

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Diplomová práce

Segmentace snímků hlasivkové štěrbiny

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta aplikovaných věd
Akademický rok: 2022/2023

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Lukáš RYPL**
Osobní číslo: **A21N0032P**
Studijní program: **N3902 Inženýrská informatika**
Studijní obor: **Medicínská informatika**
Téma práce: **Segmentace snímků hlasivkové štěrbiny**
Zadávající katedra: **Katedra informatiky a výpočetní techniky**

Zásady pro vypracování

1. Popis a analýza vlastností videozáznamu hlasivek vysokorychlostním laryngoskopem (LHSV), přehledová studie metod pro detekci a segmentaci hlasivkové štěrbiny (glottis).
2. Testování vybraných metod používaných pro detekci glottis v záznamech LHSV. Na základě výsledků přehledové studie a testů vybraných metod bude proveden návrh vlastní metody, např. s využitím principu videokymografie.
3. Návrh a realizace programového vybavení pro zpracování videozáznamů pořízených LHSV, testování jednotlivých metod detekce glottis, archivaci a exportování výsledků pro navazující využití.
4. Podrobná analýza dosažených výsledků pro zvolený soubor záznamů z pohledu použitých metod zpracování a detekce glottis.

Rozsah diplomové práce: **doporuč. 50 s. původního textu**
Rozsah grafických prací: **dle potřeby**
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

dodá vedoucí diplomové práce

Vedoucí diplomové práce: **Ing. Pavel Nový, Ph.D.**
Katedra informatiky a výpočetní techniky

Datum zadání diplomové práce: **9. září 2022**
Termín odevzdání diplomové práce: **18. května 2023**

L.S.

Doc. Ing. Miloš Železný, Ph.D.
děkan

Doc. Ing. Přemysl Brada, MSc., Ph.D.
vedoucí katedry

V Plzni dne 11. října 2022

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 21. června 2023

Lukáš Ryppl

Poděkování

Rád bych poděkoval svému vedoucímu Ing. Pavlu Novému za ochotu, kterou projevil v rámci naší spolupráce, za všechny užitečné rady, nápady a především čas, který mi během tvorby této práce věnoval.

Abstract

This thesis deals with the design of a method for glottis segmentation of video recordings made by high-speed laryngoscopy (LHSV). In contrast to commonly used approaches, the proposed method focuses on a different approach. Glottis segmentation is performed on images that show the evolution of a single line of the video recording over time. These images are called videokymograms. The proposed method offers two variations. The first one works with whole videokymograms, while the second one focuses only on the videokymogram containing exactly one sample vocal cord oscillation.

Both methods proposed and implemented in this work are able to successfully segment the glottis in good quality video recordings. However, they fail when using automatic segmentation methods for LHSV video recordings with frequent glare (overexposure). This is because the assumptions of these methods on the distribution of luminances in the frame, i.e. on the shape of the histograms of the occurrence of each luminance in the frame, are violated.

Abstrakt

Tato diplomová práce se zabývá návrhem metody pro segmentaci hlasivkové štěrbině videozáznamu pořízeného vysokorychlostní laryngoskopií (LHSV). Na rozdíl od běžně používaných přístupů se navrhaná metoda zaměřuje na přístup odlišný. Segmentace hlasivkové štěrbině je prováděna na snímcích, které ukazují vývoj jednoho řádku videozáznamu v čase. Tyto snímky jsou nazývány videokymogramy. Navržená metoda nabízí dvě varianty. První pracuje s celými videokymogramy, zatímco druhá se zaměřuje pouze na videokymogram obsahující přesně jeden vzorový kmit hlasivek.

Obě metody, které byly navrženy a implementovány v rámci této práce, jsou schopny úspěšně segmentovat hlasivkovou štěrbinu ve videozáznamech dobré kvality. Při použití pro videozáznamy LHSV s častými odlesky (přesvícením) testované automatické prahovací metody selhávají. Je to tím, že jsou porušeny předpoklady těchto metod na rozložení jasů ve snímku, tj. na tvaru histogramů výskytu jednotlivých jasů ve snímku.

Obsah

1	Úvod	9
2	Hlasivky a poruchy hlasu	11
2.1	Poškození hlasivek	11
3	Vyšetřovací metody	14
3.1	Optické metody	14
3.1.1	Laryngoskopie	14
3.1.2	Vysokorychlostní laryngoskopie	15
3.2	Kvalita videozáznamu	16
3.3	Videokymografie	18
3.3.1	Videokymogram	18
3.3.2	Tvorba videokymogramu z videozáznamu LHSV	19
4	Metody používané pro detekci hlasivkové štěrbiny	20
4.1	Detekce ROI	20
4.2	Segmentace hlasivkové štěrbiny	21
4.2.1	Prahování	22
4.2.2	Narůstání oblastí	22
4.2.3	Přístupy pomocí strojového učení	23
5	Teoretický rozbor použitých technik	24
5.1	Zavedené značení	24
5.2	Metody segmentace pomocí analýzy histogramu	25
5.2.1	Metoda optimálního prahu (Otsu)	26
5.2.2	Kittlerova metoda nejmenší chyby	27
6	Navržené metody segmentace glottis	29
6.1	Vstupní data	29
6.2	Hlasivková frekvence	29
6.2.1	Určení hlasivkové frekvence	30
6.3	Metody aplikované na videokymogramy	34
6.4	Segmentace videokymogramu	35
6.5	Segmentace videokymogramu s využitím základní hlasivkové frekvence	36
6.5.1	Algoritmus založený na hlasivkové frekvenci	36

6.6	Postprocessing	37
6.6.1	Postup pomocí narůstání oblasti	38
6.6.2	Postup pomocí detekce ploch	38
6.6.3	Postup pro hlasivky mimo střed	39
6.7	Zpětná rekonstrukce	40
7	Program Kymogram analyzer	41
7.1	GUI	41
7.1.1	Uvítací obrazovka	41
7.1.2	Hlavní okno	41
7.2	Algoritmy pro segmentaci glottis	43
7.3	Parametry určené u štěrbině	47
7.4	Uložení výsledného videa	49
8	Výsledky práce	50
8.1	Testovací data	50
8.2	Výsledky navržených metod	51
8.2.1	Výsledky segmentace videokymogramu	51
8.2.2	Výsledky segmentace videokymogramu pomocí základní hlasivkové frekvence	56
8.3	Celkové výsledky	61
9	Závěr	63
	Literatura	65
	A Přehled zkratk	68
	B HRES ENDOCAM 5562	69
	C Uživatelská dokumentace	75
	D Obsah elektronické přílohy	76

1 Úvod

Mluvení je základem lidské komunikace. Hlas je tvořen uvnitř dýchací soustavy, avšak nejvíce se na tvorbě hlasu podílejí hlasivky, orgán uvnitř hrtanu. Jakožto každá část lidského těla, i hlasivky jsou náchylné na nejrůznější onemocnění a při jejich disfunkci je výrazně snížena životní úroveň postiženého pacienta. Pro diagnostiku poškození hlasu je využívána řada diagnostických metod, které jsou založené např. na akustické analýze lidského hlasu nebo analýze obrazové informace, která bývá získána některou z optických diagnostických metod. Tato práce se zabývá zpracováním obrazové informace v podobě záznamů videosekvencí LHSV (Laryngeal High-Speed Videendoscopy), tj. záznamů vysokorychlostní laryngoskopie. Opakovací frekvence u tohoto typu vyšetření bývá v rozsahu 2000 až 4000 snímků za sekundu [FPS], snímky obsahují pohled na hlasivky v jednotlivých fázích jejich otevírání a uzavírání. Tvar a změna hlasivkové štěrbiny (glottis), která vzniká během fonace mezi levou a pravou hlasivkovou řasou, v sobě nese důležité informace o dynamickém chování hlasivek. Pro včasnou diagnostiku poruchy hlasu je pak důležité sledování a hodnocení tohoto dynamického chování hlasivek, které lze popsat na základě řady geometrických parametrů. Proto je velmi důležité, detekovat přesně glottis v každém snímku videosekvence LHSV. Snímky videosekvencí LHSV jsou obecně zatíženy pestrým souborem poruch, které snižují kvalitu záznamů a obecně komplikují úlohu detekce glottis. Na téma detekce glottis ve videozáznamech LHSV byla publikována celá řada metod a přístupů, které jsou stručně přehledově uvedeny v této diplomové práci.

Tato práce se zaměřuje na detekci glottis využitím metod automatické segmentace na speciální případ uspořádání obrazové informace do tzv. videokymogramu. Tento přístup je odlišný od známých přístupů, kdy se pracuje s klasickým 2D zobrazením snímků uspořádaných v čase. Pro segmentaci videokymogramu jsou použity dvě „klasické“ automatické prahovací techniky, tzv. metoda Otsu a Kittler, které vychází z vlastností relativního histogramu videokymogramu pro definovaný počet snímků. Tyto dva přístupy jsou dále upraveny o odhad základní hlasivkové frekvence F_0 , která je odvozena ze změn jasů v obrazovém bodě, který odpovídá anatomické struktuře hlasivek, a o odhad osy symetrie hlasivek, která je využita v „postprocessingu“ segmentovaných videokymogramů.

Pro testování výsledků byl použit dataset videosekvencí LHSV, který je podmnožinou datového korpusu „692“, obsahujícího 692 anonymizova-

ných videozáznamů LHSV ze systému HRES ENDOCAM 5562, ORL klinika FN Plzeň. Dataset obsahuje výběr takových videosekvencí, kdy byla glottis úspěšně detekována jinými metodami, např. shlukovou analýzou (K-means). V analytické části práce je provedena analýza výsledků, zejména jsou hledány příčiny neúspěšné segmentace. Výsledky jsou prezentovány na několika kazuistikách.

2 Hlasivky a poruchy hlasu

Hlasivky jsou orgán nacházející se v krku. Jsou především zodpovědné za produkci zvuku při mluvení nebo zpěvu (tzv. fonace). Základní strukturu hlasivek tvoří dvě přepážky z měkké tkáně, které se nazývají hlasové vazy, ale v praxi jsou běžně nazývány hlasivkami.

Hlasivky se nacházejí uvnitř hrtanu (lat. larynx), konkrétně na jeho zadní stěně, a jsou chráněny chrupavčitými oblouky. Hlasivky jsou tvořeny různými vrstvami tkání, které se postupně stávají pevnějšími směrem k povrchu. Vnitřní vrstva tvoří sliznici, dále je pak vrstva svaloviny a na povrchu jsou kryty tenkou vrstvou epitelu. Umístění hlasivek lze vidět na řezu na Obrázku 2.1.

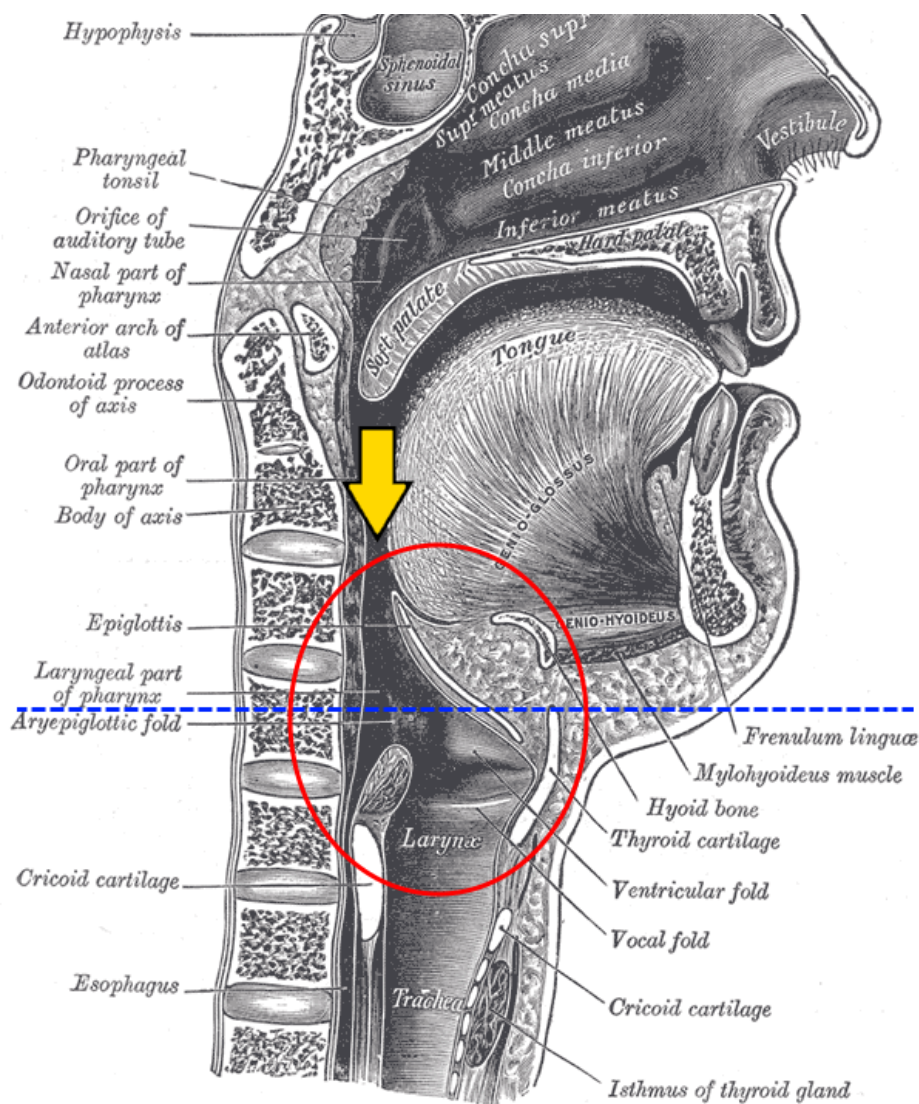
Svalovina hlasivek se skládá ze dvou hlavních svalů, které při normální funkčnosti umožňují nejrůznější úpravy tónu a výšky hlasu. Při normálním dýchání hlasivky spočívají v otevřené poloze, kdy vzduch volně prochází hrtanem z a do plic. Při fonaci se však hlasivky napínají a zužují tím průchod pro vzduch. To způsobuje vibrace a zapříčiňuje vznik zvuku - hlasu. Tón a kvalita zvuku jsou ovlivněny především napětím a polohou hlasivek, vliv mají ale i další faktory jako například rezonance v ústní a nosní dutině.

Tato práce se bude zabývat především hlasivkovou štěrbinou a její lokalizací. Hlasivková štěrbina (lat. glottis) je označení pro již zmíněný otevřený prostor mezi hlasivkami v hrtanu. Jedná se o úzký průchod pro vzduch, který vzniká mezi vnitřními hranami hlasivek. Hlasivková štěrbina je klíčová pro vznik zvuku. Všechny výše popsané útvary tvořící hlasivky lze vidět shora na Obrázku 2.2. Obrázek odpovídá pohledu poskytnutému laryngoskopickou kamerou při vyšetření hlasivek a je využit ve vstupních datech pro tuto diplomovou práci.

Velikost a tvar hlasivkové štěrbině se obvykle liší v závislosti na výšce a intenzitě produkovaného zvuku. Při produkci vysokých tónů se glottis zúží, dojde ke zvýšení napětí hlasivek a tím je zvýšena i frekvence vibrací. Při produkci nízkých tónů se naopak glottis rozšiřuje, je sníženo napětí hlasivek a frekvence vibrací.

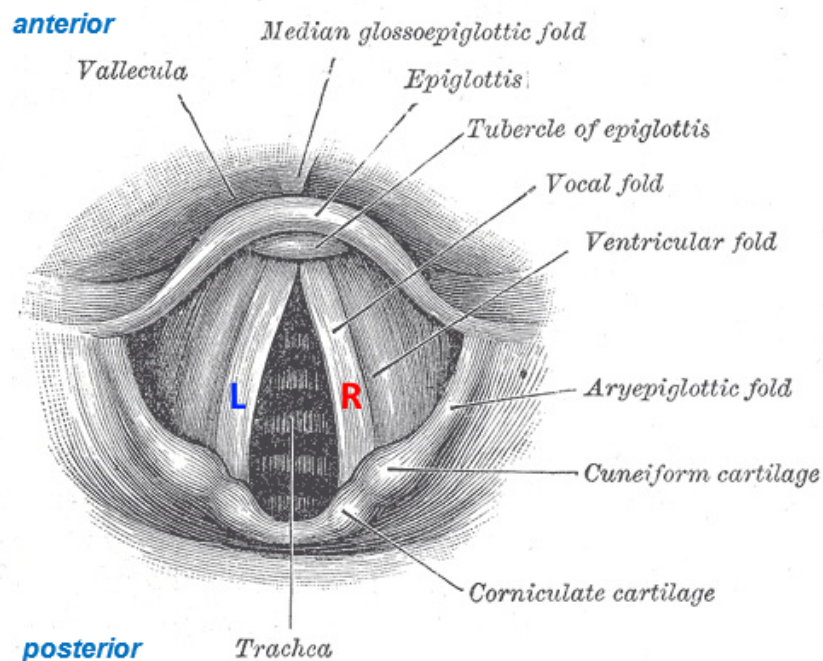
2.1 Poškození hlasivek

V průběhu života může samozřejmě dojít k poškození hlasivek. Poškození může být způsobeno především ze dvou důvodů - patologickou změnou na hlasivkách nebo funkční změnou [4, 7]. Pro účely této práce jsou funkční



Obrázek 2.1: Řez dýchacích cest. Sledovaná oblast označena červeně, modře je znázorněna rovina, kterou sledujeme pomocí kamery umístěné v místě šipky [8, 10].

změny (jako například únava po přemáhání) nepodstatné, neboť nejsou způsobeny viditelnými změnami na hlasivce. Patologické jevy na druhou stranu na hlasivce viditelné jsou a lze je tudíž rozpoznat pomocí počítačového vidění.



Obrázek 2.2: Pohled na hlasivky pomocí laryngoskopu (kamery) [8, 10].

Mezi nejobvyklejší patologické změny způsobující zhoršení hlasu patří:

- polyp - výrůstek na sliznici vyplněný vazivem
- uzlík - vazovité ztluštění namáhaného místa na hlasivkách
- cysta - zapouzdřená dutina na tkáni často vyplněná tekutinou
- obrna hlasivek - pohyb hlasivek je zajišťován zvratným nervem a při jeho poškození je pohyblivost omezena
- nádory - zhoubné i nezhojivé

Díky povaze poruch a jejich dopadu na běžný život pacienta je obvykle problém detekován poměrně brzy (např. u nádorů přímo na hlasivkách). Díky včasnému vyšetření a léčbě lze často problém odstranit či alespoň zastavit jeho rozvoj.

3 Vyšetřovací metody

Vyšetřování hlasivek má poměrně dlouhou minulost a jeho počátky se datují již k přelomu 18. a 19. století. Tehdy byly hlasivky poprvé nepřímo pozorovány, popsány a prezentovány vědecké společnosti. Původní metoda sledování hlasivek byla poměrně primitivní, jednalo se pouze o soustavu dvou zrcátek bez jakéhokoliv přidaného osvětlení (pro nasvícení muselo být využito slunečních paprsků) [23].

V průběhu let se metody vyšetření hlasivek stále více rozvíjely, nacházela se nová využití (např. pro zlepšení úspěšnosti intubace pacienta). Vznikl tak samostatný vědní obor pro studium hlasivek zvaný laryngoskopie.

