

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

Diplomová práce

Systém vyhodnocení parametrů jednoho kmitu hlasivek

Prohlášení

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

V Plzni dne 9. května 2015

Lukáš Kroupa

Poděkování

Rád bych poděkoval vedoucímu diplomové práce Ing. Pavlu Novému Ph.D. za odborné vedení, cenné rady a čas, který mi věnoval při řešení dané problematiky. Dále děkuji Ing. Jiřímu Peštovi a MUDr. Monice Vohlídkové z ORL kliniky Fakultní nemocnice Plzeň za konzultace metodiky vyšetření hlasivek a poskytnuté informace z oblasti hlasových poruch. Velký dík patří také rodině za podporu během studia.

Abstrakt

Tato práce se věnuje návrhu metod pro stanovení standardního kmitu hlasivek v záznamu dlouhodobé fonace s cílem detekce odlišných kmitů v rámci této fonace. Dále se zabývá návrhem parametrů pro hodnocení shody, resp. odlišnosti jednotlivých period kmitů od tohoto standardního kmitu. Dílčím problémem je určení základní frekvence hrtanového tónu a detekce period fonace. Odlišnost kmitů je hodnocena podle parametrů jednotlivých period signálu ve frekvenční a časové oblasti. Navržené metody jsou implementovány v experimentální aplikaci pro vizualizaci a zpracování zaznamenaných dat. Aplikace umožňuje zobrazit a porovnat zvukový signál, EGG signál, videosekvenci snímků hlasivek a videokymogram.

Klíčová slova: analýza signálu, standardní kmit hlasivek, hlasivky, porucha hlasu, elektroglottograf, dlouhodobá fonace

Abstract

Aim of this work is to devise methods for selection of standard vocal fold oscillation in audio recording of prolonged phonation in order to detect irregular oscillations. Next are defined parameters for evaluation of differences of vocal fold oscillation periods from standard oscillation. Part of the problem is estimation of fundamental frequency of larynx tone and detection of phonation periods. Difference rating of vocal fold oscillations is based on parameters of signal period in time and frequency domain. Proposed methods are implemented in experimental software to display and process recorded data. The software allows to display and compare audio signal, EGG signal, video recording and videokymogram.

Key words: signal analysis, standard vocal fold oscillation, vocal cords, voice disorder, electroglottography, prolonged phonation

Obsah

1	Úvod a cíle práce	1
2	Hlasové ústrojí a tvorba hlasu	2
2.1	Anatomie hlasivek	2
2.2	Tvorba hlasu	3
2.3	Poruchy hlasu	4
2.3.1	Přehled hlasových poruch	5
2.4	Metody vyšetření	6
2.4.1	Aerodynamické metody	6
2.4.2	Akustické metody	7
2.4.3	Elektrofyzilogické metody	8
2.4.4	Optické metody	9
2.5	Metody používané na ORL FN Plzeň	10
2.5.1	Standardní metody vyšetření	10
2.5.2	Speciální metody vyšetření	10
3	Data a zaznamenávací zařízení	11
3.1	Určení tvaru periody kmitu	12
3.1.1	Audiozáznam	12
3.1.2	Elektroglottogram	14
3.2	Synchronizace záznamů	16
3.3	Vlastnosti zaznamenaných dat	17
3.3.1	Audiozáznam	17
3.3.2	Elektroglottogram	20
3.3.3	HSV a Videokymogram	22
4	Analýza záznamu	23
4.1	Detekce kmitu hlasivek	23
4.1.1	Určení základní frekvence	23
4.1.2	Algoritmus detekce periody kmitu	27

5	Určení standardního kmitu fonace	29
5.1	Převzorkování period	29
5.2	Standardní perioda v časové oblasti	31
5.2.1	Průměrná hodnota vzorků	31
5.2.2	Medián vzorků	31
5.3	Standardní perioda ve frekvenční oblasti	32
5.3.1	Minimální eukleidovská vzdálenost mediánu amplitudového spektra	33
5.3.2	Skóre četnosti amplitudového spektra	34
6	Detekce nestandardních period	35
6.1	Detekce v časové oblasti	35
6.1.1	Průměrná odchylka vzorků periody	35
6.1.2	Vzdálenost vzorků periody	35
6.1.3	Skóre vzorků mimo pásmo σ	36
6.2	Detekce ve frekvenční oblasti	37
6.2.1	Eukleidovská vzdálenost amplitudového spektra	37
6.2.2	Vzdálenost Fourierových deskriptorů	37
6.3	Výpočet dalších parametrů kmitu	39
6.3.1	Kontaktní kvocient EGG signálu	39
6.3.2	Výpočet SCORE	41
7	Hodnocení výsledků	43
7.1	Určení základní frekvence	43
7.2	Detekce kmitu	46
7.3	Metoda SCORE	48
7.4	Kontaktní kvocient hlasivek	51
7.5	Detekce nestandardních kmitů	53
7.5.1	Záznam č. 1	55
7.5.2	Záznam č. 2	59
7.5.3	Záznam č. 3	63
7.5.4	Záznam č. 4	67
7.5.5	Záznam č. 5	71
7.5.6	Záznam č. 6	75
7.5.7	Diskuze výsledků metod detekce	78
7.6	Další možné rozšíření práce	79
8	Závěr	80

Přílohy diplomové práce	82
A Adresářová struktura DVD	83
B Uživatelský manuál	84
C Struktura databáze pacientů a vyšetření	89
D Dekódování souboru BLD	90
D.1 Analýza souboru bld	91
D.2 Dekódování souboru bld	92
D.2.1 Rekonstrukce snímku v rozlišení 256×256	95
D.3 Hodnocení výstupu	97

1 Úvod a cíle práce

Hlas jako prostředek verbální komunikace umožňuje člověku zapojení do společnosti. Hlasová porucha znemožňující komunikaci tak má zásadní vliv na kvalitu života a je proto důležité hlasové poruchy včas diagnostikovat. Záměrem této práce je přispět do systému včasné diagnostiky hlasových poruch a poskytnout nový pohled na data z běžných vyšetření, který by doplňoval existující metody a systémy užívané k určení diagnózy.

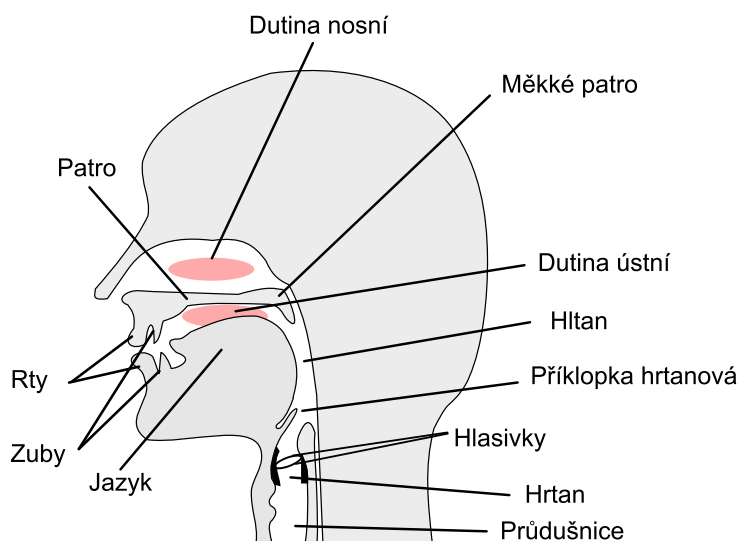
Cílovou skupinou pacientů tohoto systému jsou lidé s problémy, u kterých nelze běžnými vyšetřovacími postupy určit jednoznačně jejich příčinu nebo ji zachytit pozorováním pohybu hlasivek. Charakteristické pro tyto problémy je, že se při vyšetření projevují například náhodným zakmitáním hlasivek nebo občasnou asymetrií pohybu levé a pravé hlasivky.

Cílem práce je využít data z vyšetření pomocí integrovaného systému zaznamenávajícího současně zvukový záznam fonace, EGG záznam a vysokorychlostní videozáznam pohybu hlasivek. Projevy změn v chování hlasivek lze primárně určit podle změn v akustickém signálu při fonaci vokálů, především „a“, „e“, „i“. Doplnující funkci plní následný rozbor záznamu elektroglossografu a videosekvence HSV na pozici, která odpovídá detekované odlišnosti v chování hlasivek. Součástí práce je implementace a ověření navrhovaného přístupu, založeného na analýze a hodnocení parametrů jednotlivých kmitů a následném určení standardního kmitu statistickými metodami. Porovnáním se standardním kmitem bude možné nalézt a vyhodnotit odchylky jednotlivých kmitů v záznamu a určit časové úseky, v nichž dochází ke změně kinematiky hlasivek.

Dílčí problém práce je návrh a ověření postupu rozdělení záznamu fonace na jednotlivé periody kmitů hlasivek a s tím spojený výběr vhodné metody určení základní frekvence fonace.

2 Hlasové ústrojí a tvorba hlasu

Lidský hlas vzniká průchodem vzduchu hlasovým ústrojím a je ovlivněn zejména stavbou hrtanu a rezonančních dutin horních cest dýchacích. V hrtanu je kmitáním hlasivek produkován hrtanový tón, který je dále zesílen a upraven v rezonančních dutinách, čímž vzniká charakteristický hlas každého člověka.



Obrázek 2.1: Hlasové ústrojí.

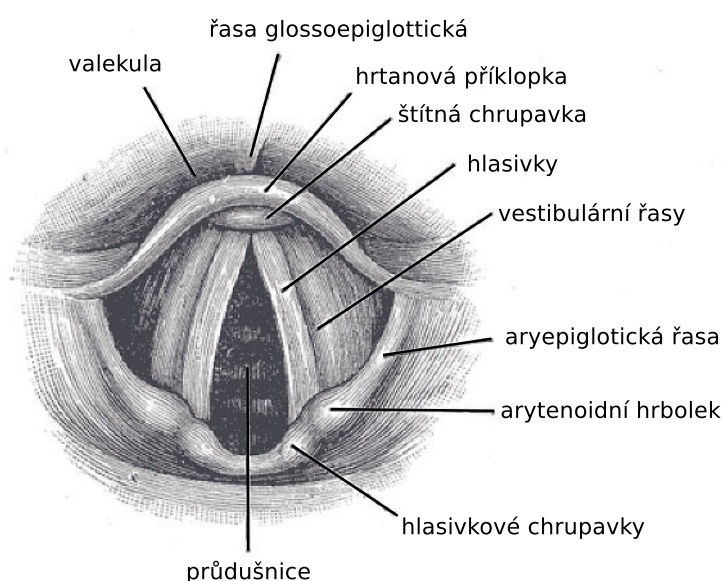
2.1 Anatomie hlasivek

Hlasivky jsou součástí hrtanu, což je chrupavkami vyztužená trubice dlouhá 5 až 7 centimetrů, uložená nad průdušnicí, viz ilustrace 2.1. Hrtan má mimo tvorby hlasu také funkci dýchací a ochrannou, kterou zajišťuje produkcí hlenu, při podráždění cizím tělesem vyvolává kašlací reflex pro uvolnění dýchacích cest. Vstup do hrtanu je chráněn hrtanovou příklopkou, která při polykání brání vstupu potravy do průdušnice. [16]

Střední část hrtanu je zúžena dvěma páry řas, řasami vestibulárními, někdy také nazývané nepravé hlasivky, níže uložený pár řas jsou hlasivky, mezi kterými je hlasivková štěrbina. Hlasivka je upnuta vpředu na chrupavce

štítné, vzadu na chrupavce hlasivkové. Hlasivky jsou složeny z vazivové části, podél níž jsou hlasivkové svaly, které se podílejí na napínání hlasivek spolu s posuny chrupavek tvořících hrtan a chrupavek hlasivkových, na povrchu jsou kryty sliznicí. Umístění hlasivek je znázorněno na obrázku 2.2.

Délka hlasivek určuje výšku hlasu, průměrná délka hlasivek u žen je 18 mm, u mužů 22 mm. Při běžné řeči hlasivky kmitají s frekvencí přibližně 110 Hz u mužů a 200 Hz u žen, napnutím hlasivek pomocí svalů lze při zpěvu dosáhnout frekvence až 480 Hz (vysoký ženský hlas při zpěvu). [1]



Obrázek 2.2: Laryngoskopický pohled na hlasivky. [5]

2.2 Tvorba hlasu

Proces fonace lze rozdělit do tří kroků postupně probíhajících při průchodu vzduchu hlasovým ústrojím ([1, 13]).

1. **Sloupec vzduchu je vytlačován k hlasivkám.** - Vzduch je vytlačován z plic koordinovaným stahem bránice, břišních svalů, hrudních svalů a hrudního koše.

