

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

Diplomová práce

Frekvenční analýza kmitů hlasivkové štěrbiny

Místo této strany bude
zadání práce.

Prohlášení

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

V Plzni dne 27. června 2018

Christine Baierová

Abstract

This work aims to analyse vocal fold oscillation in HSV (High Speed Video) recordings. Work focuses on methods analysing the frequency behaviour of the vocal fold oscillations. The luminance signal is extracted from video sequence out of selected pixel and further analyzed. One of the problems is to detect the correct fundamental frequency of the luminance signal. Based on gained parameters the experimental visualisation application was developed. Methods allow to display the behaviour of vocal fold oscillations in a different manner. Next part of this work analyses differences in shape between selected periods of the signal.

Keywords: glottis, voice disorder, signal analysis, HSV, laryngotopography

Abstrakt

Práce se zabývá analýzou kmitání hlasivek ze záznamu z vysokorychlostní kamery HSV (High Speed Video). Zaměřuje se na oblast frekvenčních metod a zkoumá průběh extrahovaného jasového signálu ze zvoleného obrazového bodu napříč snímky z videozáznamu. Dílčím problémem je určení základní frekvence změny jasu a detekce period. Na základě získaných parametrů byly v experimentální aplikaci implementovány vizualizační metody, jež umožňují zobrazit chování pohybu hlasivek z jiného úhlu. Další část práce pak zkoumá dílčí kmity signálu, zejména pak odlišnost tvaru.

Klíčová slova: hlasivky, porucha hlasu, analýza signálu, HSV, laryngotopografie

Obsah

| | | |
|----------|--|-----------|
| 1 | Úvod | 7 |
| 2 | Hlasové ústrojí a tvorba hlasu | 9 |
| 2.1 | Tvorba hlasu | 10 |
| 2.2 | Poruchy hlasu | 11 |
| 2.3 | Metody vyšetření hlasivek | 12 |
| 2.3.1 | Akustické metody | 12 |
| 2.3.2 | Aerodynamické metody | 13 |
| 2.3.3 | Elektrofyzilogické metody | 14 |
| 2.3.4 | Optické metody | 15 |
| 3 | Vysokofrekvenční snímání | 17 |
| 3.1 | Metoda vyšetření vysokorychlostní kamerou | 17 |
| 3.2 | Záznamy | 18 |
| 3.2.1 | Videozáznam | 18 |
| 3.2.2 | Audiozáznam | 20 |
| 3.2.3 | Kvalita záznamu | 20 |
| 3.3 | Metody analýzy záznamu | 21 |
| 3.4 | Metody založené na segmentaci glotální oblasti | 22 |
| 3.4.1 | Segmentace hlasivkové štěrbiny | 23 |
| 3.5 | Metody založené na redukcii dimenzionality | 24 |
| 3.5.1 | Kymogram | 24 |
| 3.5.2 | Frekvenční metody | 25 |
| 4 | Frekvenční analýza videosekvence HSV | 26 |
| 4.1 | Přehled použitých metod | 26 |
| 4.2 | Návrh systému | 29 |
| 4.3 | Určení základní frekvence | 32 |
| 4.3.1 | Audiozáznam | 32 |
| 4.3.2 | Videozáznam | 36 |

| | | |
|----------|--|-----------|
| 4.3.3 | Mapování základní frekvence jasového a zvukového signálu | 38 |
| 4.4 | Analýza videosekvence | 40 |
| 4.5 | Analýza jednoho kmitu hlasivek | 43 |
| 5 | Implementace | 51 |
| 5.1 | Algoritmy | 51 |
| 5.1.1 | Rychlá Fourierova transformace | 51 |
| 5.1.2 | Detekce kmitu hlasivek | 51 |
| 5.2 | Užité technologie | 52 |
| 6 | Výsledky | 54 |
| 6.1 | Experiment č. 1: zdravé hlasivky | 56 |
| 6.1.1 | Zhodnocení | 58 |
| 6.2 | Experiment č. 2: polyp | 59 |
| 6.2.1 | Zhodnocení | 61 |
| 6.3 | Experiment č. 3: paréza | 62 |
| 6.3.1 | Zhodnocení | 64 |
| 6.4 | Experiment č. 4: cysta | 65 |
| 6.4.1 | Zhodnocení | 67 |
| 7 | Závěr | 69 |
| | Literatura | 71 |

Kapitola 1

Úvod

Hlas je u člověka odjakživa prostředkem verbální komunikace. Jeho studiem se zabývá široká škála vědních disciplín. Význam spolupráce odborníků různých specializací je při výzkumu hlasové problematiky nezbytný pro vývoj léčebných a diagnostických metod při poškození hlasu. Hlasová porucha má bezpochyby zásadní vliv na kvalitu života, u níž platí, že čím dříve je diagnostikována, tím účinnější může být její léčba.

Kromě dýchacího (plíce) a artikulačního ústrojí (ústa, jazyk), se na tvorbě hlasu podílí i ústrojí fonační, jehož nejdůležitější součástí tvoří hlasivky. Problémy, na které je práce zaměřena, se týkají právě oblasti hlasivek. Typicky se patologie projevuje například arytmií pohybu levé či pravé hlasivky nebo náhodným zakmitáním.

Hlasivky lze zachytit specializovanou, vysokorychlostní kamerou, jejíž výstupy v podobě videosekvencí zpracovává právě tato práce. Smyslem práce je přispět do oblasti diagnostiky hlasových poruch a s použitím různých metod získat informace, které mohou pomoci k včasné diagnostice chorob.

První část práce se zabývá hlasovým ústrojím a tvorbou hlasu. Součástí je i výčet metod užívaných k vyšetření hlasivek včetně vysokofrekvenčního snímání, kterému se práce podrobněji věnuje ve třetí kapitole, kde jsou rozebírány různé způsoby zpracování HSV záznamu.

Ve čtvrté kapitole je pak vlastní řešení práce spolu s popsányi metodami užitými k řešení. Kapitola od sebe odděluje dvě hlavní části této práce: analýza jasového signálu a dále analýza dílčích kmitů.

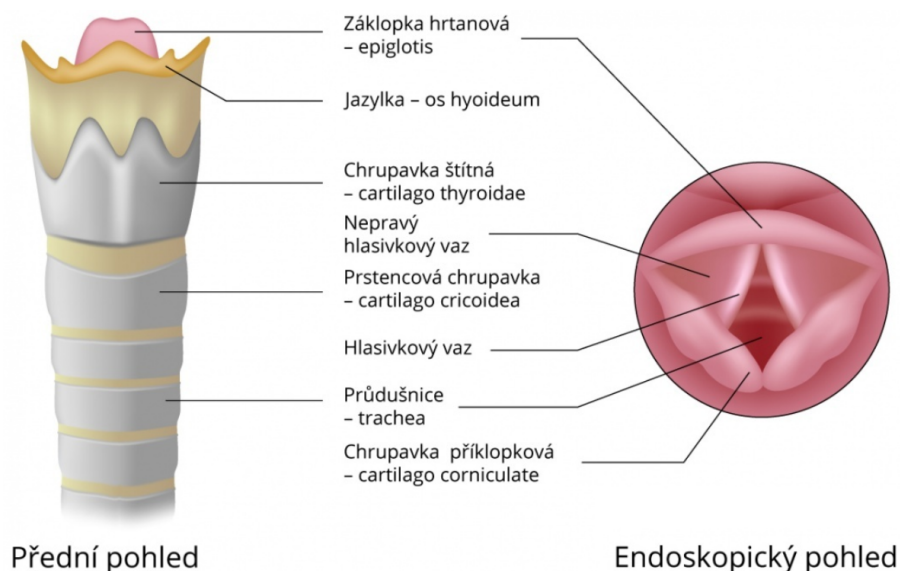
Pátá kapitola popisuje implementační záležitosti a šestá kapitola podrobně shrnuje dosažené výsledky na reálných datech. Data byla vybrána tak, aby jasně demonstrovala aplikaci metod jak na zdravých hlasivkách, tak na hlasivkách s patologickým nálezem.

Kapitola 2

Hlasové ústrojí a tvorba hlasu

Na začátku dolních cest dýchacích se nachází hrtan – nepárový dutý orgán sloužící například k dýchání (respiraci) a tvorbě zvuků (fonaci) (obr. 2.1). Vstup do hrtanu je chráněn hrtanovou příklopkou, jež brání při polykání vstup potravy do průdušnice (obr. 2.2) [22].

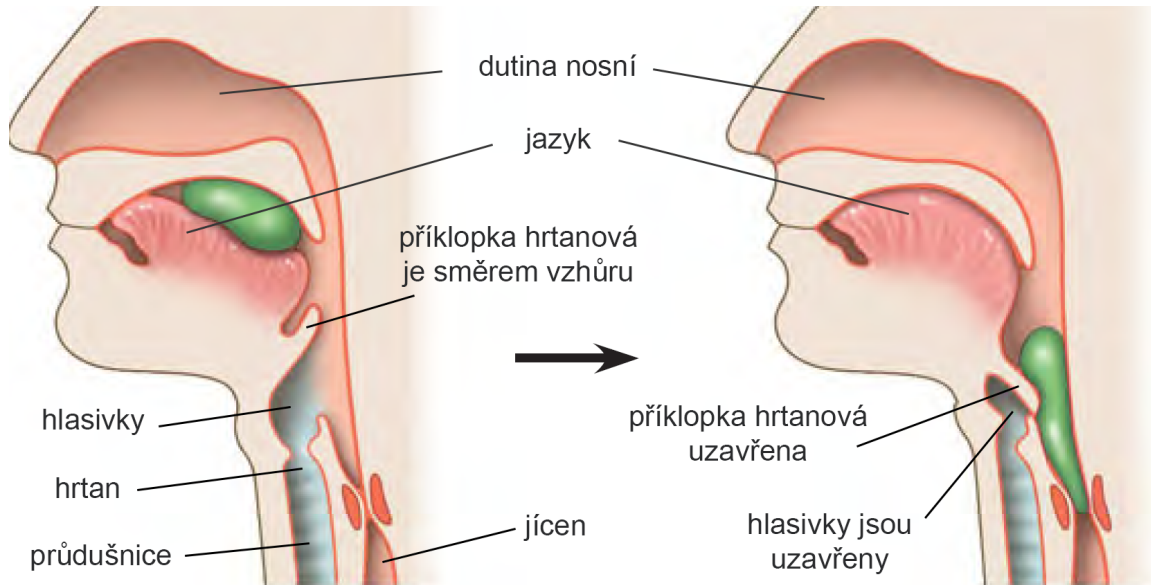
Střední část hrtanu je zúžena párem řas, řasami vestibulárními, nebo-li nepravými hlasivkami. Pod nimi je uložen pár hlasivek s hlasivkovou štěrbínou. Hlasivka je vpředu upnuta na chrupavce štítné a vzadu na chrupavce hlasivkové [9].



Obrázek 2.1: Anatomie hrtanu a hlasivek [10]

Příčně pruhované svaly hrtanu ovládají pohyby chrupavek hrtanu. Určují

tak napětí hlasivkových vazů a šířku štěrbiny mezi nimi. Hlasivky lze vidět z ústní dutiny za hrtanovou příklopkou, což se využívá při jejich vizuální analýze [9].



Obrázek 2.2: Při dýchání je příklopka hrtanová otevřena, stejně tak i hlasivky (vlevo). Při polykání příklopka uzavírá vstup do hrtanu a brání tak potravě ve vstupu do průdušnice (vpravo) [22].

2.1 Tvorba hlasu

V klidovém stavu jsou hlasivky otevřeny, což umožňuje dýchání. Při mluvení se uzavírají a svaly tak do různé míry napínají hlasivkové vazy. Ty se proudem vydechaného vzduchu rozechvívají a vzniká tón. Výška hlasu je přímo úměrná napětí hlasivkových vazů a nepřímo úměrná délce. Muži mají větší hrtan a delší hlasivkové vazy (24 mm muži, ženy 20 mm). Mužský hlas je položen níže [9].

Orgány, jež se podílejí na tvorbě hlasu, jsou následující:

1. **bránice a plíce** – generátor proudu vzduchu
2. **hrtan** – generátor hlasu
3. **násadní trubice** – (hltan, dutina ústní a nosní, vedlejší dutiny nosní) modulátor hlasu

Na začátku fonace dochází k přiblížení hlasivek do tzv. fonačního postavení, jež umožní zvýšení subglotického tlaku. Tento tlak roste působením stahu bránice, až otevře glottis díky jejím elastickým vlastnostem. Proudění vzduchu skrze glottis vyvolá tzv. Bernoulliho efekt, jenž spolu s poklesem tlaku subgloticky, nárůstem tlaku supragloticky, opět glottis uzavře. Děj se opakuje pravidelně, čímž vzniká zvuk – hrtanový tón. Hlasivky kmitají v horizontální rovině, zatímco sliznice vertikálně unduluje (způsobuje proudění vzduchu glottis). Změny v elasticitě sliznice nebo špatné postavení hlasivek narušuje správné kmitání sliznice, čímž vzniká nežádoucí chrapot. Změny v nadhrtanových prostorách mohou rovněž způsobit změnu hlasu. [15].

Vznikající zvuk je slabý a řezavý. Rezonancí v dutinách a nad hrtanem (hltan, dutina ústní a nosní, vedlejší dutiny nosní) je pak vzniklý tón formován do barvy lidského hlasu. Síla hlasu je závislá na síle, kterou je vzduch vydechován. Vlastní řeč – tvorba samohlásek a souhlásek je ovšem záležitostí dutiny ústní a hltanu. Samohlásky jsou výsledkem rezonance kdy každé z nich odpovídá určitý tvar dutiny ústní (vzniklý polohou jazyka a postavením rtů). Souhlásky vznikají průchodem vzduchu zúženými místy, například mezi zuby (s,z,t) či mezi jazykem a patrem (l,r) [9].

2.2 Poruchy hlasu

Ačkoliv je vnímání kvality hlasu čistě subjektivní a neexistuje stanovená norma hlasu, odborník dokáže snadno posoudit, zda u pacienta došlo k patologickým změnám či nikoliv. Patologická změna kvality hlasu se nazývá *dysphonia*, tedy chrapot, a může vzniknout [15]:

- změnou hmoty hlasivek (nepravidelné kmitání)
- změnou závěru hlasivkové štěrbiny (nedomykavost glottis) a únik vzduchu při fonaci

Příčinou chrapotu může být onemocnění, zranění i nadměrná zátěž hlasu. Hlasové poruchy jsou děleny na organické, jež jsou způsobeny patologicko-anatomickými změnami na hrtanových strukturách, a dále na ty funkční, kdy je hrtan v pořádku, ale narušena fonační funkce. Pro vizuální pozorování jsou podstatné pouze organické poruchy, které lze identifikovat při analýze snímku [27].

Přehled některých hlasových poruch

Hlasových poruch existuje značné množství. Pro informaci byl uveden krátký výčet možných organických poruch dle [27]:

- **Uzlík** – slizniční zbytnění hrany hlasivky, které se většinou vyskytuje na obou hlasivkách a znemožňuje jejich úplné uzavření.
- **Polyp** – Většinou jednostranný útvar ve tvaru stopky nasedající na hlasivku. Způsobuje nedomykavost glottis. Pravděpodobně vzniká extrémní mechanickou zátěží kmitáním hlasivek, jež vede k porušení cév.
- **Edém** – Výrazný otok hlasivek dosahujících i větších rozměrů, při kterém dochází k překryvu hlasivek.
- **Cysta** – Zapouzdřený útvar uvnitř hlasivky, díky němuž vzniká nerovný okraj.
- **Nádory hrtanu** – zhoubné i nezhojbné nádory hrtanu. Typicky důsledky závisejí na velikosti a umístění útvaru.
- **Poruchy inervace** – neurologická postižení v oblasti hrtanu. Může dojít k obrně nervů či k trvalým stahům svalů.

Úplný výčet možných poruch lze dohledat např. v [17].

2.3 Metody vyšetření hlasivek

Oblast medicíny, která se zabývá poruchami a zraněními hrtanu se nazývá laryngologie. Ještě však existuje obor foniatrie, jenž se zabývá vyšetřováním fyziologií, patofyziologií, diagnostikou i léčbou tvorby hlasu. Metody zabývající se vyšetřením hlasivek lze rozdělit do čtyř kategorií: akustické, aerodynamické, elektrofyziologické a optické.

2.3.1 Akustické metody

Hlasové pole

Metoda spočívá v záznamu nejtiššího a nejhlasitějšího hlasu pacienta v celém jeho frekvenčním rozsahu a je určena k získávání kvantitativních parametrů hlasu. Hodnoty jsou nanášeny do grafu, přičemž na vodorovné ose je frekvence a na svislé intenzita měřená v dB [25].

Multidimenzionální analýza

Metoda je založena na analytickém programu *Multidimensional voice program* firmy Kay Elemetrics¹ umožňující výpočet 33 parametrů z prodloužené fonace hlásky „á“, kdy doporučená doba bývá 4 vteřiny. Jeho výsledky mohou být porovnávány s naměřenými parametry populace bez hlasových problémů. Z parametrů lze jmenovat dva základní: **jitter** (popisuje stupeň poruchy periodicity základní frekvence) a **shimmer** (porucha amplitudové stability) [13].

2.3.2 Aerodynamické metody

Hlasivkovou štěrbinou prochází při fonaci určité množství vzduchu s danou rychlostí. Aerodynamické metody jsou založeny na měření množství či rychlosti procházejícího vzduchu nebo při sledování dýchání.

Pneumografie

Jde o starší metodu, jež se zabývá zkoumáním pohybu hrudníku a břicha při dýchání a fonaci. Například s pomocí mechanických snímačů upevněnými na elastických páscích je měřena změna tlaku při mechanických podnětech, tedy zmenšení a zvětšení obvodu hrudníku v průběhu dechového cyklu. Ty jsou převáděny na elektrický signál, který je zesilován a zobrazen na monitoru [25].

Pneumotachografie

Jde o metodu měřící množství a rychlost proudu vzduchu v hlasivkové štěrbině či subglotického tlaku. Pacient má při vyšetření ucpané nosní dírky a na ústech přiloženou masku. Senzitivním snímačem je pak zaznamenáván proud vzduchu a fonace mikrofonom. Měření subglotického tlaku je však problematické:

- zavedením jehly do subglotického prostoru, tedy do oblasti prstencové chrupavky (obr. 2.1). Jde však o invazivní metodu, jež není kladně přijímána pacientem ani lékařem.
- zavedením katetru a měřícího balonku do subglotické oblasti, nicméně balonek narušuje normální fyziologickou funkci při tvorbě hlasu

¹Kay Elemetrics Corp. je americká firma vyrábějící medicínský hardware a software. <http://kayelemetrics.com>

- umístěním hadičky na jazyk je možné změřit subglotický tlak nepřímo při fonaci slabiky „pá“ (obvykle jde o fonaci souhlásky „p“ následovanou samohláskou). Při fonaci této konkrétní slabiky je vzduchový sloupec od plic až po rty nepřerušen, tlak v ústech je tedy totožný s tlakem v oblasti pod hlasivkami [24].

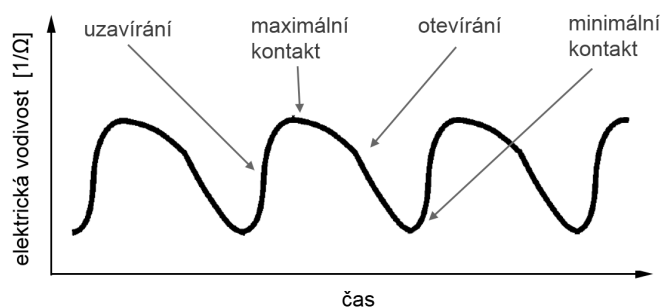
2.3.3 Elektrofyziologické metody

Další skupina se zaměřuje na změny elektrických vlastností při samotné fonaci.

Elektroglottografie (EGG)

Na kůži pokrývající oblast štítné chrupavky (obr. 2.1) se umístí snímací elektrody. Metoda se zabývá sledováním změn elektrického odporu (vodivosti), vzniklým kmitáním hlasivek a změnou tvaru hlasivkové štěrby [25]. Tato metoda sice nedokáže diagnostikovat konkrétní typy hlasových poruch, nicméně může být užita jakožto doplňující metoda, například při snímání s vysokorychlostní kamerou HSV.

Perioda EGG záznamu odpovídá základní hlasivkové frekvenci. Křivka signálu vyjadřuje plochu kontaktu hlasivek při fonaci. Popis jednotlivých fází pohybu hlasivky a vyjádření kontaktní plochy lze vidět na obrázku 2.3.



Obrázek 2.3: Popis elektroglottogramu – jednotlivé fáze pohybu a kontaktní plocha glottis

Elektromyografie (EMG)

Metoda spočívá v zavedení jehlových či povrchových elektrod. Poskytuje informace o aktivitách vnitřních hrtanových svalů v průběhu fonace. Zpravidla se využívá převážně u pacientu s obrnou zvrátého nervu, kdy se například

jedna hlasivková řasa pohybuje výrazně pomaleji oproti druhé řase. Je sledována aktivita svalů během fonace a také je ověřováno, zda došlo k přerušení zvratného nervu při obrně. Metoda je ve větší míře využívána v jiných lékařských oborech, ale i ve foniatrii si našla využití. Je měřena aktivita dílčích svalů a velikost amplitud potenciálů.

2.3.4 Optické metody

Poslední skupinu zaujímají optické metody. Obvykle je nutné se k hlasivkám co nejvíce přiblížit zasunutím přizpůsobeného zařízení do oblasti ústní dutiny.

Laryngoskopie

Při laryngoskopii se vyšetřují anatomické změny v hrtanu a funkce nitra hrtanu. Rozlišujeme *nepřímou* laryngoskopii, kdy je hrtan spolu s hlasivkami sledován pomocí zrcátka, nebo *přímou*, kdy se využívá optických vláken (např. flexibilní laryngoskop) [15].

