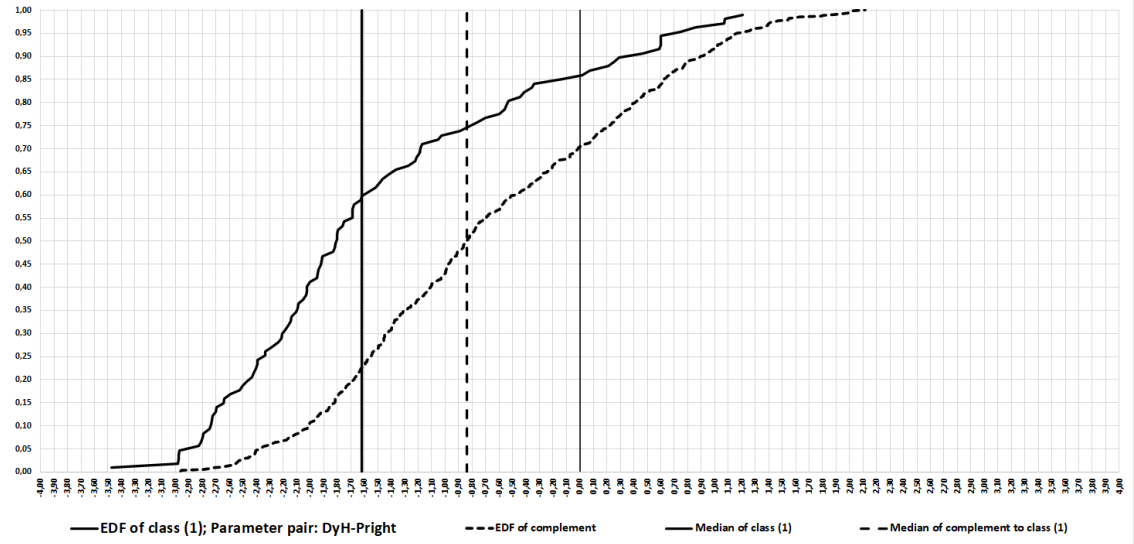


Figure 7.14: *Empirical distribution function (EDF) of the parameter pair $D_yH - P_{right}$. The EDF of class 1 is "on the left side" from the EDF of the complement to this class. Thus these EDFs would have been stochastically comparable.*

[4]In order for EDF to be an estimate of the actual distribution function, it is necessary that the observed data form a random sample (iid = equally distributed and independent observations). That may not be the case here

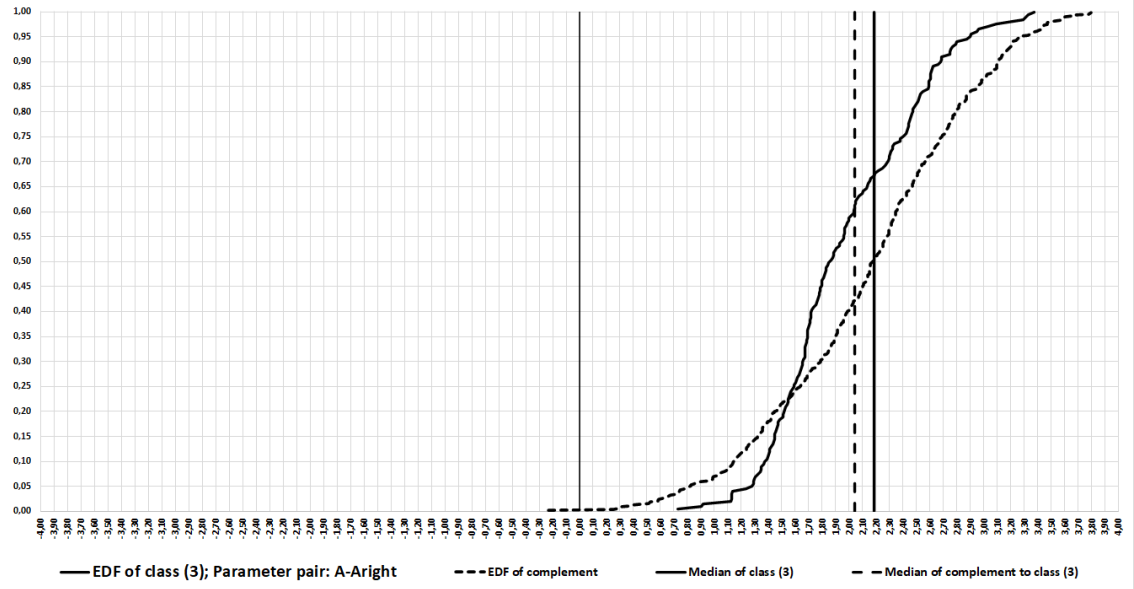


Figure 7.15: *Empirical distribution function (EDF) of the parameter pair* $A - A_{right}$. *The EDF of class 3 and the EDF of the complement to this class intersect and the mutual position cannot be simply determined. Thus these EDFs would have not been stochastically comparable.*

For the "on the right side" case of medians position (median of the class is to the right from the complement median), the basic estimate of the false positive error will be the ratio of the number of observations of the class to the left from the averages of both medians to all observations of that class. The basic estimate of the probability of a false negative error will be the ratio of the number of observations of the class complement to the right of the average of the two medians to all observations of the complement of the given class. The values of the specified class are stochastically higher than the values of the complement.

For the "on the left side" case of medians position (median of the class is to the left from the complement median), the basic estimate of the false positive error will be the ratio of the number of observations of the class to the right from the averages of both medians to all observations of that class. The basic estimate of the probability of a false negative error will be the ratio of the number of observations of the class complement to the left of the average of the two medians to all observations of the complement of the given class. The values to the specified class are stochastically lower than the values to the complement.

In this way, representative values of the error probabilities of the false positive and false negative error types are determined. The values of the error estimates in a particular classification are determined in the same way as representative estimates, only the correlation value (after Fisher's transformation) of the given data pair of the classified record is used instead of the average of both medians.

7.3.4 Classification Using Oriented Areas between Class EDF and Complement

Comparing EDFs is important for error estimations, but it can be also used for another type of classification. The classification method can be adjusted by selecting the (non-exceedable in the base set) value of the false positive error estimate and searching for a parameter pair that leads to the lowest false negative error estimate value for each class while maintaining "this probability estimate" value. This concept can also be reversed. It depends on the task:

- It is more important to determine the class to which the classified record belongs.
- It is more important to determine the class(es) to which the classified record does not belong.

From these tasks, the roles of the EDF class and its complement are determined. It is obvious that stochastically more distant (E)DFs will provide better results of classification. According to the classification tasks, the “area size” between the two EDFs is important. The size of the area is unambiguous, where the two EDFs do not intersect at all (= empirically stochastically comparable). Where they intersect classification for such parameter pairs will be problematic. This property (empirical stochastic incomparability = “crossing” of the EDF) must be respected when measuring the area between the two EDFs. Concerning this, the oriented areas $A_{\Delta EDF}$ are used:

$$A_{\Delta EDF} = \left| \int_{-\inf}^{+\inf} (EDF_{class}(x) - EDF_{complement}(x))\, dx \right| \tag{7.20}$$

This formula can be simplified for the numeric calculation, where the trapezoidal rule can be used:

$$A_{\Delta EDF} = \left| \int_{0}^{1} (PERCENTILE_{class}(y) - PERCENTILE_{complement}(y))\, dy \right| \tag{7.21}$$

Found Parameter Pairs

The principle is then similar to the previous method. The size of the oriented area is computed for all parameter pairs and the ones with the biggest oriented area for a class are selected for the classification. To be able to compare the correlation value with the area, the center of gravity was computed for all five largest areas. The x-value of the center of gravity is then the value that is compared with the specific correlation value, see table 7.10. This table also shows whether lower or the higher values than the center of gravity belong to the given class, so this information is also shown in table . The left means that the EDF representing values of the class are “on the left side” from the complement values so lower correlation values than the center of gravity x-value will be classified into the class. The right means that class

values are "on the right side", so higher correlation values will be classified into the class. Data in this table was taken from the training set of 396 videos, the same training corpus as in the previous method using medians.

Table 7.10: *Classes with the parameter pairs where the oriented areas between EDFs are the greatest, computed from the training set.*

class	parameter pair	oriented area size	center of gravity	class position
1	$D_yH - P_{right}$	0.837	-1.163	left
2	$P_{left} - P_{right}$	0.492	1.337	right
3	$A_{diff} - D_xH$	0.193	1.632	right
4	$P_{left} - P_{right}$	0.764	1.012	left
5	$P - P_{right}$	1.262	1.427	left

We can see that some of the parameter pairs are the same as used in the median method, but also different pairs like $P_{left} - P_{right}$ or $A_{diff} - D_xH$. Again, all pairs are related to symmetry. Figures 7.16, 7.17, 7.18, 7.19, and 7.20 show the EDFs of the values belonging to the class and the complements for each class computed from the training set. The EDFs slightly intersect for classes 3 and 4, but it is not affecting area size and we can consider EDFs as stochastically comparable in all cases.

Also, the x-values of the center of gravity are marked in the figures. According to the figures, the x-value of the center of gravity is always between the medians, but not in the middle. Because of that, similar results to the previous method are expected.