2. **Rozkmitání hlasivek.** - Hlasivky jsou svaly hrtanu staženy k sobě a dochází k hromadění vzduchu z plic. V momentě, kdy tlak vzduchu dosáhne hodnoty odpovídající rozmezí 5 - 10 cm vodního sloupce vzduch postupně rozevře spodní a horní část hlasivek. Podtlak způsobený náhlým únikem vzduchu vede ke vzniku Bernoulliho efektu, který způsobí opět postupné uzavření spodní a horní části hlasivek. Proud vzduchu je přerušen a proces kmitu se opakuje. Pohyb hlasivek při jednom kmitu je znázorněn na obrázku 2.3.

Hlasitost fonace se zvyšuje s vyšším průtokem vzduchu, který umožní větší rozevření hlasivek. Hlasivky tak zůstanou rozevřeny delší část cyklu a tím se zvýší amplituda zvukové vlny. Výška tónu fonace se zvyšuje s frekvencí kmitů hlasivek.

Tento model kmitání hlasivek popisuje *Myoelestická teorie* postulovaná v roce 1958 Van Den Bergem [15]. Dříve bylo kmitání hlasivek vysvětlováno pomocí *Neurochronaxické teorie*, podle které hlasivky vykonávají kmitavý pohyb aktivně. Tato teorie byla později vyvrácena zjištěním, že denervovaná hlasivka při jednostranné obrně nervu *n. recurrens* kmitá.[13]

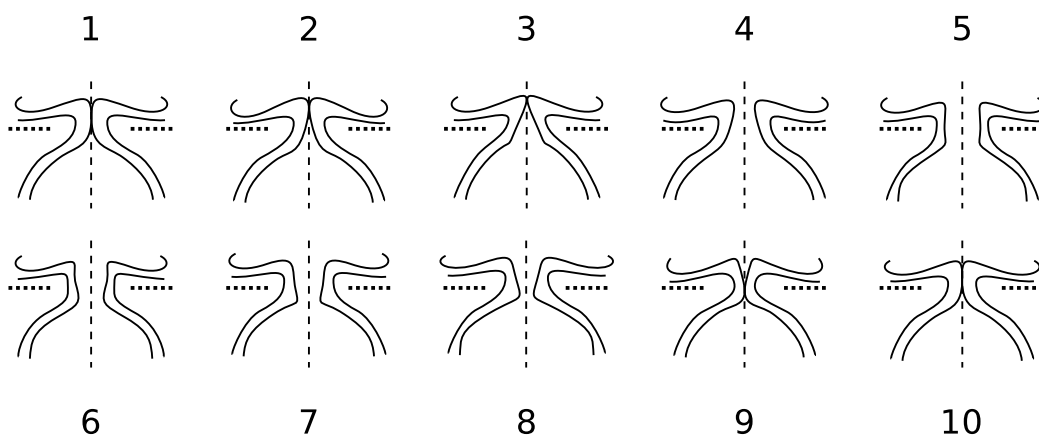
Neúplné dovření hlasivek nebo jiné poruchy kinematiky způsobují poruchy hlasu. Podrobnější popis příčin a projevů těchto patologických jevů je uveden v části 2.3.

3. **Rezonance v dutinách.** - V rezonančních dutinách, zejména hltanu, nosní a ústní dutině zvuk rezonuje a jsou zesíleny frekvence, na které je naladěna rezonance těchto dutin. Hlas tím získává specifické zabarvení.

2.3 Poruchy hlasu

Hlasové poruchy se nejčastěji rozdělují na funkční a organické. Funkční hlasové poruchy vznikají z důvodu vadné tvorby hlasu bez organických změn na hlasovém ústrojí, původ poruchy může být například psychický. Organické hlasové poruchy vznikají jako důsledek změn na struktuře hlasového ústrojí.

Z hlediska této práce jsou významné poruchy hlasu, projevující se zastřeným hlasem až chrapotem, jenž jsou způsobeny nesprávnou kinematikou hlasivek. Tento stav může být vrozený, nebo získaný a může se projevovat dočasně nebo trvale. Chrapot vzniká změnou hmoty hlasivek, nebo změnou



Obrázek 2.3: Průběh jednoho kmitu hlasivek.

závěru hlasové štěrbin. V prvním případě dochází k nepravidelnému kmitání hlasivek, které se může lišit ve frekvenci i amplitudě, v druhém případě dochází k nedomykavosti hlasivek, mezi kterými zůstává štěrbina, skrze niž uniká při fonaci vzduch.

Příčinou chrapotu může být onemocnění, zranění, či nadměrná hlasová zátěž. Zranění nebývají díky poloze hrtanu příliš častá, dochází k nim zejména při dopravní nehodě, pádu nebo obtížné intubaci. Nadměrné hlasové zátěži jsou vystaveni lidé dlouhodobě se vyskytující v hlučném prostředí, kde jsou nuceni zvyšovat hlas, dále pak učitelé, herci a často také děti. Potíže z nadměrné zátěže po hlasovém klidu odeznívají, ale mohou vést k závažnějším problémům, například vzniku hlasivkového uzlíku. Také běžná onemocnění dýchacích cest, například nachlazení, mohou vést až k paralýze hlasivkového nervu. Řada poruch nemá specifickou příčinu, dochází k samovolnému růstu útvarů na hlasivkách.[16, 14]

2.3.1 Přehled hlasových poruch

Zpěvácký uzlík - Slizniční zbytnění hrany hlasivky, uzlík nikdy nepostihuje hlubší vrstvu. Ve většině případů se vyskytuje na obou hlasivkách, znemožňuje úplné uzavření hlasivek.

Polyp hlasivky - Stopkatý útvar nasedající na hlasivku, ve většině případů jednostranný. Způsobuje neúplný uzávěr hlasivek. Pravděpodobně vzniká extrémní mechanickou zátěží kmitáním hlasivek, která vede

k porušení cév.

Reinkeho edém - Výrazný otok hlasivek, může dosahovat až enormních rozměrů, kdy se hlasivky překrývají, hlubší vrstvy hlasivek nejsou zasazeny.

Cysta hlasivky - Podslizniční výrůstek, v některých případech není přímo patrný, lze předpokládat podle rozšířených cév na horní ploše hlasivky.

Nádory hrtanu - Zhoubné i nezhooubné nádory, rozlišují se podle tkáně, z které vznikají. Velikost a umístění útvaru je značně rozmanité.

Poruchy inervace - Neurologická postižení v oblasti hrtanu. Může docházet k trvalým stahům svalů nebo naopak k obrně nervů. [9]

2.4 Metody vyšetření

2.4.1 Aerodynamické metody

Pneumografie

Tato metoda vyšetřuje dýchání během různých druhů fonace a v klidovém stavu. Výstupem měření je zvukový záznam a odpovídající hrudní a břišní křivka, na kterých je hodnocena dechová frekvence, poměr vdechu k výdechu a vztah mezi hrudní a břišní křivkou. Vyšetření bylo předepisováno u pacientů s podezřením na nesprávnou hlasovou techniku a u uchazečů o povolání označených jako hlasový profesionál.

V původní podobě byly dýchací pohyby přenášeny pásovým pneumografem a Mareyovým bubínkem na kymograf. Metoda byla několikrát zdokonalena, především použitím přesnějšího záznamového zařízení, ale postupně se přešlo na pneumotachografické vyšetření, umožňující zcela objektivně měřit funkci glottis.

Pneumotachografie

Principem tohoto vyšetření je měření rychlosti proudu a množství vzduchu procházejícího v glottis, měření subglottického tlaku a výpočtu ukazatelů

z těchto hodnot. Porovnáním naměřených hodnot zdravých a nemocných pacientů hodnotit funkci hrtanu a stupeň jeho postižení.

Pacient má při vyšetření ucpané nosní dírky a na ústech masku, citlivým snímačem je zaznamenáván proud vzduchu a současně je zaznamenávána fonace mikrofonem.

2.4.2 Akustické metody

Hlasové pole

Vyšetření hlasového pole, známé také jako VRP (Voice Range Profile), stanovuje frekvenční a dynamický rozsah hlasu. Záznamem fonace libovolné délky, při které se pacient snaží dosáhnout nejhlasitější a nejtišší fonace pro všechny frekvence získáme obraz hlasového pole. Metodou lze také hodnotit vliv hlasové zátěže, porovnáním výsledků vyšetření před a po zátěži (divadelní představení, zpěv nebo syntetické testy).

Multidimenzionální analýza

Digitální záznam a zpracování zvukového záznamu umožňuje vyhodnotit velké množství parametrů zvuku. Moderní systémy dokáží hodnotit až 33 různých parametrů hrtanového tónu a tak popsat základní kvality hlasu. Vstupem je záznam fonace vokálu „a“ v délce několik vteřin, z které jsou sledované parametry získány frekvenční analýzou, výpočtem statistických ukazatelů fonace a jednotlivých period. Získané hodnoty jsou porovnávány s normou stanovenou vyšetřením velkého množství zdravých pacientů.

Hodnocení kvality závěru hlasivek

Tato metoda umožňuje ohodnotit kvalitu závěru hlasivek známkou na stupnici 1 až 5. K hodnocení je použit jeden kmit hlasivek zvolený v záznamu fonace vokálu „a“. Ohodnocení kvality závěru poskytuje jednoznačné určení míry poruchy a lze použít k sledování úspěšnosti léčby. Metoda je založena na Fourierově analýze jedné periody kmitání hlasivek v záznamu fonace a aplikací regresního SCORE modelu, odvozeného ze znaleckého hodnocení a

výsledků vyšetřovacích metod. [12]

GRBAS

Subjektivní hodnocení hlasu expertním hodnotitelem v rozsahu 1 - 3. Je hodnoceno pět kritérií: G = grade (celkový stupeň disfonie), R = roughness (drsnot, chraptivost), B = breathness (dyšnost), A = asthenicity (slabost), S = strain (napětí, tlak).

2.4.3 Elektrofyzilogické metody

Elektroglottografie

Elektroglottografické vyšetření je založeno na snímání změn elektrického odporu během fonace, při změně tvaru hlasivkové štěrbiny. Výstupem měření je křivka glottogramu, na níž jsou hodnoceny fáze otevření a uzavření hlasivek a doba otevření hlasivek. Vyšetření je neinvazivní, lze jej provést souběžně s jiným vyšetřením. Záznam glottogramu se provádí při fonaci vokálu „a“ snímacími elektrodami umístěnými nad křídly štítné chrupavky. Do elektrod je přiváděno velice slabé střídavé napětí o vysoké frekvenci. Otevírání a zavírání hlasivek mění odpor mezi elektrodami a moduluje procházející proud. Často se používá jako signál pro synchronizaci světla stroboskopického zdroje při laryngostroboskopii. [17]

Elektromyografie

Toto vyšetření měří elektrické biosignály vycházející ze svalů při jejich aktivitě. EMG vyšetření významně přispěla k objasnění aktivity vnitřních hrtanových svalů během fonace. Vyšetření je invazivní, elektrody se vpichují do svalů a je proto omezeno především na diagnostiku obrny zvrátneho nervu. Vždy je nutné pečlivě zdokumentovat kvalitu hlasu před a po vyšetření.

2.4.4 Optické metody

Laryngoskopie

Laryngoskopie patří mezi základní vyšetřovací techniky, rozlišují se dvě metody - přímá a nepřímá. V případě nepřímé laryngoskopie lékař používá hrtanové zrcátko, které vsune do ústní dutiny, nebo flexibilní laryngoskop, složený z optických vláken, který je zaváděn obvykle nosem. Přímá laryngoskopie umožňuje lékaři přímo sledovat vnitřek dýchacích cest. Používá se při chirurgických zákrocích, pro diagnostické účely se užívá jen omezeně, jelikož se provádí v celkové anestezii. Přímá i nepřímá laryngoskopie jsou vhodné pouze k popisu základních struktur, nikoli funkce hrtanu.

Laryngostroboskopie

Aby bylo možné pozorovat pohyb hlasivek, které kmitají rychleji, než může lidské oko zaregistrovat, využívá se laryngostroboskopie. Principem stroboskopie je osvětlování kmitajících hlasivek záblesky světla s nepatrně odlišnou frekvencí než je frekvence kmitání. Hlasivky jsou tak osvětleny vždy v jiné fázi kmitu a dojde k zdánlivému zpomalení pohybu. Frekvence stroboskopu je řízena nejčastěji podle signálu elektrolottografu, moderní stroboskopy umožňují synchronizaci záblesků podle frekvence fonace.