Laryngostroboskopie

Zvláštním případem vyšetření hrtanu je zkoumání správné fonace, tedy pohybu hlasivek zvětšovací optikou při stroboskopickém světle. Díky stroboskopickému jevu dochází k optickému zpomalení kmitání hlasivek, díky čemuž je možné pozorovat jejich chování (jinak je pohyb hlasivek příliš rychlý, nepostřehnutelný lidským okem) [15]. Nevýhodou této metody je však skutečnost, že posouzení pohybu glottis je vázáno na podmínku jejich pravidelného, periodického kmitání. Pokud je hlas výrazně poškozen a tato podmínka není splněna, může vést k mylnému posuzování způsobu hlasivkových kmitů [26].

Videokymografie

Metoda spočívá v použití videokymografické kamery, jež dokáže fungovat ve 2 modech: jako běžná kamera (s frekvencí 50-ti pulsů za vteřinu) nebo vysokofrekvenční kamera. Princip vysokofrekvenčního modu (videokymografického) spočívá v přepnutí snímacího režimu kamery z plošného do řádkového modu. Na úkor vypuštěné prostorové informace pak kamera zvýší frekvenci snímání až na téměř 8000 snímků /s. Obrazy z vybraného řádku jsou skládány za sebe a poskytují tak informaci o způsobu chvění hlasivek v daném místě [26].

Vysokofrekvenční snímání

Vysokofrekvenční snímání (*High Speed Video – HSV*), je metoda, jejímž účelem je sledovat chování hlasivek kamerou, která je schopna zaznamenat více než 1000 snímků za vteřinu. Zaznamenáván je tak celý obraz hlasivek, na rozdíl od videokymografie. Metoda se objevila s vývojem technologie snímání obrazu. Podrobnější popis se nachází v následující kapitole č. 3).

Kapitola 3

Vysokofrekvenční snímání

Vyšetření optickou metodou za pomoci vysokorychlostní kamery produkuje záznamy, jež jsou vhodné k další analýze. Kapitola nejprve popisuje získání takového záznamu, jeho vlastnosti a poté metody jeho zpracování.

3.1 Metoda vyšetření vysokorychlostní kamerou

Záznamy zpracovávané v této práci byly pořízeny kamerou HRES ENDO-CAM 5562 od firmy Richard Wolf ¹ na ORL klinice ve Fakultní nemocnici Plzeň Bory. Tato kamera je schopna zachytit až 4000 snímků za vteřinu. Systém rovněž spolu s videem zaznamenává zvukový záznam fonace. Jednou z výhod tohoto systému je tedy možnost hodnocení stavu hlasivek a jejich vlastnosti ve stejném okamžiku odlišnou metodou. Začátek i konec záznamu je synchronizován.

¹Richard Wolf je německá firma zabývající se výrobou softwaru a hardwaru v oblasti zdravotnictví.



Obrázek 3.1: Vysokorychlostní kamera HRES ENDOCAM 5562 vyrobená firmou Richard Wolf

Při examinaci pacienta je kamera vložena do úst a zaznamenávána oblast hlasivek. Kamera je opatřena světelným zdrojem, jenž osvětluje snímáný prostor. Pacient při vyšetření vydává hlásku „á“ a kamera zaznamenává pohyb hlasivek. Vyšetření slouží k diagnostice a rovněž ke sledování vývoje hojení po operativním výkonu.

Obecně lze vyšetřením získat tyto variace výstupů:

- **HSV** – videozáznam z HSV videokamery.
- **HSV + Mikrofon** – zvukový záznam a videozáznam z HSV videokamery. Tato dvojice je získána zcela automaticky.
- **HSV + Mikrofon + EGG** – kromě zvuku a videa je navíc možné získat výstup z elektrolottografu, pokud jsou současně umístěny snímací elektrody (viz sekce 2.3.3), nicméně v praxi jde spíše o výjimečné případy.

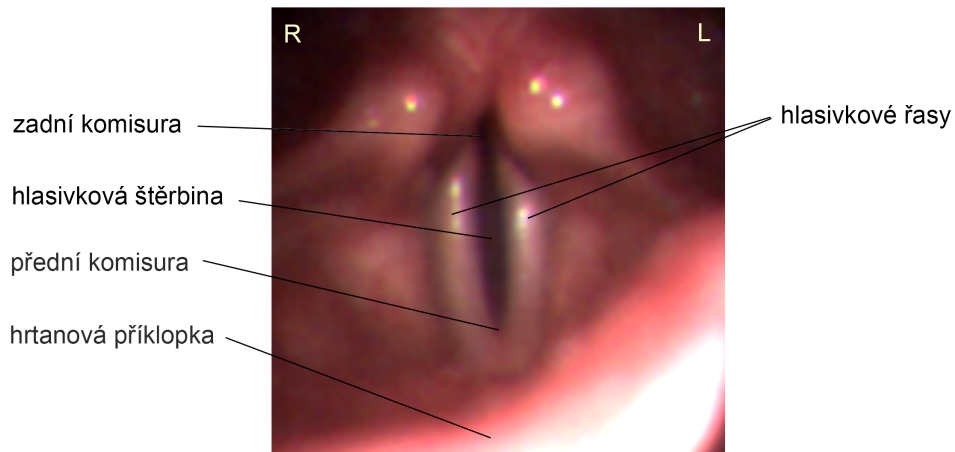
3.2 Záznamy

Videozáznam a zvukový záznam jsou běžnou součástí vyšetřovací metody, každý však sestává z jiných parametrů a oplývá jinými vlastnostmi.

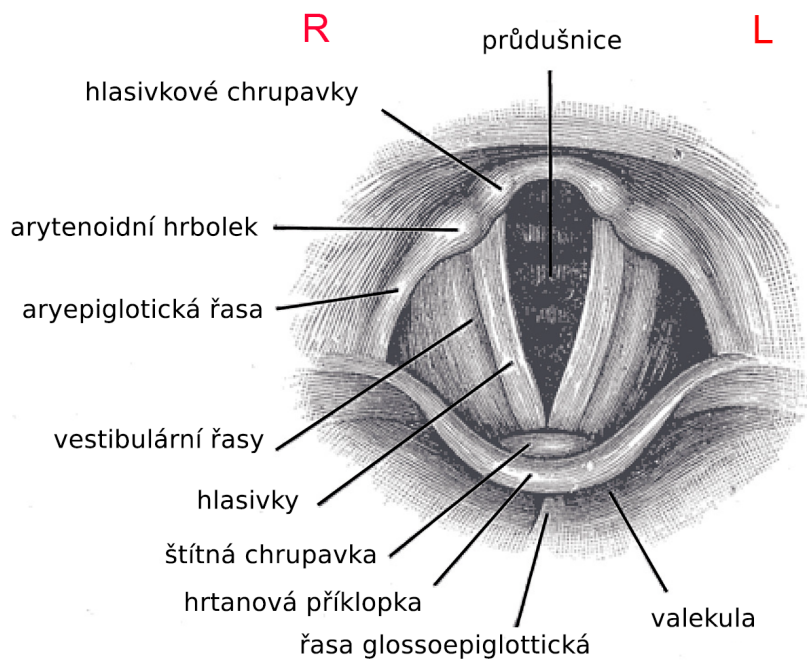
3.2.1 Videozáznam

Hlasivky mohou dle [25] kmitat zhruba od 70 Hz do 1000 Hz při fonaci. Z vysokorychlostní kamery HRES ENDOCAM (viz. Obr 3.1) tedy při 4000 snímcích za sekundu získáme až 40 snímků na jednu periodu kmitání hlasivek.

Při osvětlení hlasivek jsou hlasivkové řasy světlejší než hlasivková štěrbina, která zůstává tmavá jako na obrázku 3.2, pro srovnání byl uveden i obrázek hlasivek z anatomického slovníku, viz obr. 3.3.



Obrázek 3.2: Popis záznamu z vysokofrekvenční kamery.



Obrázek 3.3: Anatomie hrtanu – laryngoskopický pohled [7]

Parametry kamery použité pro získání záznamů v této práci jsou uvedeny v tabulce 3.1. Pro porovnání jsou uvedeny i parametry jiného výrobce vyso-

korychlostní kamery od firmy Kay Elemetrics. Vzhledem k intenzitě světla použitého pro nasvícení hlasivek je měření prováděno maximálně po dobu 4 sekund [1], [2].

| parametr | KayPentax 9710 | Wolf HREF ENDOCAM 5562 |
|---------------------------------|----------------|------------------------|
| frekvence snímání [snímky/s] | 2000 | 4000 |
| rozlišení [pixely] | 512 x 512 | 256 x 256 |

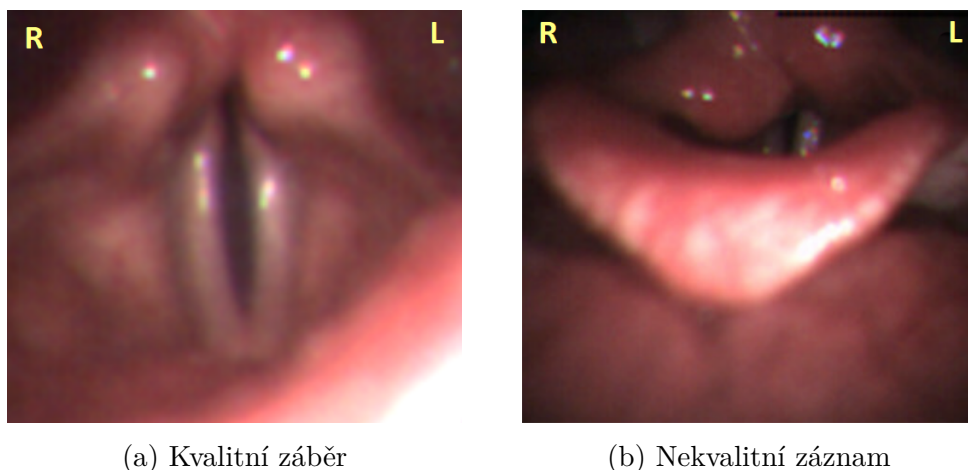
Tabulka 3.1: Porovnání vysokorychlostních kamer od dvou různých výrobců

3.2.2 Audiozáznam

Při rutinním vyšetření hlasivek s použitím prostorového mikrofonu jsou hodnoceny parametry hlasu při fonaci vokálu „á“. Při laryngoskopickém vyšetření za pomoci HSV kamery je však při tomto vokálu jazyk v nevhodné poloze. Aby došlo k žádoucímu otevření anatomického prostoru nad hlasivkami a bylo je tak možné zaměřit optikou kamery, je nutná fonace hlásky „í“. Vlivem laryngoskopu v ústech a přitlačení jazyka pak tvar akustického signálu odpovídá fonaci „e“, „e-a“, „e-i“ či neutrální hlásce „v“.

3.2.3 Kvalita záznamu

Kvalita záznamu z vysokorychlostní kamery může být negativně ovlivněna pohybem pacienta nebo kamery. Tento pohyb se projeví i při vysoké rychlosti snímání. Na čočce také může ulpět větší množství slin, což způsobí, že má obraz sníženou ostrost a kontrast. U některých pacientů se rovněž nepodaří nastavit kameru do vhodné polohy vůči hlasivkám, takže je viditelná jen v malé části záběru (obr. 3.4b).



Obrázek 3.4: Vzorčky kvality snímků při snímání vysokorychlostní kamerou.

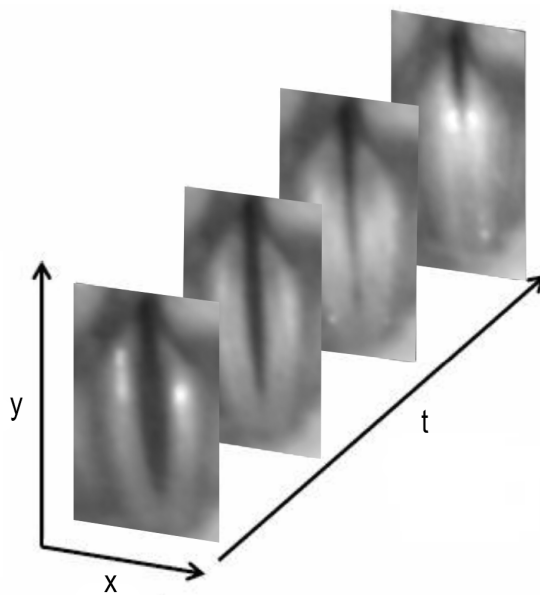
Pro účely práce jednotlivé snímky z kamery považujeme za stejnohlé, neboť je snaha při vyšetření s kamerou v ústech pacienta nehýbat. Jak již bylo zmíněno, kamera je schopná zachytit několik sekund záznamu. To však může dohromady dát 16 000 snímků. Pro naše potřeby analýzy však stačí kratší úsek frekvence trvající několik setin sekundy. Záznamy z vysokorychlostní kamery Wolf Hres Endocam, použité v této práci, jsou o rozlišení 256 x 256 pixelů a výstupní formát je primárně ukládán do souborů *BLD*². Software implementovaný v této práci však předpokládá vstup ve formátu *avi*³, který lze získat za pomoci dekodéru v práci [12].

3.3 Metody analýzy záznamu

Videozáznam je při analýze transformován do kolekce statických snímků (viz obr. 3.5). Tato kolekce uchovává nejdůležitější charakteristiky chování hlasivek.

²*BLD* je výstupní formát získaný z vyšetření kamerou HRES Endocam. Jde o ne-standardizovaný, nezdokumentovaný formát, jehož analýza je možná pouze přes zpětnou analýzu, neboli reverzní inženýrství.

³*AVI – Audio Video Interleave* je formát vyvinutý firmou Microsoft v roce 1992. Jde o multimediální kontejner, umožňující uložit více datových stop. Mohou obsahovat video, audio, efekty či text.



Obrázek 3.5: HSV záznam obsahuje dvě prostorové dimenze (x, y) a jednu časovou (t)

Neexistuje pevně stanovená hranice, která by od sebe jednotlivé metody analýzy HSV striktně oddělovala, obvykle je nezbytná jejich vzájemná kombinace a speciální interpretace výsledku, která může odhalit novou informaci. V této práci je uveden zkrácený přehled metod založených na segmentaci glotální oblasti a metod založených na redukci dimenzionality [4].

3.4 Metody založené na segmentaci glotální oblasti

Značná skupina metod analýzy HSV dat je založena na segmentaci glotální oblasti, s čímž souvisí celá škála úloh z oblasti segmentace obrazu⁴. Podstatou metod je obvykle segmentace hlasivkové štěrbiny a její osy. Po jejich vymezení lze počítat další parametry vypovídající o zdravotním stavu hlasivek. Typické je pro tyto úlohy zpracování videozáznamu snímek po snímku, čímž rostou výpočetní i paměťové nároky programu.

⁴Segmentací obrazu rozumíme skupinu metod postavených na různých principech digitálního zpracování obrazu, sloužících k automatickému rozdělení vlastního obrazu na oblasti s nějakými společnými vlastnostmi, jež mají nějaký smysluplný význam.

3.4.1 Segmentace hlasivkové štěrbiny

V HSV záznamu jsou hlasivkové řasy obarveny světlou barvou, zatímco hlasivková štěrbina zůstává tmavá. Pro účely analýzy bývá prvním krokem tuto oblast automaticky detekovat. Někdy bývá vyžadován manuální zásah uživatele, pokud automatický přístup selhává. Poté lze určit parametry odvozené z jejího tvaru, velikosti a vývoje během fáze otevírání a uzavírání hlasivek. Vyčíslené parametry jsou pak použity k detekci a klasifikaci existující poruchy hlasivek.

Samotnou detekci je možné provést několika způsoby, v práci jsou uvedeny dvě základní metody: prahování a regionální metody.

Prahování

Přímočará metoda detekce hran glottis je prahování, spočívající ve stanovení hodnoty (prahu), která vymezí, že pixely s hodnotou jasu nižší než tento práh, budou považovány za oblast hlasivkové štěrbiny (tmavší) a vyšší (světlejší) budou hlasivkové řasy. Hodnota prahu se obvykle stanovuje na základě histogramu jasů v obraze, přičemž existuje více metod, dle kterých lze jeho hodnotu stanovit. Prahování je závislé na kvalitě vstupních dat. Detailní experimenty automatizované detekce hlasivkové štěrbiny za použití různých metod prahování jsou rozebrány například v [5].

Regionální metody

Podstata regionálních metod je založena na detekci podobnosti obrazových bodů vzhledem k nějaké vlastnosti (jas, statistické vlastnosti okolí pixelu, aj.). Konstrukce postupuje zdola nahoru, tedy od jednoho pixelu po celý segment. Nejprve vybraný algoritmus rozmístí iniciální body v obraze (tzv. *seeds*), což lze například rovnoměrně či náhodně, přičemž segment pak vzniká iterativním rozrůstáním se v okolí iniciálního pixelu. Výhodou regionálních metod je odolnost vůči většímu množství šumu, což bývá u HSV záznamu přínosné. Na druhou stranu jsou závislé na umístění iniciálních bodů, čili výsledek nemusí být pokaždé stejný. U detekce hlasivkové štěrbiny je obvyklý manuální zásah uživatele.

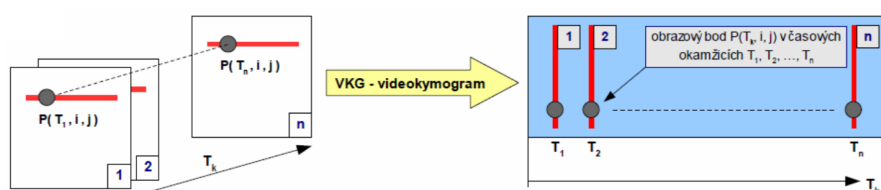
Další metody založené na segmentaci hlasivkové štěrbiny lze dohledat v literatuře, například [5].

3.5 Metody založené na redukci dimenzionality

Tato část metod analýzy HSV primárně nezávisí na segmentaci glotální oblasti, ale na redukci dimenzionality prostorově-časových dat. Temporální složka zůstává zachována, zatímco prostorová je redukována na několik obrazových bodů.

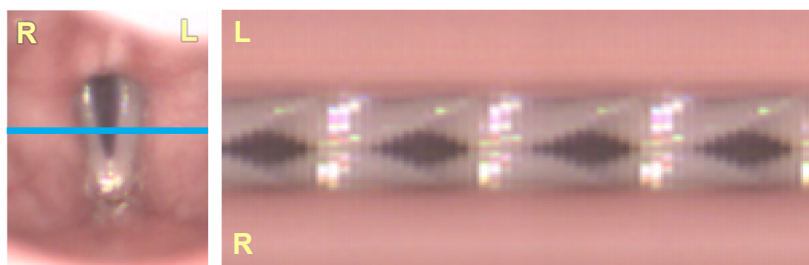
3.5.1 Kymogram

Videokymografie je popisná vyšetřovací metoda odvozená z dat získaných technikou HSV. Jejím výstupem je kymogram, který zobrazuje vývoj pohybu hlasivky v čase pod stanoveným řezem. Data lze snímat vysokorychlostní kamerou, která rovnou snímá jen jediný řádek nebo lze zachytit celé hlasivky a poté z jednotlivých snímků sekvence poskládat obraz kymogramu (obr. 3.6).

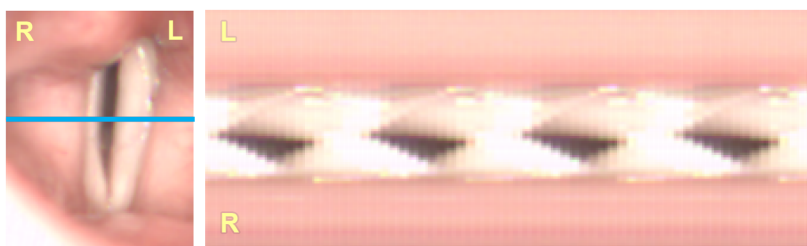


Obrázek 3.6: Princip kymografie

Zaznamenávaný řez lze umístit do libovolné pozice a vyšetřit symetrii kmitu celé hlasivky. Z kymogramu lze dobře vyčíst, zda hlasivky kmitají symetricky (obr. 3.7) či vykazují známky asymetrie (obr. 3.8). Na obrázcích je modře vyznačena rovina řezu.



Obrázek 3.7: Kymogram symetricky kmitajících hlasivek



Obrázek 3.8: Kymogram asymetricky kmitajících hlasivek

Výhodou tohoto přístupu je, že lze snadno detekovat nesymetrické pohyby hlasivek, nicméně pro automatické zpracování neposkytuje další informace.

3.5.2 Frekvenční metody

Oblast frekvenčních metod analýzy HSV záznamu spočívá v extrakci intenzity jasu obrazového bodu v daném bodě $p(x, y) \in ROI$ napříč snímky ve videosekvenci v monochromatickém zobrazení. Je zkoumán průběh signálu v čase. Metody tedy zkoumají dynamické vlastnosti pohybu glottis. Při analýze může být posuzována i změna intenzity jiné složky, uvažujeme-li snímek v barevném zobrazení, jde o hodnoty R, G, B (*Red, Green, Blue*).

Z výchozí literatury byly vybrány metody souhrnně označované jako *laryngotopografie*⁵. V zásadě jde o dva principy: *FFT Point Analysis*, kdy lze pomocí Fourierovy transformace z extrahovaného signálu získat informaci o základní změně intenzity jasu a dále o přístup *FFT Area Analysis*, kdy jsou v *ROI* obarveny nejvíce oscilující obrazové body $p(x, y)$. Toto téma je předmětem diplomové práce, a proto jsou přístupy detailně rozebrány v následující kapitole č. 4.