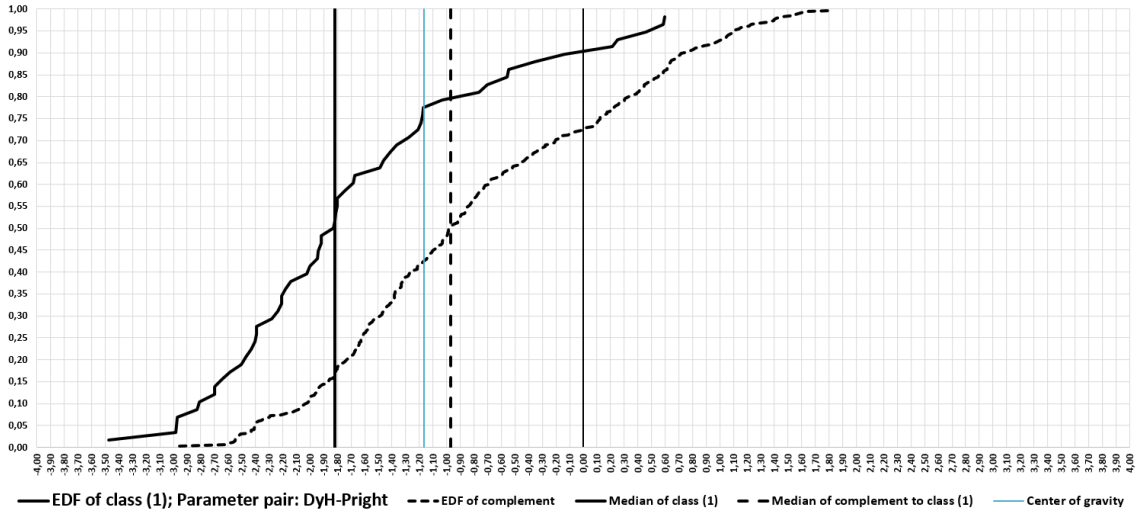


Figure 7.16: *The EDFs for the parameter pair $D_yH - P_{right}$ for class 1. On the left side, there is EDF of class 1, and on the right side, there is complement EDF. The oriented area has a size of 0.837 and is the biggest one for class 1.*

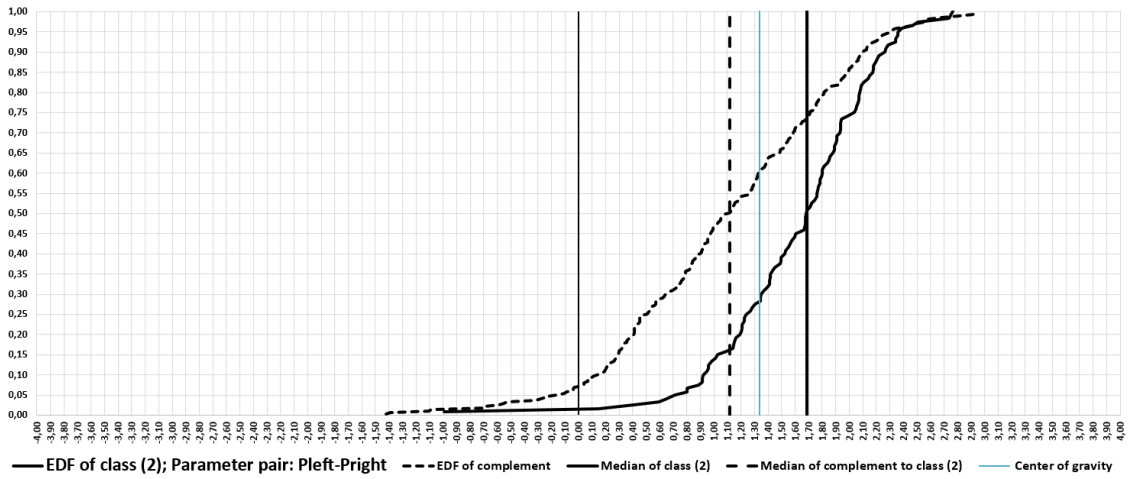


Figure 7.17: *The EDFs for the parameter pair $P_{left} - P_{right}$ for class 2. On the right side, there is EDF of class 2, and on the left side, there is complement EDF. The oriented area has a size of 0.492 and is the biggest one for class 2.*

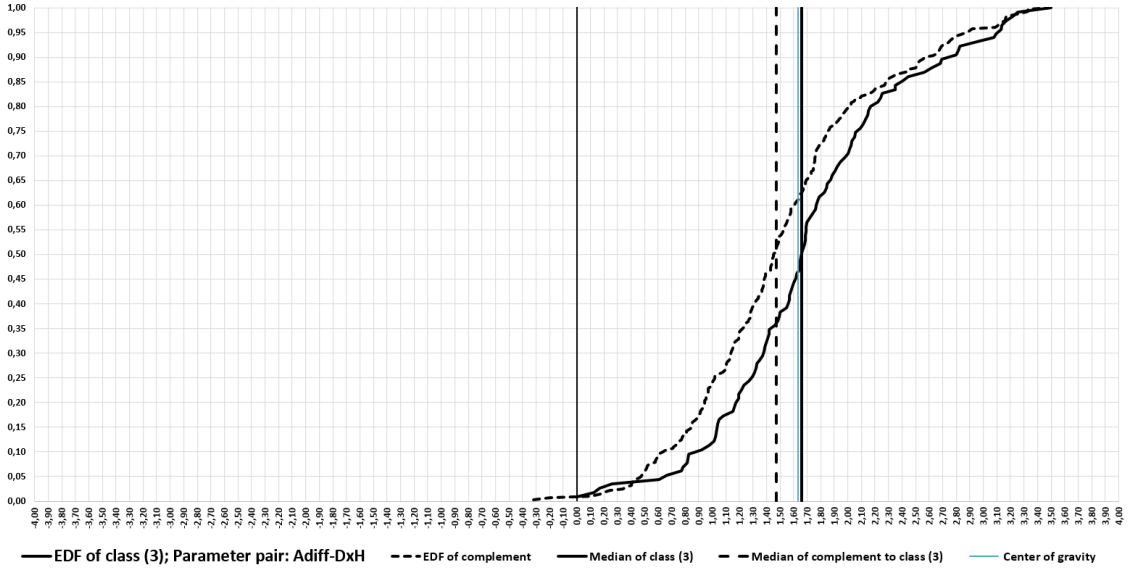


Figure 7.18: *The EDFs for the parameter pair $A_{diff} - D_xH$ for class 3. On the right side, there is EDF of class 3, and on the left side, there is complement EDF. The oriented area has a size of 0.193 and is the biggest one for class 3, but the smallest of all classes.*

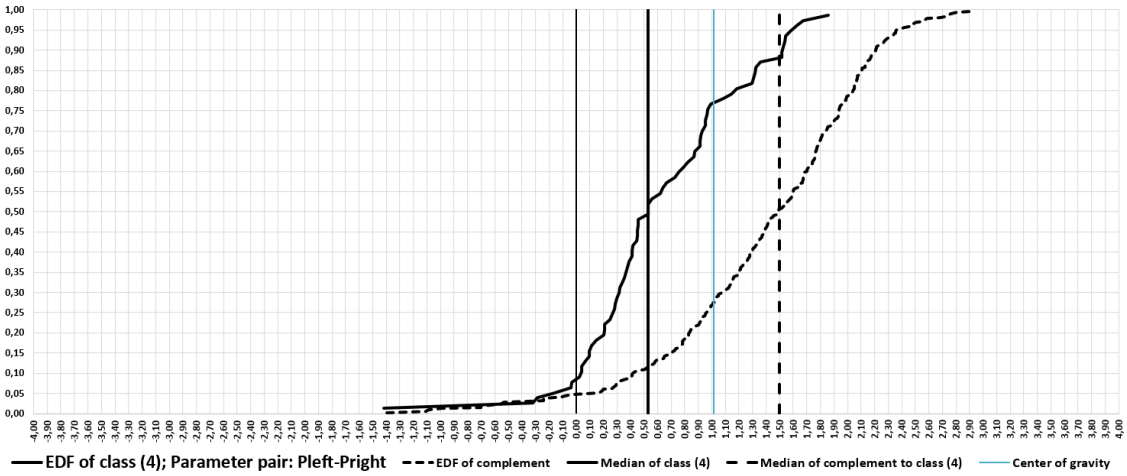


Figure 7.19: *The EDFs for the parameter pair $P_{left} - P_{right}$ for class 4. On the left side, there is EDF of class 4, and on the right side, there is complement EDF. The oriented area has a size of 0.764 and is the biggest one for class 4.*

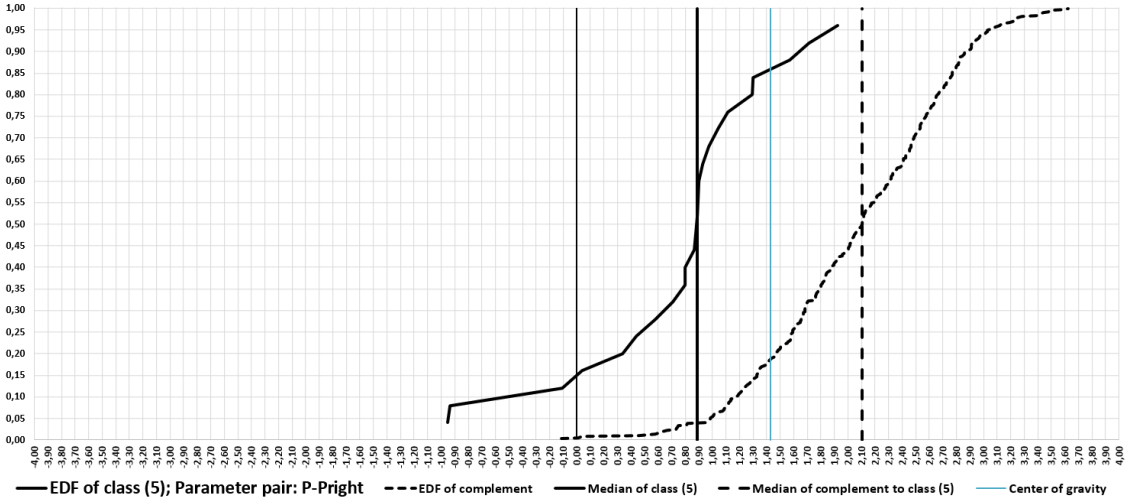


Figure 7.20: *The EDFs for the parameter pair $P - P_{right}$ for class 5. On the left side, there is EDF of class 5, and on the right side, there is complement EDF. The oriented area has a size of 1.262 and is the biggest one for class 5.*

Validation

In the same way, as the validation was done for the previous method, the training set was classified according to the found parameter pairs and the centers of gravity. Following table 7.11 shows the results of the classification.

Table 7.11: *Classification of training data set (396 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage	cumulative percentage
0	157	39.6%	39.6%
1	196	49.5%	89.1%
2	26	6.6%	95.7%
3	8	2.0%	97.7%
4	1	0.3%	98.0%
not classified	8	2.0%	100%

As can be seen in the table, the results are very similar to the previous method using median even though different parameter pairs were selected as the reference ones. There is a slightly better success rate for the difference value 1, but there is a higher number for difference values 3 and 4.