Paralelně docházelo k vývoji i mnoha jiných způsobů jak vyšetřit hlasivky pomocí jiného aparátu než zrcátek (resp. kamery). Všechny metody lze rozdělit do čtyř kategorií:

1. Akustické metody - metody založené na analýze zvukového záznamu
2. Aerodynamické metody - metody založené na sledování dýchání a průtoku vzduchu hlasivkovou štěrbínou
3. Elektrofyzilogické metody - metody založené na elektrických vlastnostech hlasivky při fonaci
4. Optické metody - metody vyšetření založené na vizuální analýze hlasivky

3.1 Optické metody

Mezi optické metody pro vyšetření hlasivek se řadí především RTG (rentgen), CT (počítačová tomografie), MR (magnetická rezonance) a laryngoskopie. Všechny tyto metody cílí na zobrazení hlasivek ať už je to pomocí radiačního záření, magnetické rezonance či kamery. Dále bude rozebrána především laryngoskopie, neboť se jí tato práce zabývá.

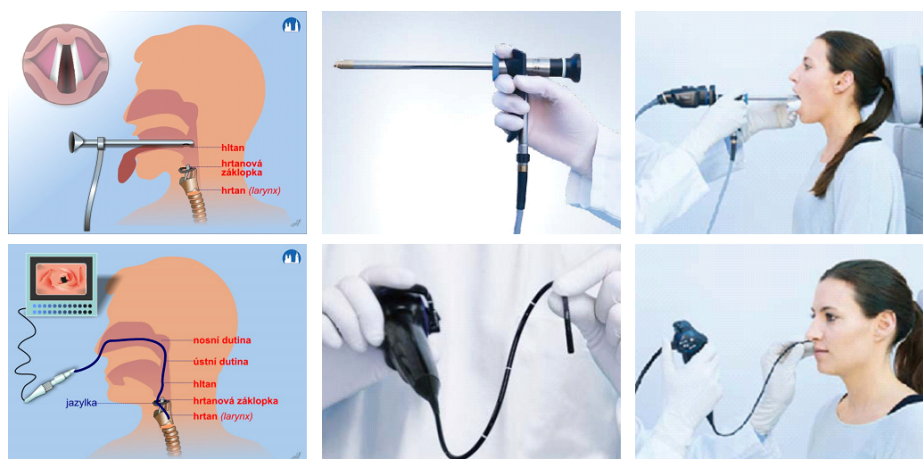
3.1.1 Laryngoskopie

Laryngoskopie je optická metoda vyšetření hlasivek, která se používá k prohlédnutí hrtanu a hlasivek pacienta. Tento postup se obvykle provádí pomocí laryngoskopu, což je speciální nástroj určený pro takovýto pohled na hlasivky.

Existují dva typy laryngoskopie: přímá a nepřímá. Nejobvyklejším druhem dnes prováděné laryngoskopie je laryngoskopie nepřímá, kdy se nedíváme na hlasivky pacienta přímo, ale využíváme odrazů světla pomocí zrcátek, případně se na hlasivku díváme pomocí kamery. Existují dvě různé techniky laryngoskopie - pevná a ohebná (viz Obrázek 3.1). Oba přístupy k vyšetření mají své zastánce, pevný laryngoskop nabízí oproti flexibilnímu například možnost hlasivky i zvětšovat, což je pro některá vyšetření výhodou. Pevný laryngoskop je zaváděn ústy, flexibilní nosní dutinou [13].

Přímá laryngoskopie, kdy se na hlasivky díváme přímo okem, se ve foniatrii příliš nevyužívá (s výjimkou chirurgických zákroků), neboť je ji třeba provádět v plné anestezii. I při chirurgických zákrocích však někteří operatři raději preferují nepřímou variantu. [6, 13]

Během nepřímé laryngoskopie pacient obvykle sedí a lékař vloží laryngoskop do úst pacienta a posouvá jej pomalu směrem k hrtanu. Světlo na konci laryngoskopu umožňuje lékaři vidět hrtan a hlasivky pacienta. Přímá laryngoskopie je prováděna vleže s tím, že pacient je nakloněn tak, aby byl umožněn přímý pohled na hlasivky.



Obrázek 3.1: Rigidní laryngoskopie (nahore) a flexibilní laryngoskopie (dole) [15].

3.1.2 Vysokorychlostní laryngoskopie

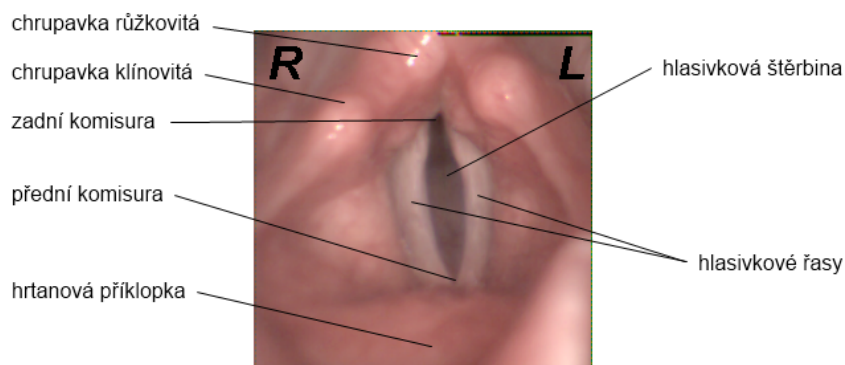
Moderní metodou vyšetření hlasivek je tzv. vysokorychlostní laryngoskopie LHSV (Laryngeal High-Speed Videoendoscopy). Jedná se o optickou metodu vyšetření hlasivek pomocí kamery podporující vysokorychlostní snímání (obvykle okolo 4000 snímků za sekundu). V rámci této diplomové práce jsem se zabýval daty získanými právě pomocí této metody.

Videozáznamy využívané v této práci byly pořízeny laryngoskopickým systémem HRES ENDOCAM 5562. Jedná se o přístroj pro nepřímou pevnou laryngoskopii, který dokáže snímat frekvencí 4000 barevných snímků za sekundu. Základní popis přístroje lze pro ilustraci nalézt v Příloze B.

V záznamu pořízeném kamerou je zachycen pohyb hlasivkových řas pohledem z hrtanu. Díky vysokému počtu snímků za sekundu je obvykle vybrán jen zajímavý úsek v němž proběhne několik period pohybu hlasivky a tento úsek je dále zpracováván.

Při záznamu je zapotřebí hlasivky přisvítit pomocí světla, které je součástí laryngoskopu. Při vyšetření je nutno dát pozor na to, aby hlasivka byla dostatečně nasvícena jinak vzniká těžko zpracovatelný záznam.

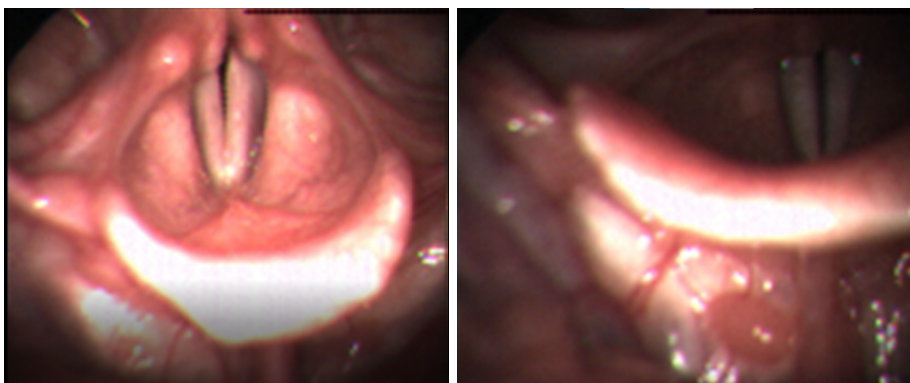
Záznam zobrazuje hlasivky shora (jako při pohledu z úst), jak lze vidět na Obrázku 3.2.



Obrázek 3.2: Pohled laryngoskopem na hlasivky [7].

3.2 Kvalita videozáznamu

Záznam z vysokorychlostní kamery samozřejmě není vždy perfektní. Pro úspěšnou automatickou detekci ať už oblasti zájmu (oblast ve které se hlasivka nachází - viz Kapitola 4.1) nebo pro samotnou segmentaci bychom si přáli, aby hlasivky byly vždy stejně nasvícené, ideálně stejně orientované a nedocházelo k žádnému pohybu kamery. Záznam je však snímán lékařem, a díky tomu je do videozáznamu často zanesena lidsky vytvořená chyba. [8]



Obrázek 3.3: Dobrý videozáznam. Obrázek 3.4: Špatně nasvícený videozáznam.

Osvícení scény

Nejčastější chybou je pravděpodobně nesprávné osvětlení celé scény (hlasivek). U osvětlení může dojít k dvěma chybám - přesvícení, kdy část hrtanu okolo hlasivek bude příliš světlá a algoritmy ji poté často budou za hlasivku nesprávně považovat nebo k nedosvícení, kdy naopak není ve videozáznamu nic dobře detekovatelné kvůli malému kontrastu.

Nejčastější příčinou špatného osvětlení scény je nesprávná manipulace se světelným zdrojem laryngoskopu - např. špatné nastavení, natočení světelného zdroje nebo vzdálenost zdroje od hlasivky.

Příklad takového problému, kdy je hlasivka špatně osvětlená lze vidět na Obrázku 3.4. Záznamy, kde se takto špatné nasvícení vyskytuje, jsou poměrně špatně automaticky zpracovatelné.

Neostrý obraz

Dalším problémem je nedoostření obrazu. Dle modelu mohou mít kamery fixní ohnisko nebo je lze doostřovat, záleží tudíž čistě na vyšetřující osobě, aby zajistila kvalitu videozáznamu správnou manipulací s kamerou.

Jiné problémy

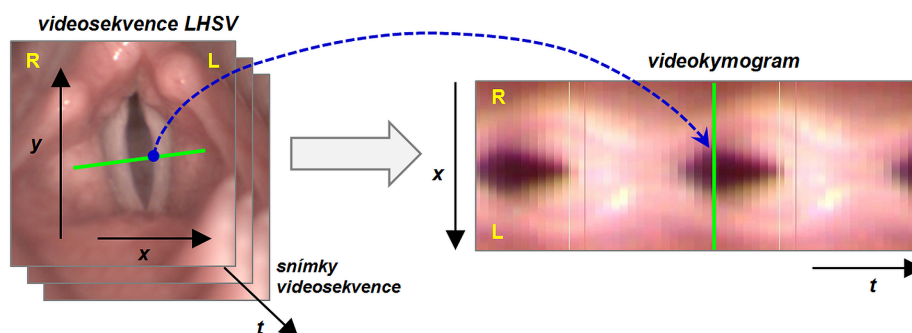
Existuje mnohem více problémů, se kterými se lze u videozáznamu z laryngoskopu setkat, například je nutno počítat s tím, že ruka vyšetřující osoby se v průběhu záznamu pohne (byť jen o velmi malý kus).

Dalším z problémů může být například výskyt sliny nebo hlenu ve videozáznamu, což opět může ovlivnit schopnost algoritmů správně detekovat štěrbinu.

V neposlední řadě se také může stát, že hlasivka ve videozáznamu není (částečně) vidět, buď z důvodu, že pohled blokuje jiná anatomická struktura (nejčastěji hrtanová příklopka) či je kamera jen špatně natočená.

3.3 Videokymografie

Videokymografie je další z optických metod pro vyšetření hlasivek. Obecně je kymografie metodou záznamu změny sledované oblasti v čase. Metodu kymografie lze uplatnit pro záznamy v mnoha různých oblastech, například pro záznam tepů srdce nebo právě pro vyšetření v akustice. Pro vyšetření hlasivek se využívá tzv. videokymografie, kdy tato metoda, vyvinutá v roce 1996 [25], využívá záznam modifikovanou kamerou, která zobrazuje jen jeden příčný řez hlasivkou v čase. Díky tomu lze dosáhnout vysoké rychlosti (8000 snímků za sekundu). Druhou možností je snímat záznam jako celek (nižší rychlostí) a řádek záznamu vždy vybrat. Metoda videokymografie zachycuje pohyb hlasivky v čase a s ním spojenou frekvenci hlasivky a amplitudu jejích kmitů pozorovatelných při otevírání a zavírání a lze na ní vidět případnou asymetrii.



Obrázek 3.5: Laryngoskopie (vlevo) vs videokymografie (vpravo), zeleně zvýrazněn řez, ze kterého je tvořen videokymogram.

3.3.1 Videokymogram

Tato práce se zabývá zpracováním výstupu především této konkrétní metody - zpracováním tzv. videokymogramu. Videokymogram je, jak již bylo zmíněno, příčný řez hlasivkou zaznamenaný v čase. Videokymogramy se obecně využívají především k analýze pohybu a změn. Díky tomu je lze využít k analýze pohyblivých částí lidského těla jako jsou například právě hlasivky. Příklad videokymogramu lze vidět na Obrázku 3.5 vpravo.

Videokymogram je buď tvořen speciálním přístrojem zvaným videokymograf, který přímo zaznamenává jen daný řez, nebo jej lze vytvořit i dodatečně z videozáznamu vyprodukovaného například pomocí vysokorychlostního laryngoskopu LHSV (viz Kapitola 3.1.2).

Oba přístupy tvorby videokymogramu produkují stejný výsledek, při dodatečném vytvoření videokymogramu je výhodou možnost pohlédnout na sledovaný útvar z více pohledů (videozáznam vs. videokymogram). Nevýhodou je to, že pro tvorbu videokymogramu z videozáznamu LHSV potřebujeme sofistikovanější plošnou kameru. Díky tomu byl videokymogram využíván již před vznikem plošných vysokorychlostních kamer.

3.3.2 Tvorba videokymogramu z videozáznamu LHSV

Tvorba videokymogramu dodatečně z vysokorychlostních videozáznamů spočívá ve vybrání libovolné úsečky, která nás bude zajímat. Nyní poskládáme za sebe vývoj vybrané úsečky v čase do jednoho obrázku a tím nám vznikne videokymogram pro danou úsečku. Častou volbou takovéto úsečky je rovnoběžka s osou (většinou x), ale postup je obecně aplikovatelný na libovolnou úsečku ve videozáznamu LHSV.

4 Metody používané pro detekci hlasivkové štěrby

Díky rozvoji metod počítačového vidění a rozvoji vysokorychlostních kamer se téma detekce glottis z videozáznamu stalo poměrně frekventovaným. Tato úloha se obvykle skládá ze dvou částí - prvně je zapotřebí určit oblast zájmu tzv. ROI a poté v rámci určeného ROI je provedena segmentace hlasivkové štěrby. Existuje několik možných přístupů k tomuto problému, které lze dělit dle zásahu uživatele poloautomatické a automatické.

Poloautomatické přístupy umožňují uživateli v průběhu segmentace do procesu manuálně zasáhnout a opravit případné nedostatky. Automatické procesy na druhou stranu nevyžadují žádný zásah uživatele. Oba přístupy mají své výhody a nevýhody, hlavní výhodou automatických přístupů je, že nevytěžují klinického specialistu, na druhou stranu výstupy z poloautomatických metod budou s velkou pravděpodobností přesnější [1].

Ukázku nejrůznějších přístupů lze nalézt ve zkratce v Tabulce 4.1 a 4.2, kde jsou vypsány různé přístupy pro detekci ROI a následně i detekci glottis. Detailněji rozebrané metody spolu s autory a odkazy na jednotlivé články lze nalézt v [1], Tab. 1 a 2.

Tabulka 4.1: Možnosti určení ROI dle [1].

bez určení ROI	bez určení ROI
	manuální
semiautomatické	pomocí odčítání
	seed body
automatické	estimace pohybu
	variace intenzit

4.1 Detekce ROI

ROI, aneb oblast zájmu, je často určena v prvním kroku před samotnou segmentací. ROI je obvykle definováno jako obdélníková oblast, která opisuje přední a zadní komisuru a celé hlasivkové vazy při maximálním otevření hlasivky.

Tabulka 4.2: Ukázka možných metod pro detekci glottis z Tab. 1 a 2 v článku [1].

semiautomatické	prahování
	narůstání oblasti
	adaptivní okno
	geometrické modely
	parametrické modely
	hranové operátory
automatické	watershed
	estimace pohybu
	morfologické operátory
	geometrické modely
	parametrické modely
	neuronové sítě
	klasifikace

Existuje více způsobů, jak lze ROI určit, hlavní dělení dle [1] je na feature-based a object-based. Feature-based přístupy se soustředí na hledání pixelů s obdobnými vlastnostmi, které u ROI předpokládáme, a na jejich následné spojení do jedné oblasti. Object-based přístupy se nesoustředí na jednotlivé pixely, zabývají se většími celky, tzn. například hledají ROI díky jeho jedinečnému tvaru a struktuře.

Mnoho článků se například také o detekci ROI vůbec nepokouší a požadují manuální zadání ROI od uživatele.

Obecně je určení ROI poměrně podstatné pro další funkci algoritmů, neboť hlasivky a hlasivková štěrbina zabírají ve videozáznamu LHSV obvykle méně než 25% celé plochy a omezením se pouze na ROI dojde k zamezení falešných detekcí během segmentace.

4.2 Segmentace hlasivkové štěrby

Pro samotnou segmentaci hlasivkové štěrby lze využít mnoha různých přístupů. V přehledové studii představují Miranda et al. [1] několik různých metod navržených a publikovaných v člancích mezi roky 1979-2019. Metody dělí právě dle interakce s uživatelem na poloautomatické a automatické. V tabulkách 1 a 2 článku [1] lze vidět, že pro poloautomatické metody je velmi často nutné definovat ROI manuálně, automatické metody naopak definici ROI dost často vynechávají úplně.

4.2.1 Prahování

Existuje více přístupů k samotné segmentaci, z těch důležitějších lze například uvést přístup pomocí prahování, příklad čehož lze nalézt v článku [3]. V tomto článku autoři popisují dvoufázový přístup k segmentaci uvnitř ROI. Autoři experimentálně stanovili dva parametry T_1 a c_2 . Nejprve provedli výpočet rozdílového snímku $DI(x, y, t)$, což je binární snímek vzniklý prahováním rozdílu snímku videozáznamu a vybraného referenčního snímku. Referenční snímek je vybrán manuálně tak, aby v něm byla co nejvíce dověřena glottis. Prahování rozdílového snímku probíhá podle stanoveného prahu T_1 . Hlavní myšlenkou tohoto prahování je to, že v $DI(x, y, t)$ zůstávají jen pohyblivé části. V tomto snímku se však mohou ještě vyskytovat nežádoucí struktury vzniklé buď špatným osvětlením či pohybem okolních struktur. Autoři článku navrhují jako řešení využití mediánového filtrování, kdy pomocí filtru dojde k rozmazání hran a tím je odstraněn šum. Po aplikaci mediánového filtru vznikne obrázek, který vstupuje do dalšího prahování, tentokrát je prahován pomocí prahu rovného součtu minimální hodnoty ve snímku a c_2 . Ve výsledku získáváme segmentovaný videozáznam.