⁵*Laryngotopografie* je dle [23] definována jako metoda pro analýzu záznamu z vysokorychlostní kamery založená na Fourierově analýze obrazového bodu časově proměnného jasového signálu napříč jednotlivými snímky.

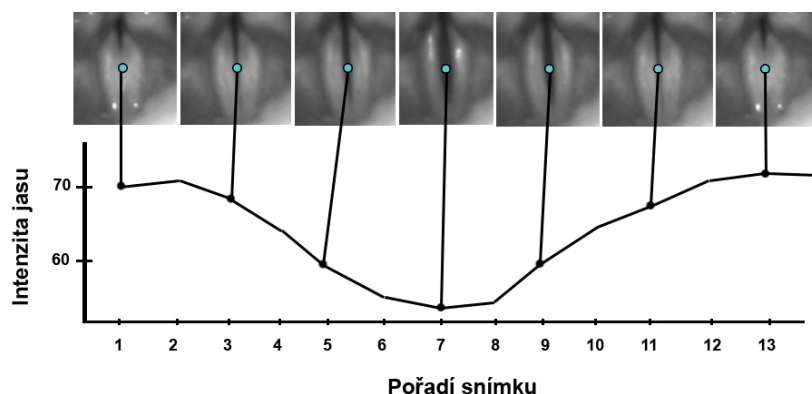
Kapitola 4

Frekvenční analýza videosekvence HSV

Tato kapitola se podrobně zabývá popisem frekvenčních metod analýzy záznamu HSV, označované také pod pojmem *laryngotopografie* (viz sekce 3.5.2). V rámci kapitoly je uveden výsledek provedené rešerše publikovaných prací na téma *laryngotopografie*, a je proveden návrh řešení frekvenční analýzy signálu HSV.

4.1 Přehled použitých metod

Kromě doporučené literatury [1] o problematice přístupů *FFT Point Analysis* a *FFT Area Analysis* byl také zpracován přehled publikací viz: [6], [23], [3] a [11]. Metody analýzy jsou však ukryty pod různými názvy a odlišují se také ve způsobu vizualizace získaných informací. V [1] je metoda *FFT Point Analysis* popsána jako extrakce hodnoty jasů ve zvoleném bodě napříč snímky ve videu viz ilustrační obr. 4.1. Dále je za pomoci Fourierovy transformace získáno amplitudové spektrum tohoto signálu (viz obr. 4.2). V metodě *FFT Area Analysis* jsou pak ve vybrané oblasti snímku obarveny pixely dle míry oscilace ve všech obrazových bodech dané oblasti (viz obr. 4.3).

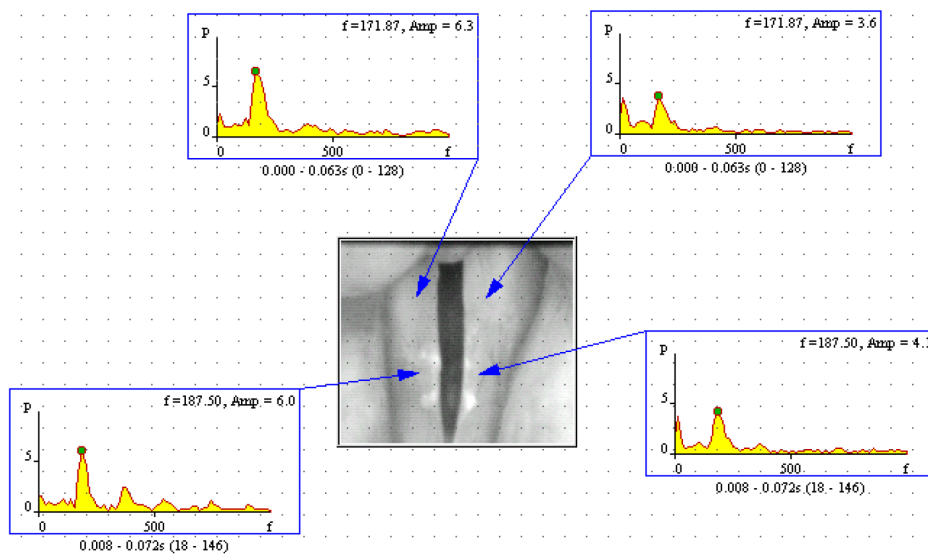


Obrázek 4.1: Průběh jasů v čase ve zvoleném bodě snímku HSV

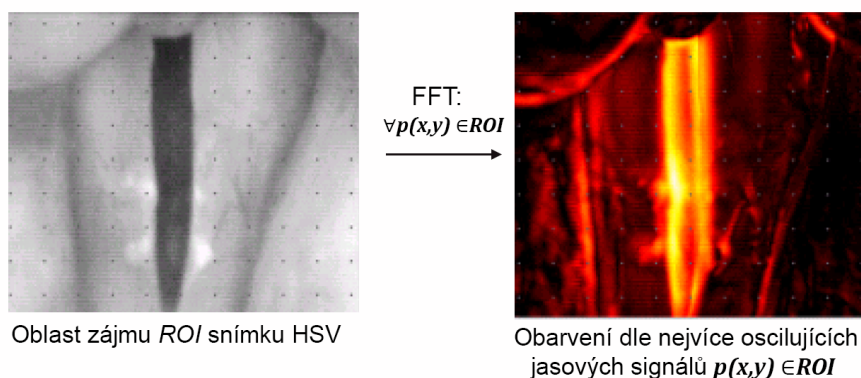
Stručný přehled publikací řešících podobnou problematiku je uveden níže:

- [6] umožňuje navíc od [1] analýzu změny fáze v hlasivkách. Článek je podrobná přehledová studie diskutovaných přístupů laryngotopografie.
- [3] se zabývá analýzou diplofonie¹, jež se projeví přítomností dvou fundamentálních frekvencí v hlase. Metoda je zde nazývána *Spectral video analysis (SVA)*.
- [23] zkoumá poruchu hlasu způsobenou přítomností subharmonických frekvencí, tedy frekvencí, jež jsou celočíselným zlomkem původní frekvence. Mezi tato onemocnění lze zařadit i diplofonii. Kromě analýzy poškozených hlasivek s pomocí spektra amplitud, frekvencí a fáze byl navrhnout model pro syntézu subharmonických hlasů. V této práci byl zaveden název *laryngotopografie*.
- [11] se výhradně zabývají změnou fáze jasových bodů ve videosekvenci v blízkosti glottis. U asymetricky se pohybujících hlasivek jsou patrné odlišnosti ve fázovém spektru. Výsledky jsou porovnávány s videokymografickým záznamem. Jsou diskutovány i výhody a nevýhody těchto přístupů. V článku mimo jiné aplikují metodu tzv. registrace obrazu, jejímž cílem je redukce důsledků pohybu kamery.

¹Diplofonie je tzv. dvojhlasí, porucha tvoření hlasu, projevující se zazníváním dvou hlasů vyskytující se nejčastěji při chronickém zánětu hrtanu.



Obrázek 4.2: *FFT Point Analysis* vybraného obrazového bodu ve videosekvenci [1]



Obrázek 4.3: Obarvení nejvíce oscilujících jasových signálů obrazových bodů dle metody *FFT Area Analysis* [1]

Články se shodují v aplikaci přístupů *FFT Point Analysis* a *FFT Area Analysis*, hlavní rozdíl spočívá v jejich odlišné vizualizaci, resp. interpretaci výsledků. Dále rozšiřují základní literaturu přidáním informace o fázi jasového signálu vybraného bodu $p(x, y) \in ROI$. Články se také odlišují ve využití těchto informací (např. pro účel hledání subharmonických frekvencí).

4.2 Návrh systému

Na základě výchozí a rozšiřující literatury popsané v podkapitole 4.1, byl navržen systém, který respektuje a aplikuje přístupy *FFT Point Analysis* a *FFT Area Analysis*. Uvedené přístupy však explicitně neřeší problém stanovení základní hlasivkové frekvence F_0 , resp. vychází z předpokladu, že průběh intenzity jasu v obrazovém bodě $p(x, y) \in ROI$ videozáznamu je nositelem informace o frekvenci F_0 nebo obecněji, odpovídá rychlosti pohybu odpovídajících anatomických struktur.

Právě v souvislosti se odhadem základní hlasivkové frekvence F_0 byly metody analýzy doplněny o detekci F_0 z audio záznamu, získaného současným měřením akustického signálu fonace mikrofonom (MIC) a videosignálu HSV. Takto získaný odhad základní hlasivkové frekvence F_0 je chápán jako globální informace o této frekvenci, $F_{0(global)}$, a doplňuje tak lokální odhady frekvence změny jasu v obrazových bodech $p(x, y) \in ROI$. Tyto lokální odhady F_0 pak dále považujeme za odhad rychlosti pohybu odpovídajících anatomických struktur (hlasivkové slizniční řasy, glottis, aj.).

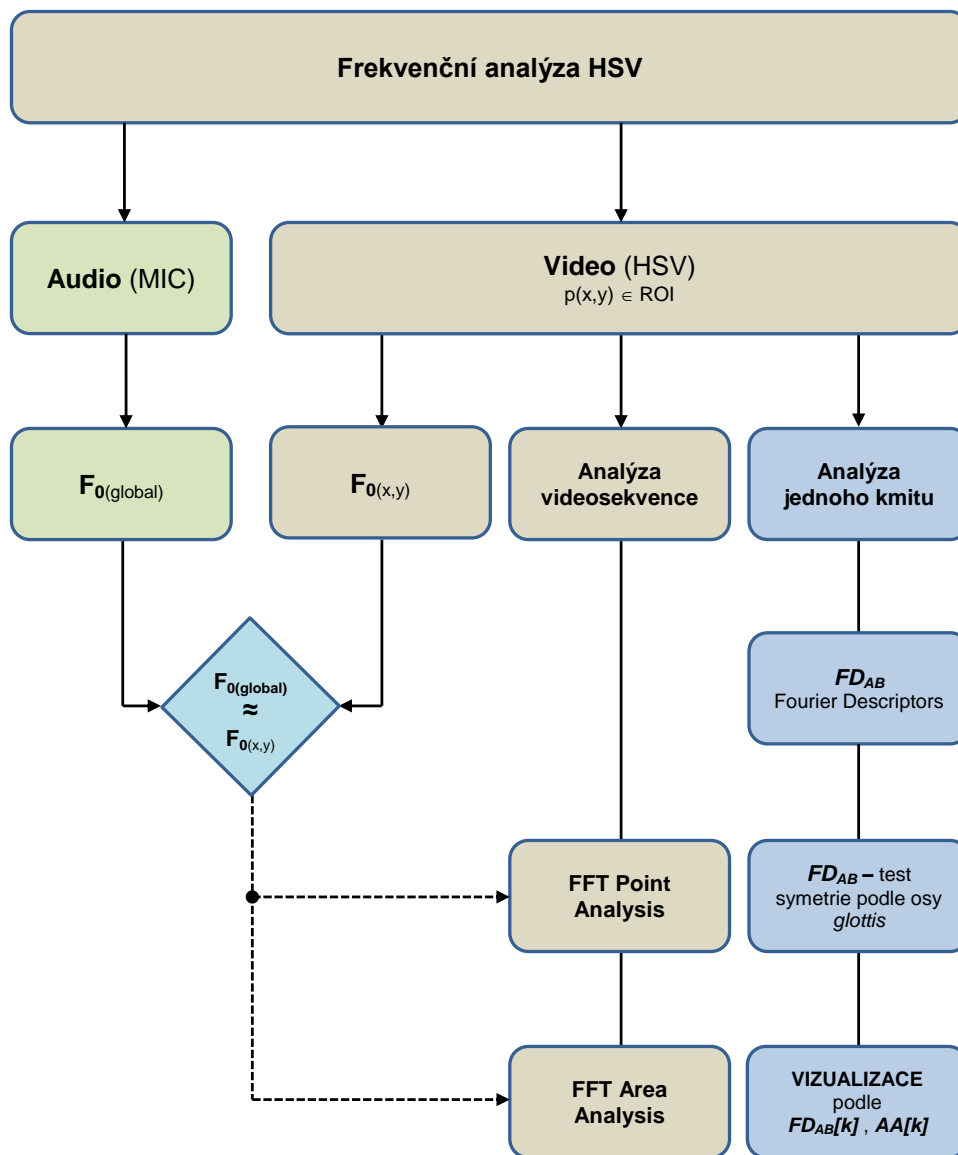
Protože je celý navrhovaný systém frekvenční analýzy navazuje na stávající soubor metod, které jsou používány k diagnostice hlasivek na ORL klinice FN Plzeň, je systém doplněn o tzv. analýzu jednoho kmitu hlasivek. Analýzou jednoho kmitu na úrovni audiosignálu jako záznamu fonace vokálu „a“ se zabývají metody SCORE, viz [18], a Standardní kmit, viz [12]. Základní myšlenka, která platí pro analýzu akustického signálu během jedné periody kmitu hlasivek (uzavřená hlasivka \rightarrow fáze otevírání \rightarrow otevřená hlasivka \rightarrow fáze uzavírání \rightarrow uzavřená hlasivka), je společná i pro průběh změny jasu v obrazovém bodě $p(x, y) \in ROI$. Jedná se o periodický signál, kdy pro jeden kmit můžeme použít popis pomocí Fourierovy řady a nad koeficienty řady můžeme definovat popisné parametry typu *Fourier Descriptors*. Tímto přístupem můžeme získat informaci o dynamickém chování hlasivek, např. symetrii hlasivek, předpokládáme-li, že jasový signál nese informaci o pohybu anatomické struktury a charakteru průběhu fází otevírání a zavírání hlasivkové štěrbin, glottis. Podrobně je tato metoda popsána v podkapitole 4.5.

Audiozáznam byl zpracováván na základě poznatků z [19] a [12], kde je detailně rozebráno stanovení základní hlasivkové frekvence F_0 i u značně poškozených hlasivek. Tento odhad hodnoty F_0 je pak dále využíván pro verifikaci, zda lokální změna jasu v obrazovém bodě $p(x, y) \in ROI$ odpovídá

rychlosti pohybu anatomických struktur hlasivek v daném bodě.

Návrh výše popsaného systému je schematicky popsán, viz obr. 4.4, kdy schéma zahrnuje nejen frekvenční analýzu HSV videozáznamu, ale také zpracování odpovídajícího audiozáznamu Audio (MIC), z něhož je vypočtena základní hlasivková frekvence $F_{0(global)}$. Systém umožňuje ověřit, zda se hodnota základní frekvence změny jasu $F_{0(x,y)}$ v testovaném obrazovém bodě $p(x, y) \in ROI$ shoduje, či odpovídá základní hlasivkové frekvenci $F_{0(global)}$, tj., zda $F_{0(global)} \approx F_{0(x,y)}$. Podobnost je posouzena výpočtem intervalu shody, protože v případě audiozáznamu Audio(MIC) a videozáznamu HSV pracujeme s odlišnými vzorkovacími frekvencemi (44100 Hz a 4000 Hz). Nezávisle na shodě $F_{0(global)} \approx F_{0(x,y)}$ je ale možné pokračovat v analýze metodami *FFT Point Analysis* a *FFT Area Analysis*. Další kategorie analýzy spočívá v analýze jednoho kmitu jasového signálu zvoleného obrazového bodu $p(x, y) \in ROI$, který popisujeme pomocí Fourierovy řady a tzv. *Fourier Descriptors* FD_{AB} . Tento popis je možné použít pro testy symetrie vzhledem k ose hlasivkové štěrbiny nebo odlišení od zvoleného referenčního obrazového bodu $ref(x, y) \in ROI$. Jako míra odlišení je používána Euklidovská metrika, viz kap. 4.5.

Návrh výše popsaného systému je schematicky zobrazen, viz obr. 4.4. Celé schéma tedy shrnuje frekvenční analýzu HSV záznamu. Kromě videozáznamu je možné do programu nahrát i příslušný audiozáznam Audio (MIC), z něhož je vypočtena základní hlasivková frekvence $F_{0(global)}$. V programu lze ověřit, zda se tato hodnota přibližuje základní frekvenci změny jasu $F_{0(x,y)}$ zvoleného obrazového bodu ve videosekvenci: $F_{0(global)} \approx F_{0(x,y)}$. Podobnost je posouzena výpočtem intervalu shody. Nezávisle na shodě je možné pokračovat v analýze za pomoci metod *FFT Point Analysis* a *FFT Area Analysis*. Další kategorie analýzy spočívá v analýze jednoho kmitu jasového signálu zvoleného obrazového bodu, popsaném za pomoci *Fourier Descriptors* FD_{AB} . První možností je otestování symetrie dle osy glottis nebo vizualizace všech synchronizovaných kmitů v *ROI*.



Obrázek 4.4: Schéma vlastního řešení práce

4.3 Určení základní frekvence

Základní hlasivková frekvence $F_0 \approx F_{0(global)}$ je důležitým parametrem tvorby hlasu a lze ji vypočítat z audiosignálu, jenž je součástí vyšetření vysokofrekvenční kamerou. Po získání základní frekvence můžeme dále vypočítat základní periodu, z níž lze vycházet při detekci hranic kmitu.

Obecně lze vztah pro určení základní frekvence F_0 vyjádřit jako poměr vzorkovací frekvence signálu F_{vz} a počtu vzorků základní periody L_0 pro diskretní signál:

$$F_0 = \frac{F_{vz}}{L_0} \quad [Hz] \quad (4.1)$$

Ze základní frekvence lze určit základní periodu T_0 , jež odpovídá délce základní periody:

$$F_0 = \frac{1}{T_0} \quad [Hz] \quad (4.2)$$

4.3.1 Audiozáznam

Při detekci základní frekvence audiosignálu $F_{0(global)}$ práce vychází z diplomových prací [19] a [12], kde je hodnota odhadována u značně poškozených hlasivek. Signál je zanesen formanty a standardní metody (např. autokorelace) tak často selhávají. Zvukový záznam disponuje vzorkovací frekvencí $F_{vz(global)} = 44100$ Hz, přičemž délka okna w , je stanovena na 8192 vzorků.

K určení základní frekvence lze přistupovat několika způsoby, přičemž je lze rozdělit do dvou skupin. První přístup spočívá v detekci základní frekvence v časové oblasti a druhý v detekci ve frekvenční oblasti. Jedná se o následující metody:

- **Detekce v časové oblasti**

 - Autokorelace (ACF)

 - Krátkodobá autokorelace (UACF)

- **Detekce ve frekvenční oblasti**

 - Spektrum součinů harmonických frekvencí (HPS)

Autokorelace

V časové oblasti lze detekovat opakující se vzory v signálu, dle kterých je možné odhadnout jejich periodicitu. Korelace je obecně chápána jako míra vyjadřující závislost v datech. V oblasti zpracování signálů určuje míru podobnosti signálů v daném vzorku (v čase). Pokud jsou proměnné ve výpočtu reprezentovány jedním a týmž signálem, hovoříme o autokorelaci. Míře podobnosti v jednom konkrétním vzorku m se říká korelační koeficient R . Množina korelačních koeficientů pak tvoří korelační funkci $R(m)$. Matematicky ji lze vyjádřit vztahem:

$$R(m) = \frac{1}{N} \sum_{k=1}^{N-1-m} x(k)x(k+m) \quad (4.3)$$

kde N je počet vzorků porovnávaného úseku, $x(k)$ je k -tý vzorek a hodnota m označuje posunutí signálu v počtu vzorků. Je-li signál periodický, bude i funkce $R(m)$ periodická. Pokud je i harmonický, lokální maxima funkce $R(m)$ budou v násobcích základní periody. Pro určení základní frekvence je nutné detekovat první vrchol $R(k)$, který následuje po maximální hodnotě autokorelační funkce $R(0)$, jež vyjadřuje energii signálu [8]:

$$R(m^*) = \max_{m=1, \dots, N-1} [R(m)] \quad (4.4)$$

Základní frekvenci pak vypočítáme vztahem:

$$F_{0(global)} = \frac{F_{vz}}{m^*} \quad [Hz] \quad (4.5)$$

Krátkodobá autokorelace

Předzpracováním vstupního signálu lze zvýšit úspěšnost autokorelační metody. Postup předzpracování byl navržen pro vyšetření hlasivek metodou hlasového pole, kdy jsou signály u pacientů s poruchou hlasu zkresleny vlivem chrapotu. Úpravy signálu mají za cíl minimalizovat nežádoucí vrcholy v autokorelační funkci způsobené např. formanty.² Jednotlivé kroky metody jsou [21]:

1. Volba mikrosegmentu $s(k)$ pro $k = 0, 1, \dots, N - 1$

²Formant je oblast lokálního maxima (špičky) v signálu vznikající rezonancí v dutinách hlasového ústrojí.

2. Filtrace FIR, Hammingovo okno w_{Hm} a dolní propust.

$$s_F(k) = \sum_{j=-N_{Hm}}^{N_{Hm}} s(k-j)w_{Hm}(j) \quad (4.6)$$

3. Posun o průměrnou hodnotu signálu v mikrosegmentu.

$$s^*(k) = s_F(k) - E(s_F) \quad (4.7)$$

$$E(s_F) = \frac{1}{N} \sum_{k=0}^{N-1} s_F(k) \quad (4.8)$$

4. Nelinearita, centrální a amplitudové omezení signálu.

$$\begin{aligned} s'(k) &= +1 \text{ pro } s^*(k) > a * \min[Max_1, Max_2, Max_3] \\ s'(k) &= -1 \text{ pro } s^*(k) < a * \max[Min_1, Min_2, Min_3] \\ s'(k) &= 0 \text{ pro ostatní } s^*(k) \end{aligned} \quad (4.9)$$

5. ACF /- test znělosti mikrosegmentu.

$$\bar{R}(m) = \frac{R'(m)}{R'(0)}, \text{ kde } R'(m) = \frac{1}{N-m} \sum_{k=0}^{N-1-m} s'(k)s'(k+m) \quad (4.10)$$

6. Po výpočtu ACF je určeno její maximum a je-li vyšší než předem stanovený práh znělosti (obvykle rozsah 0.3 až 0.4 velikosti $R(0)$), je mikrosegment označen za znělý.

$$\begin{aligned} \bar{R}(m) &= \max_{m=1, \dots, N-1} [\bar{R}(m)], \\ \bar{R}(m^*) &> \beta \Rightarrow \mu_{\text{segment - znely}} \end{aligned} \quad (4.11)$$

7. Z polohy maxima je základní frekvence:

$$F_{0(global)} = \frac{F_{vz}}{m^*} \quad [Hz] \quad (4.12)$$

kde F_{vz} je vzorkovací frekvence a m^* je poloha maxima v průběhu ACF.