Results

Following table 7.12 shows the results of the testing set. Table 7.13 then shows the results of the classification of the whole corpus no. 692 (combined training and test results).

Table 7.12: *Classification of testing data set (296 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage	cumulative percentage
0	111	37.5%	37.5%
1	137	46.3%	83.8%
2	30	10.1%	93.9%
3	5	1.7%	95.6%
4	0	0.0%	95.6%
not classified	13	4.4%	100%

Table 7.13: *Classification of complete data corpus no. 692 (692 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage	cumulative percentage
0	268	38.7%	38.7%
1	333	48.1%	86.9%
2	56	8.1%	94.9%
3	13	1.9%	96.8%
4	1	0.1%	97.0%
not classified	21	3.0%	100%

As expected, the testing results were a little bit less successful than the results of the training set. The complete results are very similar to the previous method using medians. The detailed results can be seen in the following table 7.14 and the figure 7.21.

Table 7.14: *Classification of complete data corpus no. 692 (692 recordings). The first column contains rating differences, which is an absolute difference between the ENT expert rating and computed class.*

Rating difference	number of recordings	percentage
-4	0	0.0%
-3	6	0.9%
-2	34	4.9%
-1	126	18.2%
0	268	38.7%
1	207	29.9%
2	22	3.2%
3	7	1.0%
4	1	0.1%
not classified	21	3.0%

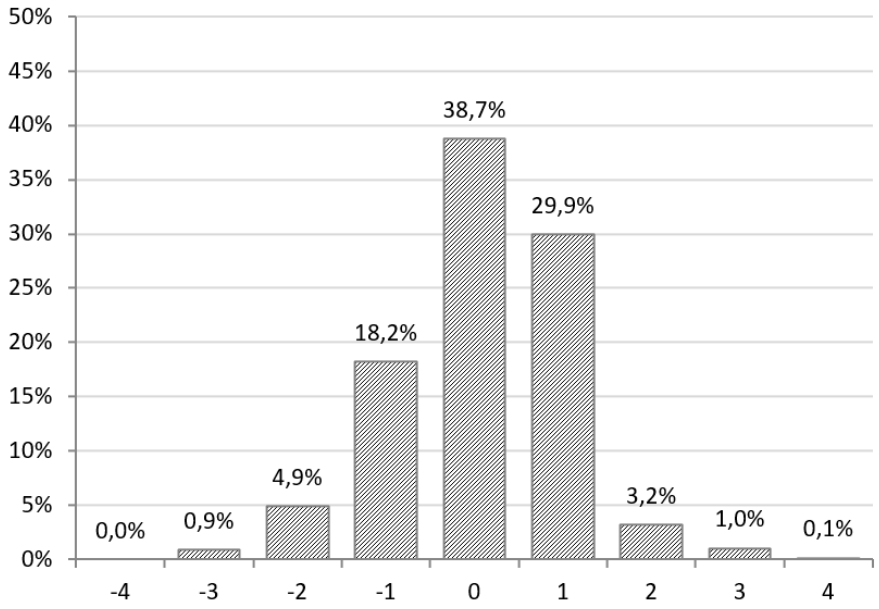


Figure 7.21: *Detailed distribution of the results of the correlation classification using EDFs and center of gravity of areas. Displayed data are the differences between the classification results and the rating of the ENT expert.*

7.3.5 Conclusion

The mentioned methods using the prior rating of vocal cords from the ENT expert show an interesting approach toward to classification of the LHSV recordings. Both methods return similar results even though different approaches and also different parameter pairs were used. Using the EDF method shows slightly better results than the method using medians, but also contains a single case of classification which is completely opposite for the rating and more LHSV recordings could not be classified.

The classification was done similarly, but the method of computing threshold values was different (average of medians in the first case, x-axis value of the center of gravity point in the second case). From figures 7.16 – 7.20, we can see that the center of gravity values are always between the median values and not too far from the median average. Although some of the parameter pairs were different, the results are very similar and we can consider both methods as validated by each other.

Table 7.15 presents a comparison of the differences, where columns represent the difference between median classification and ENT expert rating, and the rows contain differences between EDFs classification and the rating. As can be seen, the results match in the most of cases, in 410 of 692. Only in 14 cases, the classification differs by more than one class (recordings that have not been classified by one of the methods are not listed in this table).

Table 7.15: *Comparison of differences between classification results and the ENT expert rating. The results are very similar.*

Comparison		Differences in classification using medians								
		-4	-3	-2	-1	0	1	2	3	4
Differences in classification using EDFs	-4	0	0	0	0	0	0	0	0	0
	-3	0	2	3	0	0	0	0	0	0
	-2	0	6	11	12	4	0	0	0	0
	-1	0	0	4	81	35	4	0	0	0
	0	0	0	2	25	175	58	4	0	0
	1	0	0	0	0	45	128	31	0	0
	2	0	0	0	0	0	10	12	0	0
	3	0	0	0	0	0	0	4	1	0
	4	0	0	0	0	0	0	0	1	0

Using these methods based on found significant parameter pairs, an LHSV video can be classified with 85.3% probability into the same or adjacent class as rated by the ENT expert. The resulting class can be used as a scoring function. Classes 1–2 mean that there is probably no issue with the vocal cords, classes 3–4 should raise a warning, and class 5 means that there is something wrong. The classification result is a scoring function to help the doctor determine if there should be more attention given to a specific patient.

It should be noted that all resulting parameter pairs are related to symmetry, this fact proved the importance of the finding axis and comparing the behavior of the vocal cord on the left side and right side.

7.3.6 Case Studies

This section contains the results of selected case studies listed in section 6.3.3.

Case 1 – Healthy symmetrical vocal cords

Strong correlation relationships for the healthy symmetrical vocal cords in case 1 (see pic. 6.4) are shown in table 7.16. In this case, no parameter pair correlation relationship is broken and no issue is detected by the first method.

Table 7.16: *Parameter pairs selected in the first method using correlation relationships. All relationships are strong.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.999
$A - A_{right}$	0.999
$A - P$	0.938
$P - P_{left}$	0.999
$P - P_{right}$	0.999
$D_xS - D_xH$	0.960
$D_yS - D_yH$	0.996

Table 7.17 shows the results of the parameter pairs correlation relationships for the classification, table 7.18 the values for parameter pairs used for classification by EDFs, and table 7.19 contains the ENT expert rating and the results of classification.

Table 7.17: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.995	-2.984	yes
2	$P - P_{right}$	0.999	3.625	yes
3	$A - A_{right}$	0.999	3.781	no
4	$D_yS - P$	-0.991	-2.396	no
5	$P - P_{right}$	0.999	3.625	no

Table 7.18: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.995	-2.984	yes
2	$P_{left} - P_{right}$	0.994	2.933	yes
3	$A_{diff} - D_xH$	0.693	0.855	no
4	$P_{left} - P_{right}$	0.994	2.933	no
5	$P - P_{right}$	0.999	3.625	no

Table 7.19: Classification results

ENT expert	1
Median classification	2
EDFs classification	2

The computed classes are worse, but only about one acceptable mark, which is acceptable. This difference is caused by belonging to classes 1 and 2, the result is 2 in such case.

Case 2 – Non-symmetrical vocal cord with carcinoma

Lower values of correlation relationships for non-symmetrical vocal cords (see pic. 6.7) are displayed in table 7.20. Parameter pair $A - P$ has the lowest value, but no significant issue is detected by the first method.

Table 7.20: *Parameter pairs selected in the first method using correlation relationships. All relationships are strong.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.927
$A - A_{right}$	0.993
$A - P$	0.866
$P - P_{left}$	0.971
$P - P_{right}$	0.982
$D_xS - D_xH$	0.999
$D_S - D_H$	0.970

Table 7.21 shows the results of the parameter pairs correlation relationships for the classification, table 7.22 the values for parameter pairs used for classification by EDFs, and table 7.23 contains the ENT expert rating and the results of classification.

Table 7.21: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.533	-0.594	no
2	$P - P_{right}$	0.982	2.364	yes
3	$A - A_{right}$	0.993	2.811	no
4	$D_yS - P$	-0.544	-0.609	yes
5	$P - P_{right}$	0.982	2.364	no

Table 7.22: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.533	-0.593	no
2	$P_{left} - P_{right}$	0.910	1.527	yes
3	$A_{diff} - D_xH$	0.992	2.784	yes
4	$P_{left} - P_{right}$	0.910	1.527	no
5	$P - P_{right}$	0.982	2.364	no

Table 7.23: Classification results

ENT expert	4
Median classification	3
EDFs classification	3

The computed classes are better, but only about one mark, which is acceptable. This difference is caused by belonging to classes 2 and 4, resp. classes 2 and 3. There is an issue detected.

Case 3a – Non-symmetrical vocal cord with carcinoma

The correlation relationships for non-symmetrical vocal cords (see pic. 6.11) are displayed in table 7.24. Several parameter pairs have lower values indicating an issue.

Table 7.24: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.823
$A - A_{right}$	0.976
$A - P$	0.873
$P - P_{left}$	0.864
$P - P_{right}$	0.858
$D_xS - D_xH$	0.994
$D_yS - D_yH$	0.967

Table 7.25 shows the results of the parameter pairs correlation relationships for the classification, table 7.26 the values for parameter pairs used for classification by EDFs, and table 7.27 contains the ENT expert rating and the results of classification.

Table 7.25: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.758	-0.992	no
2	$P - P_{right}$	0.858	1.287	no
3	$A - A_{right}$	0.976	2.207	no
4	$D_yS - P$	-0.917	-1.573	no
5	$P - P_{right}$	0.858	1.287	yes

Table 7.26: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.758	-0.992	no
2	$P_{left} - P_{right}$	0.485	0.529	no
3	$A_{diff} - D_xH$	0.985	2.447	yes
4	$P_{left} - P_{right}$	0.485	0.529	yes
5	$P - P_{right}$	0.858	1.287	yes

Table 7.27: Classification results

ENT expert	4
Median classification	5
EDFs classification	4

The computed class for the first method is worse by one mark, and the second method matches with the ENT expert rating. Classes 4 and 5 mean that there is a severe issue, classification is correct.