Videokymografie

Videokymografie je založena na sledování kmitu hlasivek v úzkém příčném řezu. K záznamu se používá vysokorychlostní kamera, schopná zaznamenat až 8000 snímků za vteřinu, zaznamenávající pouze jediný řádek, nebo kamera zaznamenávající celé snímky hlasivek, které jsou následně digitálně upraveny do videokymogramu. Zaznamenávaný řez lze umístit do libovolné pozice a tak vyšetřit symetrii kmitu celé hlasivky. Metoda byla vyvinuta a klinicky rozšířena za spolupráce s českými spoluautory [18, 7]. Výhodou oproti laryngostroboskopii je možnost posouzení reálného kmitu hlasivek, nikoli kmitu poskládaného z mnoha jiných, a také možnost sledovat pohyb hlasivek i při velmi nepravidelném kmitání. [3]

Vysokofrekvenční snímání

Vyšetření vysokorychlostní kamerou je moderní metoda, která umožňuje zaznamenat celkový pohled na hlasivky s dostatečnou časovou přesností, aby byl pozorovatelný průběh jednotlivých kmitů. Běžně používané vysokorychlostní endokamery snímají 4000 snímků za vteřinu. Vyšetření je neinvazivní, vysokorychlostní kamerou vloženou do ústní dutiny pacienta je zaznamenán přibližně dvouvteřinový záznam, který obsáhne několik desítek kmitů hlasivek. Záznam je poté přehráván zpomaleně rychlostí 15 snímků za vteřinu.

2.5 Metody používané na ORL FN Plzeň

2.5.1 Standardní metody vyšetření

Na otorinolaryngologické klinice Fakultní nemocnice Plzeň jsou v diagnostice standardně používané následující metody vyšetření:

- Vysokofrekvenční snímání
- Hodnocení kvality závěru hlasivek metodou SCORE
- Hlasové pole
- Multidimenzionální analýza
- Elektrolottografie

2.5.2 Speciální metody vyšetření

U vybraných pacientů jsou prováděny také následující speciální metody vyšetření:

- Zátěžové testy
- Aerodynamické testy

3 Data a zaznamenávací zařízení

Cílem práce je využít data z vyšetření hlasivek pomocí integrovaného systému zaznamenávajícího současně zvukový záznam fonace, vysokorychlostní videozáznam HSV kamerou a záznam elektrolottografu. Předností tohoto systému je možnost hodnotit stav hlasivek a vlastnosti fonace z více nezávislých směrů, nález v jednom typu záznamu je možné ověřit ve stejném okamžiku jinou metodou. Počátek a konec záznamu je synchronní stejně jako vzorkování zvukového a EGG signálu. Nevýhodou integrovaného systému je omezená délka záznamu v rozsahu pouze 0,5 až 2 vteřiny a fonace vokálu „i“ z důvodu omezené viditelnosti hlasivek, podrobněji je tento problém popsán v části 3.3.

Pro ověření vlivu způsobu záznamu byly prováděny srovnávací testy s daty zaznamenanými prostorovým mikrofonom a elektrolottografem odděleně od HSV kamery. Tato data se odlišují zejména větší vzdáleností mikrofону od úst a fonovaným vokálem „a“. Výhodou je také delší doba fonace.

Konkrétní data, která jsou uvedena v této práci a která byla použita pro vývoj a ověření navrhovaných postupů jsou zaznamenána systémem Richard Wolf HRES ENDOCAM 5562 v kombinaci s glottografem EG2-PCX2 od společnosti Glottal Enterprises.



Obrázek 3.1: Endokamera HRES ENDOCAM 5562 of firmy Richard Wolf.



Obrázek 3.2: Elektrody elektrolottografu EG2-PCX2 od společnosti Glottal Enterprises

Software vytvořený v rámci této práce je navržen pro zpracování dat z výše uvedeného integrovaného systému, ale formát vstupních dat není nijak omezen a je možné zpracovat libovolná vhodná data. Pro zpracování a analýzu lze zvolit z následujících kombinací vstupních dat:

Mikrofon

Zvukový záznam z prostorového mikrofonu.

Mikrofon + EGG

Zvukový záznam prostorového mikrofonu v kombinaci s glottografem.

Mikrofon + HSV

Zvukový a videozáznam z HSV videokamery.

Mikrofon + EGG + HSV

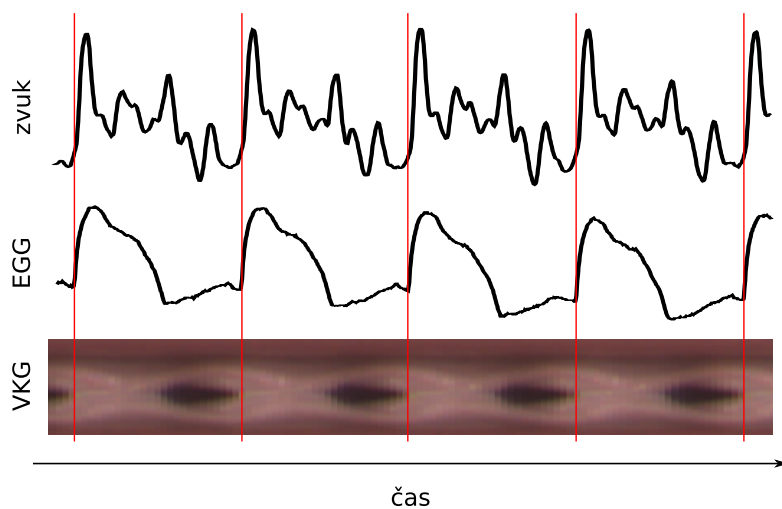
Kompletní záznam integrovaného systému HSV kamery.

3.1 Určení tvaru periody kmitu

3.1.1 Audiozáznam

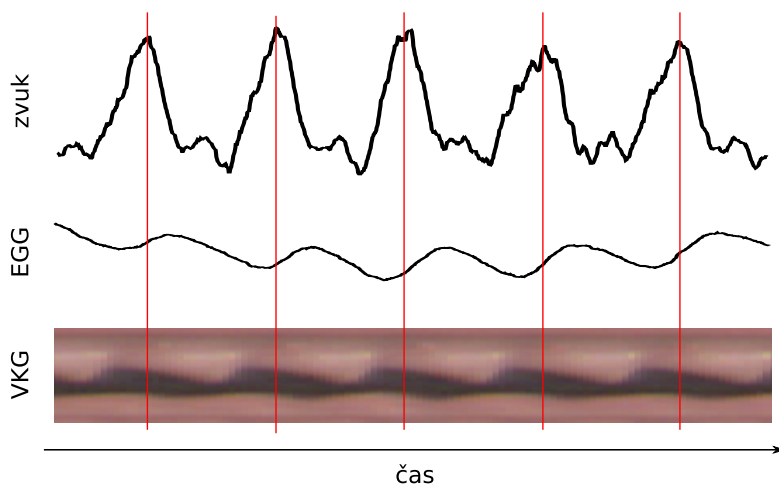
Tvar periody ve zvukovém záznamu fonace je pro každého pacienta odlišný a je proto obtížné přesně určit jednotlivé fáze kmitu. Tvar akustické vlny je možné odvodit z procesu tvorby hlasu. Hlasivky pracují jako ventil, který propouští přerušovaný proud vzduchu v podobě vzduchových tlakových pulsů. Tlakový puls je modulován v rezonančních dutinách a vychází ven ústy. Reakci rezonančních dutin na puls lze chápat jako přechodový děj [12].

V záznamu fonace se v moment uzavření hlasivek projeví největším rozkmitem signálu, který následuje několik kmitů se snižující se amplitudou. [17]. Velmi dobrý závěr glottis je uveden na obrázku 3.3.



Obrázek 3.3: Zvukový signál, EGG signál a videokymogram průběhu kmitů hlasivek.

Nedostatečný závěr hlasivek nevytvoří dostatečný puls a nevznikne tak žádný přechodový děj. Tvar periody se tak blíží sinusoidě s periodou frekvence základního hrtanového tónu. [12]. Na obrázku 3.4 je zobrazeno kmitání hlasivek s výraznou poruchou, hlasivky se v průběhu fonace nedotýkají, což je patrné také z EGG signálu a zvukové vlny.

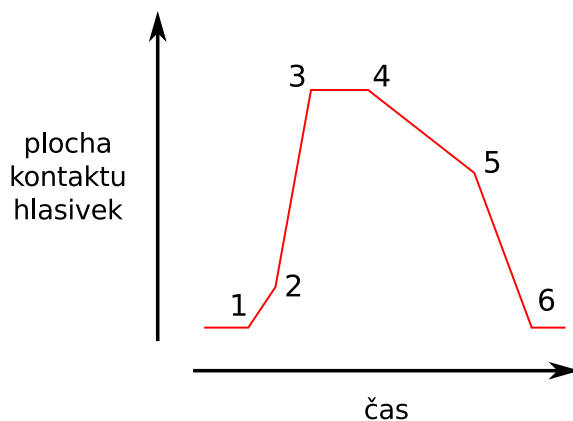


Obrázek 3.4: Zvukový signál, EGG signál a videokymogram průběhu kmitů hlasivek s nedostatečným závěrem.

3.1.2 Elektrolottogram

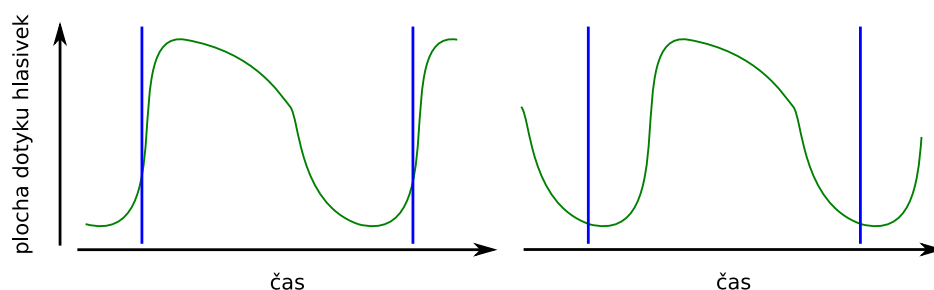
Křivka elektrolottogramu vyjadřuje plochu kontaktu hlasivek v průběhu fonace. Na standardizované křivce jedné periody glottogramu je vymezeno 6 částí, které odpovídají jednotlivým fázím pohybu hlasivek.

1. Počátek kontaktu ve spodní části hlasivek.
2. Hlasivky se začínají dotýkat celou plochou.
3. Maximální plocha doteku hlasivek. Tato fáze zřejmě odpovídá úplnému uzavření hlasivek, ale pouze z EGG signálu to nelze určit.
4. Konec maximálního doteku hlasivek. Počátek otevírání hlasivek ve spodní části.
5. Otevírání horní části hlasivek.
6. Úplné otevření hlasivek.



Obrázek 3.5: Fáze jedné periody EGG signálu.

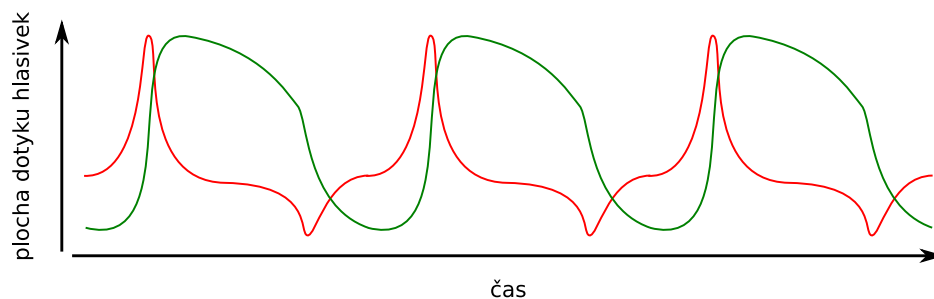
Počátek periody kmitu v EGG signálu bývá umístěn do fáze 2 - oblasti plného dotyku hlasivek, nebo do středu úseku úplného otevření hlasivek (střed fáze 6), viz obrázek 3.5. Volba počátku periody může ovlivnit výpočet kontaktního kvocientu CQ_{EGG} periody a při hodnocení CQ_{EGG} je toto nutné zohlednit. V této práci bude perioda označena prvním způsobem, periody



Obrázek 3.6: Možné určení period EGG záznamu.

zvukového a audiozáznamu tak budou popisovat shodný časový úsek a bude usnadněno určení časového posunu audio signálu.

Pro analýzu EGG signálu se také využívá jeho 1. derivace. Derivovaný EGG signál vyjadřuje rychlost změny kontaktní plochy hlasivek. Kladné vrcholy označují maximální nárůst kontaktní plochy hlasivek a záporné vrcholy maximální pokles plochy kontaktu. Vrcholy v derivovaném signálu lze použít k jednoznačnému určení hranice kmitu, nebo k výpočtu kontaktního kvocientu hlasivek.

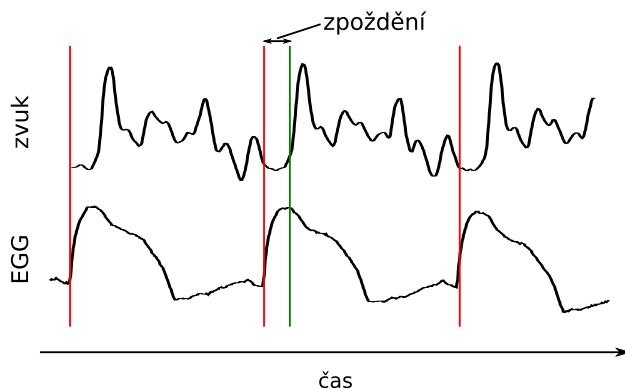


Obrázek 3.7: EGG záznam (zelená) a jeho derivace (červená).