Spektrum součinů harmonických frekvencí

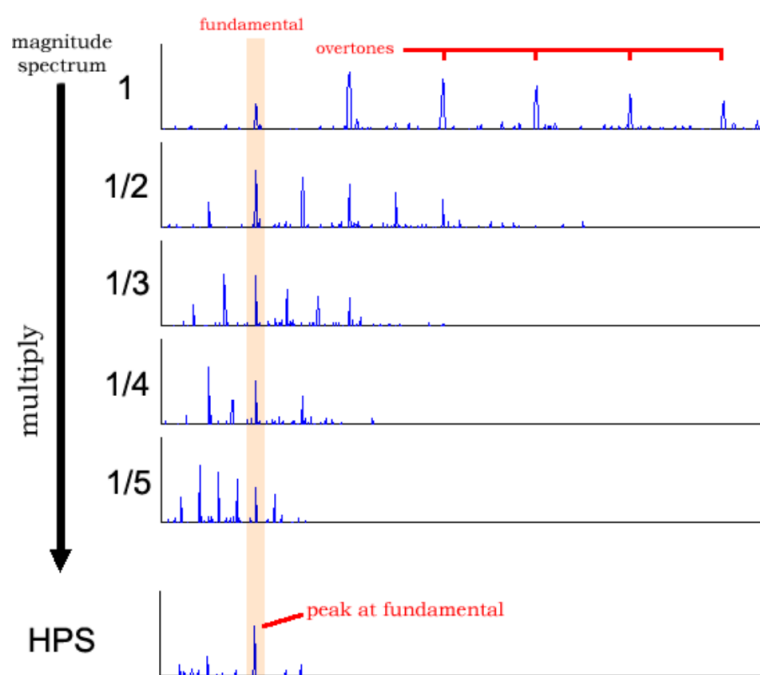
Princip detekce ve frekvenční oblasti je založen na převodu signálu z časové domény do frekvenční. Po určení úseku signálu je filtrován za pomoci okna a provedena Fourierova transformace. Přesnost detekce ve frekvenční oblasti je závislá na frekvenčním rozlišení F_R , která vyplývá z velikosti okna a vzorkovací frekvenci záznamu dle vztahu:

$$F_R = \frac{F_{vz}}{w_{DFT}} [Hz] \quad (4.13)$$

Ve frekvenční doméně je třeba detekovat, který vrchol odpovídá hledané základní frekvenci.

Spektrum součinů harmonických frekvencí (*Harmonic product spectrum, HPS*) stanovuje základní frekvenci signálu výpočtem největšího společného dělitele harmonických frekvencí.

Po vynásobení signálu Hammingovým oknem převedeme signál do frekvenční domény. Algoritmus zahrnuje dva kroky: decimaci signálu a součin. Signál je tedy několikrát podvzorkován a poté jsou tyto decimované kopie navzájem vynásobeny spolu se spektrem jako na Obrázku 4.5. Maximální hodnota tohoto spektra odpovídá největšímu společnému děliteli harmonických frekvencí, tedy hledané základní frekvenci [16].



Obrázek 4.5: Schéma algoritmu *Spektrum součinnů harmonických frekvencí (HPS)* [16]

Jedna z výhod této metody je její výpočetní nenáročnost a rezistence vůči aditivnímu a multiplikačnímu šumu. Dále je možné nahrazení poslední fáze, tedy násobení spektra za součet a snížit tak výpočetní náročnost algoritmu. Nevýhodou je závislost na velikosti okna signálu.

Uvedené metody byly implementovány a verifikovány na dostupných datech. Výsledky a hodnocení jsou uvedeny v části 6.

4.3.2 Videozáznam

Pro detekci v případě videozáznamu HSV je zásadní, jaký vybereme obrazový bod, z něhož určujeme základní frekvenci změny jasu obrazového bodu $F_{0(x,y)}$. Pro detekci byly vyzkoušeny metody uvedené v [19] a [12] a odladěny v aplikaci. Metody detekce $F_{0(x,y)}$ jsou:

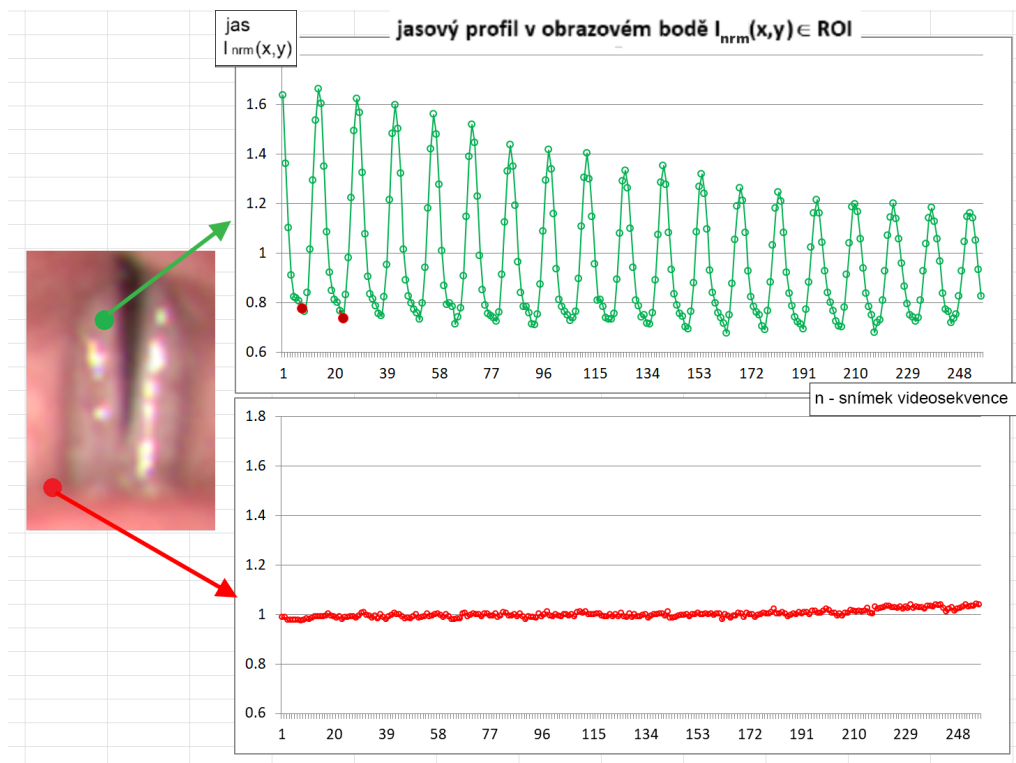
- **Detekce v časové oblasti**
 - Autokorelace (ACF)
 - Krátkodobá autokorelace (UACF)
- **Detekce ve frekvenční oblasti**
 - Spektrum součinnů harmonických frekvencí (HPS)

Postupy metod jsou shodné jako u detekce audiozáznamu (viz 4.3.1), liší se však v parametrech, protože do nich vstupují jiné signály. Na rozdíl od audia je délka videozáznamu stanovena na 256, 512, či 1024 snímků a vzorkovací frekvence $F_{vz} = 4000\text{Hz}$. Délka videozáznamu, resp. velikost okna w spolu s vzorkovací frekvencí ovlivňuje frekvenční rozlišení F_R , na kterém je závislá metoda HPS. Z tabulky č. 4.1 vyplývá, že frekvenční rozlišení roste s velikostí okna w .

| | $F_0(global)$ | $F_0(x, y)$ | | |
|---|---------------|-------------|--------|--------|
| vzorkovací frekvence F_{vz} | 44100 Hz | 4000 Hz | | |
| velikost okna w | 8 192 | 256 | 512 | 1024 |
| frekvenční rozlišení F_R | 5,38 Hz | 15,6 Hz | 7,8 Hz | 3,9 Hz |

Tabulka 4.1: Hodnoty frekvenčního rozlišení zvukového a jasového signálu v bodě $p \in ROI(x, y)$

Ukázku průběhu jasového signálu a detekce základní periody ve dvou různých obrazových bodech lze vidět na obr. 4.6.



Obrázek 4.6: Ukázka úspěšné detekce základní frekvence a kmitu (zelený signál) v oblasti hlasivkových řas a neúspěšné (červený) z důvodu nedostatečné oscilace signálu

4.3.3 Mapování základní frekvence jasového a zvukového signálu

Referenční hodnota odhadu základní hlasivkové frekvence $F_{0(global)}$ byla získána analýzou zvukového signálu. Úkolem je rozhodnout, zda vybraný obrazový bod $p(x, y) \in ROI$ vykazuje takovou změnu jasů, která se periodou změny blíží odhadu $F_{0(global)}$. Rozhodnutí provedeme na základě stanovení intervalu kolem odhadu $F_{0(x,y)}$.

Určení intervalu odhadu: ACF, UACF

Výsledkem odhadu $F_{0(x,y)}$, u metod ACF a UACF $R[m^*]$, je první maximum ze všech hodnot $R[m]$. Hodnota m^* pak určuje odhad $F_{0(x,y)}$, kde platí:

$$F_{0(x,y)} = F_{vz}/m^*, \text{ nebo také } T_{0(x,y)} = m^*/F_{vz} \quad (4.14)$$

Interval odhadu hodnot pro verifikaci, zda získaná hodnota $F_{0(x,y)}$ odpovídá $F_{0(global)}$, vychází z přesnosti vzorkovací frekvence $F_{vz(x,y)} = 4000$ Hz. Pro časovou vzdálenost mezi vzorky platí:

$$T_{vz(x,y)} = 1/F_{vz(x,y)} = 0,25 \text{ ms} \quad (4.15)$$

Tato funkce však není lineární, a proto je při odhadu intervalu postupováno následujícím způsobem:

1. Vypočítáme $F_{0(global)}$, k této hodnotě stanovíme nejbližší možnou hodnotu $F_{0(x,y)}$
2. Hodnotě $F_{0(x,y)}$ pak odpovídá $T_{0(x,y)}$:

$$T_{0(x,y)} = 1/F_{0(x,y)} \quad (4.16)$$

3. Vypočítáme sousední možné hodnoty pro $m - 1$, m , $m + 1$:

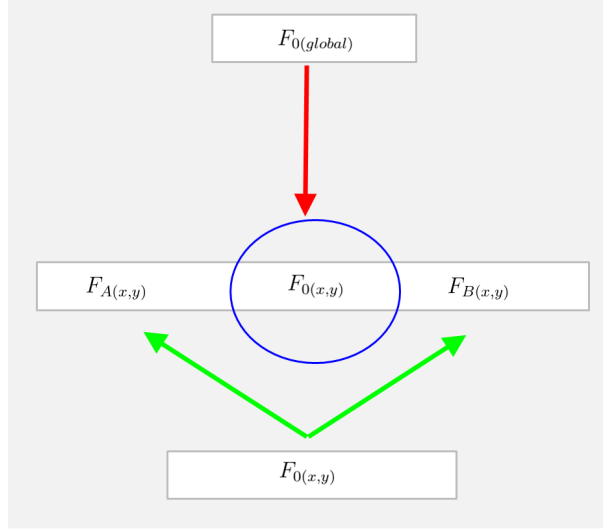
$$T_{0(x,y)} - T_{vz(x,y)} \quad a \quad T_{0(x,y)} + T_{vz(x,y)} \quad (4.17)$$

4. Pro meze kolem $F_{0(x,y)}$ platí:

$$F_{B(x,y)} = 1/(T_{0(x,y)} - T_{vz(x,y)}) \quad (4.18)$$

$$F_{A(x,y)} = 1/(T_{0(x,y)} + T_{vz(x,y)}) \quad (4.19)$$

Pokud odhadovaná hodnota frekvence změny jasu $F_{0(x,y)}$ spadá do intervalu $\langle F_{A(x,y)}, F_{B(x,y)} \rangle$, pak bod $p(x, y) \in ROI$ bude vykazovat změnu jasu, která odpovídá základní hlasivkové frekvenci hlasivek $F_{0(global)}$. Proces mapování je znázorněn na obrázku 4.7.



Obrázek 4.7: Mapování $F_{0(global)}$ na nejbližší možnou hodnotu, která odpovídá vzorkovací frekvenci a rozložení hodnot odhadů $F_{0(x,y)}$

Určení intervalu odhadu: HPS

Mapování pro metodu spektra harmonických frekvencí (HPS) je řešeno nejprve výpočtem intervalu shody:

1. Vypočítáme $F_{0(global)}$, k této hodnotě stanovíme nejbližší možnou hodnotu $F_{0(x,y)}$
2. Interval stanovíme na základě vztahu pro frekvenční rozlišení:

$$F_R = \nabla F = \frac{F_{vz}}{w_{DFT}} \quad (4.20)$$

3. Pro meze kolem $F_{0(x,y)}$ platí:

$$F_{A(x,y)} = F_{0(x,y)} - \nabla F \quad (4.21)$$

$$F_{B(x,y)} = F_{0(x,y)} + \nabla F \quad (4.22)$$

Dále pak již obdobně jako u mapování v případě metod ACF a UACF, platí: pokud odhadovaná hodnota frekvence změny jasu $F_{0(x,y)}$ spadá do intervalu $\langle F_{A(x,y)}, F_{B(x,y)} \rangle$, pak bod $p(x, y) \in ROI$ bude vykazovat změnu jasu, která odpovídá základní hlasivkové frekvenci hlasivek $F_{0(global)}$.

4.4 Analýza videosekvence

V této části je popsána metodika samotné analýzy jasového signálu – *FFT Point Analysis*. Jak již bylo zmíněno v sekci 3.5.2, pracujeme se skalární hodnotou, tedy s jasnem obrazového bodu v monochromatickém zobrazení popř. s jednotlivými barevnými složkami R, G, B . Označme tedy hodnotu jasu v obrazovém bodě $p(x, y) \in ROI$ jako $I(x, y)$, kde (x, y) je souřadnice obrazového bodu ve snímku videosekvence. Signál je normalizován:

$$I_{nrm}(n) = \frac{I(x, y)}{\sqrt{\frac{\sum_{i=1}^{N_{sig}} |I(x, y)_i|^2}{N_{sig}}}} \quad (4.23)$$

kde N_{sig} je počet vzorků signálu v časové oblasti. Průběh normovaného jasového signálu pro zvolený bod ve videosekvenci si lze prohlédnout na obr. 4.6. Takto upravený signál poté vstupuje do diskrétní Fourierovy transformace. V případě diskrétních signálů je přechod mezi signály z časové oblasti do frekvenční nazýván *diskrétní Fourierova transformace* (DFT), která je v praxi definována vztahy:

$$\begin{aligned} II(x, y)[k] = II[k] &= \sum_{n=0}^{N_{sig}-1} I_{nrm}(n) e^{-j \frac{2\pi}{N_{sig}} nk} \\ k &= 0, 1, 2, \dots, N_{harm} - 1 \\ n &= 0, 1, 2, \dots, N_{sig} - 1 \end{aligned} \quad (4.24)$$

kde $II_{(x,y)}[k]$ je označení obrazu Fourierovy transformace, N_{harm} je počet vzorků frekvenčního spektra a N_{sig} je počet vzorků signálu v časové oblasti. Protože DFT pracuje s konečnými počty hodnot v časové oblasti a konečnými posloupnostmi ve frekvenční oblasti, tj. konečnými počty vzorků frekvenčního spektra, budeme uvažovat, že signály v časové i frekvenční oblasti mají stejný počet vzorků. Lze tedy zapsat:

$$N_{harm} = N_{sig} = N \quad (4.25)$$

Poté pro přímou diskrétní Fourierovu transformaci platí:

$$II_{(x,y)}[k] = II[k] = \sum_{n=0}^{N-1} I_{nrm}(n) e^{-j \frac{2\pi}{N} nk} \quad pro \quad k = 0, 1, 2, \dots, N - 1 \quad (4.26)$$

a pro inverzní Fourierovu transformaci platí:

$$I_{nrm}(n) = \frac{1}{N} \sum_{k=0}^{N-1} II[k] e^{j \frac{2\pi}{N} nk} \quad pro \quad n = 0, 1, 2, \dots, N - 1 \quad (4.27)$$

která se také nazývá *zpětná inverzní transformace (IDFT)*.

S pomocí DFT je vypočteno N hodnot spektra $II[k]$, které jsou od sebe vzdáleny o:

$$\nabla F = \frac{F_{vz}}{N} \quad (4.28)$$

přičemž tyto hodnoty nazýváme *DFT mřížka* (angl. *DFT grid*). Hodnota ∇f bývá v literatuře nazývána *frekvenční bin* [8].

$II[k]$ je komplexním číslem, které uchovává informaci a amplitudě a fázi složky $e^{j\frac{2\pi}{N}nk}$ funkce $I_{nrm}(n)$. Výsledkem Fourierovy transformace reálné funkce je komplexní funkce:

$$II[k] = Re(II[k]) + jIm(II[k]) \quad (4.29)$$

kde $Re(II[k])$ a $Im(II[k])$ jsou reálnou a imaginární složkou $II[k]$. Velikost $|II[k]|$ nazýváme amplitudové frekvenční spektrum. Amplitudové frekvenční spektrum vypočítáme:

$$|II[k]| = \sqrt{Re(II[k])^2 + Im(II[k])^2} \quad (4.30)$$

a fázové spektrum pak:

$$\angle II[k] = \tan^{-1}\left(\frac{Im(II[k])}{Re(II[k])}\right) \quad (4.31)$$

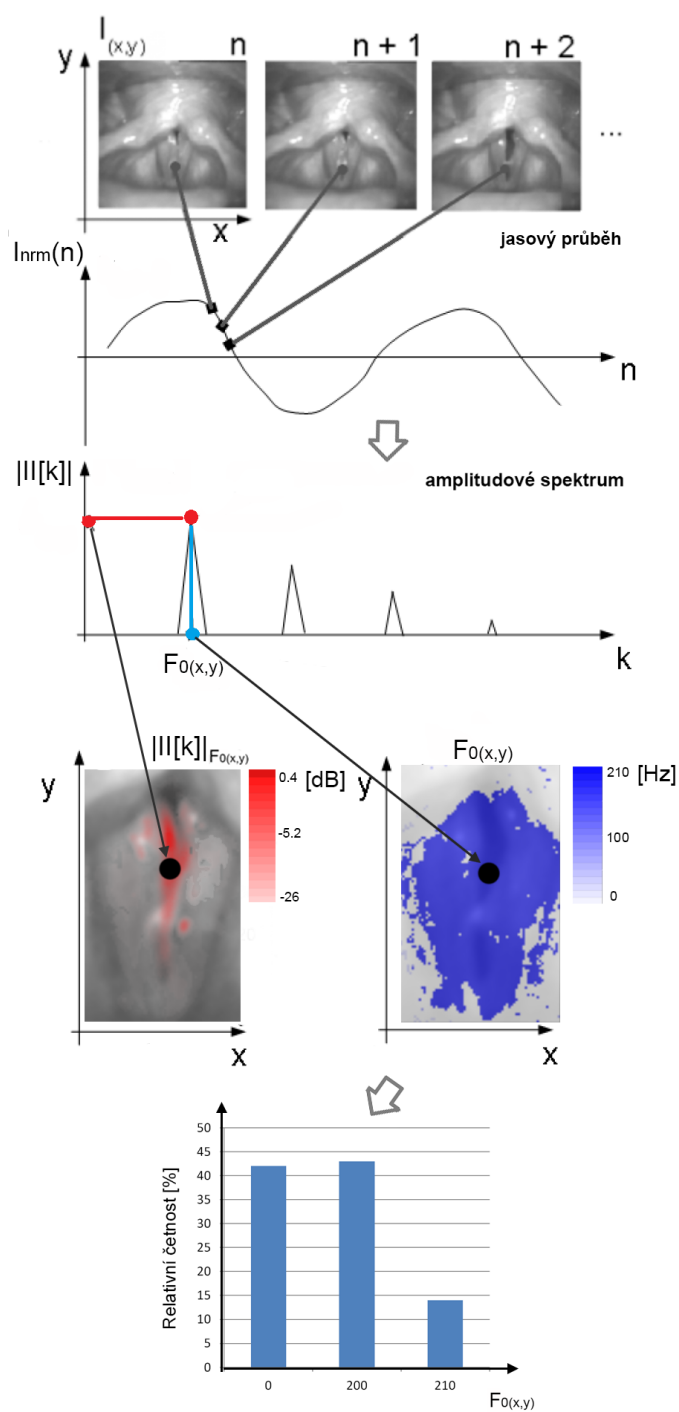
Vizualizace jednotlivých spekter je vždy provedena pro $\forall p(x, y) \in ROI$ a je znázorněna na obrázku 4.8 a 4.9. Pro každý obrazový bod je vypočítáno amplitudové i fázové spektrum a odhadnuta základní frekvence změny jasu $F_{0(x,y)}$, na jejíž pozici je zjištěna amplituda a fáze. Hodnoty jsou poté zaneseny do výstupního obrázku dle barevné škály. Hodnoty amplitudového spektra jsou převáděny do jednotek decibel, dle vztahu:

$$II[k]_{F_{0(x,y)}} [dB] = 20 \log_{10}\left(\frac{|II[k]_{F_{0(x,y)}}|}{|II[k]_{max}|}\right) \quad (4.32)$$

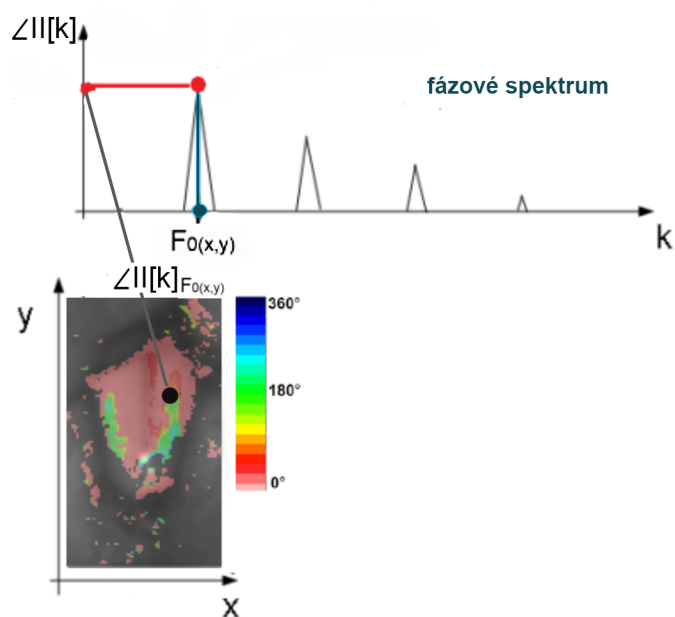
Dále je vypočítán relativní histogram výskytu četností detekovaných $F_0(x, y)$ a zanesen do grafu. Komponenty s nízkým výskytem byly vynechány (obr. 4.8).