Case 3b – Non-symmetrical vocal cord with carcinoma after microsurgery

The correlation relationships for non-symmetrical vocal cords (see pic. 6.14) are displayed in table 7.28. The relationships are stronger after microsurgery, only $A - P$ parameter pair has a lower value.

Table 7.28: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.957
$A - A_{right}$	0.982
$A - P$	0.865
$P - P_{left}$	0.951
$P - P_{right}$	0.962
$D_xS - D_xH$	0.988
$D_yS - D_yH$	0.976

Table 7.29 shows the results of the parameter pairs correlation relationships for the classification, table 7.30 the values for parameter pairs used for classification by EDFs, and table 7.31 contains the ENT expert rating and the results of classification.

Table 7.29: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.934	-1.685	yes
2	$P - P_{right}$	0.962	1.979	no
3	$A - A_{right}$	0.982	2.345	no
4	$D_yS - P$	-0.981	-2.329	no
5	$P - P_{right}$	0.962	1.979	no

Table 7.30: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.934	-1.685	yes
2	$P_{left} - P_{right}$	0.831	1.192	no
3	$A_{diff} - D_xH$	0.967	2.044	yes
4	$P_{left} - P_{right}$	0.831	1.192	no
5	$P - P_{right}$	0.962	1.979	no

Table 7.31: Classification results

ENT expert	2
Median classification	1
EDFs classification	2

The computed classes are the same or only with 1 class difference. It can be seen that the microsurgery improved the behavior of the vocal cords which corresponds with the results.

Case 4a – Non-symmetrical vocal cord with polyps before microsurgery

The correlation relationships for non-symmetrical vocal cords (see pic. 6.18) are displayed in table 7.32. The relationships are not too strong indicating a possible issue.

Table 7.32: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.927
$A - A_{right}$	0.947
$A - P$	0.931
$P - P_{left}$	0.911
$P - P_{right}$	0.888
$D_xS - D_xH$	0.987
$D_yS - D_yH$	0.976

Table 7.33 shows the results of the parameter pairs correlation relationships for the classification, table 7.34 the values for parameter pairs used for classification by EDFs, and table 7.35 contains the ENT expert rating and the results of classification.

Table 7.33: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.476	-0.518	no
2	$P - P_{right}$	0.889	1.416	no
3	$A - A_{right}$	0.947	1.800	yes
4	$D_yS - P$	-0.322	-0.334	yes
5	$P - P_{right}$	0.889	1.416	yes

Table 7.34: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.476	-0.518	no
2	$P_{left} - P_{right}$	0.621	0.727	no
3	$A_{diff} - D_xH$	0.823	1.165	no
4	$P_{left} - P_{right}$	0.621	0.727	yes
5	$P - P_{right}$	0.889	1.416	yes

Table 7.35: Classification results

ENT expert	4
Median classification	4
EDFs classification	5

The computed class are the same or only with 1 class difference. It can be seen that the vocal cords are in bad shape before microsurgery.

Case 4b – Vocal cord after microsurgery of polyps

The correlation relationships for vocal cords after microsurgery (see pic. 6.21) are displayed in table 7.36. The relationships are stronger now indicating that vocal cords are symmetrical.

Table 7.36: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.997
$A - A_{right}$	0.996
$A - P$	0.937
$P - P_{left}$	0.994
$P - P_{right}$	0.993
$D_xS - D_xH$	0.980
$D_yS - D_yH$	0.996

Table 7.37 shows the results of the parameter pairs correlation relationships for the classification, table 7.38 the values for parameter pairs used for classification by EDFs, and table 7.39 contains the ENT expert rating and the results of classification.

Table 7.37: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.969	-2.083	yes
2	$P - P_{right}$	0.993	2.859	yes
3	$A - A_{right}$	0.996	3.114	no
4	$D_yS - P$	-0.968	-2.059	no
5	$P - P_{right}$	0.993	2.859	no

Table 7.38: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.969	-2.083	yes
2	$P_{left} - P_{right}$	0.975	2.175	yes
3	$A_{diff} - D_xH$	0.867	1.322	no
4	$P_{left} - P_{right}$	0.975	2.175	no
5	$P - P_{right}$	0.993	2.859	no

Table 7.39: Classification results

ENT expert	2
Median classification	2
EDFs classification	2

The computed classes are the same as the ENT rating. The microsurgery improved the behavior of the vocal cords, where no severe issue is detected now.

Case 5a – Non-symmetrical vocal cord with cyst before microsurgery

The correlation relationships for non-symmetrical vocal cords (see pic. 6.25) are displayed in table 7.40. The relationships look near value 1 and no relationship is broken. In this case, the issue is not detected because of similar progress of the A, A_{right}, and A_{left}. The P_{diff} parameter has minimal change.

Table 7.40: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.963
$A - A_{right}$	0.966
$A - P$	0.962
$P - P_{left}$	0.989
$P - P_{right}$	0.990
$D_xS - D_xH$	0.997
$D_yS - D_yH$	0.968

Table 7.41 shows the results of the parameter pairs correlation relationships for the classification, table 7.42 the values for parameter pairs used for classification by EDFs, and table 7.43 contains the ENT expert rating and the results of classification.

Table 7.41: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	0.475	0.516	no
2	$P - P_{right}$	0.990	2.646	yes
3	$A - A_{right}$	0.966	2.033	no
4	$D_yS - P$	0.585	0.670	yes
5	$P - P_{right}$	0.990	2.646	no

Table 7.42: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	0.475	0.516	no
2	$P_{left} - P_{right}$	0.958	1.927	yes
3	$A_{diff} - D_xH$	0.921	1.599	no
4	$P_{left} - P_{right}$	0.958	1.927	no
5	$P - P_{right}$	0.990	2.646	no

Table 7.43: Classification results

ENT expert	3
Median classification	3
EDFs classification	2

This recording was rated as 3 by the ENT expert, the median classification returned the same result, and the EDFs classification returned class 2. The correlation value after Fisher transformation for parameter pair $A_{diff} - D_xH$ does not fit in class 3 only by 0.033. The computed classes are the same or only with 1 class difference, which is acceptable. An issue is detected by the median classification.

Case 5b – Vocal cord after microsurgery of cyst

The correlation relationships for vocal cords after microsurgery (see pic. 6.28) are displayed in table 7.44. But the relationships are not stronger as expected now indicating that vocal cords are still not in a good condition.

Table 7.44: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.865
$A - A_{right}$	0.932
$A - P$	0.957
$P - P_{left}$	0.942
$P - P_{right}$	0.923
$D_xS - D_xH$	0.999
$D_yS - D_yH$	0.977

Table 7.45 shows the results of the parameter pairs correlation relationships for the classification, table 7.46 the values for parameter pairs used for classification by EDFs, and table 7.47 contains the ENT expert rating and the results of classification.

Table 7.45: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	0.502	0.553	no
2	$P - P_{right}$	0.923	1.609	no
3	$A - A_{right}$	0.932	1.674	yes
4	$D_yS - P$	0.866	1.317	yes
5	$P - P_{right}$	0.923	1.609	no

Table 7.46: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	0.502	2.553	no
2	$P_{left} - P_{right}$	0.741	0.953	no
3	$A_{diff} - D_xH$	0.811	1.130	no
4	$P_{left} - P_{right}$	0.741	0.953	yes
5	$P - P_{right}$	0.923	1.609	no

Table 7.47: Classification results

ENT expert	2
Median classification	4
EDFs classification	4

In this case, the classification results are different from the ENT expert rating. A bigger difference between the left and right areas and perimeters causes lower correlation values and worse ratings than before microsurgery. After revision of the LHSV recording, the bad condition of the vocal cords was confirmed, because the vocal folds are moving irregularly like after paresis. The vocal cords were not fully recovered after the surgery yet, it seems that the results found an incorrect ENT expert rating.

Case 5c – Vocal cord after microsurgery of cyst

The correlation relationships for vocal cords after microsurgery (see pic. 6.31) are displayed in table 7.48. The relationships are much stronger than in the previous case, vocal cords seem to be in better condition.

Table 7.48: *Parameter pairs selected in the first method using correlation relationships.*

parameter pairs that meet the set criteria	
parameter pair	correlation value
$A - A_{left}$	0.983
$A - A_{right}$	0.986
$A - P$	0.969
$P - P_{left}$	0.996
$P - P_{right}$	0.996
$D_xS - D_xH$	0.993
$D_yS - D_yH$	0.997

Table 7.49 shows the results of the parameter pairs correlation relationships for the classification, table 7.50 the values for parameter pairs used for classification by EDFs, and table 7.51 contains the ENT expert rating and the results of classification.

Table 7.49: Classes with the parameter pairs from correlation classification using medians.

Classes with the parameter pair where the median distance was greatest				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.983	-2.394	yes
2	$P - P_{right}$	0.996	3.123	yes
3	$A - A_{right}$	0.986	2.460	no
4	$D_yS - P$	-0.990	-2.625	no
5	$P - P_{right}$	0.996	3.123	no

Table 7.50: Classes with the parameter pairs from correlation classification using EDFs.

Classes with the parameter pairs with the greatest oriented areas between EDFs				
class	parameter pair	correlation value	Fisher transf.	in class
1	$D_yH - P_{right}$	-0.983	-2.394	yes
2	$P_{left} - P_{right}$	0.984	2.418	yes
3	$A_{diff} - D_xH$	0.744	0.958	no
4	$P_{left} - P_{right}$	0.984	2.418	no
5	$P - P_{right}$	0.996	3.123	no

Table 7.51: Classification results

ENT expert	1
Median classification	2
EDFs classification	2

The classification results are 2 while the ENT expert rating is one. The vocal cords' condition is better from previous cases because of the successful recovery after three years. In both classifications, classes 1 and 2 were assigned so the result is 2 and there is a difference with ENT expert rating, but only one class, still indicating good condition without further severe issues.

The following figures, 7.22 and 7.23, show a comparison of correlation values of parameter pairs used for median and EDFs classifications. For each class, the indication of the classification threshold is illustrated. It can be seen that the condition of the vocal cords worsened in the second case (5b) after microsurgery and increased in the third case (5c) after three years of convalescence. In some cases, the sign was the opposite in the third case showing a significant difference in behavior from previous cases.