3.2 Synchronizace záznamů

Audio, video a EGG záznam je systémem HSV kamery, respektive sestavou prostorový mikrofón a elektrolottograf, synchronizován. Kvůli umístění mikrofónu před ústní dutinu je však nutné počítat s časovým posunutím audiozáznamu vůči EGG a HSV. Vzdálenost mikrofónu od hlasivek se pohybuje v rozmezí 15 - 30 centimetrů v závislosti na délce hlasového ústrojí a pozici mikrofónu.

Je-li rychlost šíření zvuku ve vzduchu 346,3 m/s, dosahuje zpoždění zvukového záznamu oproti EGG a videozáznamu 0,441 až 0,882 milisekund, což odpovídá posunutí o 19 až 38 vzorků při vzorkovací frekvenci 44 100 Hz.



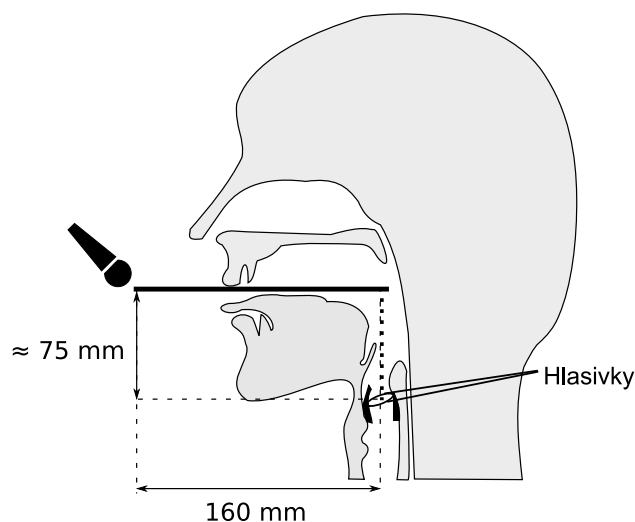
Obrázek 3.8: Zpoždění záznamu zvuku oproti EGG signálu.

Zpoždění zvuku lze určit automaticky porovnáním period kmitů EGG a audio signálu. Počátek periody v EGG signálu je určen okamžikem uzavření hlasivek. Uzavřením hlasivek je vybuzen přechodový děj, který se v audio signálu projeví výrazným rozkmitem. Počátek periody audio signálu je tedy určen okamžikem maximálního rozkmitu. Audio signál je poté možné synchronizovat s EGG signálem zarovnáním počátků period.

$$Td = \frac{r}{c} \cdot 1000 \text{ [msec]} \quad (3.1)$$

kde r je vzdálenost mikrofónu od hlasivek v metrech, c [m/s] je rychlost zvuku v metrech za sekundu. Uvedený postup pro určení zpoždění a synchronizaci zvukového signálu s EGG a video signálem byl ověřen experimentem.

Byl proveden záznam vyšetření, při kterém byla změřena vzdálenost od hlasivky k mikrofonu. Porovnáním automaticky určeného zpoždění zvukového záznamu a zpoždění určeného podle vztahu 3.1 byla potvrzena správnost algoritmu.



Obrázek 3.9: Schéma umístění mikrofonu při HSV vyšetření.

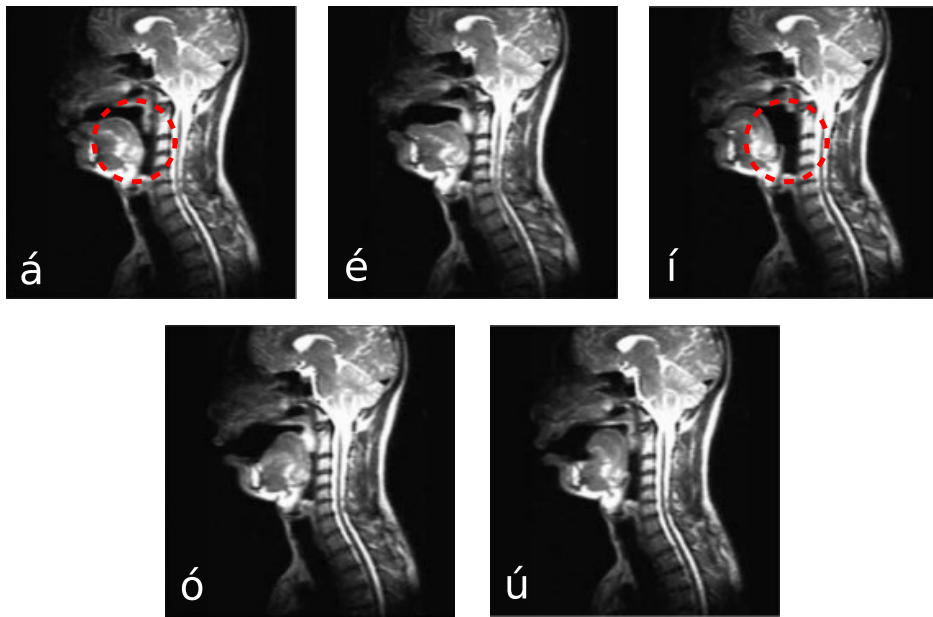
3.3 Vlastnosti zaznamenaných dat

Záznam kmitů hlasivek pomocí HSV endokamery je neinvazivní metoda, pro jejíž provedení není nutné podávat žádná sedativa či znecitlivující látky. Mnoha pacientům ovšem kamera zasunutá hluboko do ústní dutiny vyvolává dávivý reflex, čímž dochází ke stahům krčního svalstva a pohybu hlavy či celého těla. Během vyšetření lékař přidržuje vyplazený jazyk pacienta, zaznamenaná fonace tak neodpovídá běžné přirozené fonaci. Při zpracování a hodnocení dat je nutné tyto specifické vlastnosti HSV vyšetření zohlednit.

3.3.1 Audiozáznam

Při běžném vyšetření hlasivek použitím prostorového mikrofonu, případně v kombinaci s elektrolottografem, jsou hodnoceny parametry hlasu při fonaci vokálu „a“. Pro laryngoskopické vyšetření HSV kamerou však není vokál

„a“ vhodný, protože při jeho fonaci není jazyk ve vhodné poloze. Aby došlo k otevření anatomického prostoru nad hlasivkami a bylo možné hlasivky zaměřit optikou kamery, musí pacient fonovat hlásku „í“. Ve výsledku, vlivem laryngoskopu v ústech a přitlačení jazyka, tvar akustického signálu odpovídá fonaci vokálu „e“, resp. „e-i“, „e-a“ nebo spíše neutrální hlásce „v“. Hlasové ústrojí při fonaci vokálů je ilustrováno na obrázcích 3.10 z MRI.



Obrázek 3.10: Hlasové ústrojí při fonaci samohlásek. Na snímku fonace vokálu „á“ a „í“ je označena oblast otevřeného anatomického prostoru nad hlasivkami.

Typ vokálu, který vzniká při fonaci během laryngoskopického vyšetření lze určit podle polohy formantů F_1 a F_2 z celkového počtu F_1 až F_5 . Určený vokál byl ověřen metodou statistické lineární predikce LPC, pro kterou platí:

$$x(n) = a_1 \cdot x(n-1) + a_2 \cdot x(n-2) + a_3 \cdot x(n-3) + \dots + a_k \cdot x(n-k) + e(n) \quad (3.2)$$

tj. n -tý vzorek signálu je výsledkem lineární kombinace stanoveného počtu předcházejících vzorků tohoto signálu.

a_i - koeficienty LPC, které charakterizují matematický model vokálního traktu pro $i = 1, \dots, k$

k - řád LPC, který určuje aproximaci tvaru spektra hlasového signálu prokládáním pomocí polynomiálních funkcí a pro který platí 5-ti formantům F_1 až F_5 odpovídá řád LPC $k = 11$.

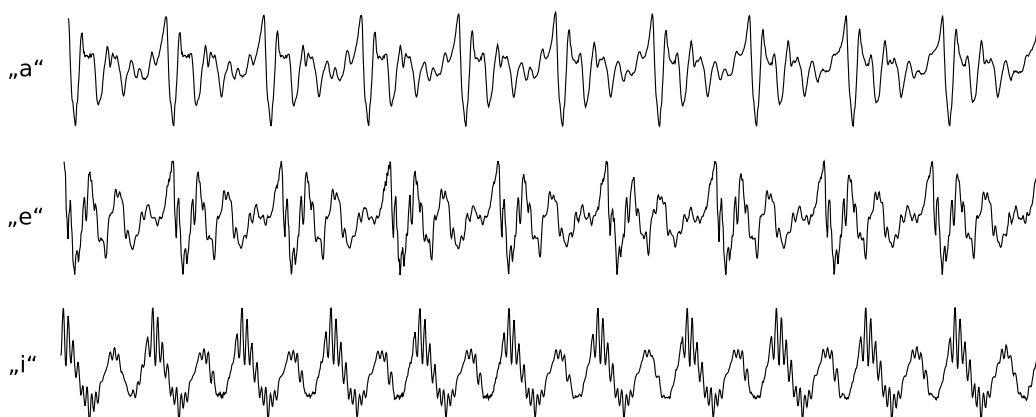
LPC umožňuje vyhlazení spektra tak, aby bylo možné jasně detekovat formanty, zejména F_1 a F_2 , které odpovídají samohláskám a-e-i-o-u. Podle rozložení těchto formantů F_1 a F_2 a tvaru frekvenčního spektra lze určit typ fonovaného vokálu.

Vokál	Četnost	
a	17	9,6%
e	134	75,2%
i	8	4,5%
e-i	8	4,5%
e-a	11	6,2%
Celkem	178	100%

Tabulka 3.1: Četnost výskytu jednotlivých vokálů v záznamech fonace.

Analýza záznamů fonace z HSV vyšetření byla provedena pomocí komerčního softwaru společnosti *Kay-PENTAX* [8], který obsahuje modul LPC-Analysis. Většina záznamů odpovídá fonaci vokálu „e“, ze sady 178 záznamů 75% obsahuje fonaci vokálu „e“ a 10,7% přechod mezi vokály „e-a“, případně „e-i“. Zbývající záznamy obsahují fonaci „a“ nebo „i“. Příklad výstupu analýzy fonací vokálů je uveden v elektronické příloze.

Tvar periody kmitu je při fonaci různých vokálů odlišný, v signálu se při fonaci „e“ a „i“ objevují vyšší frekvence, jak je ilustrováno na obrázku 3.11.



Obrázek 3.11: Záznam fonace vokálu „a“, „e“, „i“

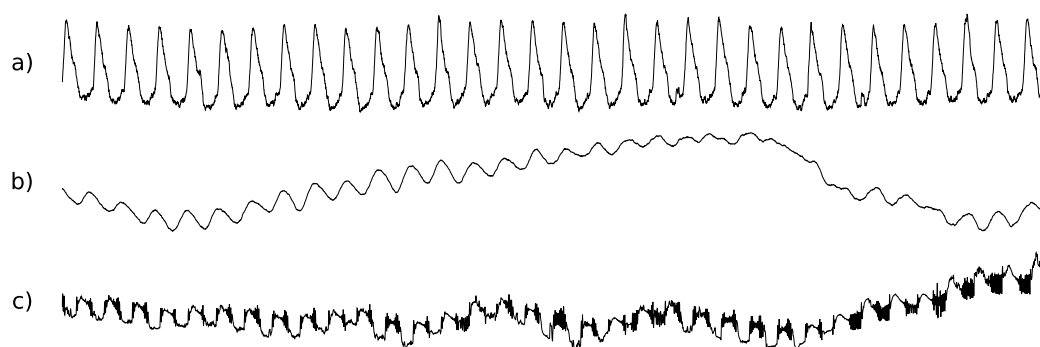
Přesné určení produkováného vokálu je důležité pro určení kvality závěru hlasivek metodou SCORE, jejíž regresní model je založen na hodnocení tvaru periody. Aby metoda SCORE správně hodnotila záznamy z HSV endokamery, je nutné znovu určit koeficienty modelu podle vzorku dat z HSV vyšetření. V záznamu fonace vokálů „e“, nebo „i“ je kvůli vyšším frekvencím a z toho plynoucím větším počtu vrcholů v signálu problematická detekce hranice period.

3.3.2 Elektrolottogram

Pro kvalitní záznam elektrolottogramu je důležité správné umístění elektrod elektrolottografu. Elektrody musí mít dobrý kontakt s pokožkou pacienta a musí být umístěny ve stejné úrovni nad křídly štítné chrupavky.