Fázové spektrum je interpretováno ve stupních dle vztahu:

$$II[k]_{F_{0(x,y)}} [\text{deg}] = \angle II[k]_{F_{0(x,y)}} \cdot 180/\pi \quad (4.33)$$



Obrázek 4.8: Schéma průběhu jasu ve zvoleném bodě $p(x, y) \in ROI$, výpočet amplitudového spektra, vizualizace složky $F_0(x, y)$ a amplitudy



Obrázek 4.9: Fázové spektrum a vizualizace fáze v bodě $p(x, y) \in ROI$

4.5 Analýza jednoho kmitu hlasivek

Analýza jednoho kmitu hlasivek vychází z předpokladu, že změna jasu obrazového signálu v daném obrazovém bodě $p(x, y) \in ROI$ monochromatického signálu či hodnota složek R, G, B v barevném režimu, nese informaci o charakteru průběhu fáze otevírání a uzavírání hlasivkové štěrbiny (glottis). Může být nositelem informace o symetrii v dynamice hlasivek. Tato část diplomové práce vychází z dřívějších prací, jež se zabývají analýzou jednoho kmitu hlasivek. V jednom případě se jedná o analýzu audiozáznamu hrtanového tónu s cílem popsat a klasifikovat průběh odezvy audiosignálu v okamžiku uzavřené glottis. Metoda je popsána v publikaci [18].

Dalším typem analýzy je definování tzv. standardního kmitu hlasivek na úrovni analýzy audiosignálu. Jinými slovy se jedná o odhad průběhu audiosignálu, který je charakteristický (resp. statisticky významný) pro daného pacienta a aktuální stav jeho hlasového ústrojí, viz práce [12].

V této práci je však analýza audiosignálu nahrazena analýzou změny jasu v obrazovém bodě $p(x, y) \in ROI$. Jas v obrazovém bodě je popsán v předchozí kapitole 4.4. Pro dílčí snímky ve videosekvenci pak píšeme $I(x, y, n)$, kde:

- (x, y) je souřadnice obrazového bodu ve snímku videosekvence
- n je pořadí snímku ve videosekvenci, kde $n = 0, 1, 2, \dots, N_{sig} - 1$
- N_{sig} je počet snímků z videosekvence použitý pro analýzu

Signál je dále normalizován pomocí vztahu č. 4.23, zjednodušeně můžeme zapsat:

- $I(x, y)$ je původní hodnota jasu (nenormovaná)
- $I_{nrm}(x, y)$ je hodnota jasu normovaného signálu

V signálu $I_{nrm}(x, y)$ dále detekujeme jeden vybraný kmit, který odpovídá fázím:

1. otevřená hlasivka
2. fáze uzavírání hlasivky
3. uzavřená hlasivka
4. fáze otevírání hlasivky
5. uzavřená hlasivka

Jasový signál, který odpovídá vybranému kmitu označíme jako $I_L(x, y, L)$, kde (x, y) je souřadnice obrazového bodu ve snímku videosekvence a L je počet vzorků, které odpovídají jednomu kmitu hlasivek.

Pro parametry daného HSV zařízení (viz část 3.1) pak platí:

- $F_{vz} = 4000$ Hz
- $T_{vz} = 0,25$ ms
- $T_{period} = L \cdot T_{vz} = 0,25 \cdot L$
- $F_0 = 1/T_{period} = F_{vz}/L = 4000/L$ [Hz]

Pro analýzu jednoho kmitu hlasivek byl zvolen signálový popis rozvoje jedné periody signálu změny jasu pomocí trigonometrické řady (tzv. Fourierův rozvoj) s definovaným počtem jejich členů, viz literatura [14]. Pro zvolený signálový popis budeme psát:

$$I_{(x,y,L)}[k] = I_{(L)}[k] = \sum_{i=0}^{HARM} a_F[i] \cos(i \frac{2\pi}{L} k) + b_F[i] \sin(i \frac{2\pi}{L} k) \quad (4.34)$$

Po úpravě získáváme jednotlivé koeficienty Fourierovy řady:

$$a_F[i] = \frac{2}{L} \sum_{k=1}^L (I_{(L)} \cos(i \frac{2\pi}{L} k)) \quad (4.35)$$

$$b_F[i] = \frac{2}{L} \sum_{k=1}^L (I_{(L)} \sin(i \frac{2\pi}{L} k)) \quad (4.36)$$

Z koeficientů Fourierovy řady lze vytvořit tzv. invarianty, popisující tvar křivky změny jasu relativně k jejímu posunutí a hodnotě stejnosměrné složky signálu. Pro tento účel byl formulován vztah pro amplitudové spektrum:

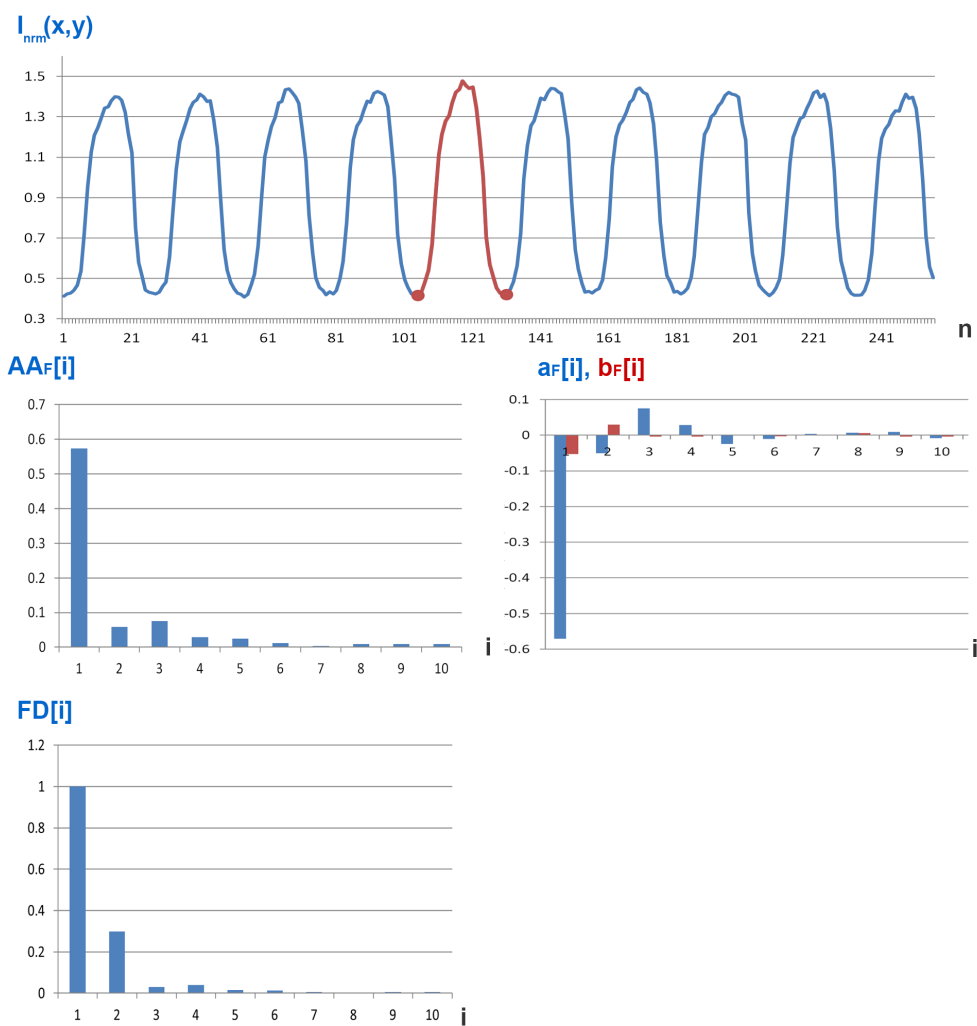
$$AA_F[i] = \sqrt{(a_F[i])^2 + (b_F[i])^2}, \quad \forall i = 1, 2, 3, \dots, HARM \quad (4.37)$$

Na základě koeficientů amplitudového spektra $AA_F[i]$, definujeme *Fourier Descriptor*, který je invariantní ke změně měřítka:

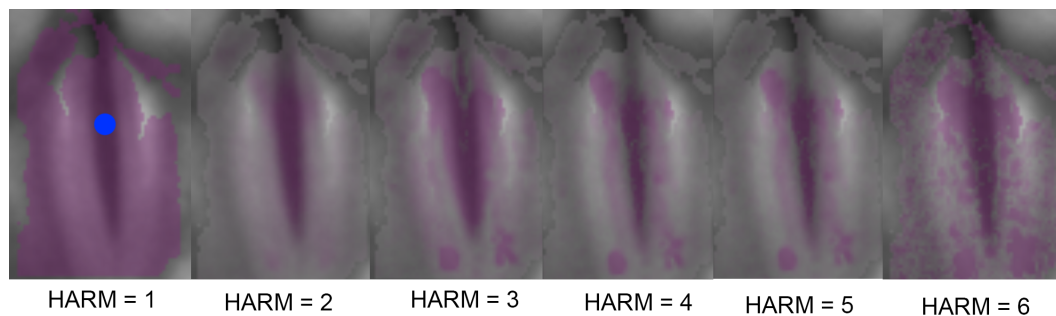
$$FD[j] = \frac{AA_F[j]}{AA_F[1]} = \frac{\sqrt{((a_F[j])^2 + (b_F[j])^2)}}{\sqrt{((a_F[1])^2 + (b_F[1])^2)}}, \quad (4.38)$$

pro $j = 2, 3, 4, \dots, HARM$, když $FD[1] = 1$.

Výše zmíněné hodnoty byly vypočítány pro zvolený kmit v jasovém signálu $I(x, y)$ a zaneseny do grafů (viz obr. 4.10). Pro $\forall p(x, y) \in ROI$ jsou zobrazeny postupně naškálované hodnoty $FD[j]$, tj. je zobrazováno spektrum podle jednotlivých harmonických (ukázka viz obr. 4.11).



Obrázek 4.10: Průběh jasového signálu $I(x, y)$ s vyznačeným analyzovaným kmitem. Dále jsou vypočítané grafy hodnot $AA_F[i]$, $a_F[i]$, $b_F[i]$ a $FD[i]$



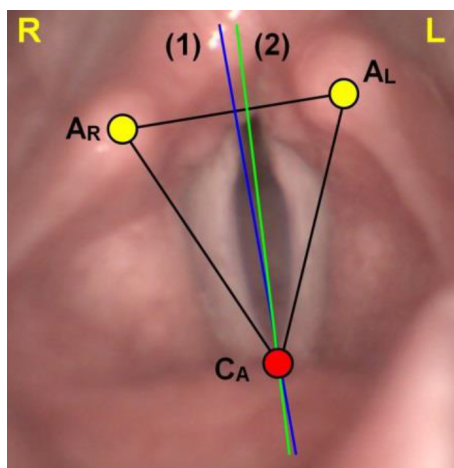
Obrázek 4.11: FD spektrum dle jednotlivých harmonických \approx zobrazení $\forall p(x, y) \in ROI$, u kterých je detekována (resp. je možné určit) $F_{0(x,y)}$

Takto definovaný popis jednoho kmitu hlasivek na úrovni jasové změny v definovaném obrazovém bodě videosekvence $p(x, y) \in ROI$ lze využít pro řadu analýz. Jako základní se nabízí analýza symetrie hlasivek, tj. porovnávání dvou obrazových bodů $p_A(x, y) \in ROI$ a $p_B(x, y) \in ROI$, které jsou vzájemně symetrické vzhledem k ose symetrie hlasivkové štěrbině glottis. Jako srovnávací parametr je pak použit definovaný *Fourier Descriptor* $FD[j]$.

Z hlediska anatomie a kinematiky hlasivek jsou používány dvě definice tzv. osy hlasivek, viz [18]:

- hlavní osa hlasivek je definována jako spojnice středů přední a zadní komisury
- anatomická osa hlasivek vychází z vlastní definice glottis, kdy glottis chápeme jako trojúhelníkovou štěrbinu otevřenou při dýchání a uzavřenou při fonaci, která je ohraničena laterálně hlasivkovými řasami *plīcaes vocalis*, a dorzálně interarytenoidním prostorem. Anatomická osa je pak těžnice v trojúhelníku s vrcholem ve středu přední komisury hlasivek a středem vzdálenosti mezi arytenoidními hrboly.

Osa symetrie je pak odhadem polohy anatomické osy hlasivek. Metodami odhadu polohy anatomické osy hlasivek se detailně zabývají v [18] a [5]. V této diplomové práci dále předpokládáme, že je osa symetrie glottis zadána parametricky, ve vlastní aplikaci pro testování navržených postupů a metod je zadávána ručně společně s definováním *ROI*.



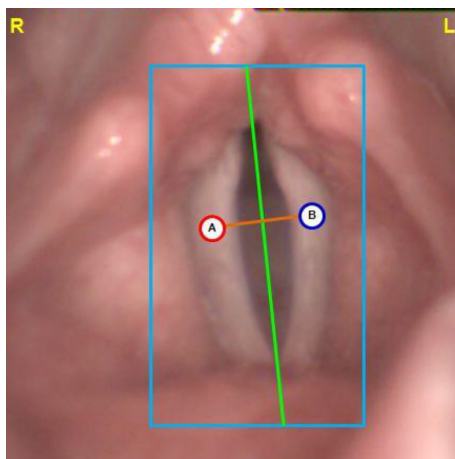
Obrázek 4.12: Snímek glotické oblasti zachycený vysokorychlostní kamerou [18]. Hlasivková štěrbina je zde ve fázi maximálního otevření během fonace vokálu i . Dále jsou v obrázku vykresleny tyto značky: (1) – anatomická osa hlasivky, (2) – osa symetrie hlasivkové štěrbiny, A_L a A_R – levý a pravý arytenoidní hrbol, C_A – přední komísura hlasivek.

Pro kvantifikaci odlišnosti tvaru jasového profilu jednoho kmitu u dvou symetrických obrazových bodů $p_A(x, y) \in ROI$ a $p_B(x, y) \in ROI$ (viz obr. 4.13), tj. $I_{(L)}^A(x, y)$ a $I_{(L)}^B(x', y')$, použijeme absolutní hodnotu rozdílu jednotlivých *Fourier Descriptors* $FD_A[j]$ patřících $I_{(L)}^A(x, y)$ a $FD_B[j]$ patřících $I_{(L)}^B(x', y')$. Dále pak:

$$FD_{(AB)}[j] = |FD_A[j] - FD_B[j]| \quad (4.39)$$

Celkovou míru odlišení vypočítáme jako:

$$\Delta_{(AB)}^H = \sum_{j=2}^{HARM} FD_{AB}[j] = \sum_{j=2}^{HARM} |FD_A[j] - FD_B[j]|, \quad \forall j = 2, 3, \dots, HARM. \quad (4.40)$$

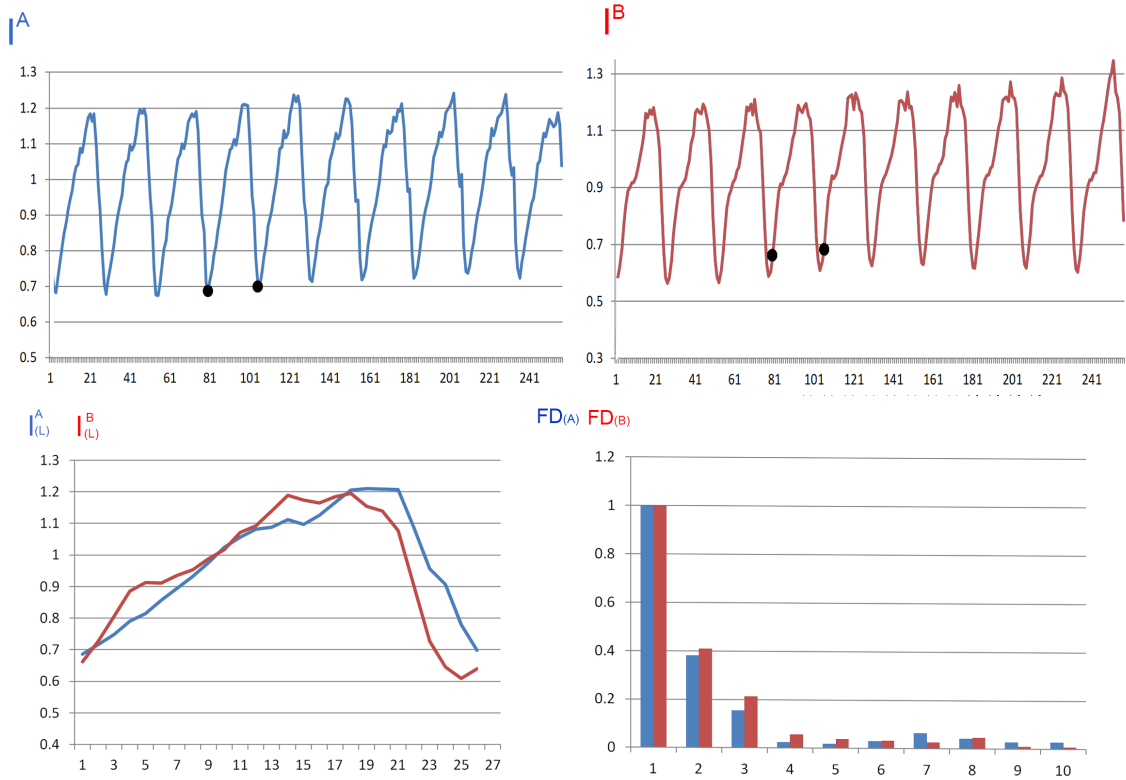


Obrázek 4.13: Test odlišnosti zvolených kmitů ve vyznačených bodech A , B by měl u symetricky kmitajících hlasivek být co nejmenší

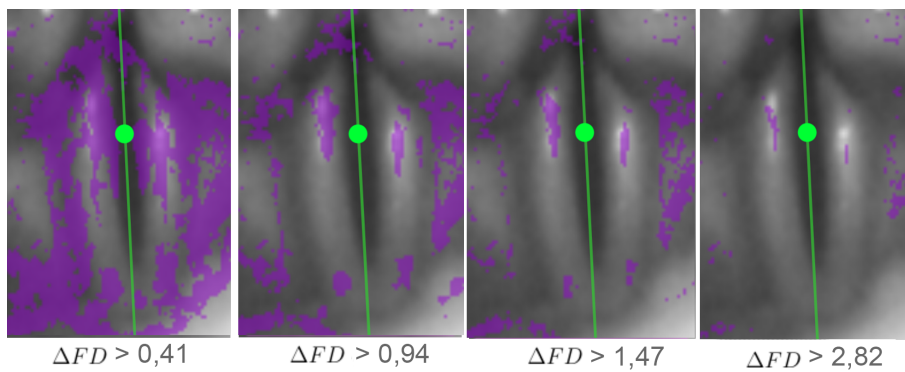
Výsledky ze vztahu č. 4.40 jsou vizualizovány dvěma způsoby:

1. vzájemně symetrické obrazové body $p(x, y) \in ROI$ dle osy glottis, jejichž míra odlišení $\Delta FD > T_{\Delta FD}$, jsou zvýrazněny s cílem poukázat na anomálie v chování hlasivek na úrovni tvaru jasového profilu. Analogicky je možné zobrazit body pod stanovenou mezí.
2. pro $\forall p(x, y) \in ROI$, jejichž míra odlišení $\Delta FD > T_{\Delta FD}$ vůči zvolenému referenčnímu $ref[x, y]$, jsou zvýrazněny pro posouzení odlišnosti tvaru kmitů ostatních bodů vůči porovnávanému bodu. Analogicky je možné zobrazit body pod stanovenou mezí.

Ukázka odlišnosti tvaru jasového profilu jednoho kmitu u dvou symetrických obrazových bodů $p_A(x, y) \in ROI$ a $p_B(x, y)$ vyznačena na obrázku 4.14.



Obrázek 4.14: Jasové profily I^A , I^B a jejich periody $I_{(L)}^A$, $I_{(L)}^B$ u dvou symetricky zvolených bodů. Míra odlišnosti *Fourier Descriptors* FD_A a FD_B : $\Delta_{(AB)}^{II} = 0,217$.