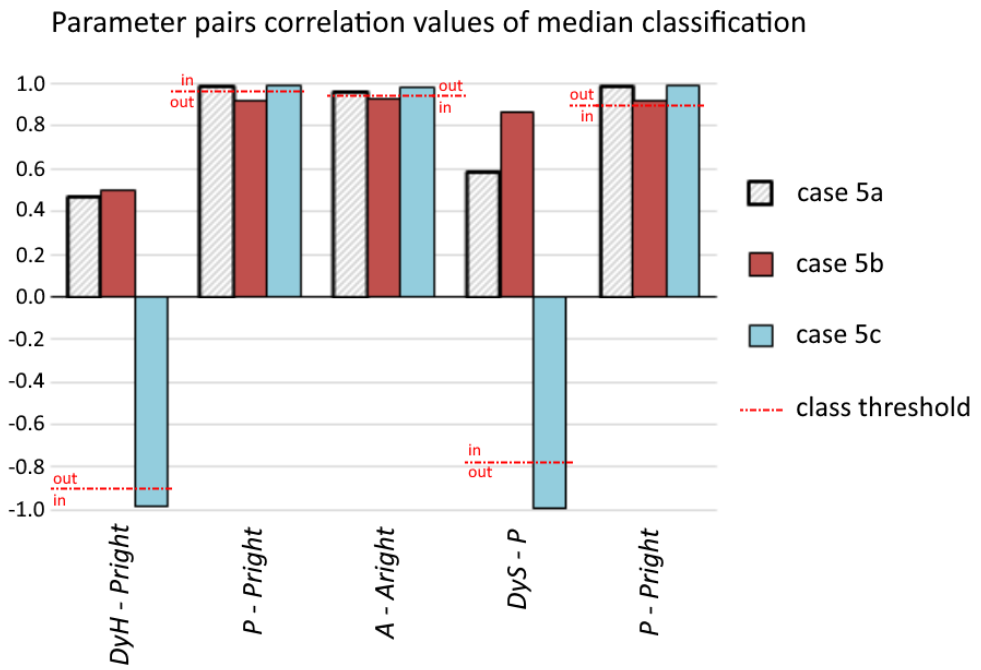


Figure 7.22: *Development of the correlation relationship values for cases 5a, 5b, and 5c with the indication of the threshold values of each class of median classification. Classes 1 and 4 show opposite signs in the third case as the correlation values were very different.*

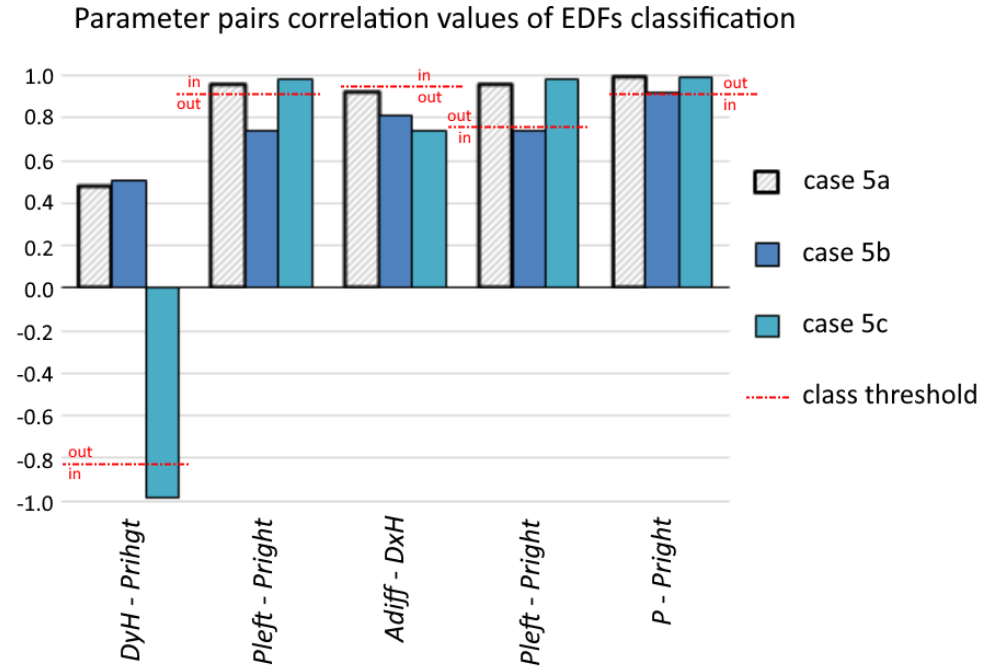


Figure 7.23: *Development of the correlation relationship values for cases 5a, 5b, and 5c with the indication of the threshold values of each class of EDFs classification. Class 1 shows the opposite sign in the third case as the correlation value was very different.*

8 Conclusion

This work summarizes the knowledge and results of the processing of laryngoscopy data obtained by the Laryngeal High-Speed Videoendoscopy (LHSV) system. The work is focused on the evaluation of the quality of the vocal cords' behavior based on parameters computed from 2D image information contained in a sequence of laryngoscopic images.

To clarify the basis of the entire work, the first part of the thesis describes the anatomical structures of the vocal cords in connection with the physiology of voice production and images of LHSV sequences. The LHSV system HRES Endocam 5562, used for obtaining recordings for this work, has been described and presented in context with other investigative methods in the field of phoniatrics, functional or descriptive diagnostics, and compared with other available high-speed laryngoscopic systems. For testing methods of processing and evaluating the quality of vocal cords behavior, the anonymized data corpus of LHSV recordings was used, which contained 692 recordings of various diagnoses, provided by the ENT department of the University Hospital in Pilsen. Since the data corpus was not primarily created for the purpose of testing image processing methods, it consists of video recordings of different quality. All video recordings are also rated by an ENT expert from the point of view of the quality of the kinematics of the vocal cords to validate and compare final results.

As part of a comprehensive solution to the defined goals, the next part of the work, chapter 4, describes the issue of image data processing from the LHSV system to detect the vocal gap (glottis) in each frame of the recordings. This task was divided into two steps. The first step was to find the Region of Interest (ROI) in the heterogeneous structure of the image. The ROI is the part of the image with the vocal cords without as many surroundings as possible to avoid interference for further processing. For this purpose, two original methods were introduced, the first based on the thresholding of a difference image, and the second, which uses the principle of the brightness change of pixels and frequency analysis. The methods were trained on the smaller data corpus and detection success was tested on the full data corpus of 692 recordings. The thresholding method shows a success rate of 70.95%, the DFT method has a success rate of 85.69% in the case of the general variant and 90.17% (624 of 692 recordings) for the geometrized variant (8×8), which was used for further processing.

The second step, the detailed glottis segmentation within the ROI, is then described. Two different methods were designed, implemented, and tested here, the thresholding method and the cluster analysis method. In the case of the thresholding method, the MAX-MIN thresholding approach and other consecutive processes were used, in the case of the cluster analysis method, the K-means variant was selected. The MAX-MIN thresholding method was not successful enough for the lower quality images within the entire corpus of 692 recordings, the success detection was performed only in 20.2% of LHSV recordings with successfully determined ROI (126

out of 624 recordings). For the cluster analysis method, for the same assumptions and K-means with adaptive parameters implementation, the success rate of glottis estimation was achieved in 60.4% of cases (377 out of 624 recordings). According to the results, further processing uses the K-means cluster analysis method with adaptive parameters for glottis estimation.

The next part of the work, chapters 5 and 6, contains a solution to the description of the segmented glottis using selected parameters. Based on the analysis and visualization, it is possible to contribute to decisions regarding the quality of the vocal cords' behavior and to provide additional information contained in the LHSV video sequences but difficult to recognize it by mere observation. One of the basic characteristics of the so-called correct behavior of the vocal cords is the symmetry of the glottis. An original method for estimating the axis of symmetry of the glottis was created and validated on 233 LHSV videos with different types and degrees of pathological issues where the glottis was successfully detected. To assess the accuracy of the estimate of the glottis symmetry axis, the method for testing the conformity of two straight lines was used to compare the estimated axis of symmetry and the anatomical axis of the vocal cords, which was determined by the ENT expert for testing purposes. A conformity parameter was introduced and based on the evaluation of this parameter, 93.56% of axis estimation cases fall into the interval with axis rotation error lower than 5.7° or position estimation error lower than 5px. Besides the parameters of the area and perimeter of the segmented glottis, computed also separately for the left and right parts divided by the glottis symmetry axis (18 parameters in total), the position of the center of gravity of the area and the border of the glottis proved to be beneficial parameters. Movements of the center of gravity in the direction of the symmetry axis and the normal line direction contribute to the detection of asymmetric behavior of the vocal cords and thus significantly supplement the possibilities of interpreting the measured data. A total of 8 parameters were derived from the position of the center of gravity of the area and the border of the glottis. The relationships between all mentioned parameters have been mutually analyzed and the results were demonstrated in selected case studies. For the used corpus of 692 recordings, 26 basic parameters were calculated for each video recording, with other derived parameters, the complete set consisted of 94 parameters.

The last part of the thesis, chapter 7, deals with the issue of evaluation of the vocal cords' behavior based on the calculated values of an extensive set of parameters. This set was analyzed and narrowed down to 14 parameters of the area, perimeter, and center of gravity of the segmented glottis. The overall vocal cords evaluation approach was further based on the assumption that the selected parameters have a geometric basis and certain pairs of these parameters should show strong mutual correlations for healthy vocal cords. Violation of these correlation relationships is then an indicator with potential for diagnosis decision-making. All possible combinations of parameter pairs (91 pairs of 14 parameters) were analyzed by setting thresholds to identify ones with significant and exceptional correlation values. In this way, 7 parameter pairs were selected with the most significance for assessing the violation of the correlation relationships according to the established criteria.