4.2.2 Narůstání oblastí

Z dalších přístupů lze například uvést přístup pomocí narůstání oblasti publikovaný v článku [19]. Jedná se opět o poloautomatickou metodu, která na počátku vyžaduje po uživateli výběr několika bodů uvnitř regionu glottis. Pomocí těchto bodů je určena střední hodnota uvnitř regionu glottis a s její pomocí je oblast glottis segmentována. V binarizovaném obraze vzniklém prahováním je posléze nalezen největší segment (nazývaný blob) a je určen centroid tohoto blobu. Tento centroid pak vstupuje jako seed pro algoritmus narůstání oblasti, kdy se oblast okolo centroidu zvětšuje, za předpokladu, že je splněno kritérium homogenity. Toto kritérium je popsáno jako rozdíl intenzit okolních pixelů a střední hodnoty intenzity segmentované oblasti. Pixely, které mají rozdíl menší než uživatelem definované maximum, jsou poté přidány do výsledné oblasti. Algoritmus probíhá dokud se nedostane do bodu, kdy už nejde do oblasti přidat žádný nový bod. Autoři uvádějí, že pro své testování použili toleranci 15% - tzn. hodnota jasů pixelů se mohla lišit o 115%.

4.2.3 Přístupy pomocí strojového učení

Poslední kategorie, která stojí za zmínku především v dnešní době neustálého poměrně rychlého rozvoje strojového učení, je přístup právě pomocí takovýchto metod.

Shlukování

Shlukování je jednou z možností, jak se lze pokusit o nalezení hlasivkové štěrbin. Jedná se o metodu strojového učení bez učitele, kdy je určen předem daný počet tříd (objektů), na které má být obraz rozdělen na základě podobnosti parametrů, viz [8]. V této práci jsou popisovány parametry, které tvoří tento prostor pro shlukování.

V práci [8] jsou popsány parametry, které tvoří takovýto prostor pro shlukování. Jako parametr je zvolena především barva (RGB), pozice v obraze (x , y), dále rozdíl R a B složky a vzdálenost od středu. Tím vzniká sada parametrů, podle kterých jsou body videozáznamu nadále shlukovány. Je ukázáno, že shlukování funguje i na snímky s horší kvalitou (např. kdy je část glottis zakryta jinými anatomickými strukturami).

Neuronové sítě

V neposlední řadě lze pro detekci glottis využít i nejrůznějších neuronových sítí.

Například autoři článku [9] analyzovali 18 různých nastavení různých neuronových sítí a snažili se najít ideální nastavení, které by fungovalo nad jejich datasetem. Jako nejlepší označili LSTM síť (long short-term memory), což je druh rekurentní neuronové sítě, který efektivně řeší problém neuronových sítí podobného typu týkající se zapomínání dlouhodobých vztahů mezi jednotlivými vstupy [5, 22].

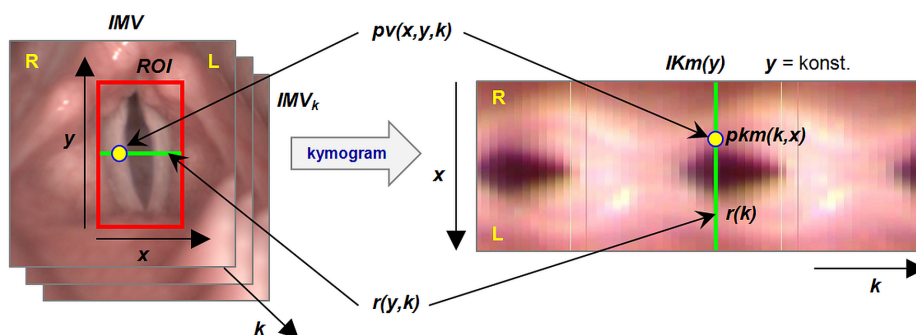
Největší problém, který se naskytá při práci s videozáznamy hlasivek pomocí neuronových sítí, je dostupnost dat. Díky povaze dat není jednoduché nalézt dostatečně velký volně přístupný dataset nad kterým by šlo provést správné trénování. Autoři článku [11] popisují problémy s menšími datasey. Dále také zveřejňují větší dataset právě pro trénink neuronových sítí a popisují vlastní přístup k verifikaci výstupů z analýzy hlasivek pomocí neuronových sítí.

5 Teoretický rozbor použitých technik

V rámci této diplomové práce jsem se zabýval návrhem a ověřením metody využívající odlišného přístupu než metody zmíněné v Kapitole 4. V této kapitole budou rozebrány obecné algoritmy využitě v navrženém přístupu a v následující kapitole bude představen samotný algoritmus navržený v rámci této práce.

5.1 Zavedené značení

Pro následující kapitoly je zavedeno značení (viz Obrázek 5.1):



Obrázek 5.1: Zavedeného značení bodů ve videozáznamu a ve videokymogramu.

- IMV ... videosekvence snímků z LHSV;
- IMV_k ... jeden snímek sekvence IMV , $k = 0, 1, 2, \dots, N - 1$;
- N ... počet snímků IMV .

V rámci snímků IMV předpokládáme, že je definována oblast zájmu, tzv. ROI . ROI je oblast, pro náš účel pravidelný pravoúhelník, který ohraničuje anatomické struktury hlasivek ve snímku. Tyto anatomické struktury jsou vymezeny polohou přední a zadní komisury a vnějším okrajem levé a pravé hlasivkové řasy v okamžiku maximálního otevření hlasivek. ROI je v rámci této práce považováno za vstupní parametr. Byť každý IMV_k má v rámci

této práce stejné ROI , lze označit ROI ve snímku pro případné upřesnění jako ROI_k . Pak budeme psát:

ROI ... definovaná oblast zájmu v IMV , platí $ROI \in IMV$;

ROI_k ... definovaná oblast zájmu pro snímek $IMV_k \in IMV$.

V rámci každého ROI_k lze definovat jednotlivé obrazové body, které budou dále označovány jako $pv(x, y, k)$. x a y jsou souřadnice bodu v ROI_k nabývající hodnot $x = 0, 1, \dots, X - 1$ a $y = 0, 1, \dots, Y - 1$, kde X je počet sloupců a Y je počet řádků ROI_k . Bod bude tudíž definován:

$pv(x, y, k)$... bod IMV_k patřící do ROI_k ;

X ... počet sloupců ROI_k ;

Y ... počet řádků ROI_k .

V Kapitole 3.3.1 byl dále definován videokymogram. Jedná se o zobrazení jednotlivých řádků $r(y, k)$ příslušných ROI_k pro $\forall k = 0, \dots, N - 1$. Můžeme značit:

$IKm(y)$... videokymogram sestavený jako průběh řádků $r(y, k)$ v čase (přes všechna k).

$pkm(k, x)$... je bod videokymogramu $IKm(y)$

Obrazový bod $pkm(k, x)$ videokymogramu $IKm(y)$ odpovídá obrazovému bodu $pv(x, y, k)$ z původní videosekvence IMV .

Body $pkm(k, x)$ a $pv(x, y, k)$ se skládají ze tří barevných složek R , G , B . Pro účely dalších algoritmů předpokládejme, že v potaz bereme vždy hodnotu pouze jedné složky (nejčastěji R , neboť anatomické struktury jsou prokrvené, a tudíž červené) nebo bereme hodnotu jasu (Y).

5.2 Metody segmentace pomocí analýzy histogramu

Hlavním úkolem této práce je segmentace hlasivkové štěrbiny. Je proto nutné rozebrat některé algoritmy, které budou dále v této práci využívány. V rámci této práce jsem vybral dvě metody prahování na dva prahy, které budou používány, možných metod však existuje mnohem více.

Prahování je metoda využívaná na rozdělení snímku IM na dvě (nebo více) navzájem disjunktní podmnožiny obrazu IM pokrývající celý IM . Tato práce se nadále bude zabývat pouze dělením obrazové scény na dvě množiny pomocí dvou metod automatického prahování - metodou optimálního prahu (Otsu) a metodou minimální chyby (Kittler).

5.2.1 Metoda optimálního prahu (Otsu)

Metoda optimálního prahu [16, 18, 24] je jednou z automatických metod pro odhad prahu. Algoritmus Otsu se definuje práh, který histogram zadaného obrázku rozdělí tak, aby došlo k maximální separaci dvou tříd (obvykle označovány jako popředí a pozadí) vzniklých prahováním, případně se hledá práh, kde je dosaženo minimálního rozptylu v obou třídách.

Nejprve je potřeba definovat relativní četnosti výskytu obrazového bodu $pkm(k, x)$ s jasem i (hodnota buď Y , R , G nebo B) v $IKm(y)$ jako:

$$p_i = \frac{NN_i}{NN} \geq 0, \quad \text{ä platí, že } \sum_{i=1}^n p_i = 1 \quad (5.1)$$

pro $\forall i = 1, \dots, n$, kde NN_i je počet bodů s jasem i v $IKm(y)$ a NN je celkový počet bodů. Dále jsou definovány navzájem disjunktní třídy C_0 a C_1 pro které platí:

$$\begin{aligned} C_0 &= \{1, 2, 3, \dots, T\}, & T \text{ je hledaný práh} \\ C_1 &= \{T + 1, \dots, n\} \end{aligned} \quad (5.2)$$

Nyní jsou definovány pravděpodobnosti výskytu obrazového bodu IKm v jednotlivých třídách jako:

$$\begin{aligned} C_0 : \quad \omega_0(T) &= \sum_{i=0}^T p_i \\ C_1 : \quad \omega_1(T) &= \sum_{i=T+1}^n p_i; \quad \omega_0(T) + \omega_1(T) = 1 \end{aligned} \quad (5.3)$$

Dále jsou dány střední hodnoty pro jednotlivé třídy a celý histogram jako:

$$\begin{aligned} C_0 : \quad \mu_0(T) &= \frac{1}{\omega_0(T)} \sum_{i=0}^T ip_i \\ C_1 : \quad \mu_1(T) &= \frac{1}{\omega_1(T)} \sum_{i=T+1}^n ip_i; \quad \mu_T = \sum_{i=1}^n ip_i \end{aligned} \quad (5.4)$$

A také rozptyly ve třídách a histogramu:

$$\begin{aligned}
C_0 : \quad \sigma_0^2(T) &= \frac{1}{\omega_0(T)} \sum_{i=0}^T [i - \mu_0(T)]^2 p_i \\
C_1 : \quad \sigma_1^2(T) &= \frac{1}{\omega_1(T)} \sum_{i=T+1}^n [i - \mu_1(T)]^2 p_i; \quad \sigma_T^2 = \sum_{i=1}^n [i - \mu_T]^2 p_i
\end{aligned} \tag{5.5}$$

Kritérium optimality pro stanovení prahu T^* je pak formulováno jako hledání maxima separability $\sigma_B^2(T)$ tříd C_0 a C_1 :

$$\sigma_b^2(T^*) = \max_{1 < T < n} \{\sigma_b^2(T)\}, \tag{5.6}$$

kde

$$\sigma_b^2(T) = \omega_0(T)\omega_1(T)[\mu_0(T) - \mu_1(T)]^2 \tag{5.7}$$

5.2.2 Kittlerova metoda nejmenší chyby

Druhou metodou pro určení prahu je Kittlerova metoda nejmenší chyby [12, 14]. Tato metoda je opět automatickou metodou. Metoda je obdobná metodě optimálního prahu, také vychází definuje pravděpodobnosti p_i :

$$p_i = \frac{NN_i}{NN} \geq 0, \quad \text{ä platí, že } \sum_{i=1}^n p_i = 1 \tag{5.8}$$

pro $\forall i = 1, \dots, n$, kde NN_i je počet bodů s jasem i v $IKm(y)$ a NN je celkový počet bodů a také třídy C_0 a C_1 , jejich střední hodnoty a rozptyly:

$$\begin{aligned}
C_0 : \quad \omega_0(T) &= \sum_{i=0}^T p_i \\
\mu_0(T) &= \frac{1}{\omega_0(T)} \sum_{i=0}^T i p_i \\
\sigma_0^2(T) &= \frac{1}{\omega_0(T)} \sum_{i=0}^T [i - \mu_0(T)]^2 p_i
\end{aligned} \tag{5.9}$$

$$\begin{aligned}
C_1 : \quad \omega_1(T) &= \sum_{i=T+1}^n p_i \\
\mu_1(T) &= \frac{1}{\omega_1(T)} \sum_{i=T+1}^n i p_i \\
\sigma_1^2(T) &= \frac{1}{\omega_1(T)} \sum_{i=T+1}^n [i - \mu_1(T)]^2 p_i
\end{aligned}$$

Nyní je však histogram je chápán jako míra hustoty pravděpodobnosti dvou statistických rozdělání. Jedná se o odhad pravděpodobnosti funkce $PP(i)$, která je směsí statistického souboru úrovní vybrané barvy, který patří objektům a pozadí. Předpokládáme, že každá z komponent $PP(i|C_0)$ a $PP(i|C_1)$ funkce $PP(i)$, tj. směsi statistického souboru, bude mít normální rozdělání s parametry: $[\mu_0(T); \sigma_0(T)]$, $[\mu_1(T); \sigma_1(T)]$, dále zkráceně $[\mu_0; \sigma_0]$, $[\mu_1; \sigma_1]$ a že známe apriorní pravděpodobnosti patření do třídy C_0 a C_1 tj. $\omega_0(T)$ a $\omega_1(T)$ (zkráceně ω_0 a ω_1).

Potom pro odhad hustoty pravděpodobnosti směsi $PP(i)$ platí:

$$\begin{aligned}
 PP(i) &= \omega_0 PP(i|C_0) + \omega_1 PP(i|C_1), \text{ kde} \\
 PP(i|C_0) &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(i - \mu_0)^2}{2\sigma_0^2}\right], \\
 PP(i|C_1) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{(i - \mu_1)^2}{2\sigma_1^2}\right]
 \end{aligned} \tag{5.10}$$

Formulujeme kriteriální funkci $J(T)$, která má po úpravách výsledný tvar:

$$J(T) = 1 + 2[P_0 \ln \sigma_0(T) + P_1 \ln \sigma_1(T)] - 2[P_0 \ln P_0 + P_1 \ln P_1] \tag{5.11}$$

Algoritmus hledání optimálního prahu T^* ve smyslu minimální chyby vychází z podmínky hledání minima kriteriální funkce $J(T)$:

$$J(T^*) = \min_T \{J(T)\} = \min_{\forall T \in (1, n)} \{J(T)\} \tag{5.12}$$

6 Navržené metody segmentace glottis

Jak již bylo zmíněno, cílem této práce je ze vstupních dat, videosekvence ve formátu *avi* (případně *bld*, což je speciální formát vytvářený laryngoskopickými zařízeními) pocházejících z vyšetření hlasivek laryngoskopem segmentovat hlasivkovou šterbinu (glottis). Nyní budou představeny dvě metody, které byly v rámci této práce za tímto účelem navrženy. Na Obrázku 6.1 lze nalézt diagram těchto metod, které budou dále do detailu rozebrány.

6.1 Vstupní data

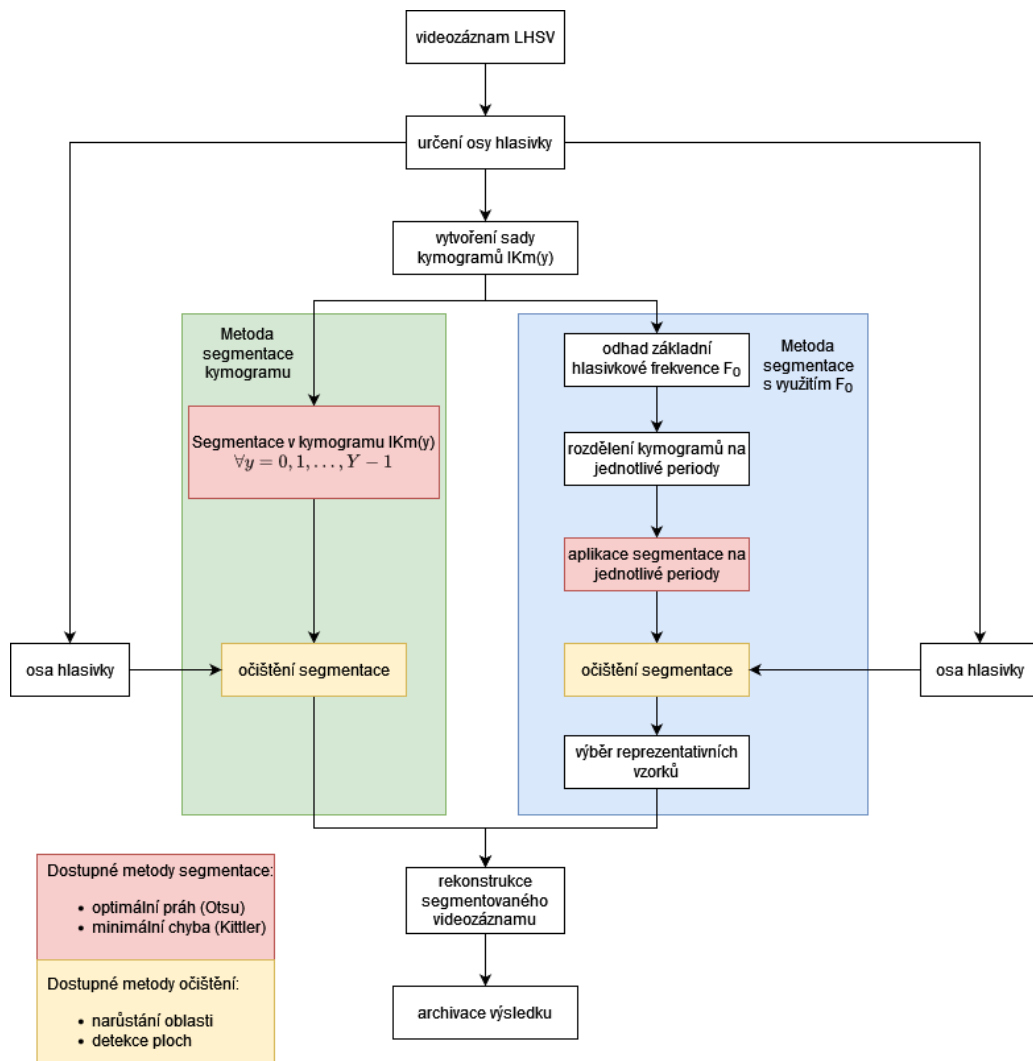
Jako vstupní data pro metody navržené v rámci této diplomové práce jsou videozáznamy vysokorychlostní kamerou LHSV (viz Kapitola 3.1.2). Tyto videozáznamy mají rozlišení 256 na 256 pixelů a jsou snímány frekvencí 4000 snímků za vteřinu. Videozáznam je obvykle tvořen 500-1000 snímky (tudíž se nejedná ani o celou vteřinu videozáznamu). Tento počet snímků je dostačující a kratší délkou je omezen vliv třesu ruky vyšetřujícího a je také maximalizováno pohodlí pacienta.

Videozáznamy LHSV byly pořízeny ve spolupráci s ORL FN Plzeň a jedná se anonymizovaný dataset, který obsahuje jak zdravé tak různě postižené hlasivky. Jak již bylo výše popsáno, obvykle je během vysokorychlostní laryngoskopie pořízen záznam zvuku. V rámci této práce bylo mým úkolem předpokládat, že tyto hlasové záznamy nejsou dostupné a parametry, které je možné určit z akustického záznamu, budou odvozeny z videozáznamu.

6.2 Hlasivková frekvence

Hlasivková frekvence, označme F_0 je jedním z parametrů, který je obvykle při vyšetření určován. Jedná se o počet kmitů hlasivek za sekundu během fonace.