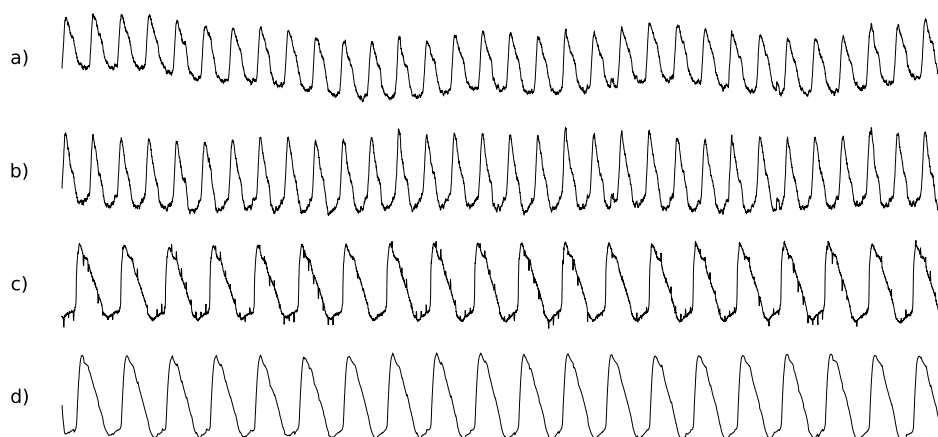
Během záznamu může dojít k posunu elektrod, pohybu hrtanu nebo snížení kontaktu, což se projeví jako nízkofrekvenční šum v zaznamenaném signálu. Vzhledem k vysoké citlivosti přístroje může signál také obsahovat vysokofrekvenční šum způsobený napájecím zdrojem. Rušení od zdroje napájení je běžně odstraněno použitím kvalitního zdroje či baterie.

Defekty v EGG záznamu ztěžují určení hranice period kmitů hlasivek a tím i stanovení parametrů jednotlivých period. V případě, kdy hranice period určit lze, defekt signálu zkresluje hodnoty parametrů period. Příklady defektů signálu jsou uvedeny na obrázku 3.12.



Obrázek 3.12: Příklad záznamu EGG signálu. a) kvalitní EGG signál b) nekvalitní záznam EGG - pohyb hrtanu c) nekvalitní záznam - vnější rušení

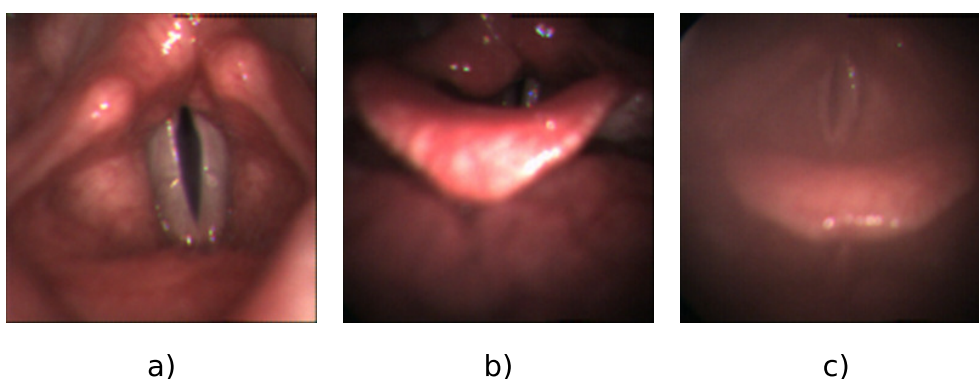
Kvalitu EGG signálu lze upravit filtrací horní, dolní, případně pásmovou propustí. Frekvence EGG signálu odpovídá základní hlasivkové frekvenci, dolní mezní frekvenci pásmové propusti tak lze zvolit například 40 Hz a horní mezní frekvenci 1000 Hz, aby byly zachovány významné znaky signálu. Filtrováním signálu lze zpřesnit například výpočet kvocientu kontaktu hlasivek a přesněji určit statistické vlastnosti jednotlivých kmitů. Efekt filtrace na různým způsobem zašuměné signály je uveden na obrázku 3.13.



Obrázek 3.13: Vliv filtrace na nekvalitní EGG signál. a) původní signál s nízkofrekvenčním šumem b) signál filtrovaný horní propustí c) původní signál s vysokofrekvenčním šumem d) signál filtrovaný dolní propustí

3.3.3 HSV a Videokymogram

Kvalita záznamu vysokorychlostní endokamery může být ovlivněna pohybem pacienta nebo kamery, který se projeví i při vysoké rychlosti snímání. Použití takového videozáznamu je problematické například pro tvorbu videokymogramu. Jestliže na čočce kamery ulpí větší množství slin, může být záznam neostrý a nekонтastní. U některých pacientů se nepodaří kameru nastavit do vhodné polohy a hlasivka je tak viditelná pouze v malé části záběru. Rozdílnou kvalitu záznamů ilustrují příklady na obrázku 3.14.



Obrázek 3.14: Příklad kvality HSV videozáznamu. a) kvalitní videozáznam b) nepřesné umístění kamery c) znečištěná čočka kamery

Výstupní formát pro videozáznam produkovaný kamerou je ukládán do souborů *BLD*, tento formát je proprietární formát výrobce a pro jeho zpracování bylo nutné vytvořit dekodér. Popis formátu a implementovaného dekodéru je uveden v tištěné příloze v části D.

4 Analýza záznamu

4.1 Detekce kmitu hlasivek

Aby bylo možné stanovit a počítat parametry jednotlivých kmitů hlasivek, je nutné spolehlivě nalézt hranice kmitů. Délka analyzovaného záznamu je v případě vyšetření *mikrofon + EGG* v řádu jednotek vteřin, v případě *HSV vyšetření* v délce do jedné vteřiny. Záznam obsahuje desítky kmitů, ruční označení jednotlivých kmitů hlasivek je v tomto případě sice možné, ale časově náročné a nepraktické. Proto je nutné nalézt automatizovaný přístup detekce kmitu hlasivek.

Jednotlivé kmity se mohou lišit svojí délkou a vlivem různých poruch hlasu také svým průběhem. Detekce kmitů musí být dostatečně robustní, aby nalezla i značně nepravidelné kmity. Pokud je fonace natolik zdeformovaná hlasovou poruchou, že není možné určit hranice kmitu, zřejmě nebude možné ani nalézt standardní periodu, která by vhodně reprezentovala fonaci, porucha je v tom případě jasně zřetelná a další detailní analýza typu detekce odlišného kmitu nemá význam.

Ztěžujícím faktorem pro detekci kmitu může být také kvalita záznamu. V případě audiozáznamu jde především o slabou fonaci či kolísající frekvenci a amplitudu fonace. V případě elektroglottogramu je problematické špatné umístění elektrod, nebo jejich pohyb v průběhu záznamu. Podrobněji byla data a jejich defekty popsány v části 3.3.

4.1.1 Určení základní frekvence

Základní hlasivková frekvence odpovídá mechanické rychlosti otevírání a zavírání hlasivek. Ze základní frekvence fonace lze odvodit základní periodu, z které lze vycházet při detekci hranic periody.

Ze základní hlasivkové frekvence lze určit základní periodu fonace podle vztahu 4.1, resp. 4.2 pro diskrétní signál. F_0 označuje základní frekvenci hrtanového tónu, T_0 odpovídá délce základní periody, L_0 počtu vzorků základní periody a f_{vz} vzorkovací frekvenci záznamu.

$$F_0 = \frac{1}{T_0} [Hz] \quad (4.1)$$

$$F_0 = \frac{f_{vz}}{L_0} [Hz] \quad (4.2)$$

K určení základní frekvence lze zvolit několik přístupů spadajících do dvou hlavních kategorií [2]:

Detekce v časové oblasti

Detekce v časové oblasti vychází z jednoduché myšlenky sledovat opakující se vzory v signálu, podle kterých je možné odhadnout jejich periodicitu.

1. **Metoda průchodu nulou** - Jednoduchá metoda, založená na počítání průchodu signálu nulovou úrovní. Tato metoda není příliš přesná, zejména je-li záznam zatížen šumem. Filtrací signálu lze přesnost zlepšit.
2. **Autokorelace** - Touto metodou hledáme shodu úseku signálu s úsekem posunutým o rostoucí počet vzorků. Matematicky lze autokorelaci pro omezený úsek signálu vyjádřit vzorcem:

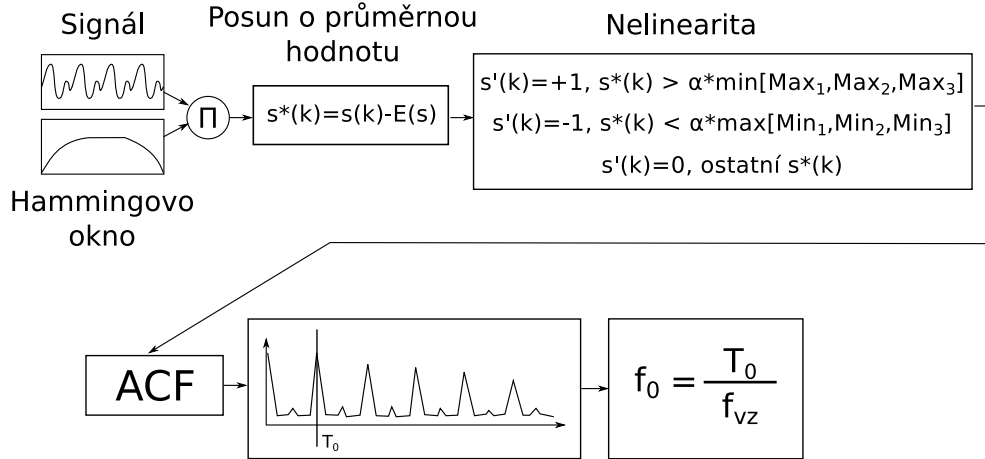
$$R(m) = \frac{1}{N} \sum_{k=1}^{N-1-m} x(k)x(k+m) \quad (4.3)$$

kde N označuje počet vzorků porovnávaného úseku, $x(k)$ je k -tý vzorek a hodnota m označuje posunutí signálu v počtu vzorků. Pokud je signál periodický, funkce $R(m)$ bude také periodická a pokud je signál harmonický, lokální maxima funkce $R(n)$ budou v násobcích základní periody. K určení základní frekvence musíme detekovat první vrchol $R(k)$, který následuje po maximální hodnotě autokorelační funkce $R(0)$, vyjadřující energii signálu:

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

3. **Upravená autokorelace** - Úspěšnost detekce autokorelační metody lze zvýšit vhodným předzpracováním signálu. Uvedený postup předzpracování signálu byl navržen pro vyšetření hlasivek s poruchou hlasu metodou hlasového pole [11]. Autokorelační funkce je upravena pro aplikaci na signály zkreslené vlivem chrapotu. Úpravy signálu mají za cíl

minimalizovat nežádoucí vrcholy v autokorelační funkci, které jsou způsobené například formanty. Postup zpracování signálu je uveden na obrázku 4.1



Obrázek 4.1: Schéma algoritmu autokorelační funkce s předzpracováním signálu.

Detekce ve frekvenční oblasti

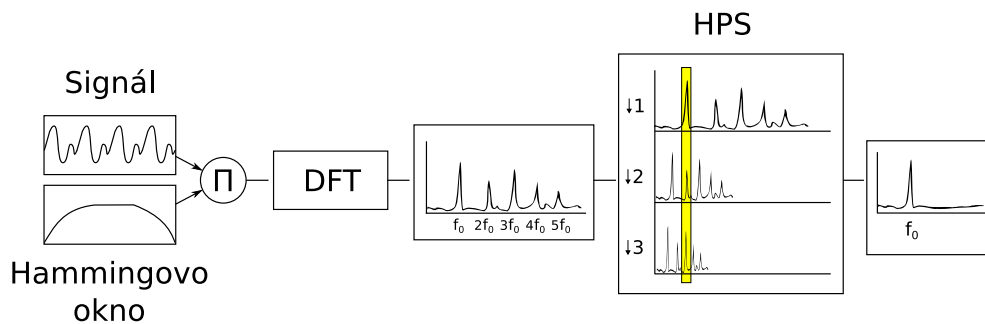
Tento přístup je založen na převedení signálu do frekvenční oblasti. Postup většiny metod začíná určením krátkého úseku signálu, filtrací pomocí okna a provedení Fourierovy transformace. Periodický signál se projeví vrcholy v násobcích základní frekvence.

Přesnost detekce ve frekvenční oblasti je závislá na frekvenčním rozlišení f_R vycházejícím z velikosti okna a vzorkovací frekvenci záznamu dle vztahu 4.5. Pokud je například vzorkovací frekvence $f_{vz} = 44100$ a velikost okna je $w_{DFT} = 8192$ vzorků signálu, rozlišení bude 5,38 Hz.