Obrázek 4.15: Odlišení $ref(x, y) \in ROI$ vzhledem k $p(x, y) \in ROI$, u kterých je definována $F_{0(x,y)}$

Kapitola 5

Implementace

V kapitole jsou popsány algoritmy užívané v této práci a implementační detaily.

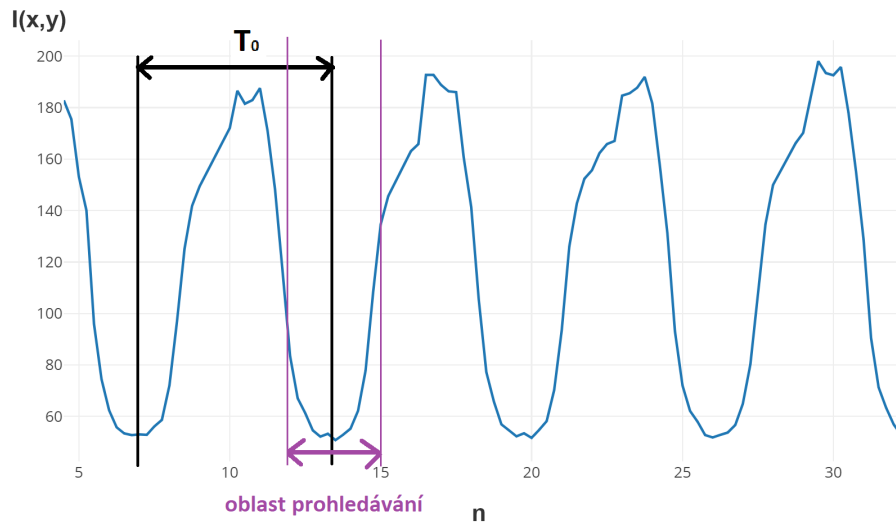
5.1 Algoritmy

5.1.1 Rychlá Fourierova transformace

Výpočetní složitost diskrétní Fourierovy transformace (DFT) pro N bodů dle vztahu č. 4.24 je dána jako $O(N^2)$ [20]. Existuje však optimalizovaný výpočetní algoritmus o složitosti $O(N \log_2 N)$, nazývaný rychlá Fourierova transformace (FFT) zpracovávající pole hodnot o velikosti 2^N vzorků. V práci byla pro urychlení výpočtu využita verze Danielson-Lanzcos algoritmu publikovaná v [20].

5.1.2 Detekce kmitu hlasivek

Detekce hranic kmitů jasového signálu užitá při výpočtech z části č. 4.5 je založena na znalosti základní frekvence změny jasu. Počátek kmitu je odvozen od pozice lokálního minima v signálu, jenž odpovídá okamžiku maximálnímu otevření hlasivek. Následující hranice je vyhledávána ve vzdálenosti základní periody. V signálu dochází k frekvenčnímu kolísání, a proto je hranice kmitu vyhledávána v okolí předpokládané pozice.



Obrázek 5.1: Perioda jasového signálu $I(x, y)$ a oblast hledání hranice periody

Rozsah prohledávání lze pevně stanovit na daný počet vzorků, nicméně vzhledem k rozdílným uvažovaným základním frekvencím změny jasu je lepší stanovit oblast prohledávání například na $1/4$ vzorků periody. V případě, že bude perioda čítat $L = 28$ vzorků, bude oblast prohledávání 8 vzorků. Formálně lze popsat algoritmus následovně:

$$\begin{aligned}
 p[0] &= n, \min_{0 \leq n \leq L} \\
 p^*[k] &= p[k] + L \\
 p[k] &= n, \min_{p^*[k]-8 \leq n \leq p^*[k]+8} I_{nrm}(n)
 \end{aligned} \tag{5.1}$$

kde $p[k]$ je pozice prvního vzorku k -té periody signálu, L je počet vzorků periody a $I_{nrm}(n)$ značí n -tý vzorek záznamu.

5.2 Užité technologie

Programovací jazyk užitý k tvorbě aplikace je $C\#$. Design byl kompletně navržen v novější technologii *Windows Presentation Foundation (WPF)*, oproti starší technologii *Windows Forms*. Aplikace vyžaduje verzi *.NET* č. 4.5. Vývoj proběhl v software *Microsoft Visual Studio 2017* a otestován na stroji s operačním systémem *Microsoft Windows 10, 64-bit* s parametry: *Intel(R) Core(TM) i5-4200U CPU; 4,0GB RAM*.

Knihovny, které vytvořené programové vybavení využívá, jsou přibaleny k instalačnímu balíku. Stručný popis knihoven spolu s využívanou funkcionalitou je uveden níže:

- *Live Charts*¹ – grafický výstup dat, tj. diagramy a grafy
- *MahApps*² – grafický vzhled aplikace
- *AForge.Net*³ – načtení a předzpracování videozáznamu

¹Live Charts je knihovna obsahující mnoho nástrojů pro vizualizaci dat, real-time grafů aj. Více na: <https://lvcharts.net/>

²MahApps je knihovna pro tvorbu grafického uživatelského rozhraní ve stylu Metro – v designu s jednoduchými geometrickými prvky specifické pro Windows aplikace. Domovská stránka projektu: <https://mahapps.com/>

³AForge.NET je framework určený především pro vývoj a výzkum v oblasti počítačového vidění, umělou inteligenci, např. zpracování obrazu, neuronové sítě, strojové učení aj. Více informací na: <http://www.aforgenet.com/>

Kapitola 6

Výsledky

V rámci této práce bylo provedeno testování výsledků nad vybraným souborem videozáznamů z vysokorychlostní kamery spolu se sesynchronizovaným zvukovým záznamem. Soubor byl sestaven tak, aby zahrnoval různé druhy videozáznamů dle kvality snímku, typu onemocnění a délky záznamu (počet snímků). Zahrnuty byly jak záznamy se zdravými hlasivkami, pohybujícími se symetricky, tak záznamy s asymetrií. Asymetrie mohla být způsobena například obrnou zvratného nervu nebo díky vyrůstající hmotě na hlasivce (polyp). Popis jednotlivých záznamů je uveden v tabulce 6.1.

| HSV (video) | HSV (mic) | pohlaví | věk | Dg | $F_{0(HSV)}$ [Hz] | $F_{0(global)}$ [Hz] | počet snímků |
|----------------|--------------|---------|-----|---------------------|----------------------|-------------------------|-----------------|
| 01.avi | 01.wav | žena | 50 | polyp | 204.80 | 206.00 | 686 |
| 02.avi | 02.wav | žena | 57 | polyp | 194.00 | 194.00 | 732 |
| 03.avi | 03.wav | žena | 19 | zdráv | 272.00 | 283.00 | 1098 |
| 04.avi | 04.wav | muž | 49 | polyp | 145.00 | 145.00 | 549 |
| 05.avi | 05.wav | žena | 22 | zdráv | 215.58 | 219.00 | 640 |
| 06.avi | 06.wav | muž | 24 | zdráv | 156.29 | 156.00 | 641 |
| 07.avi | 07.wav | žena | 58 | paréza | 235.80 | 246.00 | 1693 |
| 08.avi | 08.wav | žena | 53 | paréza | 207.50 | 204.00 | 870 |
| 09.avi | 09.wav | žena | 19 | paréza | 250.00 | 230.00 | 1830 |
| 10.avi | 10.wav | žena | 19 | paréza po léčbě | 285.64 | 286.00 | 458 |
| 11.avi | 11.wav | muž | 66 | karcinom | 129.35 | 125.00 | 869 |
| 12.avi | 12.wav | muž | 20 | zdráv | 149.00 | 149.00 | 686 |
| 13.avi | 13.wav | žena | 41 | cysta vlevo | 222.00 | 217.00 | 687 |
| 14.avi | 14.wav | žena | 38 | polyp | 190.00 | 186.00 | 686 |
| 15.avi | 15.wav | žena | 38 | polyp po zákroku | 200.00 | 202.00 | 641 |

Tabulka 6.1: Testovací data s výsledky detekce základní frekvence

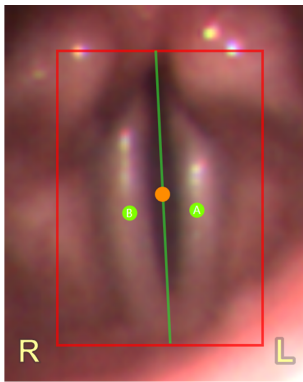
V práci jsou prezentovány 4 příklady jako kazuistiky, na kterých je patrná schopnost nástroje detekce problému symetrie/nesymetrie. Výstupy experimentů jsou vždy ve stejném pořadí, které souhrnně popíšeme následovně:

- (a) *ROI* s označením osy symetrie glottis, $ref[x, y]$, osově symetrických bodů A a B
- (b) zobrazení obrazových bodů $p(x, y) \in ROI$, s vyznačením detekované základní frekvence změny jasu F_0
- (c) histogram četnosti obrazových bodů $p(x, y) \in ROI$, podle detekované F_0
- (d) *FFT analysis* – amplitudové spektrum
- (e) *FFT analysis* – fázové spektrum
- (f) analýza jednoho kmitu – rozložení hodnot FD pro jednotlivé harmonické Fourierovy řady

- (g) analýza jednoho kmitu – srovnávací analýza dvou osově symetrických bodů A a B
- (h) analýza jednoho kmitu – zobrazení dvojice osově symetrických bodů $\in ROI$, jejichž míra odlišení ΔFD přesáhla stanovenou mez $T_{\Delta FD}$, kde $\Delta FD \in \langle 0, \Delta FD_{max} \rangle$, analogicky lze zobrazit dvojice, jež jsou pod stanovenou mezí
- (i) analýza jednoho kmitu – zobrazení bodů $\in ROI$ vůči $ref[x, y]$, jejichž míra odlišení ΔFD přesáhla stanovenou mez $T_{\Delta FD}$, kde $\Delta FD \in \langle 0, \Delta FD_{max} \rangle$, analogicky jsou zobrazeny body nižší než stanovená mez

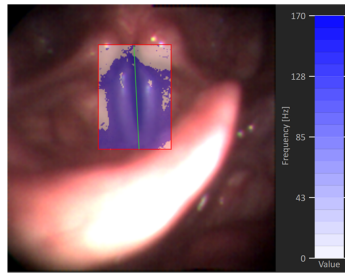
6.1 Experiment č. 1: zdravé hlasivky

Nejprve byl vybrán záznam se zdravými hlasivkami, viz obr. 6.1.

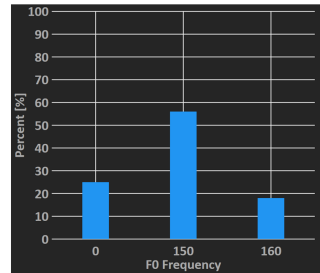


a)

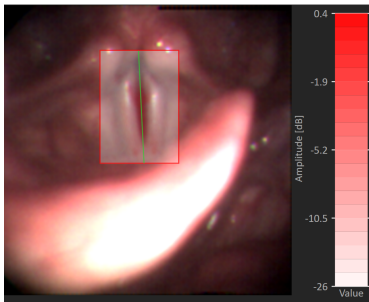
muž, 24 let, dg: zdrav



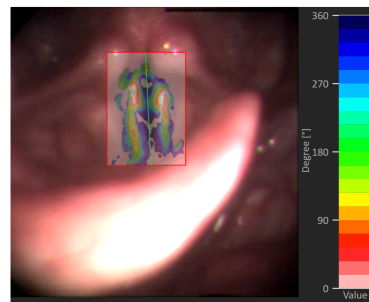
b)



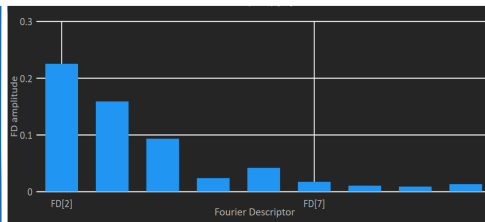
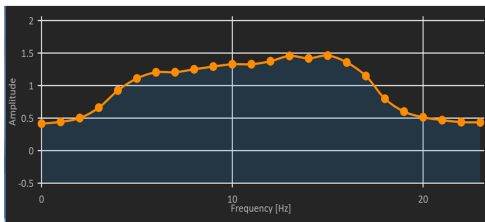
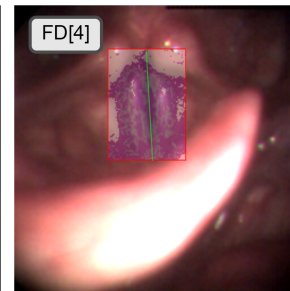
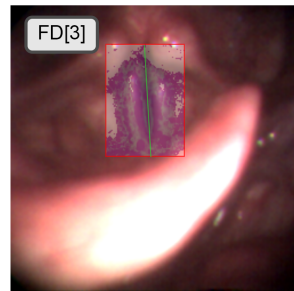
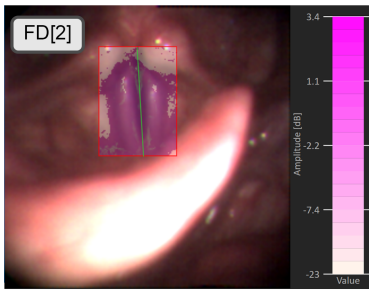
c)



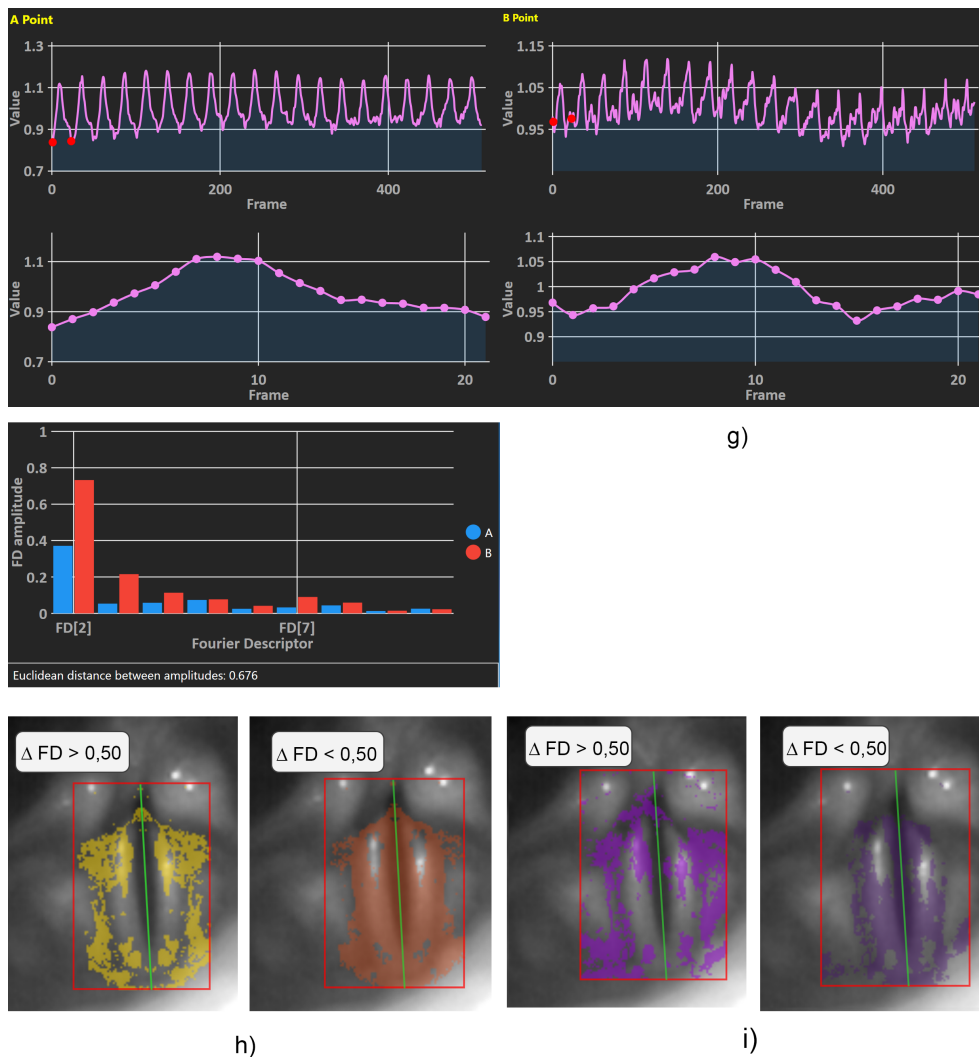
d)



e)



f)



Obrázek 6.1: Popis výsledků vyšetření experimentu č. 1 – zdravé hlasivky

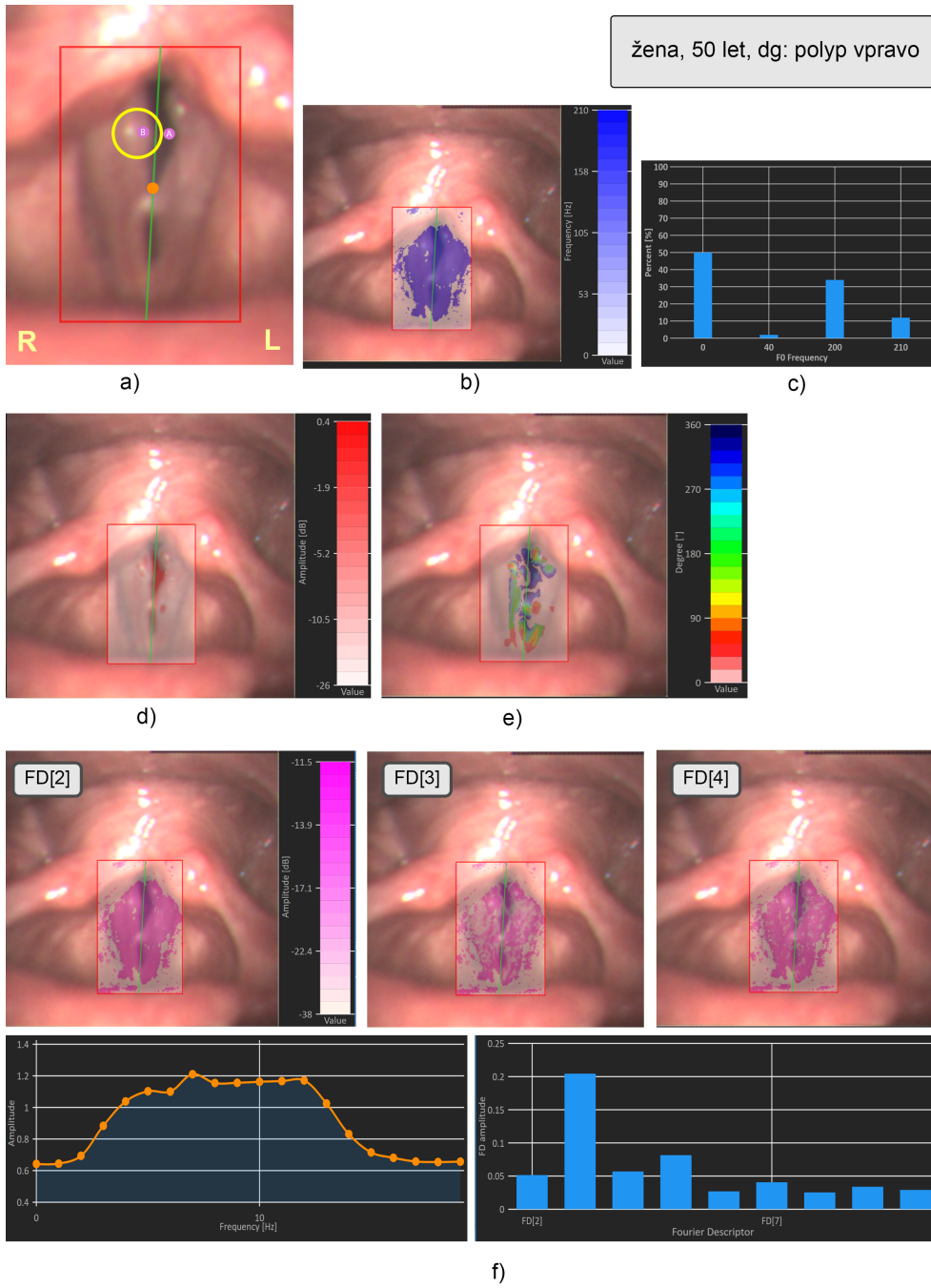
6.1.1 Zhodnocení

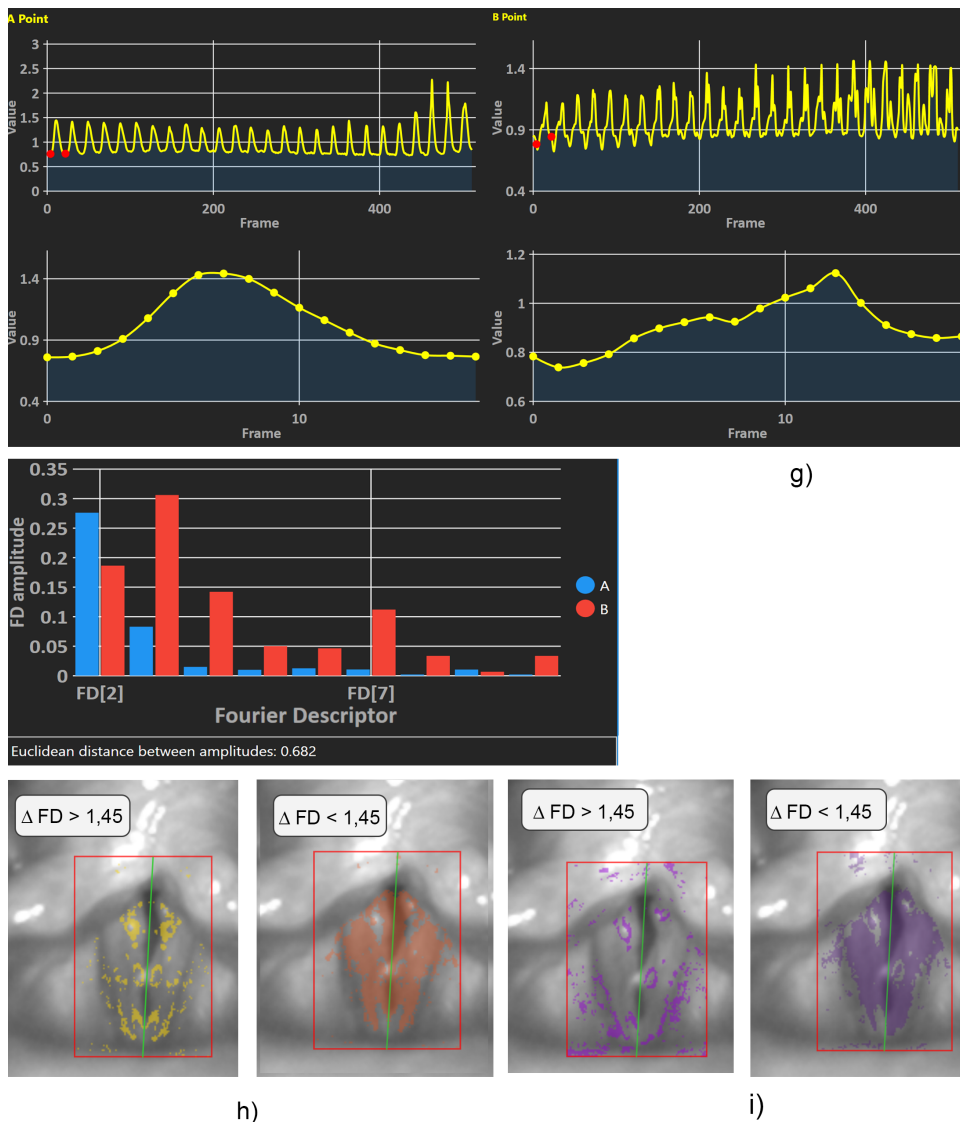
- frekvenční spektrum neodhaluje známky asymetrického chování (b), jen je zvýrazněna oblast světelných artefaktů, kde nebylo možné detekovat $F_{0(x,y)}$ pro příslušné $p(x,y)$; histogram ukazuje nejvyšší četnost výskytu detekované $F_{0(x,y)} = 150$ Hz (c);
- rozložení amplitudového a fázového spektra ukazuje na symetrii hlasivkové štěrbině glottis (d), (e);
- analýza jednoho kmitu – rozložení hodnot FD pro jednotlivé harmonické Fourierovy řady vykazuje pro $FD[2]$, $FD[3]$, $FD[4]$ rovnoměrné rozložení po oblasti glottis, kromě míst světelných artefaktů (f)