Obvykle je tento parametr určován ze zvukového záznamu, který je snímán současně s videozáznamem, ale lze jej určit také například z pohybu anatomických struktur hlasivek, který je zachycen ve videozáznamu. Protože tato práce předpokládá nedostupnost audiozáznamu, bude dále rozebráno několik způsobů jak určit hlasivkovou frekvenci právě tímto způsobem.



Obrázek 6.1: Metoda navržená v rámci této práce. Zeleně označenou metodu segmentace videokymogramu (větev vlevo) lze najít v Kapitole 6.4, modře označenou metodu segmentace s využitím hlasivkové frekvence F_0 (větev vpravo) lze nalézt v Kapitole 6.5.

6.2.1 Určení hlasivkové frekvence

Pro určení hlasivkové frekvence z videozáznamu pohybu hlasivek je více možností. V rámci této práce jsem se zabýval třemi přístupy pro detekci frekvence:

- Detekce pomocí zadaného bodu $pv(x, y) \in ROI$
- Detekce pomocí videokymogramu $IKm(y)$
- Detekce uvnitř ROI

Detekce dle zadaného bodu

První metodou je detekce frekvence z pohybu anatomických struktur pod zadaným bodem $pv(x, y)$ uvnitř ROI . Vývoj jasu (případně složky R, G nebo B) takového zadaného bodu $pv(x, y)$ v čase (k) tvoří 1D signál $x(n)$.

Na takovýto signál lze aplikovat diskrétní Fourierovu transformaci (DFT), což je transformace signálu z časové do frekvenční oblasti. Jedná se o transformaci mezi posloupnostmi $\{x(n)\}_{n=0}^{N-1}$ a $\{X(k)\}_{k=0}^{N-1}$ s N vzorky a lze ji popsat vztahem:

$$PV_{(x,y)}(m) = \sum_{k=0}^{N-1} pv(x, y, k) e^{-j \frac{2\pi}{N} km},$$

$$|PV(m)|_{(x,y)} = \sqrt{\{Re [PV_{(x,y)}(m)]\}^2 + \{Im [PV_{(x,y)}(m)]\}^2}$$

pro $\forall m = 0, 1, 2, \dots, N - 1$;

$$\forall (x, y) \in IMV_k \text{ a } IMV_k \in IMV \text{ pro } \forall k, \text{ kde } k = 0, 1, 2, \dots, N - 1$$
(6.1)

$PV_{(x,y)}(m)$ se nazývá fourierovo spektrum a $|PV(m)|_{(x,y)}$ necht' je nazýváno DFT-amplitudové spektrum.

V dalším kroku lze pak určit frekvence signálu (zde frekvence změny jasu barevných komponent) tím, že hledáme maximum DFT-amplitudového spektra pro $\forall m = 1, 2, \dots, N/2$; hodnota DFT-amplitudového spektra pro $m = 0$ odpovídá stejnosměrné složce signálu, volba $N/2$ odpovídá vlastnostem spektra reálné posloupnosti $pv(x, y, k)$.

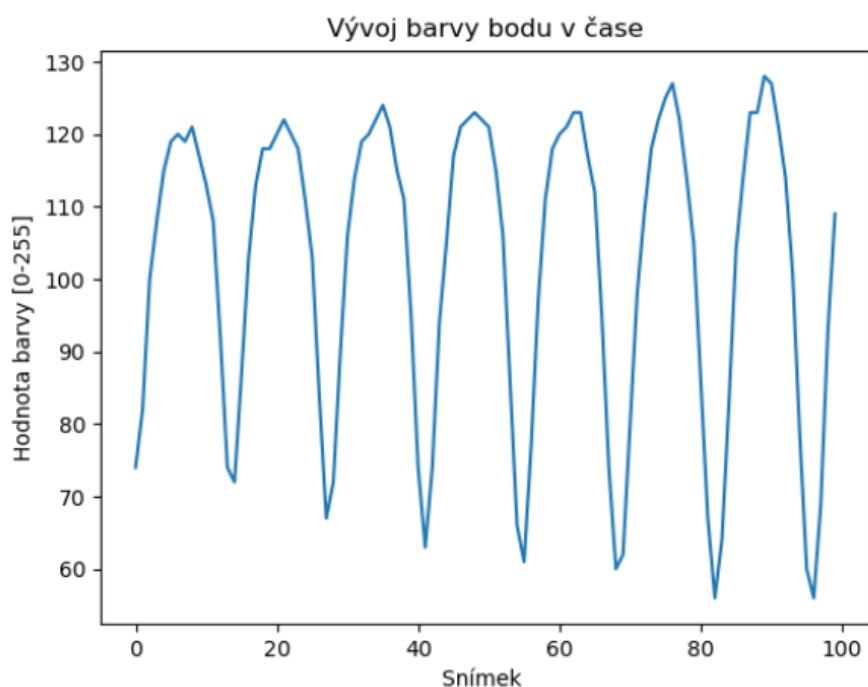
Stanovení indexu m max pak můžeme popsat vztahem:

$$|PV(m_{max})|_{(x,y)} = \max_{\forall m=1,2,\dots,N/2} \{|PV(m)|_{(x,y)}\}$$
(6.2)

m_{max} pak odpovídá frekvenci změny jasu Y (barevných komponent R,G,B) v bodě $pv(x, y, k)$. Pro výpočet frekvence $f_{sig}(x, y)$ [Hz] změny jasu, tj. odhadu F_0 , použijeme vztah:

$$F_0 \approx f_{sig}(x, y) = m_{max} \frac{f_{sampl}}{N} = m_{max} \Delta f$$
(6.3)

kde Δf je DFT grid (frekvenční bin) a f_{sampl} je vzorkovací frekvence systému LHSV (FPS).



Obrázek 6.2: Ukázka vývoje barvy (R) v čase v manuálně zvoleném bodě na hlasivce.

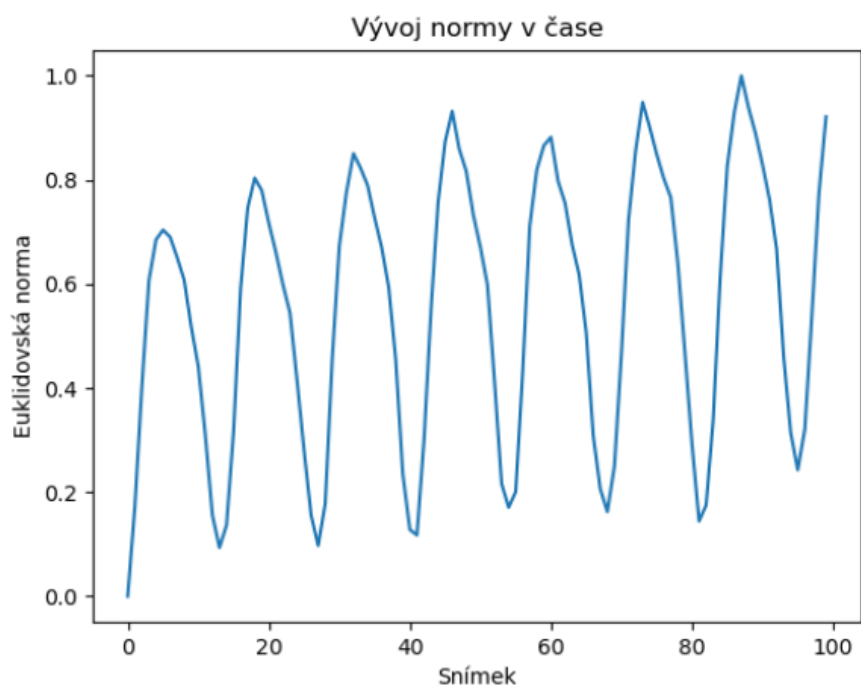
Detekce ve videokymogramu

Druhou metodou je určení frekvence z jednoho daného videokymogramu $IKm(y)$, tj. podle definovaného řádku $y \in ROI$. Tato metoda probíhá následovně:

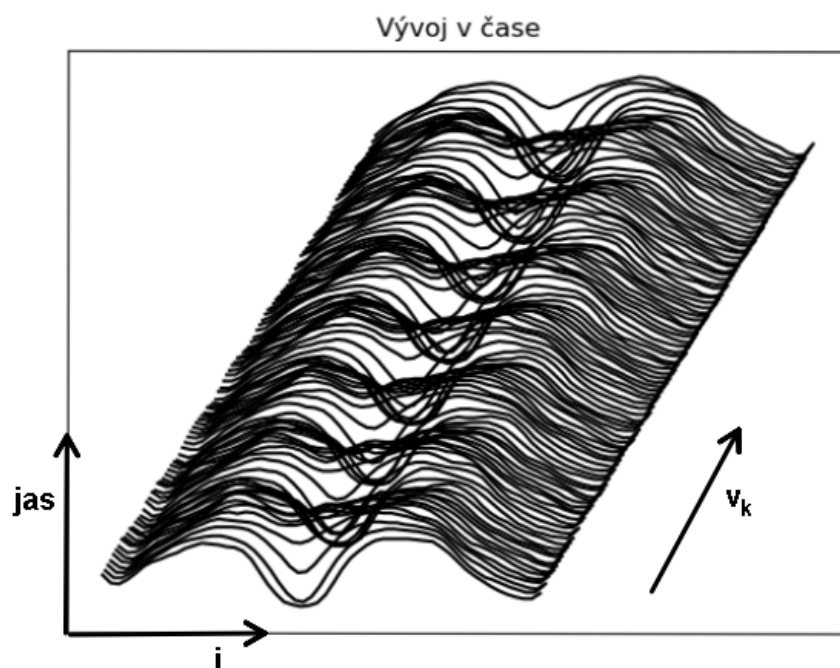
1. Každý sloupec $IKm(y)$ (tzn. řádek $r(y)$ z původního ROI_k) je vzat jako vektor v_k s X prvky. Toto je zopakováno pro všech k sloupců
2. Je spočtena euklidovská norma každého vektoru v_k :

$$d_k = \sqrt{\sum_{i=0}^X v_{k_i}^2} \quad (6.4)$$

3. Hodnoty d_k jsou vzaty jako 1D signál (osa y - d_k , osa x - k neboli čas) viz Obrázek 6.3
4. Fourierova transformace takto vzniklého signálu a nalezení frekvence stejným stylem jako pro signál vytvořený pomocí jasů bodu v Kapitole 6.2.1



Obrázek 6.3: Ukázka výsledného signálu vytvořeného v Kapitole 6.2.1.

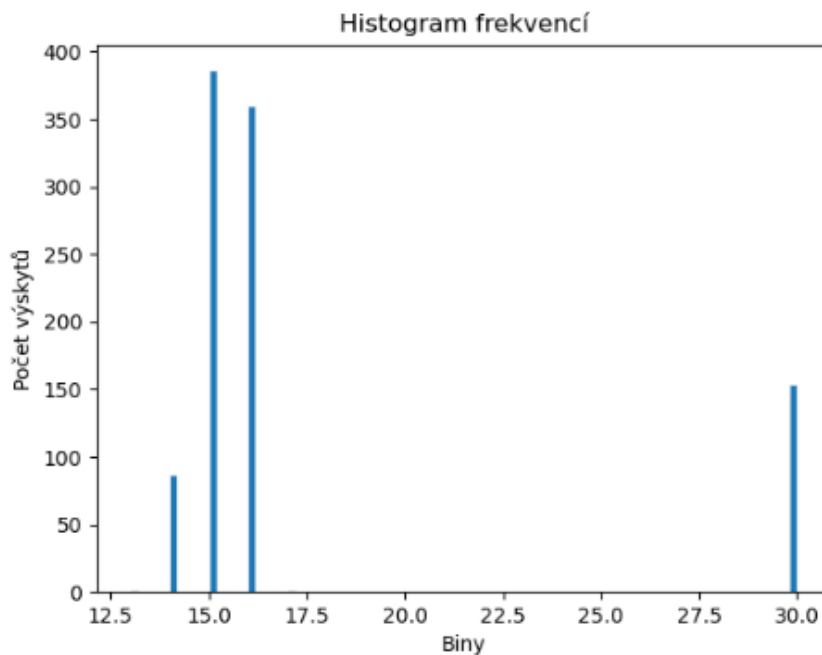


Obrázek 6.4: Ukázka vývoje vektoru v v čase.

Předpokladem této metody bylo to, že se snímky periodicky opakují a tudíž jejich vzdálenost (např. zde Euklidovská) od počátku bude víceméně stejná v každé periodě. Díky tomu bude perioda zachycena ve výsledném signálu z těchto vzdáleností. Průběh vývoje vektoru v v čase (osa z) lze pro ilustraci vidět na Obrázku 6.4 a vývoj vzdáleností pro ten samý příklad je na Obrázku 6.3.

Detekce uvnitř ROI

Poslední způsob jak určit frekvenci, je výpočet F_0 pro každý bod v ROI . Jedná se o aplikaci algoritmu z Kapitoly 6.2.1 na každý bod $pv(x, y)$ vybrané oblasti ROI . Tím vzniká 2D pole frekvencí F , kde pro každý bod $pv(x, y)$ je určena jeho frekvence $F_{(x,y)}$. Z tohoto pole frekvencí F je dále vybrán medián, neboť hlasivková frekvence F_0 je převládající frekvencí uvnitř ROI [2].

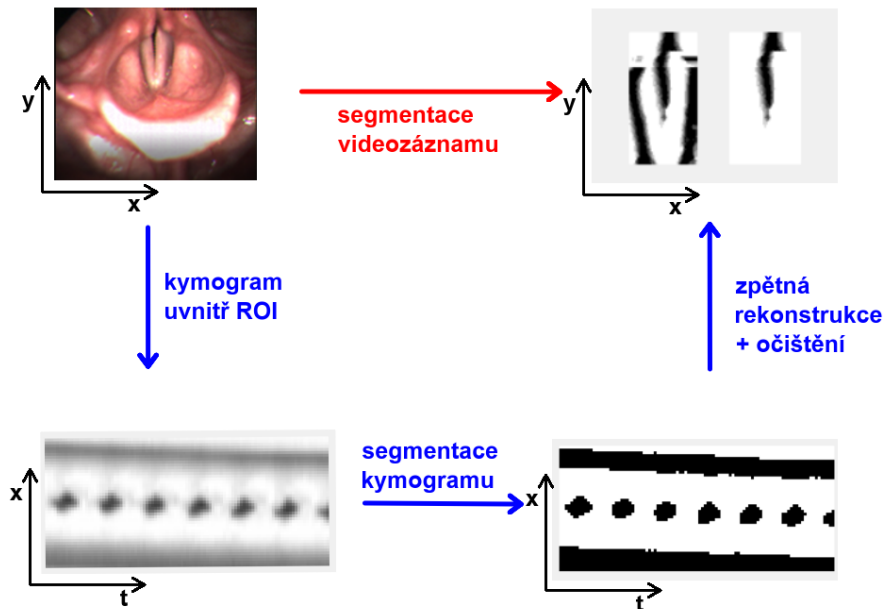


Obrázek 6.5: Ukázka četností frekvencí pro ROI .

6.3 Metody aplikované na videokymogramy

Prvním krokem ve zpracování je převod všech snímků sekvence IMV na jednotlivé videokymogramy $IKm(y)$ pro každé y z ROI . Tím vznikne sada Y

videokymogramů (kde Y je počet řádků ROI). Označme tuto sadu IKm . Sada videokymogramů IKm je dále zpracovávána dvěma způsoby - viz Obrázek 6.1.



Obrázek 6.6: Srovnání navrženého algoritmu (zpracovávajícího videokymogram - modré šipky) a metody, kdy je segmentace aplikována přímo na videozáznam (červená šipka).

6.4 Segmentace videokymogramu

Jednou z možností jak sadu IKm využít pro segmentaci glottis, je na každý videokymogram $IKm(y)$ aplikovat algoritmus pro automatickou segmentaci (viz Kapitola 5.2.1 a 5.2.2). Myšlenka spočívá v tom, že hlasivková štěrbina i ve videokymogramu bude poměrně odlišná od ostatních anatomických struktur a automatická metoda prahování najde právě takový práh, který oblast glottis označí za objekty popředí a ostatní anatomické struktury budou označeny za objekty pozadí.

Zároveň však segmentace neprobíhá v jednotlivých ROI_k , ale právě nad videokymogramem $IKm(y)$, což zachycuje vývoj štěrbiny v čase a prahovací metody díky tomu nastavují automatický práh za základě odlišných parametrů.

Jak již bylo zmíněno, metod pro automatické určení správného prahu existuje mnoho, v rámci této práce je využito metod Otsu a Kittler, obě

jsou popsány v Kapitolách 5.2.1 a 5.2.2.

Po aplikaci těchto metod získáváme pro každý videokymogram $IKm(y)$ práh T^* , pomocí kterého lze pixely videokymogramu $pkm(k, x)$ transformovat na výsledné pixely po segmentaci $pkm_s(k, x)$.

$$pkm_s(k, x) = \begin{cases} 255, & \text{pro } pkm(k, x) \geq T^* \\ 0, & \text{pro } pkm(k, x) < T^* \end{cases} \quad (6.5)$$

Pixely $pkm_s(k, x)$ tvoří segmentovaný videokymogram, který pro další zpracování v Kapitole 6.6 označíme jako $IKm_s(y)$.

6.5 Segmentace videokymogramu s využitím základní hlasivkové frekvence

Druhou možností, kterou se tato práce zabývala v rámci práce se sadou vytvořených videokymogramů IKm , bylo rozdělení každého z videokymogramů $IKm(y)$ na I menších částí $IKm_i(y)$ (pro $i = 0, 1, \dots, I$) pomocí hlasivkové frekvence F_0 , takových, že každá $IKm_i(y)$ obsahuje právě jednu periodu pohybu hlasivky. Perioda hlasivky je definována tak, aby obsahovala vždy celý cyklus od otevření do zavření a opětovného otevření hlasivek.

6.5.1 Algoritmus založený na hlasivkové frekvenci

Pro určení periody (což je obrácená hodnota hlasivkové frekvence F_0) bylo popsáno několik metod v Kapitole 6.2.1. Pomocí jakékoliv z těchto metod je tudíž určen počet snímků (p), ze kterých se skládá jedna perioda.

Vezmeme tudíž videokymogram $IKm(y)$ (rozměr $N \times X$) a rozdělíme ho po p sloupcích na menší snímky $IKm_i(y)$, které mají rozměr $p \times X$. Poslední snímek nemusí mít dostatečný rozměr v x-ové ose (pokud p nebylo soudělné s N) a tento snímek bude v tomto případě zahozen pro neúplnost - neobsahuje celou periodu.

Na jednotlivé snímky $IKm_i(y)$ můžeme nyní opět aplikovat algoritmy automatického prahování (Otsu a Kittler) a získat naprahované snímky $IKm_{i_s}(y)$ (stejně jako v Kapitole 6.4).

Z těchto snímků pro snazší zpracování uděláme jeden reprezentativní snímek a to třemi způsoby:

$IKm_{max}(y)$... maximum přes $IKm_{i_s}(y)$ po pixelech $pkm(k, x)$

$IKm_{min}(y)$... minimum přes $IKm_{i_s}(y)$ po pixelech $pkm(k, x)$

$IKm_{mean}(y)$... průměr přes $IKm_{i_s}(y)$ po pixelech $pkm(k, x)$

Tyto tři reprezentativní snímky vytvoříme pro každý videokymogram ze sady IKm . Tyto snímky budou dále zpracovány v postprocessingu a ve výsledku z nich bude opět složen segmentovaný videozáznam.