Another direction of the research was the use of the above-mentioned findings regarding the correlations between the parameters of the segmented glottis together with the ENT expert rating where every LHSV recording was assigned to 1 of 5 classes according to the quality of the vocal cords closure. The motivation was to provide an independent evaluation of the vocal cords' behavior as additional information for diagnosis decision-making or use this system as a tool for warning in case of non-standard vocal cords' behavior. This evaluation process was based on the statistical methods using Fisher transformation of the correlation values and classification of whether an LHSV recording belongs to one of five classes. Two methods were designed and developed for this classification task, the first was based on comparing correlation values of a recording with the computed class and complement medians, and the second method determined the classification thresholds based on the center of gravity of the empirical distribution functions differences. The medians and thresholds were set according to the analysis of the training set containing 396 LHSV recordings with the ENT expert rating. Then complete corpus of 692 recordings was tested with the result of 85.3% agreement median classification with the ENT expert rating, and 86.9% in the case of EDFs classification when the accepted difference was ± 1 class number. In the majority of cases, the classification agreed or was more strict than the rating of the ENT expert which can be understood positively from the point of view of the warning system. The results were presented in several case studies with a comparison between the introduced methods.

The output of this work consists of methods for glottis segmentation of LHSV recordings, the set of parameters for visualization of the glottis parameters especially the ones related to the glottis symmetry, the set of specific parameter pairs with expected strong correlation relationships, and the classification methods returning a single value in the range of 1-5. All these results can be used for an implementation of a tool that can contribute to the decision-making process during diagnosis determination during the LHSV examination of vocal cords.

8.1 Achieved Goals

The goal of this work was to contribute to the improvement of voice examination and the decision-making process regarding the diagnosis of vocal cord diseases. The solution to the problem is based on the processing and evaluation of image information obtained by the descriptive laryngoscopic examination called Laryngeal High-Speed Videoendoscopy (LHSV). It includes the issue of image processing, specifically segmentation tasks, detection of the glottis area and its boundaries, description of the glottis shape using defined parameters, and solving a classification task for evaluating the quality of the dynamic behavior of the vocal cords. This evaluation of the dynamic behavior quality of the vocal cords is based on the calculated parameters of the glottis and professional expertise. This comprehensive goal was therefore divided into three objectives, which were fulfilled in the following way.

The first objective was to solve the problem of image segmentation and detection of

the region that delimits the glottis in each frame of the LHSV recording. Since the images obtained through LHSV contain various anatomical structures, disturbing visual artifacts and noise, and are in general of various quality, it was necessary to first solve the task of closer localization of the vocal cords to delimit the area for further image analysis, called Region of Interest (ROI). Anatomically, this is defined by the space between the anterior and posterior commissures and the range of moving vocal folds. Two fundamentally different methods were used and gradually modified to define the ROI. The first method is based on thresholding the difference image and using the Connected Component Labeling method, see [4], [21], [52], and the DFT method, which is based on the frequency analysis of brightness change in the image pixels. The DFT method was implemented in two versions, see [48]. Both methods are original in their particular implementation and according to known published ROI detection approaches. The next step is the detection of the glottis within the defined ROI. For this purpose, two methods were proposed and implemented too. The first, MAX-MIN Thresholding, belongs to the category of thresholding methods and is a sequence of consecutive processes, see [4], [21], [52], the second method is based on the principle of cluster analysis, K-means specifically, see [21], [50], [52]. Although thresholding methods belong to one large class of methods for glottis segmentation, the chosen sequence of processes for the MAX-MIN Thresholding method is original within the known published results. Similarly, the K-means cluster analysis method is well known in general but it is originally used for glottis segmentation.

The second objective was to analyze the kinematics of the vocal folds to extract information contained in the LHSV recordings, which is not normally observable. Because the symmetry of vocal cords is an important feature for correct functionality, a method of estimating the symmetry axis of the glottis was created. The method is original and is described, including a terminological discussion, in the publication [52]. The set of parameters derived from the area and perimeter of the segmented glottis was further extended by the parameters of the center of gravity position, specifically the center of gravity of the segmented glottis area and border, which extends the possibilities of vocal cord kinematics analysis. The importance of the center of gravity parameters is described and demonstrated in several case studies and also in [21], [52], [53]. The complete set of parameters was further analyzed and 14 parameters based on area size, perimeter length, and center of gravity position were selected for further use. The subsequent direction of the work was the statistical analysis of parameter pairs and correlations between them. This was based on the idea that the segmented glottis parameters have a geometric basis, i.e., in the normal state of healthy vocal cords, strong correlations between them were assumed. Violation of these correlation relationships can be a symptom with a diagnostic significance. This analysis of correlation relationships was demonstrated in selected case studies and has been published, see [55].

The third objective was the design and implementation of a classification method for evaluating the quality of vocal fold kinematics. The motivation was the method based on acoustic analysis that is currently used at the ENT clinic of University Hospital in Pilsen to evaluate the quality of the vocal cords' closure and assign

their behavior to one of five classification groups[13]. A similar approach was also chosen in this work to introduce original methods to classify LHSV recordings using selected parameter pairs and their correlation relationships. This classification task was solved by two methods using Fisher transformation for input correlation values and comparison with median values and center of gravity position of the area formed by empirical distribution functions (EDF) of trained classes. The ENT expert rating of the vocal cords was used here for method configuration. These methods were designed to help within vocal cords' examination to raise a warning in case of found irregularities or unexpected correlation values.

The individual methods designed and implemented within the three objectives of this work were tested on many LHSV recordings, the final data corpus contains 692 LHSV recordings including evaluation from ENT experts.

8.2 Author's Comment

The goal was to provide a fully automatic method without human interaction. During the study, I found automatic glottis segmentation more difficult than it looked before. Every LHSV recording is different and the quality also differs. The image processing was based mainly on the brightness difference between the glottis and surroundings but in case of incorrect illumination, the edges of the glottis are not easily detectable. It was found that processing a smaller area led to better results, thus I decided to preprocess all recordings by localizing the ROI first. The thresholding method based on subtraction of maximally open and maximally closed vocal cords works well but can be easily distracted by moving fluid or mucus or by light reflections. The second method using frequency analysis and DFT provided better results where the correct ROI was found in 90% of LHSV recordings.

The glottis segmentation itself was a little bit tricky, especially in the case of recordings containing very dark areas or blurriness. Because the minimum error thresholding method for threshold detection works well for high-quality recordings only, the method using cluster analysis was invented. This approach of classification based on pixel properties shows interesting results, but it was very difficult to find the proper weight configuration. The working configuration for several recordings was not working well for others. It led to the creation of a method of self-adapting parameter weights based on ROI size and histogram. All configurations were done heuristically and during extensive manual testing. Improving this area could provide even better results than the current 60% success rate.

Then I focused on the analysis of computed parameters. To have a complete set of parameters from all LHSV recordings, the glottis recognition was manually adjusted to be sure parameters are corresponding to the real glottis behavior. Various analyses of the parameter behavior were performed to see their development in time. Because the parameters are based on geometry and symmetry, I assumed that there should be possible to find strong relationships between some parameters during the

"normal" vocal cords' movement. Thus correlation values were computed and explored. The differences between healthy and non-healthy vocal cords were observed and described. To provide an easily understandable rating, the classification was then implemented based on the prior ENT expert rating of the vocal cords' behavior from LHSV recordings and also other examinations. These data were then used for the training of the statistical algorithm. The classification method provided surprisingly good results with a success rate of 87% when I accepted the ±1 difference from the expert rating.

All image processing methods to obtain glottis parameters were implemented in C# application, and the statistical analysis and correlation calculations were performed in spreadsheet software. To create a fully integrated tool to be used in a hospital, further development and method implementation would be required, but it was out of the scope of this work.

8.3 Future Work

Although the glottis segmentation results from LHSV images have been improved by the introduction of a new method based on cluster analysis, the success rate could be further improved. One of the possible directions could be a more detailed analysis of the parameters used in the cluster analysis method and their weight computation, another approach could be the inclusion of more information in this process, e.g. data from the frequency analysis of the ROI detection.

This work introduces many methods of image processing, parameter computation, and statistical analysis. All these findings could be used for an implementation of a complete tool usable in the hospital during the LHSV examination to provide potential warning in case of any irregularity is detected.

Further analysis could be also done by comparing and integrating results from other vocal cords examinations which could help with complex monitoring of the voice condition of the patients.

9 Resumé

This work summarizes the findings and results of the study dealing with the processing of Laryngeal High-Speed Videoendoscopy (LHSV) data and assessment of the quality of vocal cords' kinematics. The related anatomy structures were described together with examination methods of vocal cords used during a medical examination. Further analysis and image processing were performed on the corpus with anonymized data from the ENT department of the University Hospital in Pilsen taken by high-speed camera, which contained 692 LHSV recordings with various diagnoses and quality.

Chapter 4 contains method descriptions of image data processing of LHSV recordings. The first step was to find a Region of Interest (ROI) to select part of the image with the vocal cords without as many surroundings as possible to avoid interference for further processing. Two own methods were presented, the former one based on thresholding and the second one using frequency analysis. The rest of the section contains a description of two own methods dealing with actual glottis (vocal gap) detection. The first method was based on thresholding but was not successful enough for images with worse quality. Because of that, the second method was introduced using cluster analysis of pixels as objects, which led to better segmentation results.

Chapter 5 covers a topic about glottis symmetry, which is a very good indicator of vocal cords' behavior. The method for detecting the symmetry axis is described together with the method to assess the correctness of the automatic axis determination. Chapter 6 then describes a defined set of parameters computed from the segmented glottis shape. These parameters were calculated from every frame of all used LHSV recordings creating a huge file of data used for further analysis and evaluation of the voice creation quality.

The last part of this work deals with vocal cords' evaluation using the computed parameter values. It describes the way of development of the analyses performed on correlation relationships between parameters changing in time. The correlation relationships were computed for every possible parameter pair and the first method returned several parameter pairs with an expected strong relationship. Breaking such relationships can indicate a potential issue and can alert the examiner. The second described method used a prior expert rating of the vocal cords' behavior quality for the configuration of classification of the LHSV recordings into five classes. This classification uses medians and empiric distribution functions to evaluate the vocal cords by a number in the range 1–5. This result can also lead to an early warning for the doctor. The case studies describing the results are presented for both methods.

Described methods and processes can be used as a tool during vocal cords examination and help the doctor with the decision-making process about diagnosis in the early stage of disease when the problem is difficult to recognize by mere observation or other examinations.