- analýza jednoho kmitu – indikuje v prostoru výskytu polypu rozdílnost $\Delta FD = 0.676$ vzhledem k ose hlasivkové štěrbiny, viz označené osově symetrické body $A - B$ (g);
- analýza jednoho kmitu – v odlišení osově symetrických bodů podle ΔFD , pro $\Delta FD = 0,50$ je patrná odlišnost pouze pro oblast světelných artefaktů (h), podobně pro odlišnost vůči $ref[x, y]$ (i).

6.2 Experiment č. 2: polyp

Pro druhý experiment byl vybrán záznam s diagnózou polyp na pravé hlasivkové řase, viz obr. 6.2 (a). Výsledky viz. 6.2.





Obrázek 6.2: Popis výsledků vyšetření experimentu č. 2 – polyp

6.2.1 Zhodnocení

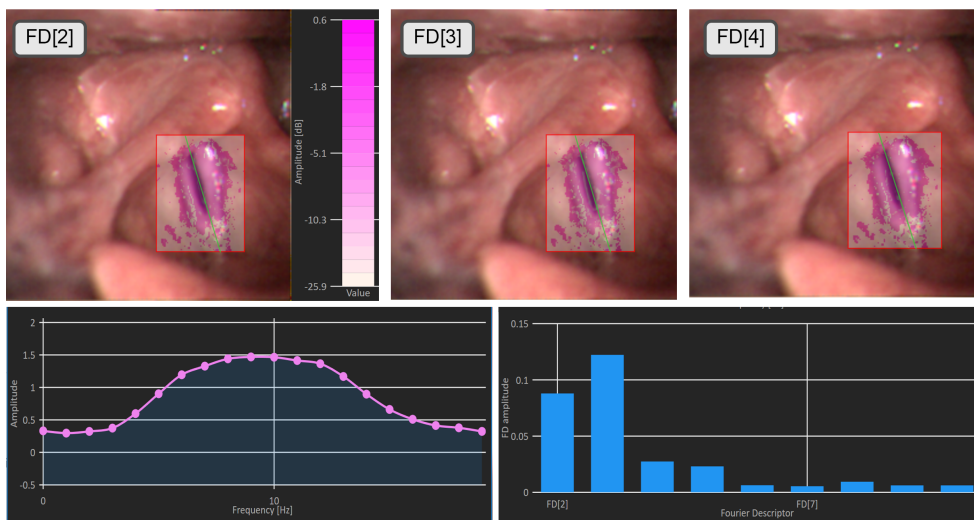
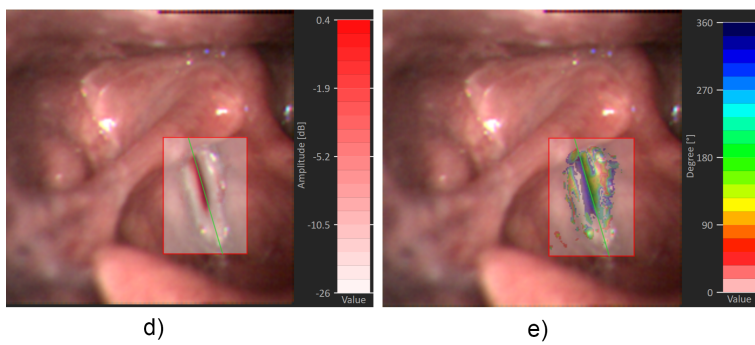
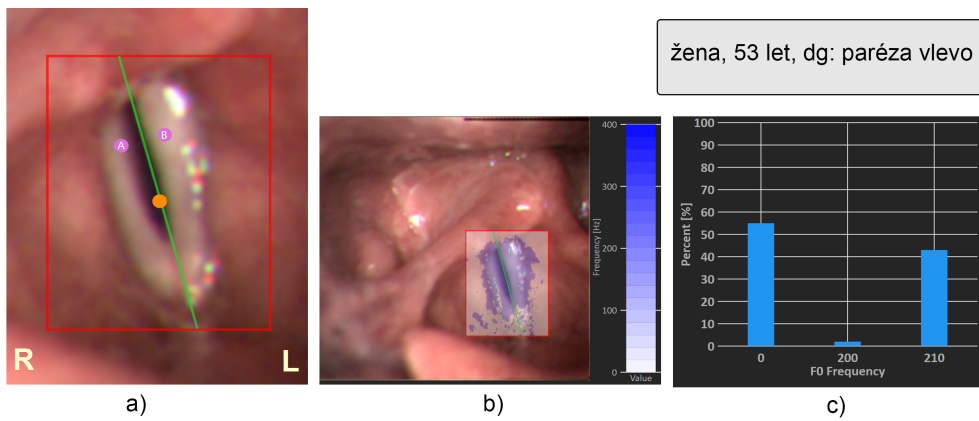
- rozložení fázového spektra naznačuje nesymetrii v oblasti hlasivkové štěrbině, stejně tak rozložení amplitudového a fázového spektra (b), (d), (e);
- amplitudové spektrum – je patrné zesílení kmitů vlevo od osy hlasivkové štěrbině (vpravo je výrazný útlum kmitů) (d);
- fázové spektrum – frekvenční analýza průběhu jasu v jednotlivých obrazových bodech zobrazuje oblasti na anatomických strukturách hlasiv-

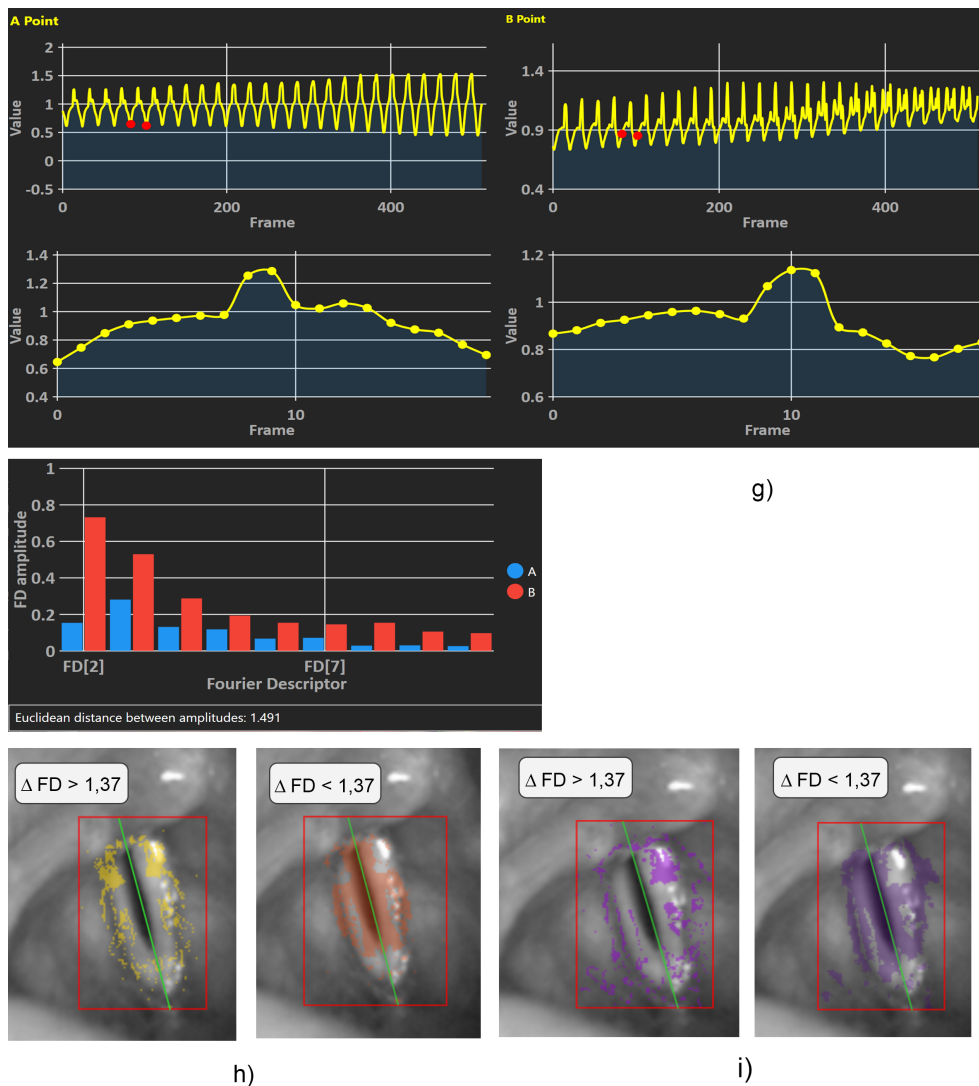
vek, které mají odlišné fázové spektrum – viz srovnání se symetrickou, zdravou hlasivkou, podkapitola 6.1 (e);

- analýza jednoho kmitu – rozložení hodnot FD pro jednotlivé harmonické Fourierovy řady vykazuje pro $FD[2]$, $FD[3]$, $FD[4]$ nesymetrické rozložení vzhledem k ose hlasivkové štěrbiny (f);
- analýza jednoho kmitu – indikuje v prostoru výskytu polypu nesymetrii vzhledem k ose hlasivkové štěrbiny, viz označené osově symetrické body $A - B$ (g);
- analýza jednoho kmitu – v odlišení osově symetrických bodů podle ΔFD , pro $\Delta FD = 1,45$ je patrná struktura výrazného odlišení pro oblast výskytu polypu vpravo (h), viz také setrvávající vysoké odlišení vůči referenčnímu $ref[x, y]$ pro $\Delta FD = 1,45$ (i).

6.3 Experiment č. 3: paréza

Uvedený záznam obsahuje hlasivky s parézou levé hlasivkové řasy, je tedy omezena její schopnost kmitání, výsledky viz obr. 6.3.





Obrázek 6.3: Popis výsledků vyšetření experimentu č. 3 – paréza

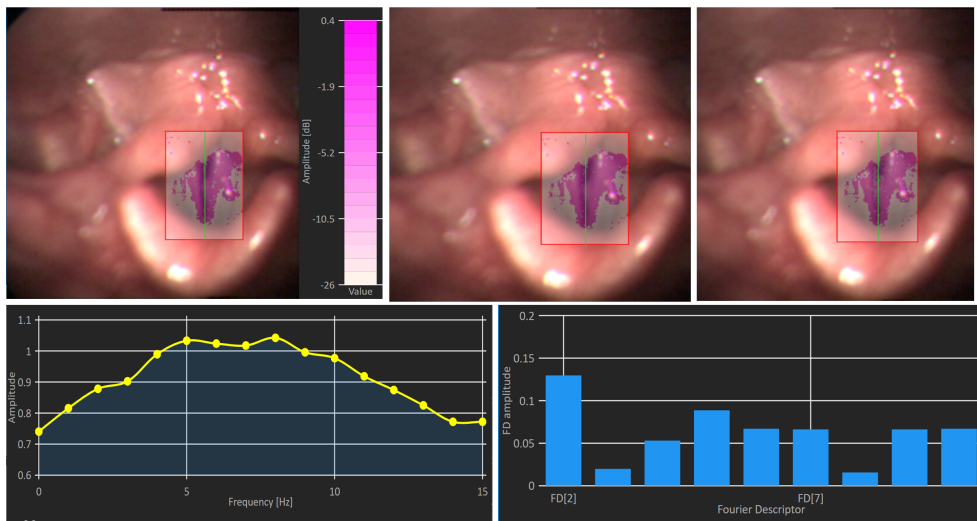
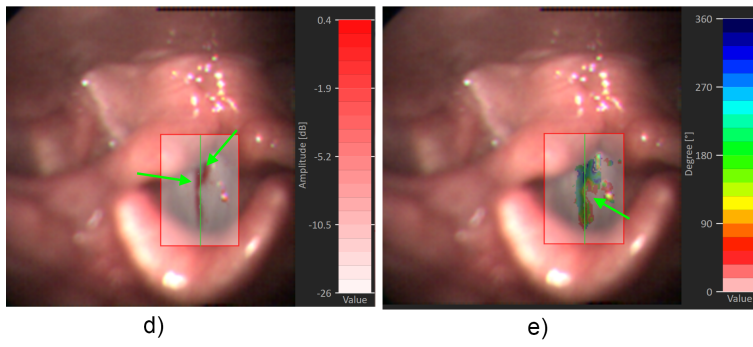
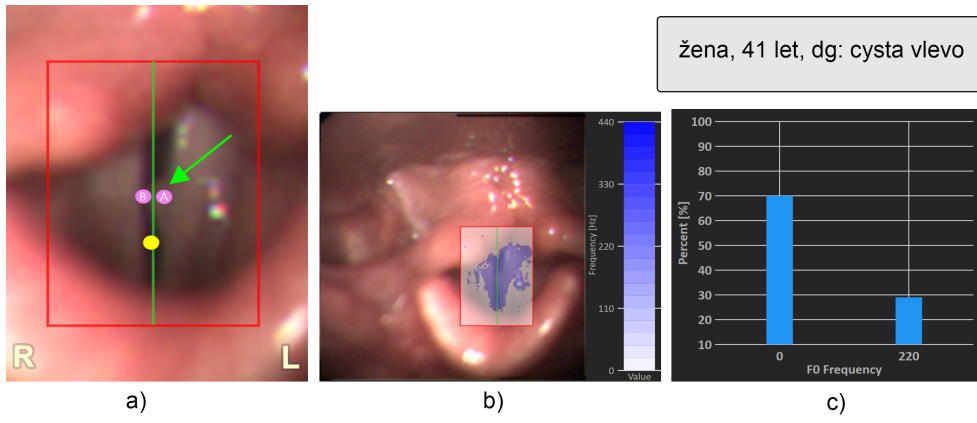
6.3.1 Zhodnocení

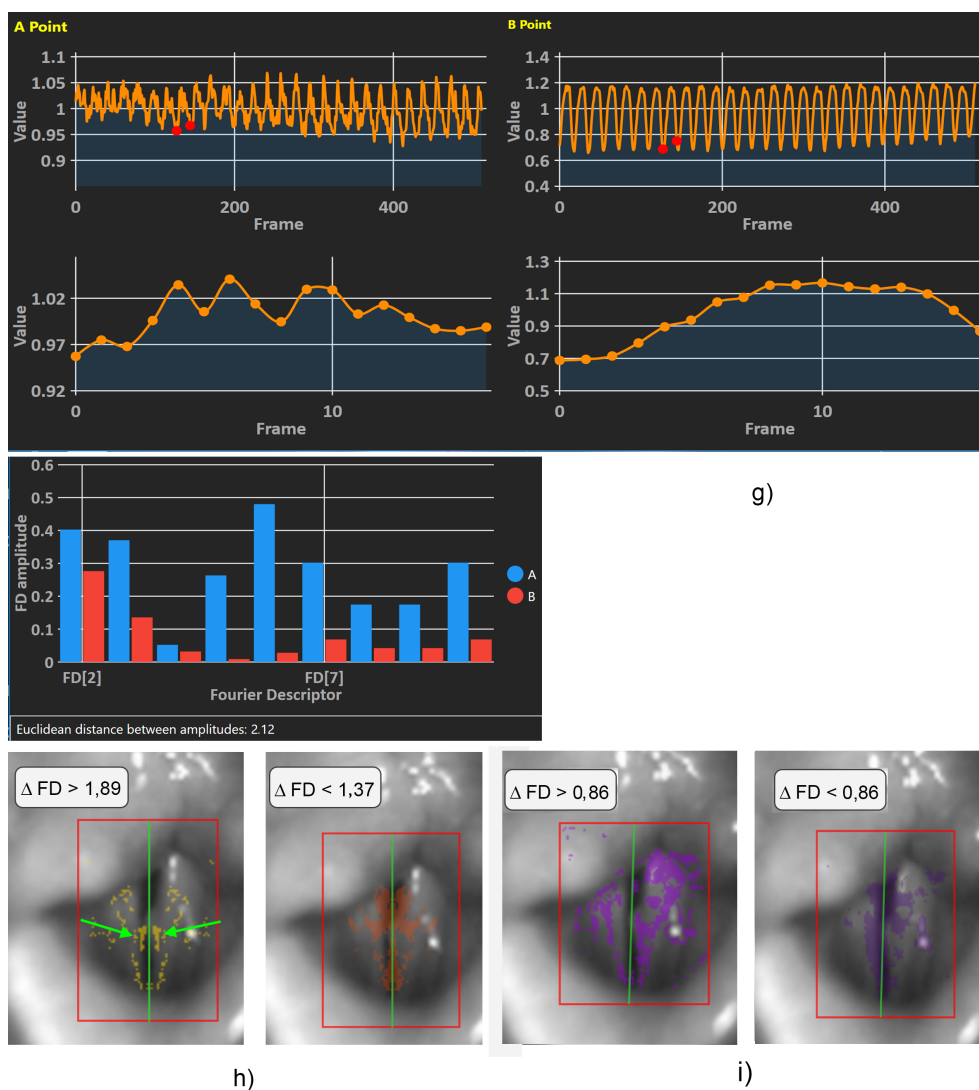
- rozložení amplitudového a fázového spektra ukazuje na nesymetrii hlasivkové štěrby glottis (c), (d);
- amplitudové spektrum – v pravé části hlasivek zaznamenáváme silné kmitání (změnu jasu), naopak levá hlasivka vykazuje minimální pohyb (d);
- fázové spektrum – pravá část hlasivek vykazuje fázi stejnou podél celé osy symetrie hlasivkové štěrby, levá hlasivka má odlišnou fázi (e);

- analýza jednoho kmitu – rozložení hodnot FD pro jednotlivé harmonické Fourierovy řady nevykazuje pro $FD[2]$, $FD[3]$, $FD[4]$ nesymetrické rozložení vzhledem k ose hlasivkové štěrbiny (f);
- analýza jednoho kmitu – odlišení pro dva osově symetrické body A–B není tak výrazné, i když lze nalézt takovou dvojici, kdy bude hodnota odlišení např. $\Delta FD = 1,49$ (g)
- analýza jednoho kmitu – celkové posouzení odlišnosti osově symetrických bodů pro $\Delta FD = 1,37$ detekuje anatomické struktury na hlasivkách směrem k zadní komisuře, pokud však provedeme bližší rozbor, je zdrojem detekované odlišnosti odlesk na anatomické struktuře hlasivkové řasy (h), (i)
- pro tuto kazuistiku lze říci, že vyšší vypovídací schopnost pro odlišení nestandardního chování hlasivek má analýza pomocí amplitudového a fázového spektra.

6.4 Experiment č. 4: cysta

Záznam experimentu obsahuje hlasivky s cystou na levé hlasivkové řase. Výsledky viz obr. 6.4.