6.6 Postprocessing

V průběhu práce se stalo poměrně jasným to, že automatické metody pro detekci prahu poměrně často selžou a práh nedetekují tak, že v jedné skupině po prahování skončí pouze glottis. Často se stalo, že s glottis se do stejné skupiny dostaly i tmavší části hlasivky, což je samozřejmě špatně.

Nastala tudíž potřeba navrhnout metodu, která by segmentovanou glottis od takovýchto špatně označených částí okolních anatomických struktur odlišila. Díky povaze dat je opět poměrně obtížné navrhnout metodu, která by zaručeně fungovala na všechny videozáznamy, které mohou být velmi nekvalitní (viz Kapitola 3.2).

Metoda, kterou jsem pro očištění glottis po segmentaci navrhl vychází z následujících experimentálně určených předpokladů:

1. hlasivková štěrbina se vždy nachází poblíž středu ROI
2. v rámci několika period nedochází k dostatečně významnému pohybu do strany ve videozáznamu
3. hlasivková štěrbina a ostatní anatomické struktury jsou od sebe odděleny částí s barvou pozadí

Ve výsledku byly navrženy tři různé postupy, které mají podobný princip. Na vstupu jsou segmentované obrázky $IKm_s(y)$, $IKm_{max}(y)$, $IKm_{min}(y)$ a $IKm_{mean}(y)$. Obecně dále budou označeny pro lepší přehlednost jen jako $IKm_s(y)$. Navrženy byly následující postupy:

6.6.1 Postup pomocí narůstání oblasti

Díky výše uvedeným předpokladům můžeme použít algoritmus narůstání oblasti. Obrázek $IKm_s(y)$ je segmentovaný se skládá ze dvou částí, označme je například A a B. Skupina A se skládá ze dvou částí - části, která obsahuje glottis a části nesprávně označené jako glottis. Skupina B je skupina s objekty pozadí. Nyní se snažíme odstranit nesprávně označenou část skupiny A (aneb to, co není glottis, ale segmentace to za glottis označila).

Postup začne tím, že se pokusíme o odhad šterbiny okolo prostředku videokymogramu a všechny pixely $pkm(k, x)$ označené jako skupina A, které okolo středu (tzn. pro $x = X/2 \pm 5$) najdeme, označíme jako skupinu C (nová skupina, která by na konci měla obsahovat už jen pouze glottis). Volba středu v $X/2$ je logická volba, která vychází z předpokladu, že ROI je umístěna středem na osu hlasivky, u které předpokládáme ideální umístění, tj. svislé. Pásmo ± 5 pixelů je experimentálně zvolená hodnota necitlivosti, která předchází posuvům osy od středu a i jejímu náklonu.

Po označení pixelů u středu projdeme celý zbytek obrazu (všechny pixely $pkm(k, x)$, které zatím nebyly navštíveny, od středu po sloupcích) a pixely skupiny A rozdělíme mezi skupiny B a C podle toho, zda sousedí s pixely skupiny C nebo nikoliv. To znamená ve výsledku, že půjdeme od středu glottis a budeme označovat pixely jako pixely glottis (C), dokud nenalezneme pixel, který není označen jako glottis. Zbylé pixely ve sloupci budou označeny jako pixely pozadí nehledě na výsledek segmentace. Po průchodu všech pixelů $IKm_s(y)$ nám zbude opět dvoubarevný obrázek $IKm_o(y)$.

Tento algoritmus pro očištění segmentace je poměrně závislý na tom, aby glottis nesplývala s okolními anatomickými strukturami při segmentaci. Pokud by k tomu došlo, není příliš možné odlišit v segmentaci mezi glottis a okolím.

6.6.2 Postup pomocí detekce ploch

Druhý přístup je využít dostupné algoritmy pro detekci ploch v obraze. Jednu z takovýchto metod lze nalézt v článku [21]. V článku je popsáno jak využít algoritmus založený na sledování hran k nalezení topologie binarizovaného obrazu. Algoritmus začne prohledáváním obrazu skládajícího se z pixelů hodnot 0 a 1. Algoritmus běží skrz celý obraz, dokud nenarazí na pixel, který splňuje jednu ze dvou podmínek. Buď se jedná o pixel vnější hranice (definovaný jako sousednost pixelů $[0, 1]$ (v tomto pořadí)) nebo o pixel hranice díry (definovaný jako sousednost pixelů $[>= 1, 0]$ (v tomto pořadí)). Jakmile takovýto počáteční pixel najde, přiřadí této nově nalezené hranici nový index (číslo > 1 , které je neustále inkrementováno pro

nové hrany). Algoritmus projde celou hranici v rámci 4/8-okolí. Pixely nejvíce vpravo v každé hranici jsou označeny negativním číslem odpovídajícím hraně, aby byly vynechány při dalším hledání počátečních pixelů. Postupně algoritmus projde všechny možné počáteční pixely hranic a během procházení tvoří strom, který zachycuje která hranice a která díra spolu souvisí v rámci obrazu.

Pro tento algoritmus je například dostupné řešení v knihovně OpenCV [17], kde je nabízena metoda `findContours` implementující algoritmus právě z výše zmíněného článku [21].

Poté, co máme nalezené plochy stejné barvy pomocí tohoto algoritmu, můžeme aplikovat obdobný postup, který byl popsán výše. Díky předpokladu, že hlasivková štěrbina se nachází poblíž středu oblasti zájmu ROI_k v ose x , se můžeme pokusit najít všechny plochy, které se u středu nacházejí. Plocha je vždy definována jako obdélník a lze tudíž zjistit, zda do něj patří střed zpracovávané oblasti zájmu ROI_k . Pokud plochy do oblasti patří, ve finálním obraze je ponecháme, a všechny pixely s barvou štěrbin, které se ve vybraných plochách nenacházejí, označíme za pixely pozadí. Tím opět vznikne očištěný dvoubarevný obraz $IKm_o(y)$.

6.6.3 Postup pro hlasivky mimo střed

Předchozí dva přístupy mají jednu zásadní vadu a tou je, silný předpoklad, že glottis leží ve středu ROI . Je možné, že pro některé videozáznamy by tento předpoklad nebyl splněn a byla proto navržena třetí metoda očištění, resp. úprava předchozích metod tak, aby byl předpoklad polohy glottis zanedbán. Tato úprava spočívá v odhadu středu pomocí osy hlasivky.

Existuje více algoritmů pro určení osy hlasivky, jedním z nich je například přístup [8], kde je navrženo, abychom vytvořili průměrný snímek štěrbin z videozáznamu IMV (průměr IMV_k pro všechny k po jednotlivých pixelech $p(x, y)$). Ohledně hlasivek máme několik předpokladů:

1. osa kopíruje hlasivku v uzavřeném stavu
2. pokud se hlasivka nedovírá, osa prochází přibližně středem nedovřené části
3. hlasivka se v krátkém úseku videozáznamu (jedna až dvě periody) nehýbe na tolik, aby to mělo vliv na výsledek

Pokud tyto předpoklady platí, můžeme poté na průměrném snímku vždy nalézt bod v řádce, který bude nejtmaší, případně střed nejtmaších bodů. Pokud toto zopakujeme pro všechny relevantní řádky glottis, získáme sadu

bodů, které můžeme proložit přímkou (např. pomocí metody nejmenších čtverců) a tuto přímku lze považovat za výslednou osu.

Pomocí takto určené osy můžeme v algoritmech z předchozích kapitol lépe zvolit odhad středu glottis než jen jako pás okolo středu obrazu. S takto lépe určeným středem pak můžeme provést již výše uvedené postupy očištění bez dalších změn a zbavit se závislosti na poloze hlasivky.

6.7 Zpětná rekonstrukce

Posledním krokem algoritmů je zpětná rekonstrukce videozáznamu. Metody popsané výše provádějí operace nad jednotlivými videokymogramy $IKm_o(y)$, avšak lepší vizualizaci poskytuje přeci jen videozáznam jako takový. Rekonstrukce však není žádný problém, je zapotřebí jen ze sady výsledných očištěných videokymogramů $IKm_o(y)$ vzít vždy jeden původní řádek videozáznamu (aneb jeden sloupec z $IKm_o(y)$) a zpětně tyto řádky složit do jednoho snímku videozáznamu. Toto lze udělat pro všechny řádky původního videozáznamu a tím vznikne segmentovaný videozáznam hlasivkové štěrbině IMV_{out} .

7 Program Kymogram analyzer

Metody segmentace glottis popsané v Kapitole 5 byly v rámci této práce implementovány a jejich výsledky jsou prezentovány v následujících kapitolách.

Jako jazyk implementace byl zvolen jazyk Python, především kvůli dostupnosti nejrůznějších algoritmů pro práci s obrazem v knihovně OpenCV [17]. Python má sice několik nevýhod, hlavní zmiňovanou je obvykle jeho rychlost, neboť se jedná o interpretovaný jazyk. Pro účely této práce, kde nejsou zpracovávány příliš velké videozáznamy (detaily viz část Kapitoly 6.1 o velikosti videozáznamu) je však naprosto postačující.

7.1 GUI

Grafické rozhraní aplikace je implementováno v knihovně tkinter [20]. Jedná se o standardní knihovnu pro práci s grafickým rozhraním pro Windows, Unix i macOS. Knihovna poskytuje základní prvky pro tvorbu grafických rozhraní s moderním vzhledem.

7.1.1 Uvítací obrazovka

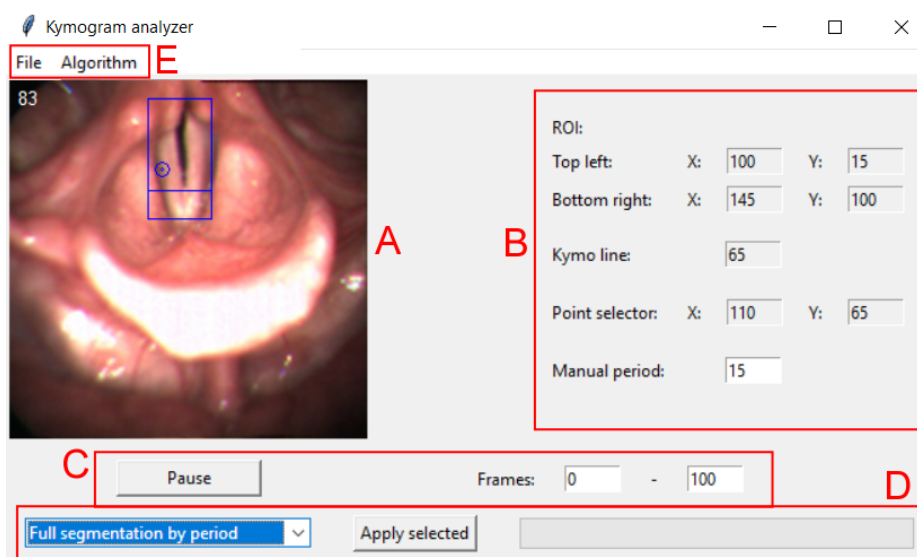
GUI aplikace se skládá z více oken. První okno, na které uživatel narazí je uvítací obrazovka programu, která uživatele vyzývá k zadání videa, které bude zpracováváno algoritmy popsanými v minulých kapitolách.

7.1.2 Hlavní okno

Po zadání zpracovávaného videa se obrazovka přepne na hlavní obrazovku aplikace. Zde je pro uživatele k dispozici hned několik možností. Na Obrázku 7.1 lze vidět všechny prvky tohoto okna.

Videopřehrávač

V části 7.1A je přehráván uživatelem zadaný videozáznam. V horním rohu videozáznamu se nachází text vypisující číslo aktuálního snímku z videozáznamu pro případnou synchronizaci či pozdější manuální dohledání.



Obrázek 7.1: Hlavní okno aplikace.

Ovládací prvky

Část 7.1B zobrazuje uživateli parametry, které lze nastavit pomocí myši uvnitř 7.1A. Potažením obdélníku lze nastavit *ROI*, které jak již bylo výše zmíněno, je potřebné pro správnou funkci všech algoritmů. Další parametry, které lze nastavit myší, jsou především pro srovnávací metody. První možností je nastavení řádky, ve které bude vytvořen videokymogram pro určení frekvence / periody dle Kapitoly 6.2.1 a pro ukázání videokymogramu v této řádce. Další, co lze nastavit je bod pro určení frekvence / periody dle kapitoly 6.2.1. Poslední parametr této sekce, který je zadáván číselně do políčka, umožňuje manuální zadání periody pro algoritmus popsany v Kapitole 6.5.

Další ovládací prvky lze vidět v části 7.1C. Zde se nachází prvky, které umožňují ovládání videozáznamu. Nachází se zde tlačítko, které umožňuje video pozastavit (a znovu poté spustit) a také pole, kde jde nastavit počáteční a koncový snímek videa. Toto je umožněno aby uživatel měl kontrolu nad délkou videozáznamů. Některé videozáznamy mohou být zbytečně dlouhé (pro většinu algoritmů stačí obvykle okolo 100 snímků) a je vhodné videozáznam zkrátit. Obvykle je důvodem především zhoršující se kvalita videozáznamu s délkou (především vliv třesu ruky vyšetřujícího), ale to je na posouzení uživatele.

Posledním ovládacím prvkem je 7.1E, kde je pomocí menu uživateli umožněno uživateli libovolně přepnout mezi zpracovávanými videi, a menu, kde je možno přepínat mezi algoritmy použitými pro očištění. Základní nastavení je algoritmus pomocí detekce ploch s automaticky odhadnutou osou.

Volba metody

Poslední částí hlavní obrazovky je combobox, který umožňuje výběr metody, která bude aplikována s aktuálním nastavením na videozáznam. Po stisku tlačítka „Apply selected“ je spuštěna vybraná metoda (detailní popis implementovaných metod viz následující kapitola 7.2). Protože některé algoritmy jsou složitější a mohou trvat několik málo sekund, nachází se zde i progress bar, který signalizuje jak metoda postupuje.

7.2 Algoritmy pro segmentaci glottis

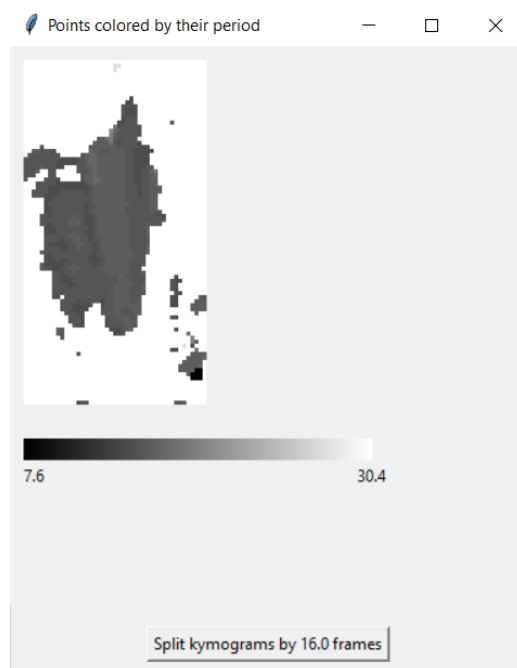
V rámci této práce byly navrženy algoritmy pro zpracování videozáznamu z vysokorychlostní kamery LHSV (viz Kapitola 5). Na hlavní obrazovce lze jednotlivé implementované metody zvolit (viz Obrázek 7.1D). Všechny metody potřebují na vstupu mít zadané ROI (které lze např. pomocí automatické metody popsané výše v Kapitole 4.1). Určení ROI nebylo předmětem této práce a proto je předpokládáno, že bude zadáno manuálně - jak již bylo popsáno - pomocí myši. Případné další potřebné argumenty jsou vždy uvedeny níže u popisu jednotlivých metod. Jednotlivé metody aplikovatelné na videozáznam jsou:

1. Show kymogram - Tato volba uživateli zobrazí videokymogram a jeho segmentaci pomocí dvou metod - pomocí Otsu prahování a prahování Kittlerovou metodou nejmenší chyby. Pro použití této metody je potřeba nejprve zvolit řádek, ve kterém bude videokymogram vytvořen.
2. Full kymogram segmentation - Automatická metoda pro segmentaci celého videokymogramu. Metoda převede celý videozáznam na jednotlivé videokymogramy, aplikuje na ně metody segmentace, očistí segmentaci a zpět složí videozáznam (přesně jak je popsáno v návrhu v Kapitole 6.4).
3. Find period from histogram - Metoda detekce periody (a frekvence) vyžadující vstup od uživatele. Pro použití je potřeba specifikovat řádek, ve kterém bude proveden algoritmus pro určení frekvence ze zadaného videokymogramu popsaný v Kapitole 6.2.1.
4. Find period by selected point - Druhá metoda detekce periody vyžadující vstup od uživatele. Perioda je určena pomocí FFT signálu tvořeného jasem myši zadaného bodu.



Obrázek 7.2: Ukázka okna, které zobrazí metoda 1. Vrchní obrázek je původní videokymogram, spodní dva jsou videokymogramy po aplikaci segmentací Kittler a Otsu.

5. Find period of whole ROI - Metoda automatického určení periody. Nevyžaduje žádný vstup kromě ROI, jež bylo zmíněno už výše. Určení frekvence je podrobně popsáno v Kapitole 6.2.1.
6. Find period of whole video - Toto je druhá verze předchozí metody. Místo prohledání celého ROI je prohledáno celé video a nalezena dominantní frekvence. Vůči předchozí metodě nemá tato volba žádnou výhodu, je náchylnější na chyby a je implementována spíše pro zajímavost.
7. Show kymogram sliced by period - Tato metoda vezme myší zadaný řádek, ve kterém vytvoří videokymogram a dále požaduje zadání periody. Poté zobrazí jak je videokymogram pomocí této periody rozsekán na více částí. Pokud je perioda určena správně, měla by každá část obsahovat jeden cyklus otevření a uzavření hlasivky.
8. Full segmentation by period - Automatická metoda pro segmentaci



Obrázek 7.3: Okno s výsledkem metod 5 a 6. Každý bod ROI je obarven dle jeho frekvence, pod ním se nachází stupnice, ukazující rozsah barev. Pomocí tlačítka vespod lze vyvolat metodu 8 s vypočtenou periodou.

videokymogramu dle zadané periody. Tato metoda opět rekonstruuje videozáznam ze segmentovaných videokymogramů (tak jak je popsáno v Kapitole 6.5).

9. Segmentation without videokymogram - Aplikuje metody segmentace (Otsu a Kittler) přímo na jednotlivé snímky videozáznamu (místo na videokymogram). Tato možnost se zde nachází především pro srovnání s metodami využívajícími videokymogram.

Metody 3-6 po svém spuštění vždy vypíší výslednou periodu a hned spustí s touto periodou metodu 7 pro ukázkou, jak videokymogram vypadá nařezaný pomocí výsledné periody. Dále je zde uživateli nabídnuta možnost spustit metodu 8 s výslednou periodou.