Resumé

Tato práce shrnuje poznatky a výsledky studie zabývající se zpracováním dat laryngeální vysokorychlostní videoendoskopie (Laryngeal High-Speed Videoendoscopy – LHSV) a hodnocením kvality kinematiky hlasivek. Byly popsány související anatomické struktury spolu s vyšetřovacími metodami hlasivek používanými při lékařském vyšetření. Další analýza a zpracování obrazu byly provedeny nad anonymizovaným datovým korpusem pořízeným vysokorychlostní kamerou na ORL oddělení Fakultní nemocnice v Plzni, který obsahuje 692 LHSV záznamů v různé kvalitě s rozličnými diagnózami.

Kapitola 4 obsahuje popisy metod zpracování obrazových dat z LHSV záznamů. Prvním krokem bylo najít oblast zájmu (Region of Interest – ROI) obsahující pouze samotné hlasivky bez okolí, aby se zabránilo rušení pro další zpracování. Pro tento účel byly představeny dvě vlastní metody, první založená na prahování a druhá využívající frekvenční analýzy. Zbytek kapitoly obsahuje popis dvou vlastních metod zabývajících se samotnou detekcí hlasivkové štěrbiny (glottis). První metoda byla založena na prahování, ale nebyla dostatečně úspěšná pro záznamy v horší kvalitě. Z tohoto důvodu byla navržena druhá metoda využívající shlukovou analýzu pixelů jako objektů, která přinesla lepší výsledky segmentace.

Následující kapitola 5 pojednává o "symetrii hlasivek", což je jeden z indikátorů chování hlasivek. Je zde popsána metoda detekce osy symetrie včetně metody pro posouzení správnosti automatického určení osy. Kapitola 6 pak popisuje sadu parametrů vycházejících z tvaru segmentované hlasivkové štěrbiny. Tyto parametry byly vypočítány z každého snímku všech záznamů LHSV, čímž vznikl obrovský soubor dat sloužících k další analýze a hodnocení kvality tvorby hlasu.

Poslední část této práce se zabývá hodnocením hlasivek pomocí vypočtených hodnot parametrů. Popisuje provedené analýzy korelačních vztahů mezi průběhem hodnot zvolených parametrů. Pro každou dvojici parametrů byly vypočteny hodnoty korelace a pomocí první metody bylo nalezeno několik dvojic parametrů, u kterých se očekávala nejvyšší hodnota korelace v testovacích datech. Případné porušení těchto korelačních vazeb může naznačovat potenciální problém hlasivek a tím upozornit vyšetřujícího lékaře na možný problém. Druhá popsaná metoda využila expertní hodnocení kvality chování hlasivek pro nastavení parametrů ke klasifikaci LHSV záznamů do pěti tříd. Tato klasifikace využívá mediány a empirické distribuční funkce jednotlivých tříd k výslednému ohodnocení hlasivek v rozsahu 1-5. Tento výsledek může také vést ke včasnému varování vyšetřujícího lékaře. Pro obě metody jsou uvedeny kazuistiky popisující ukázkové případy.

Metody a postupy popsané v této práci mohou být použity jako pomocný nástroj při vyšetření hlasivek a poskytnout lékaři další informace k rozhodování o diagnostice v raném stadiu onemocnění, kdy jsou problémy obtížně rozpoznatelné pouhým pozorováním nebo jinými metodami.

Used Abbreviations

- A – Area, the area size of the glottis
- AC – Axis conformity, parameter to evaluate the conformity of the found Axis and axis defined by ENT expert
- AOD – analysis of non-standard oscillation in voice sound recording, examination method
- B – Blue color component of RGB
- CCL – Connected component labeling, a method for searching the largest continuous area
- CP – Center point, a center of gravity of the segmented glottis
- DFT – Discrete Fourier transformation, a mathematical method for the transformation of signal to a frequency spectrum
- EDF – Empirical distribution functions
- EGG – Electroglottography System, examination method
- ENT – Ear-Nose-Throat, Another name is Otorhinolaryngology abbreviated as ORL. A surgical subspecialty within medicine that deals with the surgical and medical management of conditions of the head and neck, including vocal cords
- G – Green color component of RGB
- LHSV – Laryngeal High-Speed Videoendoscopy, examination method of vocal cords using a camera with a high frame rate
- MDVA – Multi-Dimensional Voice Analysis, examination method
- MIC – Recording of sound pressure by a microphone during vocal phonation "i:"
- NBI – Narrow Band Imaging
- P – Perimeter, the length of the glottis border
- PAS – Phonatory Aerodynamic System, examination method

- PVG – Phonovibrogram
- R – Red color component of RGB
- RGB – Red, Green, Blue color scheme to specify a color
- ROI – Region of interest, part of the image containing the interesting object
- SCORE – Quality of the Glottis Closure, evaluation method
- SD – Standard deviation
- SPI – Soft phonation index, a ratio of harmonic frequency energy in the specific frequency ranges
- SPL – Sound Pressure Level, measured in decibels
- STRESS – Stress Test of vocal cords, examination method
- VHI – Voice Handicap Index, examination method
- VKG – Videokymography, examination method
- VRP – Voice Range Profile, examination method
- Y – Brightness value computed from RGB

Bibliography

[1] Jacobson B., Johnson A., Grywalski C., Silbergleit A., Jacobson G., Benninger M.: The Voice Handicap Index (VHI): Development and Validation. American Journal of Speech-Language Pathology. 6. 66-70. 1997.

[2] Švec J.: Studium mechanicko-akustických vlastností zdroje lidského hlasu, [Studies on the mechanic-acoustic properties of the human voice. Thesis. In Czech]. Palacký University, Faculty of Natural Sciences, Department of Experimental Physics, Olomouc, 1996.

[3] Vokřál J.: Akustické parametry chraplavosti, Doktorská disertační práce. Fakulta elektrotechnická, České vysoké učení technické v Praze, 1998.

[4] Ettler, T.: Analýza vysokorychlostního záznamu kmitání hlasivek, [Analysis of Vocal Cord Oscillations from High Speed Videolaryngoscopy Recordings. Diploma Thesis. In Czech]. University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering, Pilsen, 2012.

[5] Miranda G.A., Stylianou Y., Deliyski D. D., Godino-Llorente J. I., Bernordoni N. H.: Laryngeal Image Processing of Vocal Folds Motion, Applied Sciences 2020, 10, 1556. https://www.mdpi.com/2076-3417/10/5/1556.

[6] Gray, H.: Anatomy, published in 1918. Elsevier, 2008.

[7] Čihák R.: Anatomie 2. 1. vyd. ISBN 80-060-88. Avicenum Praha, 1988.

[8] Novák A.: Foniatrie a pedaudiologie II, Poruchy hlasu - základy fyziologie hlasu, diagnostika, léčba, reedukace a rehabilitace. UNITISK, Praha, 1996.

[9] Laryngoskopie. uLékaře.cz.
URL:<http://www.ulekare.cz/clanek/laryngoskopie-1031>, 13.4.2011.

[10] Titze I.R.: I.R. Titze, Principles of Voice Production. 2nd ed., National Center of Voice and Speech: Iowa City, IA, USA, 2000, ISBN: 0-87414-122-2, pp. 87-183.

[11] Kučera M., Frič M, Halíř M.: Praktický kurz hlasové rehabilitace a reedukace. ISBN 978-8025465929. Opočno, 2010.

[12] Multi-Dimensional Voice Program (MDVP), Operations manual, Kay Elemetrics Corp. 1995.

[13] Pešta J.,Slípka J., Nový P., Vávra F.: Hodnocení kvality závěru glottis. Otorinolaryngologie a foniatrie, 4 / 2010.

[14] Kurdík, M.: Zátěžové hlasové testy. Diplomová práce, ZČU v Plzni, FAV, Katedra informatiky a výpočetní techniky, Plzeň, 2013.

[15] Puchrová, L.: Aerodynamická vyšetření ve foniatrii. Bakalářská práce, ZČU v Plzni, FAV, Katedra informatiky a výpočetní techniky, Plzeň, 2016.

[16] Kroupa, L: Detekce nestandardního kmitu hlasivek. Diplomová práce, ZČU v Plzni, FAV, Katedra informatiky a výpočetní techniky, Plzeň, 2015.

[17] Kroupa, L., Vávra, F., Nový, P.: Statistic of Quasi-periodic Signal with Random Period - First Application on Vocal Cords Oscillation. 16th Conference on Applied Mathematics Aplimat 2017, Proceedings, ISBN 978-80-227-4650-2, Institute of Mathematics and Physics, Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, Bratislava, 2017.

[18] Lohscheller J., Eysholdt U., Toy H., Dollinger M.: Phonovibrography: mapping high-speed movies of vocal fold vibrations into 2-D diagrams for visualizing and analyzing the underlying laryngeal dynamics. IEEE Trans Med Imaging. 2008 Mar;27(3):300-9. doi: 10.1109/TMI.2007.903690. PMID: 18334426.

[19] Schenk F., Aichinger P., Roesner I., Urschler M.: Automatic high-speed video glottis segmentation using salient regions and 3D geodesic active contours. Annals of the BMVA. 1-15. 2015.

[20] Dedouch K., Švec J.G., Horáček J., Kršek P., Havlík R., Vokřál J.: Akustická analýza mužského vokálního traktu pro české samohlásky, [Acoustic Analysis of Czech Vowels in a Male Vocal Tract, in Czech], Sborník. 1. Česko-slovenský foniatrický kongres a XIV. Celostátní foniatrické dny Evy Sedláčkové, Brno, 11.-13. září 2003. Audio-Fon Centr, Brno, Czech Rep.: 60-63, 2003.

[21] Ettler T.: Detekce a hodnocení videozáznamu pohybu hlasivek z vysokorychlostní kamery, [Detection and Evalution of Glottis in High Speed Video Recording. Professional work for the state doctoral exam. In Czech]. University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering, Pilsen, 2017.

[22] Otsu N.: A threshold selection method from gray-level histograms, IEEE Trans. Sys., Man., Cyber. 9 (1): 62-66, 1979. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4310076.

[23] Kittler J., Illingworth, J.: Minimum error thresholding, SERC Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, U.K, 1986.