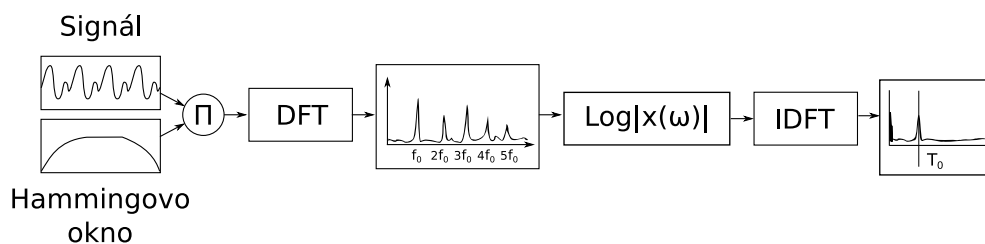
$$f_R = \frac{f_{vz}}{w_{DFT}} [Hz] \quad (4.5)$$

Určit, který vrchol ve spektru odpovídá základní frekvenci, umožňují následující metody:

1. **Spektrum součinů harmonických frekvencí** - Spektrum součinů harmonických frekvencí (Harmonic product spectrum) [2] určuje základní frekvenci signálu výpočtem největšího společného dělitele harmonických frekvencí. Součinem spektra a jeho decimovaných kopií získáme spektrum součinů harmonických frekvencí, maximální hodnota tohoto spektra odpovídá největšímu společnému děliteli harmonických frekvencí, což je základní frekvence. Předností metody je nezávislost na přítomnosti šumu a detekce základní frekvence v polyfonním signálu. Postup výpočtu pomocí HPS je uveden na obrázku 4.2.
2. **Cepstrum** - Cepstrální analýza určuje základní frekvenci z inverzní Fourierovy transformace logaritmu absolutní hodnoty spektra signálu. Pokud je signál harmonický, je jeho spektrum periodické a periodicitu je dále zvýrazněna zlogaritmováním absolutní hodnoty jednotlivých koeficientů. Provedením inverzní transformace upraveného spektra získáme signál s vrcholem v hodnotě základní periody.



Obrázek 4.2: Schéma algoritmu *Spektrum součinů harmonických frekvencí* (Harmonic product spectrum)

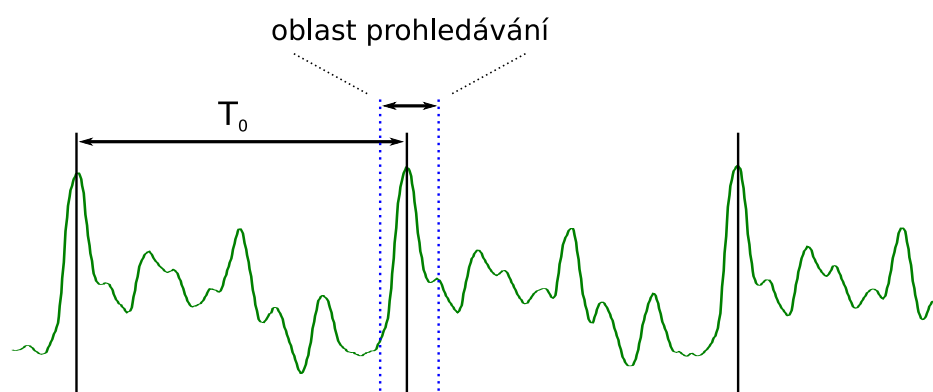


Obrázek 4.3: Schéma algoritmu *Cepstrum*

Uvedené metody byly implementovány a ověřeny na dostupných datech, hodnocení metod je uvedeno v části 7.1.

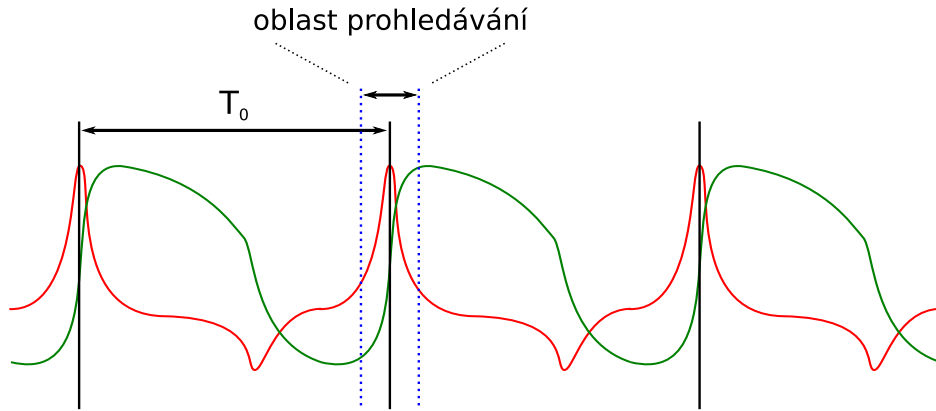
4.1.2 Algoritmus detekce periody kmitu

Detekce hranic kmitů je založena na znalosti základní frekvence, respektive základní periody fonace. Počátek kmitu je odvozen od pozice kladného lokálního maxima v signálu, který odpovídá okamžiku uzavření hlasivek a lze jej chápat jako počátek přechodového děje. Následující hranice je vyhledávána ve vzdálenosti základní periody. Během fonace dochází k frekvenčnímu kolísání (jitter) a je proto nutné hledat hranici kmitu v okolí předpokládané pozice.



Obrázek 4.4: Perioda zvukové vlny a oblast hledání hranice periody.

Při detekci hranic kmitu v elektrolottogramu stejně jako u zvukového záznamu vycházíme ze znalosti délky základní periody. Uzavření a otevření hlasivky se projeví jako lokální maximum, respektive minimum v derivaci EGG signálu, viz obrázek 4.5. Hranice kmitu je určena vždy do lokálního maxima. Algoritmus obsahuje detekci lokálního maxima derivovaného signálu ve vzdálenosti přibližně jedné základní periody. Obdobně jako u zvukového záznamu je nutné počítat s frekvenčním kolísáním a hledat hranice kmitu v okolí předpokládané pozice.



Obrázek 4.5: EGG záznam (zelená) a jeho derivace (červená).

Rozsah prohledávání lze pevně stanovit na několik desítek vzorků, ale vzhledem k frekvenčnímu kolísání (jitteru) fonace a možnému rozsahu frekvencí hlasu, viz část 2.1, je vhodnější pro každý záznam odvodit oblast prohledávání od velikosti základní periody. Základní hlasová frekvence se běžně pohybuje v rozsahu 120 Hz až 370 Hz, čemuž odpovídá počet 350 až 120 vzorků na periodu při vzorkovací frekvenci 44 100 Hz, rozsah prohledávaného okolí tak lze zvolit například 1/4 vzorků periody, v případě $L = 120$ bude oblast prohledávání 30 vzorků a algoritmus lze popsat formálně podle:

$$\begin{aligned}
 p[0] &= i, \max_{0 \leq i \leq L} x[i] \\
 p^*[k] &= p[k-1] + L \\
 p[k] &= i, \max_{p^*[k]-30 \leq i \leq p^*[k]+30} x[i],
 \end{aligned} \tag{4.6}$$

kde $p[k]$ je pozice prvního vzorku k -té periody záznamu, L je počet vzorků periody a $x[i]$ označuje i -tý vzorek záznamu.

5 Určení standardního kmitu fonace

Cílem níže uvedených metod je nalézt odhad standardní periody fonace, která je použita jako reference pro porovnání jednotlivých period a určení míry odlišnosti od běžné periody. Standardní perioda by měla být zvolena tak, aby reprezentovala tvar většiny period v záznamu fonace, ale nebyla ovlivněna ojedinělými periodami odlišného tvaru.

Pokud záznam obsahuje dvě a více přibližně shodně velkých množin period a každá množina obsahuje periody odlišného tvaru, je obtížné určit tvar standardní periody. Standardní perioda může být kombinací tvarů period ze všech množin a nereprezentovat tak ani jednu z nich, nebo může být zvolena jedna z period záznamu, která bude reprezentovat jednu množinu.

Standardní kmit lze zvolit několika metodami, které spadají do dvou hlavních kategorií: zpracování v časové a frekvenční oblasti.

Časová oblast

V časové oblasti lze pracovat s jednotlivými vzorky period fonace. Pro jejich porovnání je ale nutné zajistit shodnou délku period. To však vlivem frekvenčního kolísání není reálné a musí být použita metoda převzorkování na zvolenou reprezentativní délku periody.

Frekvenční oblast

Výhodou frekvenční oblasti je nezávislost na počtu vzorků periody. Standardní kmit je zvolen analýzou amplitudového spektra period provedením Fourierovy transformace.

5.1 Převzorkování period

Jednotlivé periody kmitu hlasivek jsou vlivem frekvenčního kolísání a hlasové poruchy rozdílně dlouhé. Aby bylo možné určit standardní periodu je nutné nejprve určit standardní délku periody. Délka standardní periody by neměla

být závislá na délkách ojedinělých extrémních period a měla by být zvolena tak, aby bylo minimalizováno zkreslení způsobené převzorkováním. Délka standardní periody je proto zvolena jako modus délek jednotlivých period.

Pro každou periodu je vypočten podíl délky aktuální periody k nejčtetnější délce periody, který odpovídá vzdálenosti sousedních bodů převzorkované periody relativně k původní vzdálenosti bodů, viz vztah 5.1.

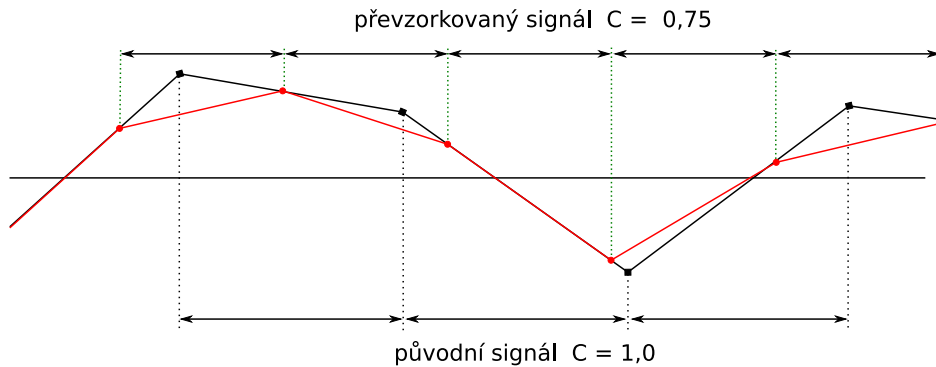
$$C_k = \frac{L_k}{L_s}, \text{ pro všechna } k = 1, 2 \dots N \quad (5.1)$$

Hodnoty nových bodů jsou určeny lineární interpolací mezi sousedními vzorky podle vztahu 5.2. Metoda lineární interpolace je v tomto případě dostatečně přesná, protože změna vzorkovací frekvence periody je pouze minimální a metody polynomiální interpolace by tak poskytly jen zcela zanedbatelné zlepšení. Postup algoritmu je znázorněn na obrázku 5.1.

$$x_k^*[0] = x_k[0]$$

$$x_k^*[i] = a \cdot (C \cdot k) - ([C \cdot k]) + b, \text{ pro } i = 1, 2 \dots L_s, k = 1, 2 \dots N$$

$$a = x_k[([C \cdot k] + 1)] - b, \quad b = x_k[([C \cdot k])] \quad (5.2)$$



Obrázek 5.1: Původní signál a převzorkovaný signál(červená).

5.2 Standardní perioda v časové oblasti

Odhad standardní periody podle níže uvedených metod je vygenerovaná perioda, která se nevyskytuje v záznamu fonace. Je vytvořena na základě statistického zpracování stejnohlých vzorků.

5.2.1 Průměrná hodnota vzorků

Základní způsob, jak určit standardní periodu fonace je zvolit pro každý vzorek standardní periody průměrnou hodnotu vzorků všech period fonace na odpovídající pozici. Průměrná hodnota vzorku je vypočítána podle vztahu 5.3, kde N je počet period fonace:

$$\bar{x}[i] = \frac{1}{N} \sum_{k=0}^N x_k[i] \quad (5.3)$$

Standardní kmit pak odpovídá vztahu 5.4, kde L_s je nejčtenější počet vzorků periody a $x_s[i]$ je i -tý vzorek standardní periody.

$$x_s[i] = \bar{x}[i], \text{ pro } i = 0, \dots, L_s \quad (5.4)$$

V průměrné hodnotě vzorků se ale projeví také extrémní hodnoty, což není pro reprezentaci standardní periody vhodné, u standardní periody je cílem maximalizovat odlišnost od extrémních tvarů period.