Metody 2 a 8 jsou metody, na které se tato práce především zaměřuje a po jejich spuštění a dokončení je uživateli zobrazeno nové okno (viz Obrázek 7.5). V tomto okně je vždy zobrazeno několik verzí výsledku dle použité metody - Otsu nebo Kittler, případně dle principu kterým byla vybrána vzorová perioda (u metody 8) - minimum, maximum nebo průměr. U každé metody se nacházejí dvě videa, která jsou podobná, neboť jediný rozdíl mezi nimi je to, že na první (levé) video nebyl aplikován žádný algoritmus pro očiš-



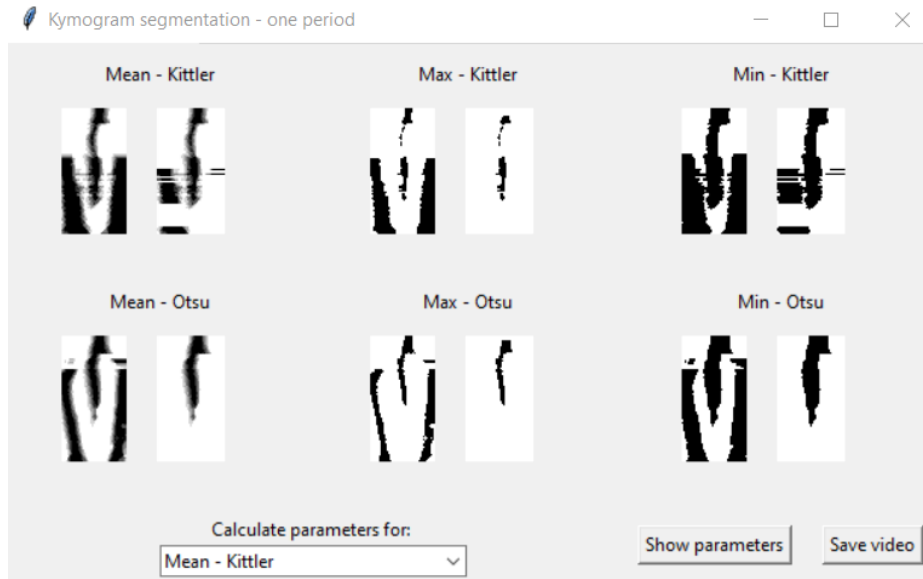
Obrázek 7.4: Okno s výsledkem metod 3, 4 a 7. Lze vidět videokymogram rozdělený pomocí zvolené (resp. vypočtené) periody. Pod ním se nachází průměrný, minimální a maximální snímek vypočtený z jednotlivých snímků.

tění glottis. Následně se zde nachází políčko, které umožňuje vybrat jeden z těchto výsledných segmentovaných videozáznamů a určit u něj nejrůznější parametry.

V Kapitole 6.6 byly navrženy dva algoritmy, jak glottis po segmentaci očistit. Oba vycházejí z podobných předpokladů a na základě testování mají oba velmi podobné výsledky. Jediné, v čem se algoritmy lišily byla jejich rychlost (a byť poměrně nepodstatně - oba dobíhají v rámci nejdéle sekundy, protože zpracováváný videozáznam není příliš velký), druhý algoritmus (6.6.2), využívající knihovnu OpenCV [17] byl přeci jen rychlejší. Z tohoto důvodu je pro všechna očištění použit právě tento algoritmus. V průběhu práce byla ještě navržena úprava těchto algoritmů očištění ta-

ková, že je využito odhadu osy pro očištění místo pouhého předpokladu, že se nachází okolo středu.

Metoda 9 nabízí podobné rozhraní jako metody 2 a 8, ale je zde především pro srovnání původní metody s metodami aplikovanými na videokymogram v následující Kapitole 8.



Obrázek 7.5: Okno s výsledkem metod 2, 8 a 9. Na obrazovce se nachází několik videí vytvořených ze segmentovaných (a očištěných) videokymogramů. Dále je zde k nalezení možnost jedno video z nich vybrat, vypočítat u něj parametry nebo ho uložit.

7.3 Parametry určené u štěrbin

Jak bylo zmíněno v předchozí kapitole, u dvou voleb metod je na konci uživateli zobrazeno segmentované video hlasivky pomocí různých metod. Pro srovnání metod je však vhodné mít možnost u glottis určit i parametry jako je například plocha hlasivkové štěrbin, její osa a jiné.

Program tuto možnost uživateli nabízí, po zobrazení okna na Obrázku 7.5 si lze povšimnout, že na spodku okna se nachází combobox umožňující výběr výsledného videa, nad kterým budou určeny parametry. Nic nebrání uživateli určit parametry pro všechna výsledná videa, ale obrazovka by byla příliš přeplněna informacemi, proto jsou parametry odsunuty na samostatnou obrazovku.

Určovanými parametry pro výslednou segmentaci jsou:

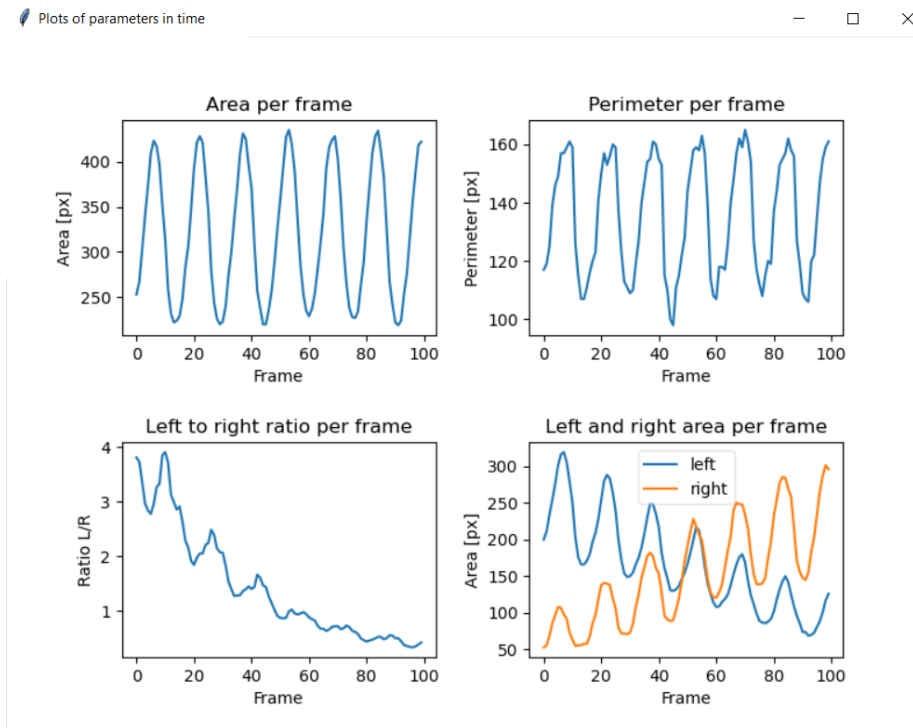
- Poloha hlasivkové štěrby (souřadnice ohraničujícího obdélníka)
- Osa hlasivkové štěrby (přímka ve tvaru $y = ax + b$)
- Obsah hlasivkové štěrby (v pixelech, neboť nemáme měřítko)
- Poměr levé a pravé strany (osa hlasivku dělí na dvě části, které lze srovnat např. ve velikosti)
- Obvod hlasivkové štěrby (opět v pixelech)



Obrázek 7.6: Okno s výslednými vypočtenými parametry - zde pro Otsu prahování. Parametry jsou vždy určeny v maximálním, minimálním a průměrném snímku z segmentované a očištěné videosekvence. Osa hlasivky je naznačena do průměrného snímku.

Okno s určenými parametry pro glottis segmentovanou metodou Otsu lze vidět na Obrázku 7.6. Zde jsou určeny parametry pro minimální, maximální a průměrný snímek. Tato sada by měla být dostatečná pro představu, jak

probíhá vývoj v čase, ale v případě, že by nestačila, se otevírá ještě druhé okno, které lze vidět na Obrázku 7.7



Obrázek 7.7: Okno s výslednými vypočtenými parametry - zde pro Otsu prahování. Parametry byly určeny v každém snímku (na rozdíl od vybraných snímků v Obrázku 7.6) a vyneseny do grafu. Grafy jsou popsány a zachycují opět plochu, obvod, poměr mezi levou a pravou částí a obsahy levé a pravé části.

7.4 Uložení výsledného videa

V neposlední řadě je důležité uživateli umožnit výsledné video uložit pro další zpracování. Uložit segmentovanou hlasivku lze na obrazovce z Obrázku 7.5. Uložený soubor má příponu avi a běží rychlostí 15 snímků za vteřinu, což však pro další načtení a zpracování stejně nemá přílišný význam. Video je poměrně malé, protože zachycuje pouze výslednou segmentaci štěrbinu uvnitř ROI.

8 Výsledky práce

Program Kymogram analyzer popsaný v předchozí kapitole umožňuje zpracování nejrůznějších videozáznamů hlasivek. Poté, co je nastaveno ROI, lze spustit jednotlivé algoritmy popsané v předchozích kapitolách a na základě jejich výsledků lze zhodnotit jejich úspěšnost detekce glottis.

Cílem bylo provést testování na vzorku dat, u kterého bylo prokázáno, že existují úspěšné metody detekce glottis. V případě neúspěchu testovaných metod je pak cílem identifikovat příčinu.

8.1 Testovací data

Pro testování implementovaných metod segmentace glottis byla využita část datového korpusu „692“, který obsahuje anonymizovaná data LHSV videozáznamů z vysokorychlostního laryngoskopu. Celkem bylo vybráno 48 videozáznamů LHSV, kde je zastoupeno několik diagnóz a videozáznamů různé kvality, viz přehledová Tabulka 8.1. Jedná se o videozáznamy, u kterých byla segmentace glottis úspěšná při použití metod K-means a prahování z [8].

Tabulka 8.1: Rozbor diagnóz pacientů ve vybraném testovacím datasetu.

LHSV datový korpus	videozáznamy LHSV	počet osob
diagnóza	48	21
cysta	3	1
chordectomie	3	2
karcinom	2	2
paréza zvratného nervu	12	5
polyp	4	2
uzlíky	7	3
zdravé hlasivky	17	6

Testy jsou prováděny nad 100 snímky videozáznamu. 100 snímků bylo vybráno především kvůli snížení vlivu posunu kamery na všechny algoritmy. 100 snímků zároveň stále zachycuje několik period, což je dostatečné pro všechny algoritmy navržené v rámci této diplomové práce.

8.2 Výsledky navržených metod

V rámci práce byly navrženy dva přístupy (viz Kapitoly 6.4 a 6.5) k segmentaci glottis ve videokymogramu. První z nich se zabýval segmentací uvnitř celého videokymogramu, aplikací metod segmentace a očištění byl videokymogram segmentován na dvoubarevný obraz a zpětně složeno výsledné video.

V druhé metodě ještě přibyl krok, kdy byla určena základní hlasivková frekvence a pomocí ní byl napřed videokymogram rozdělen na jednotlivé periody, z nich byl určen reprezentativní snímek a teprve ten byl segmentován, očištěn a zpětně složen.

8.2.1 Výsledky segmentace videokymogramu

Jak již bylo popsáno, tento algoritmus pracuje s celým videokymogramem a aplikuje na něj dvě zvolené metody. Ukázkou průběhu algoritmu i s mezivýsledky (pro jeden vybraný videokymogram) lze vidět na Obrázcích 8.1 a 8.2.

Na Obrázcích 8.1 a 8.2 lze tvidět nejprve původní *ROI*, poté videokymogram vytvořený pro jeden řádek (volený pro názornost kolem středu *ROI*) a nakonec lze vidět výstupy pro jednotlivé kombinace:

- Dle metody segmentace:
 - Metoda optimálního prahu (Otsu) - viz Kapitola 5.2.1
 - Kittlerova metoda nejmenší chyby - viz Kapitola 5.2.2
- Dle metody očištění:
 - Přístup pomocí detekce ploch - viz Kapitola 6.6.2
 - Přístup pomocí narůstání oblastí - viz Kapitola 6.6.1
 - Upravené verze dvou předchozích beroucí v potaz odhad osy hlasivky - viz Kapitola 6.6.3

Jako příklad jsou ukázány tyto dva obrázky, neboť na nich lze zpozorovat základní vlastnosti i nedostatky navržených přístupů. Na konci bude pak představena celková statistika pro celý dostupný dataset.

Podrobný rozbor metody při úspěchu

Obrázek 8.1 ukazuje případ, kdy došlo k volbě dobrého prahu. Jak lze vidět a jak již bylo popsáno v Kapitole 6.6 o očištění, ve výsledné segmentaci videokymogramu (levá část každé čtveřice obrázků) díky volbě prahu zbývá i část hlasivek. Pro algoritmy očištění (pravé části čtveřic obrázků) však není příliš velký problém takovýto nedostatek opravit a všechny čtyři kombinace (detekce ploch s / bez osy a narůstání oblasti s / bez osy) ve výsledku vracejí pouze segmentovanou glottis.

Dále lze na tomto obrázku vidět srovnání metody prahování pomocí metody optimálního prahu a pomocí metody nejmenší chyby. Pro vybraný videokymogram obě metody vracejí obdobně vypadající výsledky, ale lze si povšimnout, že byť se glottis poměrně dobře dovírá, metoda nejmenší chyby práh nastavuje výše než Otsu a tím ve výsledné segmentaci dochází k tomu, že se glottis nedovírá. Metoda nejmenší chyby obecně volí prahy hůř než metoda optimálního prahu, což ve výsledku má vliv i na její úspěšnost.

Zvolené prahy lze vidět na histogramu na Obrázku 8.1 dole. Manuálním prahováním by se pravděpodobně v tomto histogramu dal nastavit ideální práh, který by byl ještě níže než ten z metody optimálního prahu, pomocí kterého by segmentace dopadla nejlépe a pravděpodobně by nepotřebovala ani očistit. Není tomu však všude, některé snímky nemají takto dobrý kontrast barev.

Podrobný rozbor metody při selhání

Druhým příkladem je neúspěšná segmentace, která se stala na Obrázku 8.2. První, čeho si lze všimnout je, že videozáznam obsahuje záblesky světla, což výrazně rozhodí histogram (viz na Obrázku 8.2 dole). Tím, že obě metody segmentace, na které se tato práce soustředí, vychází z analýzy histogramu a četností jednotlivých barev, je zřetelné, že takovéto záblesky budou na segmentaci mít velký vliv.

V levých částech čtveřic obrázků lze vidět, že jak metoda nejmenší chyby, tak metoda optimálního prahu v tomto případě selhaly a nedošlo k nastavení správného prahu. Obě metody se zaměřily na segmentaci záblesků, neboť se skládají z více pixelů než samotná glottis.

Neboť těžce selhali již metody segmentace, metody očištění víceméně nedokáží nic s výslednou segmentací udělat, neboť na to nejsou stavěny. Lze si však povšimnout hlavního rozdílu mezi přístupem pomocí detekce ploch a pomocí narůstání oblasti. Přístup pomocí detekce ploch detekuje plochy v celém segmentovaném videokymogramu, proto po jeho aplikaci zůstal výsledek nezměněn (z pohledu algoritmu není co očistit, vše spadá do plochy



Obrázek 8.1: Průběh algoritmu segmentace videokymogramu se správným výsledkem, soubor *DP LHSV-41*.

spojené s glottis). Přístup pomocí narůstání oblasti na druhou stranu nepracuje s celým videokymogramem, ale prochází ho po sloupcích a hledá části, které s glottis nejsou spojeny. Z tohoto důvodu si lze všimnout, že algorit-

mus očistění částečně odstraní špatně segmentované kusy, tak jak by měl, ale nedokáže nic udělat s částmi, které jsou s domnělou glottis spojeny.

Nad tímto videozáznamem byla ukázána jedna z hlavních příčin, proč segmentace ve videokymogramu selhává - záblesky. Odrazy světla na hlasivkách způsobují rozhození histogramu a díky tomu ve výsledku selže segmentace a algoritmy očistění s tím už nic nezmohou.

Výsledky nad celým datasetem

Výše byly podrobně rozebrány výsledky pro jednu úspěšnou a jednu neúspěšnou segmentaci. Tyto videozáznamy byly vybrány jako reprezentativní příklady nad nimiž lze rozebrat jednotlivé vlastnosti a neduhy navrženého řešení.

V rámci této práce však proběhlo testování nad více než dvěma soubory, celkově bylo otestováno a ohodnoceno 48 videozáznamů LHSV. Segmentace byla hodnocena + nebo - dle toho zda se povedla a ve výsledné segmentaci se nachází pouze glottis. Výsledky pro celý dataset lze nalézt v Tabulce 8.2, konkrétně ve sloupcích Segmentace Otsu a Segmentace Kittler.

Tabulka 8.2: Výsledky segmentace nad celým datasetem.

Celkové skóre	Segmentace Otsu	Segmentace Kittler
+	20	3
-	28	45
	42%	6%

Jak lze vidět, segmentace pomocí metody nejmenší chyby pro videokymogramy nefunguje. Prahy jsou konstantně voleny špatně a většinou v několika videokymogramech pak dochází k tomu, že ve výsledné segmentaci je místo glottis skoro celý řádek.

Přístup pomocí optimálního prahu je na tom o něco lépe, byť také nedosahuje úspěšnosti, kterou jsme si představovali. Nejčastějším důvodem, proč segmentace Otsu selhává jsou záblesky, jejichž vliv na segmentaci byl již rozebrán výše.

Volba metody očistění na celkový výsledek vliv neměla, protože pokud selhala segmentace, metoda očistění s tím nedokáže nic udělat a pokud metoda neselhala, vybrat glottis zvládají všechny. Použití odhadu osy mělo na segmentaci vliv pouze u jedné hlasivky (3 různé videozáznamy), neboť tato hlasivka byla příliš nesymetrická pro naivní přístup, kde odhadujeme polohu glottis poblíž středu *ROI*.



Obrázek 8.3: Vliv volby *ROI* na výslednou segmentaci, širší *ROI* vpravo je zvoleno pro demonstraci vlivu volby *ROI*, správná volba je *ROI* vlevo.

matických metod. Zde pak nastává otázka, jak dobře dokáží automatické metody *ROI* zvolit, neboť každý řádek, který bude chybět či přebývat ve videokymogramu, ovlivňuje histogram, ze kterého vycházíme. Obecně i malé pohyby (zvětšení a zmenšení *ROI* o pár pixelů) mohou způsobit značné rozdíly v segmentaci, neboť počet studovaných pixelů videokymogramu se tím vcelku mění. Náznak lze vidět na Obrázku 8.3, kde bylo záměrně *ROI* vpravo zvoleno příliš široké a ve výsledku dojde ke špatné volbě prahu.

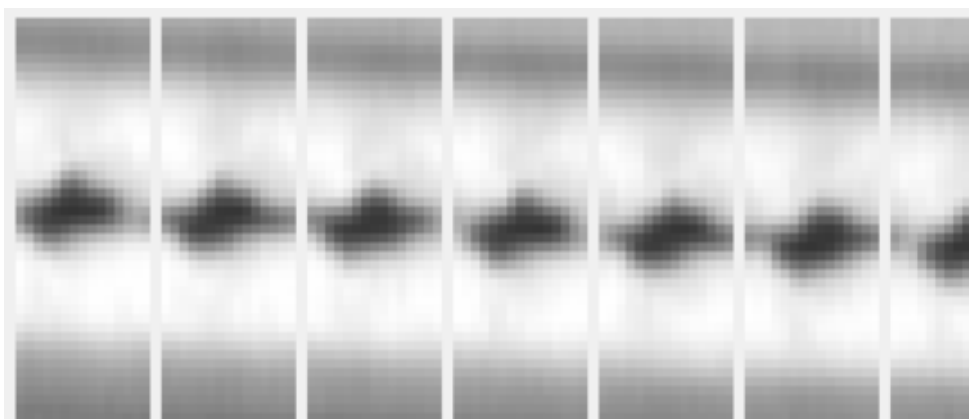
8.2.2 Výsledky segmentace videokymogramu pomocí základní hlasivkové frekvence

Druhá metoda navržená v rámci této práce se jak již bylo zmíněno zabývala prací s jedním kmitem hlasivek.