[24] Méndez A., Ismaili Alaoui E.M., García B., Ibn-Elhaj E., Ruiz I.: Glottal Space Segmentation from Motion Estimation and Gabor Filtering, 31st Annual International Conference of the IEEE EMBS. Minneapolis, Minnesota, USA, September 2-6, 2009. http://dx.doi.org/10.1109/IEMBS.2009.5332612.

[25] Gilles Degottex: Glottal source and vocal-tract separation. Signal and Image processing. Université Pierre et Marie Curie - Paris VI, 2010. English.

[26] Aghlmandi D., Faez K.: Automatic Segmentation of Glottal Space from Video Images Based on Mathematical Morphology and the Hough Transform, International Journal of Electrical and Computer Engineering (IJECE). ISSN: 2088-8708, 2012.

[27] Serra, J.: Image Analysis and Mathematical Morphology, Academic Press, New York, ISBN 0-12-637240-3, 1982.

[28] Palm C., Keysers D., Lehmann T., Spitzer K.: Gabor Filtering of Complex Hue/Saturation Images for Color Texture Classification. Proc JCIS 2000, Atlantic City, USA, pp. 45-49, 2000.

[29] Kass M., Witkin A.,Terzopoulos D.: Snakes: Active contour models. In First International Conference on Computer Vision, pages 259-268, London, June 1987.

[30] Allin S., Galeotti J., Stetten G., Dailey S. H.: Enhanced Snake Based Segmentation of Vocal Folds, 2004.

[31] Schenk F., Urscler M., Aigner C., Roesner I., Aichinger P., Bischof H.: Automatic glottis segmentation from laryngeal high-speed videos using 3D active contours. 2014.

[32] Miranda G.A., Llorente J.I.G., Velázquez L.M., García J.A.G.: An automatic method to detect and track the glottal gap from high-speed videoendoscopic images. BioMed Eng OnLine (2015) 14:100 DOI 10.1186/s12938-015-0096-3. 2015. https://doi.org/10.1186/s12938-015-0096-3.

[33] Lohscheller J., Toy H., Rosanowski F., Eysholdt U., Döllinger M.: Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. Medical Image Analysis. 2007, 11, 400-413. https://doi.org/10.1016/j.media.2007.04.005.

[34] A. Pinheiro A., Dajer M. E., Hachiya A., Montagnoli A. N., Tsuji D.: Graphical Evaluation of Vocal Fold Vibratory Patterns by High-Speed Videolaryngoscopy. J. Voice 2014, 28, 106-111. https://doi.org/10.1016/j.jvoice.2013.07.014 .

[35] Karakozoglou S. Z., Henrich N., D'Alessandro C., Stylianou Y., Automatic glottal segmentation using local-based active contours and application to glottovibrography. Speech Communication Volume 54, Issue 5, June 2012, Pages 641-654. https://doi.org/10.1016/j.specom.2011.07.010.

[36] Skalski A., Zielinki T., Deliyski D.: Analysis of vocal folds movement in high-speed videoendoscopy based on level set segmentation and image registration. In Proceedings of the International Conference on Signals and Electronic Systems (ICSES), Kraków, Poland, 2008; pp. 223-226. https://doi.org/10.1109/ICSES.2008.4673399.

[37] Blanco M., Chen X., Yan Y.: A Restricted, Adaptive Threshold Segmentation Approach for Processing High-Speed Image Sequences of the Glottis. Scientific Research, Engineering, 5, pp. 357-362, Published Online, October 2013. http://dx.doi.org/10.4236/eng.2013.510B072.

[38] Miranda G. A., Godino-Llorente J. I.: Glottal Gap tracking by a continuous background modeling using inpainting. Medical and biological engineering and computing, 55:2123-2141, 2017. https://doi.org/10.1007/s11517-017-1652-8.

[39] Ismaili Alaoui E. M., Mendez A., Ibn-Elhaj E., GarciaB.: Keyframes detection and analysis in vocal folds recordings using hierarchical motion techniques and texture information. In Proceedings of the 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7-10 November 2009; pp. 653-656. https://www.researchgate.net/publication/224114786. http://dx.doi.org/10.1109/ICIP.2009.5413745.

[40] Schenk F., Aichinger P., Roesner I., Urschler M.: Automatic high-speed video glottis segmentation using salient regions and 3D geodesic active contours. Annals of the BMVA Vol. 2015, No. 1 pp 1-15. 2015. `https://www.researchgate.net/publication/282731724_Automatic_-high-speed_video_glottis_segmentation_using_salient_regions_-and_3D_geodesic_active_contours`. www.bmva.org/annals/2015/2015-0003.pdf.

[41] Shapiro L., Stockman G.: Computer Vision, Chapter 3.4 Connected Components Labeling. Prentice Hall. pp. 69-73, 2002.

[42] KIPS Kay's Image Processing Software Documentation, Color High-Speed Video System (Model 9170), KIPS (Model 9181) [online]. KayPENTAX. [accessed 6 March 2009].

[43] Baierova Ch.: Frekvenční analýza kmitů hlasivkové štěrbiny, [Frequency Analysis of Vocal Cord Oscillations. Diploma Thesis. In Czech], University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering, Pilsen, 2018.

[44] Granqvist S.,Lindestad P.: A method of applying Fourier analysis to high-speed laryngoscopy. Journal of the Acoustical Society of America. 2001, 110, 6, pp. 3193-3197. https://doi.org/10.1121/1.1397321.

[45] Aichinger P., Roesner I., Schneider-Stickler B., Bigenzahn W., Feichster F., Fuchs A. K., Hagmüller M., Kubin G.: Spectral Analysis of Laryngeal High-Speed Videos: Case Studies on Diplophonic and Euphonic Phonation. 8th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA) pp. 81-84. 2013. http://dx.doi.org/10.13140/RG.2.2.23043.30245.

[46] Hlaváč, V., Šonka, M.: Image Processing, Analysis and Machine Vision, Chapman & Hall Computing, London, 1994.

[47] Sakakibara K. I., Imagawa H., Kimura M., Yokonishi H., Tayama N.: Modal Analysis of Vocal Fold Vibrations Using Laryngotopography. INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, 2010.

[48] Ettler T., Novy P.: Analysis of Laryngeal High-Speed Videoendoscopy Recordings - Roi Detection, Biomedical Signal Processing and Control, Volume 78, ISSN 1746-8094, 2022. https://doi.org/10.1016/j.bspc.2022.103854.

[49] Kittler J., Illingworth, J.: Minimum error thresholding, SERC Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, U.K, 1986.

[50] Ettler T. and Novy P.: Using Cluster Analysis for Image Processing in High Speed Video Laryngoscopy, 2020 International Conference on Applied Electronics (AE), 2020, pp. 1-6, doi: 10.23919/AE49394.2020.9232826.

[51] Tatiraju, S., Mehta, A.: Image Segmentation using k-means clustering, EM and Normalized Cuts. Department of EECS (2008): 1-7. 2008.

[52] Pešta, J., Slípka, J., Vohlídková. M., Ettler. T., Nový, P., Vávra, F.: Kinematika hlasivek - nové parametry hodnocení, [Vocal Cord Kinematics - New Evaluation Parameters. Journal Publication. In Czech]. Otorinolaryngologie a foniatrie, 65, c. 2, pp. 88-96, Praha, 2016.

[53] Ettler, T., Nový, P.: The parameters of the Center of gravity glottis during phonation (Parametry polohy těžiště hlasivkové štěrbiny v průběhu fonace). XVI. Česko-slovenský kongres mladých otorinolaryngologů, poster section, ISBN 978-80-87562-57-4, Česká lékařská společnost Jana Evangelisty Purkyně, Rožnov pod Radhoštěm, 2016.

[54] Hátle J., Likeš J.: Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974, pages 184, 269, 372

[55] Ettler T., Nový P.: Diagnostic meaning of correlation relationship. 19th Conference on Applied Mathematics Aplimat 2020, Proceedings, ISBN 978-80-227-4983-1, Institute of Mathematics and Physics, Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, Bratislava, 2020. http://evlm.stuba.sk/APLIMAT/indexe.htm, https://www.proceedings.com/53722.html.

[56] Rényi A.: Teorie pravděpodobnosti, ACADEMIA, Praha 1972, page 111.

Published Works

Ettler, T.: Analýza vysokorychlostního záznamu kmitání hlasivek, [Analysis of Vocal Cord Oscillations from High Speed Videolaryngoscopy Recordings. Diploma Thesis. In Czech]. University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering, Pilsen, 2012.

Ettler T.: Detekce a hodnocení videozáznamu pohybu hlasivek z vysokorychlostní kamery, [Detection and Evalution of Glottis in High Speed Video Recording. Professional work for the state doctoral exam. In Czech]. University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering, Pilsen, 2017.

Pešta, J., Slípka, J., Vohlídková. M., Ettler. T., Nový, P., Vávra, F.: Kinematika hlasivek - nové parametry hodnocení, [Vocal Cord Kinematics - New Evaluation Parameters. Journal Publication. In Czech]. Otorinolaryngologie a foniatrie, 65, c. 2, pp. 88-96, Praha, 2016.

Ettler, T., Nový, P.: The parameters of the Center of gravity glottis during phonation (Parametry polohy těžiště hlasivkové štěrbiny v průběhu fonace). XVI. Československý kongres mladých otorinolaryngologů, poster section, ISBN 978-80-87562-57-4, Česká lékařská společnost Jana Evangelisty Purkyně, Rožnov pod Radhoštěm, 2016.

Ettler T., Nový P.: Diagnostic meaning of correlation relationship. 19th Conference on Applied Mathematics Aplimat 2020, Proceedings, ISBN 978-80-227-4983-1, Institute of Mathematics and Physics, Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, Bratislava, 2020.
http://evlm.stuba.sk/APLIMAT/indexe.htm,
https://www.proceedings.com/53722.html.

Ettler T. and Novy P.: Using Cluster Analysis for Image Processing in High Speed Video Laryngoscopy, 2020 International Conference on Applied Electronics (AE), 2020, pp. 1-6, doi: 10.23919/AE49394.2020.9232826.

Ettler T., Novy P.: Analysis of Laryngeal High-Speed Videoendoscopy Recordings - Roi Detection, Biomedical Signal Processing and Control, Volume 78, ISSN 1746-8094, 2022. https://doi.org/10.1016/j.bspc.2022.103854.