5.2.2 Medián vzorků

Vhodnější přístup, jak určit standardní periodu audiozáznamu, je určit medián hodnoty vzorku na každé pozici periody podle vztahu 5.5. Vlastností mediánu je nezávislost na extrémních hodnotách, odvozená standardní perioda tak není ovlivněna extrémními tvary period.

$$x_s[i] = \tilde{x}[i], \text{ pro } i = 0, \dots, L_s \quad (5.5)$$

5.3 Standardní perioda ve frekvenční oblasti

Ve frekvenční oblasti je standardní perioda zvolena z existujících period fonace podle dané kritériální funkce. Periody jsou převedeny do frekvenční oblasti výpočtem koeficientů Fourierova rozvoje a analýzou amplitudového spektra je vybrána standardní perioda.

Fourierův rozvoj

K popisu tvaru jedné periody hrtanového tónu je použit Fourierův rozvoj pro definovaný konečný počet členů Fourierovy řady. Dostatečně přesný tvar periody lze rekonstruovat již z 15 členů Fourierovy řady, pro hodnocení odlišnosti či hledání standardních period jsou v této práci určeny členy do úrovně $HARM = 20$ harmonických frekvencí, vyšší frekvence se ve fonaci téměř nevyskytují a jsou pro porovnání tvaru period již nepodstatné, viz [12]. Vztah pro výpočet koeficientů Fourierovy řady je v následujícím tvaru:

$$x[i] = \sum_{j=0}^{HARM} (a_x[j] \cos(\frac{2\pi j i}{L}) + b_x[j] \sin(\frac{2\pi j i}{L})) \quad (5.6)$$

kde význam proměnných odpovídá:

$x[i]$... i -tý vzorek periodické křivky

L ... délka aktuální periody

j ... j -tý člen Fourierovy řady

$a_x[i]$ $b_x[i]$... koeficienty Fourierovy řady

Koeficienty Fourierovy řady jsou po úpravě vypočítány podle vztahu:

$$a_x[j] = \frac{2}{L} \sum_{k=1}^L (X[k] \cos(\frac{2\pi j k}{L})), \quad b_x[j] = \frac{2}{L} \sum_{k=1}^L (X[k] \sin(\frac{2\pi j k}{L})) \quad (5.7)$$

Ke koeficientům řady lze vytvořit tzv. invarianty, které popisují tvar křivky nezávisle na její poloze a relativně k jejímu posunutí. Invarianty jsou vytvořeny pomocí amplitudového spektra, viz 5.8.

$$AA_x[j] = \sqrt{(a_x[j])^2 + (b_x[j])^2}, \text{ pro všechna } j = 1, 2, \dots, HARM \quad (5.8)$$

5.3.1 Minimální eukleidovská vzdálenost mediánu amplitudového spektra

Známe-li amplitudová spektra všech period, je možné určit medián amplitudy pro každou harmonickou frekvenci. Hodnoty amplitudového spektra jsou nejprve normovány, aby byla vyloučena variabilita způsobená intenzitou fonace podle vztahu 5.9.

$$AAn[j] = \frac{AA_x[j]}{\sqrt{\sum_{i=1}^{HARM} (AA_x[i])^2}}; \quad j = 1, 2, \dots, HARM \quad (5.9)$$

Standardní periodu lze poté vybrat podle minimální eukleidovské vzdálenosti od těchto mediánových hodnot. Medián amplitud harmonických frekvencí odpovídají vztahu 5.10:

$$AA_{med}[j] = \widetilde{AAn}[j], \text{ pro } j = 0, \dots, HARM \quad (5.10)$$

kde $\widetilde{AA}[i]$ je medián amplitudy i -té harmonické frekvence periody fonace. Zvolená perioda by měla reprezentovat běžné periody fonace z hlediska skladby harmonických frekvencí. Volbou mediánu amplitud není standardní perioda ovlivněna ojedinělými extrémy ve fonaci. Eukleidovská vzdálenost amplitudových spekter je určena podle vztahu 5.11, kde $m_e[k]$ je vzdálenost amplitudového spektra k -té periody:

$$m_e[k] = \sqrt{\sum_{j=0}^{HARM} (AA_{med}[j] - AA_k[j])^2}, \text{ pro } k = 1, \dots, N \quad (5.11)$$

Standardní perioda pak odpovídá existující periodě fonace s nejmenší eukleidovskou vzdáleností:

$$L_s = L_k, \min_{0 \leq k \leq N} m_e[k]. \quad (5.12)$$

5.3.2 Skóre četnosti amplitudového spektra

Referenční hodnoty amplitudového spektra pro nalezení standardní periody je možné určit také podle četnosti výskytu. Hodnoty amplitudového spektra jsou obdobně jako v předchozí metodě normovány, aby byl vyloučen vliv intenzity fonace podle vztahu 5.9. Normovaná amplituda dané harmonické frekvence pro každou periodu je zanesena do histogramu s vhodnou velikostí intervalu třídy, například 0,01. Každá perioda je poté ohodnocena součtem četností třídy histogramu odpovídající amplitudám dané periody.

$$S_T[k] = \sum_{j=0}^{HARM} h_j[AAbin], \text{ kde } AAn_k[j] \approx AAbin \quad (5.13)$$

Jako standardní perioda fonace je zvolena perioda s nejvyšší hodnotou $S_T[i]$, tedy perioda, která nejvíce odpovídá nejčetnějším amplitudám harmonických frekvencí.

$$L_s = L_k, \max_{0 \leq k \leq N} S_T[k]. \quad (5.14)$$

6 Detekce nestandardních period

6.1 Detekce v časové oblasti

6.1.1 Průměrná odchylka vzorků periody

Základním přístupem je určení odlišnosti kmitů podle průměrné absolutní hodnoty rozdílu jednotlivých vzorků periody. Hodnota vzorků se pohybuje v rozsahu $x[i] \in \langle -1, 1 \rangle$, průměrná odchylka tak může nabývat hodnot z intervalu $D_s \in \langle 0, 2 \rangle$

$$D_s[k] = \frac{\sum_{i=0}^L |x_s[i] - x_k[i]|}{L}, \text{ pro } k = 1, \dots, N \quad (6.1)$$

Z vlastností metody vyplývá, že periody s malou odchylkou ve více vzorcích budou hodnoceny stejně jako periody s výraznou odchylkou v malém počtu vzorků. Pro hodnocení odlišnosti je ale vhodnější výrazné odchylce přikládat vyšší váhu.

6.1.2 Vzdálenost vzorků periody

Další způsob jak porovnat periody v časové oblasti na úrovni jednotlivých vzorků je určit eukleidovskou vzdálenost vzorků. Velký rozdíl hodnot vzorků bude v tomto případě mít větší váhu než malá odchylka ve více vzorcích.

$$D_{sd}[k] = \sqrt{\sum_{i=0}^L (x_s[i] - x_k[i])^2}, \text{ pro } k = 1, \dots, N \quad (6.2)$$

6.1.3 Skóre vzorků mimo pásmo σ

Určením směrodatné odchylky pro každý vzorek všech period získáme pásmo σ kopírující tvar průměrné periody kmitu fonace.

$$\sigma_i = \sqrt{(x_k[i] - \bar{x}[i])^2} \quad (6.3)$$

kde $x_k[i]$ je i -tý vzorek k -té periody a $\bar{x}[i]$ je průměrná hodnota vzorků na i -té pozici podle vztahu 5.3.

Odlišnost periody lze určit podle počtu vzorků, jejichž hodnota je mimo stanovené pásmo. Tímto způsobem ale nejsou odlišeny vzorky, které přesahují pásmo σ minimálně a vzorky, které přesahují pásmo σ výrazně. Přesněji lze odlišnost periody určit započtením vzorku tolikrát, kolik pásem σ přesahuje, maximálně do pásma 3σ . Míra odlišnosti je určena podle vztahu 6.4.

$$D_\sigma[k] = \frac{1}{L} \sum_{i=0}^L y_k[i], \text{ pro } k = 1, \dots, N,$$

$$\text{kde pro } y_k[i] \text{ platí: } y[i] = 0, |x_k[i]| < \sigma$$

$$y_k[i] = 1, \sigma \leq |x_k[i]| < 2\sigma$$

$$y_k[i] = 2, 2\sigma \leq |x_k[i]| < 3\sigma$$

$$y_k[i] = 3, |x_k[i]| > 3\sigma \quad (6.4)$$

kde platí: L počet vzorků periody, $y_k[i]$ míra odlišnosti vzorku, $x_k[i]$ i -tý vzorek k -té periody. Periody, jejichž všechny vzorky leží v pásmu σ mají míru odlišnosti nulovou. Maximální míra odlišnosti periody je 3.

6.2 Detekce ve frekvenční oblasti

6.2.1 Eukleidovská vzdálenost amplitudového spektra

Oproti metodám v časové oblasti lze výpočtem amplitudového spektra hodnotit odlišnosti tvaru period na úrovni harmonických frekvencí. Amplitudové spektrum je pro každou periodu určeno podle vztahu 5.8. Odlišnost period je určena jako Eukleidovská vzdálenost spekter od standardní periody podle vztahu 6.5.

$$D_{AA}[k] = \sqrt{\sum_{i=0}^{HARM} (AA_s[i] - AA_k[i])^2}, \text{ pro } k = 1, \dots, N \quad (6.5)$$

6.2.2 Vzdálenost Fourierových deskriptorů

Metoda Fourierových deskriptorů je publikována jako popis tvaru dvourozměrného objektu pomocí Fourierovy analýzy hranice objektu. Výhodou Fourierových deskriptorů je nezávislost na posunutí, rotaci a změně měřítka objektu. Postup výpočtu Fourierových deskriptorů lze rozdělit na následující kroky [10]:

1. Souřadnice bodů hranice objektu jsou rozvedeny do Fourierových řad:

$$x(k) = \frac{1}{2}a_{x0} + \sum_{i=0}^n (a_x[i] \cos(\frac{2\pi ik}{L}) + b_x[i] \sin(\frac{2\pi ik}{L}))$$

$$y(k) = \frac{1}{2}a_{y0} + \sum_{i=0}^n (a_y[i] \cos(\frac{2\pi ik}{L}) + b_y[i] \sin(\frac{2\pi ik}{L}))$$

Je možné použít i alternativní reprezentaci hranice objektu, například řetězovým kódem.

2. Vypuštěním koeficientů a_{x0} a a_{y0} se popis stane invariantní vůči posunutí.
3. Nezávislost na změně měřítka je zajištěna škálováním pomocí $|A_{x1}|$, respektive $|A_{y1}|$:

$$Axx_i = \frac{\sqrt{(a_{xi})^2 + (b_{xi})^2}}{\sqrt{(a_{x1})^2 + (b_{x1})^2}}, \quad Ayy_i = \frac{\sqrt{(a_{yi})^2 + (b_{yi})^2}}{\sqrt{(a_{y1})^2 + (b_{y1})^2}}$$

4. Fourierovy deskriptory jsou invariantní vůči otočení úpravou hodnot podle vztahu:

$$D_i = \sqrt{Axx_i^2 + Ayy_i^2}$$

Pro účely úlohy detekce odlišného kmitu byla metoda Fourierových deskriptorů zjednodušena. Pro popis jednorozměrného signálu není nutná nezávislost na otočení a výše uvedený 4. krok je možné vynechat.

Pro každou periodu signálu je proveden rozvoj do Fourierovy řady a poté je určeno amplitudové spektrum podle vztahu 5.8. Hodnota $AA_x[0]$ je vypuštěna a zbytek spektra je škálován pomocí $AA_x[1]$. Pro celý záznam poté platí:

$$FD_k[i] = \frac{AA_x[i]}{AA_x[1]}, \text{ pro } i = 1, \dots, N, \quad k = 0, \dots, K \quad (6.6)$$

Odlišnost period kmitů lze hodnotit podle Eukleidovské vzdálenosti Fourierových deskriptorů standardního a hodnoceného kmitu.

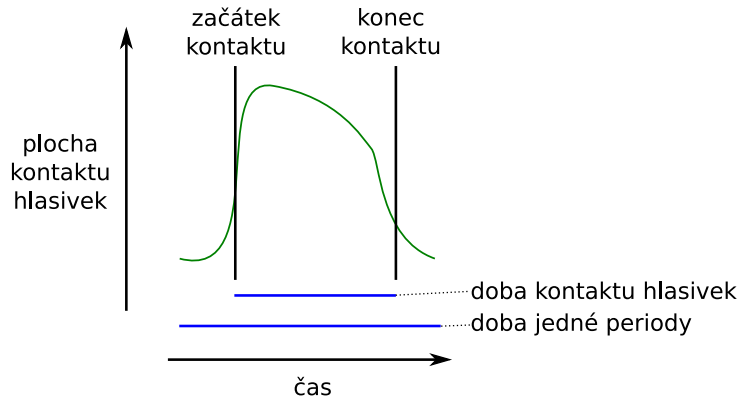
$$D_{fd}[k] = \sqrt{\sum_{i=0}^{HARM} (D_s[i] - D_k[i])^2}, \text{ pro } k = 1, \dots, N \quad (6.7)$$

6.3 Výpočet dalších parametrů kmitu

Jednotlivé detekované periody kmitu hlasivek lze ohodnotit funkcí SCORE, určit kontaktní kvocient a získat tak další parametr pro porovnání kmitů. Tyto parametry ale příliš nevypovídají o tvaru periody a nejsou proto vhodné k nalezení patologických period fonace, které se projevují nestandardním tvarem průběhu signálu. Z těchto parametrů lze odvodit celkovou kvalitu hlasu, podle které je například možné lépe rozhodnout o významu odlišnosti jednotlivých period.