Obrázek 6.4: Popis výsledků vyšetření experimentu č. 4 – cysta

6.4.1 Zhodnocení

- rozložení amplitudového a fázového spektra ukazuje na nesymetrii hlasivkové štěrbině glottis vlevo, kde se vyskytuje cysta (b), (d);
- amplitudové spektrum – v pravé části hlasivky podél osy symetrie hlasivkové štěrbině je vidět rozkmit hlasivek, vlevo je pak maximální rozkmit směrem k zadní komiseře, v místě cysty je hodnota amplitudového spektra minimální (d);
- fázové spektrum – pravá část podél osy symetrie hlasivkové štěrbině má shodnou fázi změny jasu s anatomickými strukturami hlasivkové

řasy vlevo směrem k zadní komisuře (e);

- analýza jednoho kmitu – rozložení hodnot FD pro jednotlivé harmonické Fourierovy řady nevykazuje pro $FD[2]$, $FD[3]$, $FD[4]$ nesymetrii vzhledem k ose symetrie hlasivkové štěrbině (f);
- analýza jednoho kmitu – indikuje v prostoru u vybraných osově symetrických bodů $A - B$ výraznější odlišení (bod A je umístěn do zdravé tkáně, B je cysta) $\Delta FD = 2, 12$ (g);
- pokud se zaměříme na celkové odlišení podle ΔFD , pak je vidět odlišení u několika struktur, které odpovídají cystě vlevo (ale jen části) a zdravé tkáni vpravo (h), (i);
- u této kazuistiky se doplňuje metoda jednoho kmitu a FFT analýza v detekci nežádoucích struktur na hlasivkách.

Kapitola 7

Závěr

Tato práce se soustřeďuje na frekvenční analýzu obrazových bodů v záznamu. Dle základní i rozšiřující literatury bylo vytvořeno programové vybavení implementující metody *FFT Point Analysis* a *FFT Area Analysis* vizualizující dynamické změny pohybu hlasivek v záznamu HSV.

Práce navíc kromě základní frekvence změny jasu obrazového bodu $F_{0(x,y)}$ zahrnuje funkcionalitu pro výpočet základní hlasivkové frekvence $F_{0(global)}$ z audiozáznamu, jenž je běžnou součástí vyšetření HSV kamerou. V realizační části se proto práce zabývá metodami, jak tuto frekvenci spolehlivě odhadnout a současně stanovuje tzv. interval odhadu, tj. zda se sobě tyto dvě frekvence přibližují.

V další části realizace se práce soustřeďuje na analýzu jednoho kmitu. Tato část byla zahrnuta nad rámec a nevyskytuje se v dostupné literatuře věnující se zpracování HSV záznamu. Testována byla odlišnost dílčích kmitů jak u zdravých hlasivek, tak u hlasivek s patologií. Následně bylo implementováno několik vizualizačních metod, například odlišnost kmitů dle osy glottis, odlišnost zvoleného referenčního kmitu vůči ostatním $p(x, y) \in ROI$. Dále bylo zahrnuto zobrazení spekter dle jednotlivých harmonických.

Vytvořená experimentální aplikace je určena pro medicínskou praxi zaměřující se na včasnou detekci nesymetrie pohybu hlasivek. Aplikace obsahuje metody, jež jsou schopny detekovat nesymetrické chování, nicméně může být výstup zanesen artefakty z důvodů odlesků na sliznici glottis nebo důsledky vzniklé při pohybu kamery. Vhodným rozšířením v oblasti zpracování HSV videozáznamu i této práce by mohlo být právě předzpracování HSV videozáznamu metodami registrace obrazu. U některých snímků je patrný pohyb

endokamery do stran, ale i přiblížení a natočení vůči snímaným hlasivkám. Práce redukující tyto důsledky posunu endokamery by byla žádoucí v této oblasti.

Literatura

- [1] *Dokumentace Color High-Speed Video System (Model 9170), KIPS (Model 9181)* [online]. KayPENTAX. [cit.]. Dostupné z: <http://www.kayelemetrics.com/>.
- [2] *Dokumentace HRES ENDOCAM (Model 5562)* [online]. Richard Wolf. [cit. 1.3.2018]. Dostupné z: http://www.kungshusen.se/endoskopi/pdf/Oron_nasa_hals/High_speed.pdf.
- [3] AICHINGER, P. et al. *Spectral Analysis of Laryngeal High-Speed Videos: Case Studies on Diplophonic and Euphonic Phonation*, s. 81–84. ., 2013.
- [4] CHEN, G. – KREIMAN, J. – ALWAN, A. The glottaltopogram: a method of analyzing high-speed images of the vocal folds. *Computer speech and language*. 2014, 28, 5, s. 1156–1169. doi: 10.1016/j.csl.2013.11.006.
- [5] ETTLER, T. *Detekce a hodnocení videozáznamu pohybu hlasivek z vysokorychlostní kamery. Odborná práce ke státní závěrečné zkoušce*. KIV-FAV-ZČU Plzeň, 2017.
- [6] GRANQVIST, S. – LINDESTAD, P. A method of applying Fourier analysis to high-speed laryngoscopy. *Journal of the Acoustical Society of America*. 2001, 110, 6, s. 3193–3197.
- [7] GRAY, H. *Anatomy of the human body*. Lea & Febiger, 1918. ISBN ISBN 1-58734-102-6.
- [8] HLAVÁČ, V. – SEDLÁČEK, M. *Zpracování signálů a obrazů*. Praha: Vydavatelství ČVUT, 2001. ISBN ISBN 80-01-02114-9.
- [9] ČIHÁK, R. *Anatomie 2, 1. vyd.* Avicenum Praha, 1988. ISBN 80-060-88.
- [10] KNOTKOVÁ, S. *Hlasivky* [online]. Oracle, 2014. [cit. 2017/25/12]. Dostupné z: <http://www.symptomy.cz/>.
- [11] KRENMAYR, A. et al. Visualizing phase relations of the vocal folds by means of high-speed videoendoscopy. *Journal of Speech, Language, and*

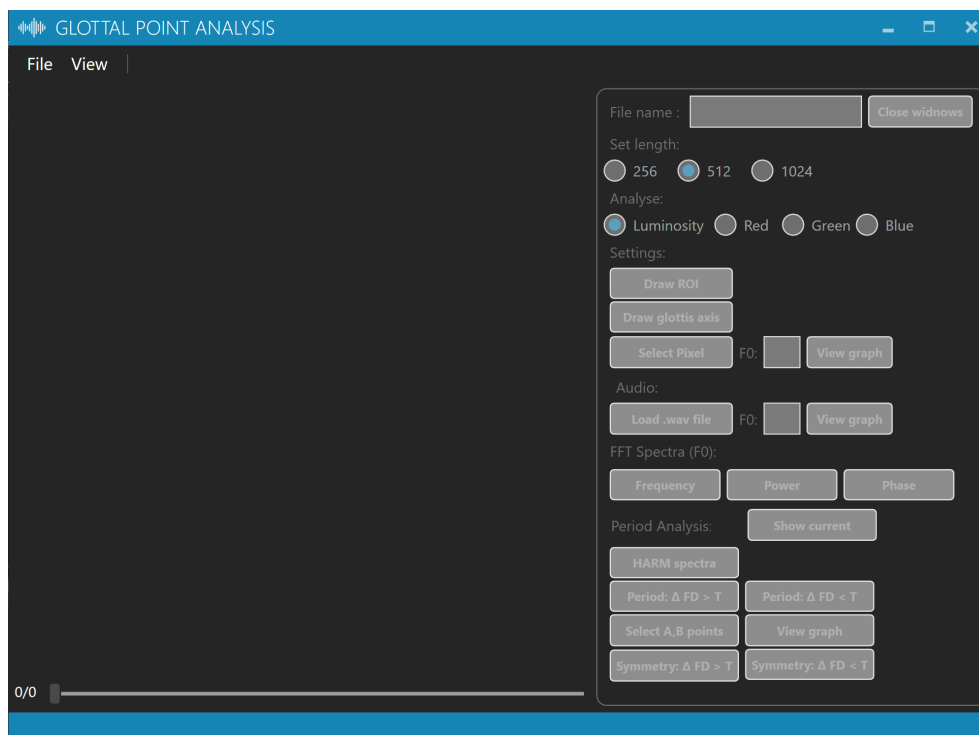
- Hearing Research : JSLHR*. 2011, 58, 5, s. 1488–1492. doi:
10.1044/2015_JSLHR-S-14-0253.
- [12] KROUPA, L. *Systém vyhodnocení parametrů jednoho kmitu hlasivek, diplomová práce*. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, 2015.
- [13] KUČERA, M. – FRÍČ, M. – HALÍŘ, M. *Praktický kurz hlasové rehabilitace a reedukace*. Opočno, 2010.
- [14] KUFNER, A. *Fourierovy řady*. ACADEMIA Praha, 1969.
- [15] ŠLAPÁK, I. – JANEČEK, D. – LAVIČKA, L. *Základy otorinolaryngologie a foniatrie pro studenty speciální pedagogiky*. Elportál, Brno : Masarykova univerzita, 2009.
- [16] MIDDLETON, G. Pitch detection algorithms. *Produced by The Connexions Project and licensed under the Creative Commons Attribution License*. 2003, s. 1–6.
- [17] NOVAK, A. *Foniatrie a pedaudiologie II*. UNITISK spol. s. r. o., 1996.
- [18] PEŠTA, J. et al. Vocal Cord Kinematics - New Evaluation Parameters. *Otorinolaryngologie a foniatrie*. 2016, 65, s. 88 – 96. ISSN 1210-7867.
- [19] PÁNEK, L. *Informační systém pro analýzu hlasového pole, diplomová práce*. Západočeská univerzita v Plzni, Fakulta aplikovaných věd, 2012.
- [20] PRESS, W. H. et al. *Numerical Recipes in C*. Cambridge University Press, 1992.
- [21] PSUTKA, J. *Komunikace s počítačem mluvenou řečí*. Praha: Academia, 1996. ISBN 80-200-0203-0.
- [22] REECE, J. B. et al. *Campbell biology, 9. vyd.* Benjamin Cummings, 2010. ISBN 0321558235.
- [23] SAKAKIBARA, K.-I. et al. Modal analysis of vocal fold vibrations using laryngotopography, 01 2010.
- [24] SCHAEFFER, N. Pre- and Poststimulation Study on the Phonatory Aerodynamic System on Participants with Dysphonia. *Journal of Voice*. 2017, 31, 2, s. 254.e1 – 254.e9. ISSN 0892-1997. doi:
<https://doi.org/10.1016/j.jvoice.2016.06.020>. Dostupné z: <http://www.sciencedirect.com/science/article/pii/S0892199716301254>.

- [25] ŠVEC, J. *Studium mechanicko-akustických vlastností zdroje lidského hlasu, doktorská dizertační práce.* Univerzita Palackého v Olomouci, Přírodovědná fakulta, 1996.
- [26] ŠVEC, J. – ŠRAM, F. – SCHUTTE, H. Videokymografie: Nová vysokofrekvenční metoda vyšetřování kmitů hlasivek. 01 1999, 48, s. 155–162.
- [27] VOKŘÁL, J. *Akustické parametry chraptivosti, doktorská dizertační práce.* České vysoké učení technické v Praze, Fakulta elektrotechnická, 1998.

Příloha A

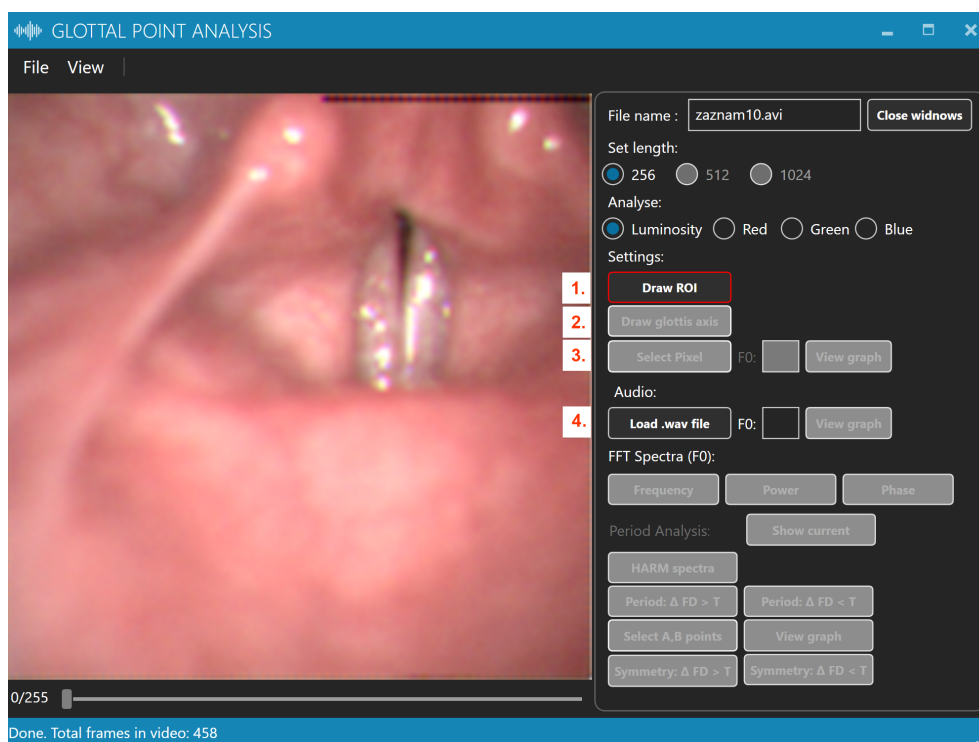
Uživatelská dokumentace k aplikaci

Spuštěním aplikace se zobrazí úvodní okno (viz obr. 7.1).



Obrázek 7.1: Úvodní okno aplikace

Načtení videozáznamu provedeme v horní liště **File** → **Open HSV**. V nově otevřeném okně vybereme videozáznam z přiložené testovací množiny záznamů. Načtený videozáznam v aplikaci je na obr. 7.2.



Obrázek 7.2: Aplikace s načteným videozáznamem

K další analýze je nutné provést první 3. kroky (viz obr. 7.2):

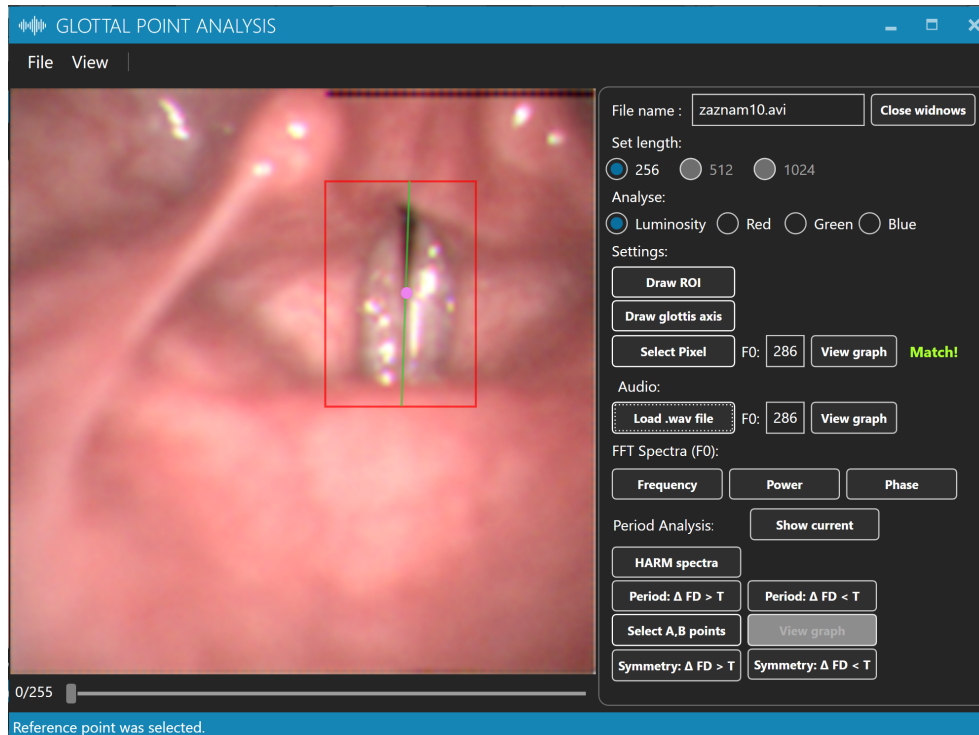
1. Draw ROI – nakreslit oblast zájmu *ROI*
2. Draw glottis axis – vyznačit osu glottis
3. Select Pixel – výběr referenčního bodu $ref[x, i]$
4. Load wav. file – (volitelné) nahrání příslušného zvukového záznamu k porovnání základní frekvence změny jasu obrazového bodu $F_{0(x,y)}$ a základní hlasivkové frekvence $F_0(global)$ zvukového záznamu.

Po provedení popsaných kroků lze přistoupit k hlavním výpočtům aplikace (viz obr. 7.3).

Dále je možné procházet jednotlivé snímky posuvníkem pod zobrazeným snímkem videa a změnit snímek aktuální analýzy, resp. dochází k posouvání výběru kmitů.

Délku analyzované videosekvence lze nastavit zaškrtnutím délky v panelu **Set length**: na volbu: 256, 512, nebo 1024, pokud je však videozáznam dostatečně dlouhý.

Chceme-li změnit analyzovaný signál z jasu monochromatického zobrazení, zaškrtneme volbu R , G , B v **Analyse** pro analýzu červené, zelené či modré barevné složky.



Obrázek 7.3: Provedené kroky nezbytné k další analýze

Popis jednotlivých tlačítek a jejich funkcí

Analýza oblasti *ROI*:

- **Frequency** – frekvenční spektrum s možností vykreslení histogramu
- **Power** – amplitudové spektrum
- **Phase** – fázové spektrum

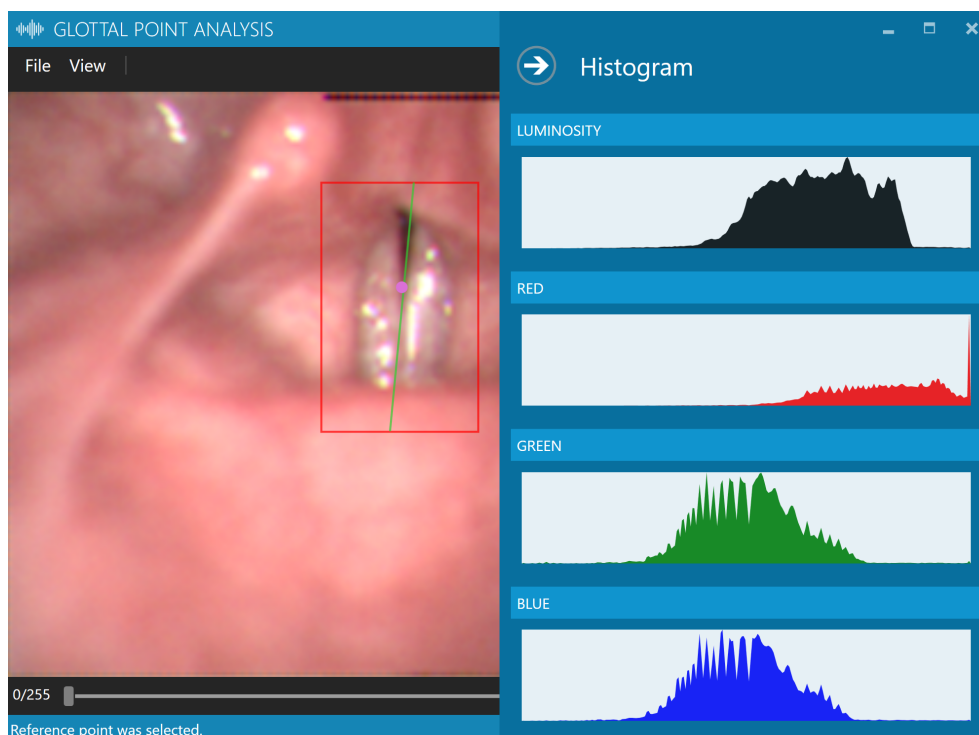
Analýza kmitu:

- **Show current** – zobrazení kmitu pro aktuálně vybraný snímek video-
sekvence spolu s vypočítanými *Fourier Descriptors*
- **HARM spectra** – rozložení hodnot FD pro jednotlivé harmonické Fou-
rierovy řady

- **Period: $\Delta FD > T$** – zobrazení bodů $\in ROI$ vůči $ref[x, y]$, jejichž míra odlišení ΔFD přesáhla stanovenou mez $T_{\Delta FD}$
- **Period: $\Delta FD < T$** – zobrazení bodů $\in ROI$ vůči $ref[x, y]$, jejichž míra odlišení ΔFD nepřesahuje stanovenou mez $T_{\Delta FD}$
- **Select A,B points** – srovnávací analýza dvou osově symetrických bodů A a $B \in ROI$, tlačítko **View grap** zobrazí vybrané kmity a výsledek analýzy pro aktuálně vybraný snímek videosekvence
- **Symmetry: $\Delta FD > T$** – zobrazení dvojice osově symetrických bodů $\in ROI$, jejichž míra odlišení ΔFD přesáhla stanovenou mez $T_{\Delta FD}$
- **Symmetry: $\Delta FD < T$** – zobrazení dvojice osově symetrických bodů $\in ROI$, jejichž míra odlišení ΔFD je pod stanovenou mezí $T_{\Delta FD}$

Všechna otevřená okna s výpočty lze zavřít poklepáním **Close windows** nahore v panelu.

Histogram četností výskytu složek R,G,B a jasů v monochromatickém zobrazení lze zobrazit pro aktuální snímek volbou **View** \rightarrow **Histogram** (viz obr. 7.4).



Obrázek 7.4: Histogram aktuálního snímku videosekvence

Struktura přiloženého DVD

Na přiloženém DVD se nachází:

- Instalační soubor pro spuštění aplikace k zpracování videozáznamů včetně zdrojových kódů
- Text této práce ve formátu PDF a zdrojová data
- Testovací soubory dat: videosekvence a příslušný zvukový záznam
- Vybrané citované zdroje v práci v elektronické podobě