Určení frekvence

Nejprve bylo potřeba určit základní hlasivkovou frekvenci F_0 . V rámci této práce byl navržen přístup, jak takovouto hlasivkovou frekvenci určit uvnitř *ROI* - viz Kapitola 6.2.1. Pomocí F_0 lze pak jednotlivé videokymogramy rozdělit na periody, jak lze vidět například na Obrázku 8.4. Lze si také povšimnout, že poslední perioda není celá a jak bylo popsáno v kapitole s návrhem metody, tato necelá perioda bude v průběhu algoritmu zanedbána.

Algoritmus určení frekvence úspěšně určuje frekvenci pro 43 z 48 videozáznamů s tím, že u 4 je perioda určena o 1 snímek (neboli sloupec videokymogramu) mimo a u jednoho videozáznamu se liší o 2.



Obrázek 8.4: Ukázka videokymogramu rozděleného na jednotlivé periody.

Perioda ve videokymogramech obecně není celočíselná, ale na rozdělení videokymogramu celé číslo potřebujeme. Proto může docházet k posunu o několik málo pixelů v posledních periodách (nasčítaná chyba zaokrouhlení). Jednotlivé segmenty dále průměrujeme (resp. bereme jejich maximum nebo minimum), a proto musí být všechny úseky stejně dlouhé. Tato chyba spojená se zaokrouhlením by se výrazněji mohla projevit při zpracovávání delších videozáznamů, ale tato práce se záměrně omezuje na videozáznamy kratší délky (100 snímků), kde k vlivu této chyby nedochází.

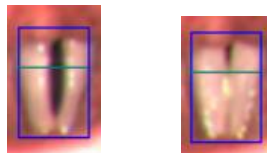
Podrobný rozbor metody při úspěchu

Nyní se lze již podívat na výsledky samotného algoritmu poté, co získáme videokymogram rozdělený pomocí hlasivkové frekvence. Opět byly vybrány dva vzorové videozáznamy, nad kterými lze ukázat vlastnosti tohoto přístupu a na konci bude opět rozebrána statistika pro celý dostupný dataset.

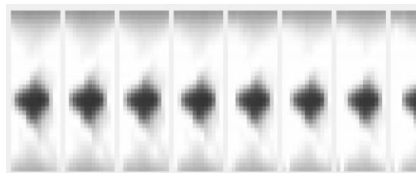
První příklad výstupu tohoto algoritmu (opět ilustračně pro jeden videokymogram) lze vidět na Obrázku 8.5. Na obrázku lze opět vidět původní zvolené *ROI* v momentech největšího a nejmenšího otevření, dále také videokymogram rozdělený pomocí periody vycházející ze základní hlasivkové frekvence - perioda tohoto videozáznamu je rovna 13 snímkům.

Dále jsou na obrázku výstupy obou segmentací, metoda nejmenší chyby je vlevo, metoda optimálního prahu je vpravo. Lze si na první pohled všimnout, že metoda nejmenší chyby opět nefunguje, tak jak bychom si představovali a volí práh špatně, obdobně jako tomu bylo u předchozí metody zkoumající celý videokymogram. Práh je opět zvolen příliš vysoko a většina i mírně šedých oblastí videokymogramu skončí ve skupině, kde by měla zůstat pouze glottis.

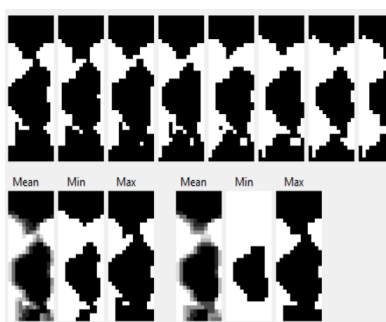
ROI v bodě největšího a nejmenšího otevření



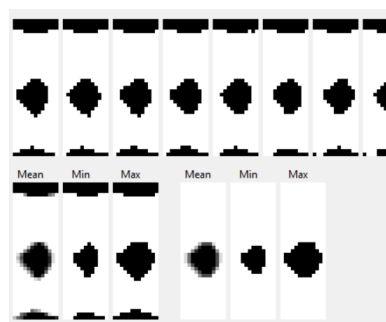
Průběh pro jeden kymogram



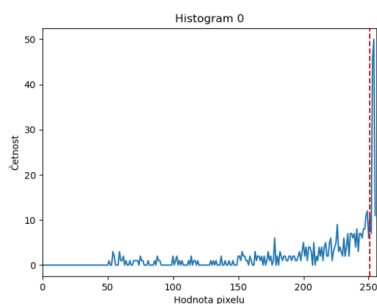
Segmentace, výběr reprezentativního snímku a jeho očištění – metoda nejmenší chyby



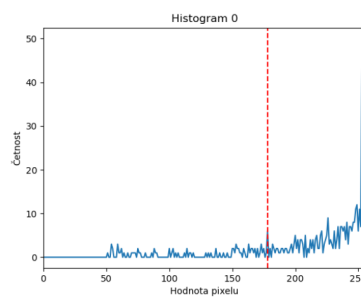
Segmentace, výběr reprezentativního snímku a jeho očištění – metoda optimálního prahu



Histogram pro první segment – práh zvolen metodou nejmenší chyby



Histogram pro první segment – práh zvolen metodou optimálního prahu



Obrázek 8.5: Průběh algoritmu segmentace videokymogramu pomocí základní hlasivkové frekvence se správným výsledkem, soubor *DP LHSV-38*.

Metoda segmentace Otsu na druhou stranu nalezne práh správně, pouze zde opět dochází k ponechání částí hlasivek v segmentaci, což, jak již bylo ukázáno, není žádný problém pro algoritmus očištění.

Vliv volby algoritmu očištění byl již podrobně ukázán v předchozí kapitole, opět platí, že nehledě na volbu algoritmu očištění, očištění segmentované glottis funguje, pokud před ním neselhala samotná segmentace.

V neposlední řadě lze na Obrázku 8.5 vidět vybrané reprezentativní snímky, ze kterých bude posléze složen segmentovaný videozáznam jednoho kmitu. Reprezentativní snímky jsou získány jako průměr / minimum / maximum po pixelech jednotlivých period.

Nejzajímavější obecně je průměrný snímek, neboť zachycuje rozdíly mezi jednotlivými periodami. To znamená, že pokud bychom měli hlasivku, kde dochází k nepravidelnému pohybu třeba každý čtvrtý kmit, byly by tyto kmity v průměrném videozáznamu vidět jako šedé kusy místo černých. Dále zde lze nalézt i minimum a maximum (díky tomu, že glottis je černou barvou (hodnota 0), minimum je největší glottis a maximum nejmenší). Z minimálního snímku lze například nalézt jak moc se hlasivka otevírá v průběhu videozáznamu. Maximální snímek na druhou stranu ukazuje část hlasivky nacházející se v každém snímku.

Nakonec jsou na Obrázku 8.5 uvedeny histogramy pro jednu periodu (zde pro první). Je opět vidět, že zvolené prahy jsou poměrně odlišné a tomu odpovídá i výsledná segmentace.

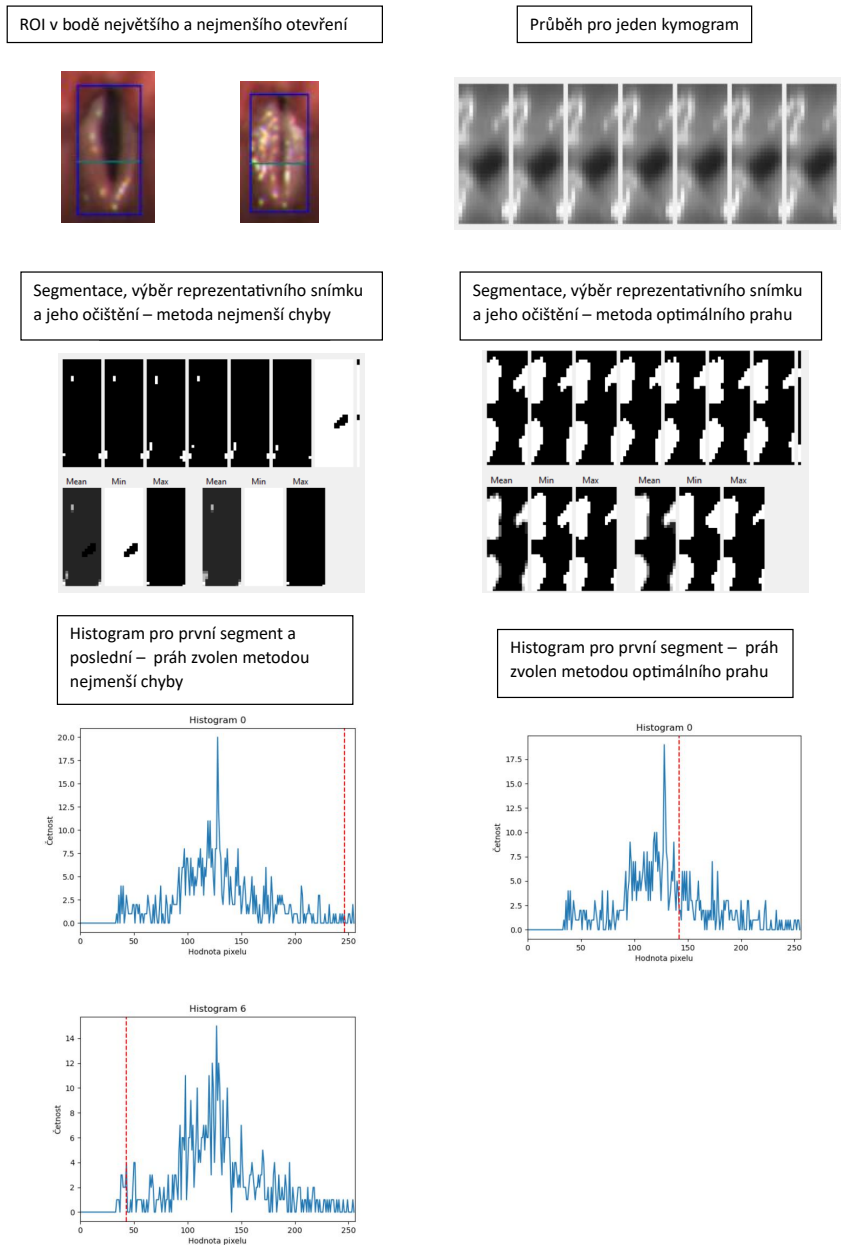
Podrobný rozbor metody při selhání

Druhým ukázkovým vstupem je vstup z Obrázku 8.6. Jedná se opět o méně kvalitní vstup obsahující již zmíněné záblesky, které lze zřetelně vidět jako bílé oblasti na rozděleném videokymogramu na Obrázku 8.6 nahoře.

Obě metody prahování zde opět selhávají. Co je na tomto videozáznamu nejzajímavější, je ukázka toho, jak je metoda nejmenší chyby nestabilní, co se volby prahu týče. Na obrázku lze vidět dva histogramy, horní je pro první periodu, dolní je pro poslední (celou) periodu. Byť jsou periody na první pohled poměrně podobné a i jejich histogramy se příliš neliší tvarem, metoda nejmenší chyby volí prahy úplně v jiných místech - mimochodem oba špatně, byť ten druhý se blíží ideálnímu prahu. Tento problém nastává poměrně často a je to jeden z předních důvodů proč metoda nejmenší chyby obecně selhává. Volba prahu je nekonzistentní, což vede k rozdílům mezi poměrně podobnými videozáznamy a ve výsledku to vede ke vzniku černých pruhů ve výsledném segmentovaném videu.

Metoda Otsu, byť zde selhala, tak prahy volí konzistentně, jak lze vidět ze segmentovaných úseků, což je i jeden z důvodů, proč je obecně úspěšnější než metoda nejmenší chyby.

Obdobně jako u předchozí metody nedokáží metody očištění napravit chyby segmentace, neboť na to nejsou stavěny.



Obrázek 8.6: Průběh algoritmu segmentace videokymogramu pomocí základní hlasivkové frekvence při selhání metod segmentace, soubor *DP LHSV-26*.

Výsledky nad celým datasetem

Výše byly opět podrobně rozebrány výsledky pro jeden videozáznam, kde došlo k úspěchu, a pro jeden, kde došlo k neúspěchu metody. V rámci této práce probíhalo testování nad již zmíněným datasetem 48 videozáznamů LHSV a výsledky ve stejném formátu jako v předchozí kapitole (hodnocení + / -) lze najít v následující Tabulce 8.3.

Tabulka 8.3: Výsledky segmentace pomocí hlasivkové frekvence nad celým datasetem.

Celkové skóre	Otsu + F_0	Kittler + F_0	Určena F_0
+	20	3	48
-	28	45	0
	42%	6%	100%

Výsledky jsou stejné jako pro předchozí metodu, segmentace uvnitř jednotlivých period trpí naprosto stejnými problémy, jaké byly uvedeny u předchozí metody a zmenšení zpracovávaného videozáznamu nemá příliš velký dopad. Záblesky jsou i zde velký problém, který brání určení optimálního prahu.

Jak již bylo zmíněno, navržené určení frekvence je funkční pro všechny dostupné videozáznamy.

8.3 Celkové výsledky

V Tabulce 8.4 lze najít výsledky pro celý anonymizovaný dataset 48 videozáznamů LHSV.

První tři sloupce tabulky obsahují název datového souboru, diagnózu pacienta a základní hlasivkovou frekvenci určenou z audiozáznamu.

Poté následují sloupce, které ukazují hodnocení výsledků metody segmentace videokymogramů pro metody segmentace Otsu a Kittler. Hodnocení +/- zde značí úspěch / neúspěch metody segmentace.

Dále následují sloupce, které ukazují určenou frekvenci uvnitř ROI, z ní určenou periodu v počtu snímků a poté opět výsledky segmentace pomocí metod Otsu a Kittler.

Poslední sloupec ukazuje vliv určení osy při aplikaci algoritmu očištění. Plus zde značí, že určení osy vliv mělo, mínus značí, že určení osy výslednou segmentaci neovlivnilo.

V rámci tabulky jsou navíc barevně označeny 4 záznamy, které byly představeny v rámci rozborů výsledků uvnitř předešlé Kapitoly 8.2.

Tabulka 8.4: Výsledky nad anonymizovaným datasetem.

Metody segmentace glottis			Segmentace kymogramu		Segmentace kymogramu s využitím základní hlasivkové frekvence				Mělo vliv určení osy symetrie hlasivek?
Datový soubor	Dg	LHSV-F0 [Hz]	Otsu	Kittler	F0 [Hz]	perioda v ROI [snímky]	Otsu	Kittler	
DP LHSV-1	cys	201	-	-	197	20	-	-	-
DP LHSV-2	cys	222	+	-	237	17	+	-	-
DP LHSV-3	cys	249	-	-	276	14	-	-	-
DP LHSV-4	chor	231	-	-	250	16	-	-	+
DP LHSV-5	chor	230	+	-	250	16	+	-	+
DP LHSV-6	chor	219	-	-	237	17	-	-	+
DP LHSV-7	chor	240	-	-	237	17	-	-	-
DP LHSV-8	kar	298	+	-	250	16	+	-	-
DP LHSV-9	kar	125	-	-	131	30	-	-	-
DP LHSV-10	par	202	-	-	210	19	-	-	-
DP LHSV-11	par	150	+	-	158	25	+	-	-
DP LHSV-12	par	187	-	-	184	22	-	-	-
DP LHSV-13	par	190	-	-	197	20	-	-	-
DP LHSV-14	par	137	-	-	131	30	-	-	-
DP LHSV-15	par	225	+	-	237	17	+	-	-
DP LHSV-16	par	228	+	-	237	17	+	-	-
DP LHSV-17	par	238	+	-	237	17	+	-	-
DP LHSV-18	par	215	+	-	237	17	+	-	-
DP LHSV-19	par	204	+	-	197	20	+	-	-
DP LHSV-20	par	205	+	+	197	20	+	+	-
DP LHSV-21	pol	136	-	-	131	30	-	-	-
DP LHSV-22	pol	120	-	-	131	30	-	-	-
DP LHSV-23	pol	152	-	-	158	25	-	-	-
DP LHSV-24	pol	129	-	-	131	30	-	-	-
DP LHSV-25	uzl	198	-	-	197	20	-	-	-
DP LHSV-26	uzl	279	-	-	276	14	-	-	-
DP LHSV-27	uzl	248	+	-	250	16	+	-	-
DP LHSV-28	uzl	192	-	-	184	22	-	-	-
DP LHSV-29	uzl	191	-	-	197	20	-	-	-
DP LHSV-30	uzl	161	-	-	158	25	-	-	-
DP LHSV-31	uzl	205	-	-	197	20	-	-	-
DP LHSV-32	zdr	215	-	-	210	19	-	-	-
DP LHSV-33	zdr	221	+	-	250	16	+	-	-
DP LHSV-34	zdr	252	-	-	250	16	-	-	-
DP LHSV-35	zdr	227	+	+	237	17	+	+	-
DP LHSV-36	zdr	236	-	-	250	16	-	-	-
DP LHSV-37	zdr	252	-	-	276	14	-	-	-
DP LHSV-38	zdr	310	+	-	302	13	+	-	-
DP LHSV-39	zdr	425	-	-	434	9	-	-	-
DP LHSV-40	zdr	272	-	-	302	13	-	-	-
DP LHSV-41	zdr	271	+	-	250	16	+	-	-
DP LHSV-42	zdr	290	+	+	302	13	+	+	-
DP LHSV-43	zdr	219	-	-	237	17	-	-	-
DP LHSV-44	zdr	226	+	-	237	17	+	-	-
DP LHSV-45	zdr	227	+	-	237	17	+	-	-
DP LHSV-46	zdr	334	+	-	328	12	+	-	-
DP LHSV-47	zdr	277	-	-	276	14	-	-	-
DP LHSV-48	zdr	365	+	-	368	11	+	-	-

9 Závěr

V rámci této diplomové práce jsem se zabýval zpracováním videozáznamu z vysokorychlostní kamery. Úkolem bylo ve videozáznamu hlasivek vytvořeného laryngoskopem nalézt hlasivkovou štěrbinu (neboli glottis). Navržené metody zpracovávají místo jednotlivých snímků videozáznamu videokymogram, což je průběh jednoho řádku videozáznamu seskládaný v čase do obrázku. Až nad takovýmto obrázkem byly aplikovány metody segmentace. V rámci práce byly navrženy dvě metody segmentace - první byla aplikována přímo na videokymogram bez jakéhokoliv předzpracování, druhá metoda nejprve určila automaticky hlasivkovou frekvenci a pomocí té videokymogram rozdělila na několik částí, z nichž každá zachycovala jednu periodu. Z těchto period byl dále vytvořen reprezentativní videozáznam jedné periody a na ten byly teprve aplikovány metody segmentace. Po aplikaci metod segmentace bylo potřeba výsledný videozáznam očistit díky předpokladu, že se glottis nachází cca uprostřed videozáznamu. Po očištění lze u výsledné segmentace spočítat mnoho parametrů, v rámci této práce byla především sledována osa hlasivky, obvod a obsah glottis a poměr mezi levou a pravou částí hlasivky.

V první části této práce byla popsána anatomie hlasivek a různé problémy spojené se vznikem hlasu. Následně byly rozebrány nejrůznější metody vyšetření hlasivek od akustických až po optické, na jejichž výstup se tato práce soustředila.

V další části byly probrány různé existující algoritmy pro detekci objektů v obraze, speciálně glottis.

Ve páté kapitole bylo rozebráno zavedené značení pro další kapitoly. Jako hlavní metody prahování, na které se práce soustředí, byly představeny metody prahování Otsu a Kittlerova metoda nejmenší chyby.

Šestá kapitola se zabývá samotným návrhem metod na zpracování videokymogramu. Je zde představen vstup navržených metod a posléze samotné metody segmentace, očištění a zpětné rekonstrukce.

Následující kapitola představuje program „Kymogram analyzer“ vytvořený v rámci této diplomové práce. Program je napsán v Pythonu, GUI je vytvořeno v knihovně Tkinter a pro algoritmy segmentace je využito knihoven Numpy a OpenCV. Program umožňuje zpracování videozáznamů ve formátu *.avi* pomocí navržených algoritmů, určení parametrů u výstupu či uložení výstupu.

V poslední kapitole před závěrem jsou rozebrány vlastnosti jednotlivých algoritmů nad různými videozáznamy a celková úspěšnost nad datasetem 48 videozáznamů LHSV.

Obě metody jsou náchylné na tmavá data, především na různé stíny okolo glottis, které jsou obvykle přidány do oblasti označené jako glottis v segmentaci a není možno se jich zbavit bez lidského zásahu. Druhý častý důvod pro selhání metody segmentace jsou záblesky světla, které rozhazují histogram, ze kterého metody prahování vycházejí.

Obecně fungovaly lépe metody využívající metodu optimálního prahu pro prahování (42%). Metoda nejmenší chyby se ukázala být nad poskytnutým datasetem nefunkční až na několik málo výjimek.

V rámci testování jsem se pokoušel předejít zábleskům pomocí mediánového filtru, který odstraňuje alespoň částečně šum a také pomocí úpravy kontrastu a jasu. Oba tyto přístupy se neosvědčily, záblesky jsou příliš velké pro odstranění filtrem a navíc ještě dochází k rozmazání glottis, neboť zpracovávané *ROI* je poměrně malé.

Dále došlo k pokusům s manuální segmentací, ve velké většině neúspěšných segmentací lze v histogramu videokymogramu nalézt manuálně lepší práh, než našly metody automatické segmentace, ale manuálně by bylo nutno takovýto práh najít pro každý řádek *ROI*, což je nemožné z pohledu uživatele.

Program vytvořený v rámci této práce umožňuje export segmentované glottis pro další zpracování, případně existuje možnost program libovolně rozšiřovat.

Závěrem je nutné poznamenat, že dvě základní techniky automatické segmentace, tj. metoda Otsu a Kittler, které byly v této práci použity pro segmentaci videokymogramu, nejsou jedinou možností segmentace. Jejich poměrně nízká úspěšnost, zejména u metody „Kittler“, neznamená nutně definitivní uzavření cesty testování tímto směrem segmentování glottis z videokymogramu. Segmentace glottis je posloupnost na sebe navazujících kroků (preprocessing, segmentace a postprocessing) a je možné použité algoritmy zaměnit za jiné a otestovat jejich výsledky.

Literatura

- [1] ANDRADE-MIRANDA, G. et al. Laryngeal Image Processing of Vocal Folds Motion. *Applied Sciences*. 2020, 10, 5. ISSN 2076-3417. doi: 10.3390/app10051556. Dostupné z: <https://www.mdpi.com/2076-3417/10/5/1556>.
- [2] BAIEROVÁ, C. *Frekvenční analýza kmitů hlasivkové štěrbinyk* [online]. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2018. [cit. 2023/05/15]. Vedoucí: Ing. Pavel Nový, Ph.D. Dostupné z: <http://hdl.handle.net/11025/31793>.
- [3] BLANCO, M. – CHEN, X. – YAN, Y. A Restricted, Adaptive Threshold Segmentation Approach for Processing High-Speed Image Sequences of the Glottis. *Engineering*. 01 2013, 05, s. 357–362. doi: 10.4236/eng.2013.510B072.
- [4] BŘESKÁ, H. *Poruchy hlasu z pohledu hlasových profesionálů* [online]. Masarykova univerzita, pedagogická fakulta, Katedra speciální a inkluzivní pedagogiky, Brno, 2019. [cit. 2023/03/12]. Dostupné z: https://is.muni.cz/th/gfmgf/DP_Breska.pdf.
- [5] COLAH. *Understanding LSTM Networks* [online]. 2015. Dostupné z: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- [6] COLLINS, S. R. Direct and Indirect Laryngoscopy: Equipment and Techniques Discussion. *Respiratory Care*. 2014, 59, 6, s. 850–864. ISSN 0020-1324. doi: 10.4187/respcare.03033. Dostupné z: <https://rc.rcjournal.com/content/59/6/850>.
- [7] ETTLER, T. *Analýza vysokorychlostního záznamu kmitání hlasivek* [online]. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2012. [cit. 2023/03/10]. Vedoucí: Ing. Pavel Nový, Ph.D. Dostupné z: <https://theses.cz/id/izuug5/>.
- [8] ETTLER, T. *Detekce a hodnocení videozáznamu pohybu hlasivek z vysokorychlostní kamery* [online]. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, Plzeň, 2017. [cit. 2023/03/25]. Vedoucí: Ing. Pavel Nový, Ph.D. Dostupné z: <https://www.kiv.zcu.cz/~novyp/foniatry/lhsv%20analysis%20origo.pdf>.

- [9] FEHLING, M. K. et al. Fully automatic segmentation of glottis and vocal folds in endoscopic laryngeal high-speed videos using a deep Convolutional LSTM Network. *PLOS ONE*. 02 2020, 15, 2, s. 1–29. doi: 10.1371/journal.pone.0227791.
- [10] GRAY, H. *Anatomy*. Elsevier, 2008. původně publikováno 1918.
- [11] GÓMEZ, P. et al. BAGLS, a multihospital Benchmark for Automatic Glottis Segmentation. *Sci Data*. 2020, 7, s. 186. doi: doi.org/10.1038/s41597-020-0526-3.
- [12] KITTLER, J. – ILLINGWORTH, J. Minimum error thresholding. *Pattern Recognition*. 1986, 19, 1, s. 41–47. ISSN 0031-3203. doi: https://doi.org/10.1016/0031-3203(86)90030-0.
- [13] NOVÁK, A. *Foniatrie a pedaudiologie. Poruchy hlasu - základy fyziologie hlasu, diagnostika, léčba , reedukace a rehabilitace*. Unitisk Praha, 1996.
- [14] NOVÝ, P. Analýza obrazové scény. Učební text k přednáškám předmětu KIV/AOS, . Katedra informatiky a výpočetní techniky, FAV ZČU v Plzni, 2022.
- [15] NOVÝ, P. Základy obecné fyziologie člověka. Učební text k přednáškám předmětu KIV/ZOF, . Katedra informatiky a výpočetní techniky, FAV ZČU v Plzni, 2022.
- [16] NOVÝ, P. Zpracování vizuální informace. Učební text k přednáškám předmětu KIV/ZVI, . Katedra informatiky a výpočetní techniky, FAV ZČU v Plzni, 2022.
- [17] OPENCV. *Knihovna OpenCV* [online]. OpenCV. [cit. 2023/03/16]. Dostupné z: <https://docs.opencv.org/4.x/index.html>.
- [18] OTSU, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. 1979, 9, 1, s. 62–66. doi: 10.1109/TSMC.1979.4310076.
- [19] PINHEIRO, A. P. et al. Graphical Evaluation of Vocal Fold Vibratory Patterns by High-Speed Videolaryngoscopy. *Journal of Voice*. 2014, 28, 1, s. 106–111. ISSN 0892-1997. doi: https://doi.org/10.1016/j.jvoice.2013.07.014. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0892199713001501>.
- [20] PYTHON SOFTWARE FOUNDATION. *tkinter* [online]. Dostupné z: <https://docs.python.org/3/library/tkinter.html>.

- [21] SUZUKI, S. – BE, K. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*. 1985, 30, 1, s. 32–46. ISSN 0734-189X. doi: [https://doi.org/10.1016/0734-189X\(85\)90016-7](https://doi.org/10.1016/0734-189X(85)90016-7).
- [22] WIKIPEDIA. *Long short-term memory* — *Wikipedia, The Free Encyclopedia* [online]. 2023. Dostupné z: https://en.wikipedia.org/w/index.php?title=Long_short-term_memory&oldid=1148032239.
- [23] WIKIPEDIA. *Laryngoscopy* — *Wikipedia, The Free Encyclopedia* [online]. 2023. Dostupné z: <https://en.wikipedia.org/wiki/Laryngoscopy>.
- [24] WIKIPEDIA. *Otsu's method* [online]. 2023. Dostupné z: https://en.wikipedia.org/w/index.php?title=Otsu%27s_method.
- [25] ŠVEC, J. G. – SCHUTTE, H. K. Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*. 1996, 10, 2, s. 201–205. ISSN 0892-1997. doi: [https://doi.org/10.1016/S0892-1997\(96\)80047-6](https://doi.org/10.1016/S0892-1997(96)80047-6).
Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0892199796800476>.

A Přehled zkratk

LHSV - Laryngeal High-Speed Videoendoscopy

ROI - Region of interest

FPS - snímky za sekundu

DFT - diskrétní Fourierova transformace

RTG - rentgen

CT - počítačová tomografie

MR - magnetická rezonance

B HRES ENDOCAM 5562



The visible voice



HRES ENDOCAM 5562

Laryngoscopic diagnosis system

High speed for the doctors – High speed for the patients

Digital high-speed videophotography in diagnosis and therapy

With its increasing number of professions relying on voice, modern society places ever greater demands on our speech organs. As a result, dysphonia is on the rise in a wide variety of professions.

The voice is the acoustic product of the larynx function and the structure of the phonation organs. In addition to morphological evaluation, the functional examination of the larynx is therefore of special diagnostic importance. Effective diagnostics is thus an essential prerequisite for selective, individual therapy to restore or improve speech quality and performance. According to the recommendations of the European Laryngological Society as well, research and evaluation of the speech organs form an essential part of voice diagnostics.

Endo-stroboscopy has also been established for some time as a routine clinical procedure. The stroboscopic

principle is based on a visual illusion and facilitates detailed and nuanced evaluation of vocal cord oscillations. The essential diagnostic significance lies especially in the differentiation of organic dysphonias. However, this process is subject to limits imposed by its temporal resolution in recording and interpreting irregular vocal cord oscillations.

While an analysis of only 25 full images per second is possible with the aid of stroboscopy, high-speed videophotography as presented in this brochure has been able to capture 4000 or more images per second. It is therefore superior to clinical laryngostroboscopy in many areas of speech diagnostics. The faster resolution of digital high-speed videophotography makes it possible to record periodic deviations in individual vocal cord movements precisely and to observe opening and closing phases in the motion process of the vocal cords

without phonation. Examining the use of voice has a great diagnostic importance, especially in evaluating functional dysphonias. In addition, color display options allow not only a functional, but also a detailed morphological evaluation.

Combined with a special laryngoscope, a continuous light source, and a PC processing system, the WOLF HreS ENDOCAM 5562 provides researchers with a complete diagnostic system that allows more precise and extended evaluation of physiological and pathological voice parameters at a reasonable price.

**University Professor
Dr. Patrick Zorowka
Medical University of Innsbruck**

- Up to 4000 colour images per second
- High-resolution mode

- Processing in the form of video kymogram
- Processing in the form of glottogram

- Retrospective examination timeframe



See more than is possible



The new high-speed technologies from WOLF allow you to record movements that simply cannot be perceived by the human eye.

Up to 4000 individual colour images per second can be captured and replayed in slow motion. This shows a dynamic resolution of 20 images per vocal cord oscillation (at phonation of 200 Hz) - irrespective of the microphone signal. For the first time ever, this enables the recording of vocal cord vibrations, allowing observation of frequency fluctuations and aperiodic vibrations. All in colour. A significant advance in comparison with video stroboscopy.

Up to 4000 colour images per second

The decisive seconds



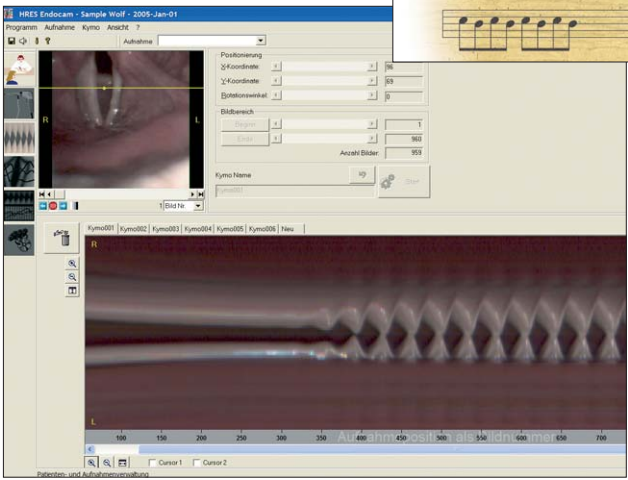
Continuous recording and storage of the last two seconds. In examination situations, that provides you with an advantage you cannot underestimate. If you spot an interesting sequence, simply stop the recording in full knowledge that what you have seen will have been recorded.

This saves your patient from having to perform repetitions for your observations, significantly shortening the examination period. And you can rely on the results.

Retrospective examination timeframe



From start to peak



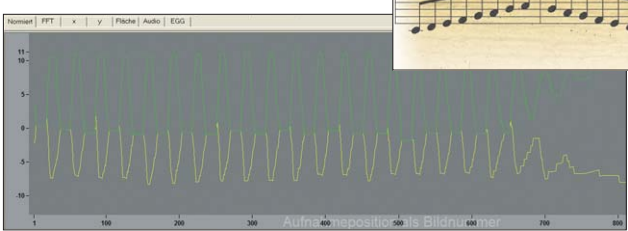
As soon as the recording has been made, the software part of WOLF's ENDOCAM solution starts to process the data. The latest developments in the video software allow it to process the recorded image information, immediately transforming it into diagnostic views.

The software allows you to define a line across any point of the glottis, for which it then transposes the sampled movement into a digital kymogram, displayed on the screen.

The large number of images means that, for the first time, you can record, view and assess the initial vibrations of the vocal cords. Special anatomical features, malformation, or pathological structures can be observed in the process of movement, providing information about air flow during speech. This especially improves functional assessment in phonation phases or aphonia. There is no need for a microphone signal during the examination, allowing a much better assessment of "hoarseness".

Processing in the form of video kymogram

Movement analysis

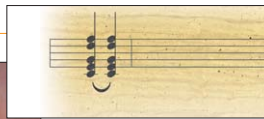


The movement of the vocal cords can also be analysed in the form of a glottogram. The user defines regions of interest, the dynamics of which are then plotted as a curve.

Processing in the form of glottogram

A system for doctors and patients

Sensitive in detail



In addition to its high-speed functionality, the WOLF ENDOCAM is also able to provide high-resolution sequences. This makes all essential diagnostic processes for the motion sequences of the vocal cords available in one system: High-speed recording of complex motion processes and high-resolution image diagnostics.

High-resolution mode

The various functions of the WOLF H_{RE}S ENDOCAM 5562 are combined with a trailblazing diagnostic system for special evaluation of the larynx function, especially in terms of how speech comes into being at the level of the glottis.

For doctor and patient alike, the ease of use of this system makes the entire examination process more comfortable and improves the validation of diagnoses significantly.

Numerous features in the software make managing and archiving data simple, with commands available at the click of the mouse.

- User-friendly system just press the button on the camera
- Modern patient management system with administration tools for appropriate information
- Differentiated archiving and documentation functions
- Voice-independent recording procedures
- High-speed recordings with up to 4000 images/second
- Up to 4 seconds retrospective observation time frame in high-speed mode

- Replay in slow motion or as individual images, directly after recording on the video monitor
- Generation of digital kymograms
- Visualise movements in the form of a glottogram
- Interactive evaluation of various visualisation forms
- Compact, space-saving system

HRES ENDOCAM 5562

Overview



	Type
HRES ENDOCAM 5562 Camera Controller	
with standard interfaces	
Colour system PAL	5562.101
Colour system NTSC	5562.701
High-performance light source AUTO LP 5132	5132.0021
incl. 300 W Xenon light module (2431.101)	
HRES ENDOCAM 5562 Camera Head	5562.901
Integrated RIVO zoom lens	
HRES Laryngoscope	
Special laryngoscope for use	
with HRES Endocam 5562, ø 9.1 mm, WL 167 mm	
HRES Laryngoscope angle of view 70°	8454.003
HRES Laryngoscope angle of view 90°	8454.002
Integrated microphone	5052.801
Air guide tube	8454.155
Removable for use with HRES laryngoscope	
Thermal protective cap	
for use with laryngoscope 70°	8454.154
for use with laryngoscope 90°	8454.156

	Type
HRES ENDOCAM 5562 Complete System	
Desktop PC incl. mouse, keyboard, Windows XP	
German	5562.401
English	5562.402
PC monitor 17" TFT	5370.003
SONY LCD monitor 14"	5370.014
Pro-Cart ENT II Unit Trolley	32113.001
only approved for Europe (CE-tested);	
base chassis, keyboard drawer with mouse pad,	
storage shelf with handle.	
Two monitor brackets for for LCD monitors,	
separate 1200 VA transformer with 6-way European safety socket block	
and 5 European safety socket cables.	
Camera head holder for video trolley	32113.513
RIVOmobil^{smart} unit trolley	31114.001

For further information and more accessories, please consult our main catalogue.

C Uživatelská dokumentace

Aplikace je do detailu popsána v Kapitole 7. Pro spuštění aplikace není potřeba žádné instalace, aplikace vytvořená v rámci této práce je dodána jako stand-alone spustitelný *.exe* soubor.

Pro videozáznam ve formátu *.avi* je možné spustit samotný program *Kymogram analyzer.exe*. Po spuštění je nutné nejprve zadat vstupní soubor s videem, a poté je již zobrazena hlavní obrazovka, která uživateli nabízí všechny metody implementované v rámci této práce. Před spuštěním metod je vhodné zvolit správné *ROI*, což je děláno pomocí myši popotažením zvýrazněné oblasti. Pro relevantní metody je také důležité zvolit pomocí myši řádek, ve kterém se bude zkoumaný kymogram tvořit (popřípadě bod). Opět je vše možné uskutečnit popotažením objektu myší.

D Obsah elektronické přílohy

- Aplikace_a_knihovny
 - Kymogram_analyzer - výsledný *.exe* soubor spolu s potřebnými soubory
 - Zdrojove_soubory - soubory *.py*
- Poster
 - Poster ve formátech *.pub* a *.pdf*
- Text_prace
 - Text práce a *.tex* soubory projekty
- Vstupni_data
 - Vstupní data pro program Kymogram analyzer
- Vysledky
 - Výsledky ve formátu *.xlsx*
- Readme.txt
 - Soubor popisující strukturu složky