6.3.1 Kontaktní kvocient EGG signálu

Kvocient kontaktu EGG vyjadřuje poměr kontaktní fáze k celkovému cyklu otevření a uzavření hlasivky (obrázek 6.1). Pomocí CQ_{EGG} lze popsat stav a pohyb hlasivek v průběhu fonace. Hodnota CQ_{EGG} u zdravého člověka se pohybuje v rozsahu 0,4 až 0,6 z intervalu $CQ_{EGG} \in \langle 0, 1 \rangle$. Hodnoty nižší než 0,4 diagnostikují hypoadukci, vyšší než 0,6 hyperaddukci.



Obrázek 6.1: Podíl kontaktní fáze v jednom cyklu hlasivek.

Hodnota kontaktního kvocientu EGG signálu je vypočítána podle vzorce 6.8, kde $CQ_{EGG}[i]$ je kontaktní kvocient i -tého kmitu, T_i je počet vzorků i -té periody, $x_i[\text{otevření}]$ a $x_i[\text{uzavření}]$ označují vzorky, kdy došlo k otevření, respektive uzavření hlasivek.

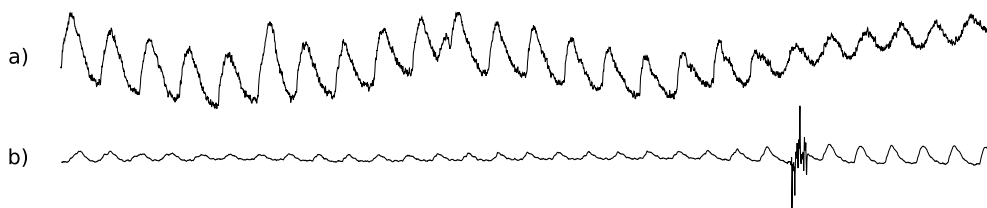
$$CQ_{EGG}[i] = \frac{x_i[\text{otevření}] - x_i[\text{uzavření}]}{T_i} \quad (6.8)$$

Metody výpočtu CQ_{EGG} jsou závislé na přesném určení okamžiku uzavření a otevření hlasivek. Kvůli složité struktuře a kinematice hlasivek neexistuje shoda na přesné definici uzavřeného a otevřeného stavu. Existuje několik přístupů, jak tyto okamžiky určit podle hodnoty nebo první derivace signálu. [6] Jednotlivé metody jsou rozdílně citlivé na různé defekty signálu, například šum, který lze dostatečně redukovat filtrací pouze v omezeném množství vhodných záznamů.

Metoda prahu určuje okamžik uzavření a otevření podle průchodu signálu stanoveným prahem. Používané hodnoty prahu se pohybují v rozmezí 20% až 50% z amplitudy od vrcholu k vrcholu. Tato metoda selhává obsahuje-li signál nízkofrekvenční šum. Je-li prahová hodnota určena pro každou periodu, metoda bude fungovat i pro signál s nízkofrekvenčním šumem, ale může dojít k zkreslení výsledků pokud se mezi periodami fonace mění plocha dotyku hlasivek a tím klesá amplituda kmitů.

Metoda Decom využívá první derivace EGG signálu (DEGG). Označíme-li kladné vrcholy v DEGG jako okamžiky uzavření a záporné vrcholy jako okamžik otevření hlasivek, lze z těchto bodů určit CQ_{EGG} . Tato metoda podává nepřesné výsledky vyskytne-li se v derivovaném signálu více vrcholů v jedné periodě, například je-li signál zatížen vysokofrekvenčním šumem, nebo pokud pohyb hlasivky není plynulý.

Hybridní metoda je kombinace obou metod, okamžik uzavření hlasivek je určen z derivace EGG nalezením kladného vrcholu, okamžik uzavření je určen podle stanoveného prahu. Hodnota prahu může být konstantní pro celý záznam, nebo určena z hodnoty signálu v okamžiku uzavření hlasivek.



Obrázek 6.2: Problematické EGG signály pro určení kontaktního kvocientu. a) Signál obsahuje vysokofrekvenční a nízkofrekvenční šum, tvar kmitů je nepravidelný, amplituda fonace klesá. b) Signál obsahuje výrazný defekt.

Porovnání uvedených metod na dostupných datech je uvedeno v části 7.4.

6.3.2 Výpočet SCORE

Metoda SCORE ohodnocuje kvalitu závěru glottis analýzou jednoho kmitu [12]. Určením hodnoty SCORE pro každý kmit je možné sledovat její vývoj v průběhu fonace a nalézt oblast nestandardního pohybu hlasivek. Hodnota SCORE je určena analýzou Fourierova rozvoje jedné periody a aplikací váhové funkce. Fourierův rozvoj je vypočten postupem uvedeným v části 5.3. Podle vztahu 5.8 je určeno amplitudové spektrum, které je normováno podle vztahu 6.9, aby byla vyloučena variabilita způsobená intenzitou fonace.[12].

$$a[i] = \frac{AA_x[i]}{\sqrt{\sum_{j=1}^{15} (AA_x[j])^2}}; \quad i = 1, 2, \dots, 15 \quad (6.9)$$

Váhová funkce

Matematický popis jedné periody je následně nutné hodnotit vhodným kvalitativním kritériem. K hodnocení je použita váhová funkce, jejíž parametry byly určeny na základě trénovací množiny záznamů fonace vokálu „a“, které byly ohodnoceny znalcem ORL podle definované stupnice kvality závěru glottis.

Klasifikace je založena na výsledcích vyšetření MDVA, vyšetření HSV, hodnocení hlasového projevu pacienta a tvaru periody podle zkušenosti experta. Výsledná stupnice získaná kvantifikací výsledků je v rozsahu 1 až 5, přičemž hodnocení 1 označuje kvalitní závěr glottis bez hlasové poruchy. Hodnota 5 reprezentuje stav, kdy se hlasivky zřetelně nedovírají po celé délce.

Následně byl sestaven regresní model, založený na prvních šesti normovaných amplitudách koeficientů Fourierovy řady ve tvaru:

$$SCORE = \alpha_0 + \sum_{i=1}^6 \alpha_i \log(a[i]) \quad (6.10)$$

Výsledný model SCORE byl sestaven pro 500 znalecky ohodnocených záznamů [12]:

$$\begin{aligned} SCORE = & 1,0789 * \log(a[1]) - 0,0864 * \log(a[2]) - 0,4306 * \log(a[3]) \\ & - 0,2736 * \log(a[4]) - 0,0971 * \log(a[5]) - 0,1905 * \log(a[6]) + 0,7629 \end{aligned} \quad (6.11)$$

Váhová funkce pro HSV vyšetření

Záznamy fonace pořízené během HSV vyšetření odpovídají u většiny pacientů vokálu „e“, jak bylo uvedeno v části 3.3. Tvar periody jednoho kmitu se při fonaci různých vokálů významně liší a je proto nutné znovu odvodit koeficienty váhové funkce. Koeficienty pro HSV vyšetření jsou určeny pro trénovací množinu 200 znalecky ohodnocených záznamů.

$$\begin{aligned} SCORE = & 2,1652 * \log(a[1]) + 1,0526 * \log(a[2]) - 0,1485 * \log(a[3]) \\ & - 0,3734 * \log(a[4]) - 0,2222 * \log(a[5]) - 0,1052 * \log(a[6]) + 0,0389 \end{aligned} \quad (6.12)$$

V případě vokálu „e“ se jako přesnější ukázal zjednodušený lineární model ve tvaru:

$$\begin{aligned} SCORE = & 2,8120 * a[1] + 1,9548 * a[2] - 0,4362 * a[3] \\ & - 1,5435 * a[4] - 1,4976 * a[5] - 0,8486 * a[6] - 0,5274 \end{aligned} \quad (6.13)$$

Hodnocení přesnosti uvedených modelů SCORE v porovnání s hodnocením experta je uvedeno v části 7.

7 Hodnocení výsledků

Všechny metody detekce a výpočty parametrů jednotlivých kmitů jsou ověřeny na záznamech HSV vyšetření poskytnutých foniatrickým oddělením ORL kliniky FN Plzeň. Záznamy obsahují fonace s hlasovou poruchou v rozsahu od zdravého hlasu po výrazný chrapot.

V testovací sadě záznamů jsou vzhledem k dostupnosti dat z vyšetření integrovaným systémem HSV zastoupeny záznamy s diagnózou: karcinom, polyp, uzlík, Reinkeho edém a záznamy bez nálezu. Pro ověření navržených metod detekce nestandardních kmitů a prezentaci výsledků jsou použity vybrané kazuistiky, které reprezentují výše uvedené diagnózy.

Testované záznamy obsahují výřez fonace v délce maximálně jednotek vteřin. Pokud není uvedeno jinak, záznamy nejsou nijak předzpracovány či filtrovány.

7.1 Určení základní frekvence

Určení základní frekvence F_0 je kromě získání parametru popisujícího hlas také základ pro správnou detekci jednotlivých kmitů hlasivek. Algoritmy detekce základní frekvence uvedené v části 4.1.1 byly implementovány a ověřeny nejprve na vygenerovaných signálech se známou základní frekvencí, čímž byla ověřena správnost implementace.

F_0	F_0 ACF	F_0 UACF	F_0 CEPSTRUM	F_0 HPS
120 Hz	119 Hz	124 Hz	122 Hz	119 Hz
200 Hz	200 Hz	198 Hz	204 Hz	200 Hz

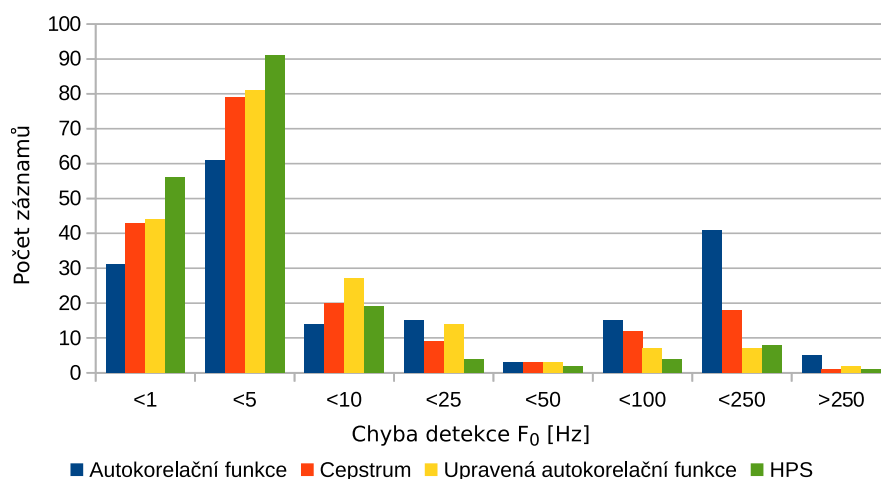
Tabulka 7.1: Detekovaná základní frekvence generovaných signálů.

Dále byly metody ověřeny na 185 záznamech fonace z HSV vyšetření. Záznamy obsahují fonace zdravých i nemocných pacientů s různou diagnózou onemocnění hlasivek. Referenční hodnota základní frekvence byla pro každý záznam určena vykreslením spektra a určením maximální hodnoty. V případech, kdy nebyla ze spektra jasně patrná základní frekvence, byl záznam poslechem porovnán s vygenerovanými tóny odpovídajícím možným základním frekvencím.

Rozdíl F_0 oproti referenci [Hz]	Počet záznamů			
	ACF	UACF	Cepstrum	HPS
<1	31	44	43	56
<5	61	81	79	91
<10	14	27	20	19
<25	15	14	9	4
<50	3	3	3	2
<100	15	7	12	4
<250	41	7	18	8
>250	5	2	1	1
celkem	185	185	185	185

Tabulka 7.2: Četnosti chyb detekce F_0 na reálných záznamech fonace.

Podle očekávání nejméně spolehlivě detekovala základní frekvenci záznamů klasická autokorelační funkce, ze všech uvedených metod dosáhla největšího počtu chyb s hodnotou více než 50 Hz. Ostatní metody poskytují oproti autokorelační funkci vyrovnané výsledky, nejpřesněji detekovala základní frekvenci metoda HPS, velmi dobrých výsledků dosáhla také upravená autokorelační funkce.

Obrázek 7.1: Histogram chyb detekce F_0 v audio signálu.

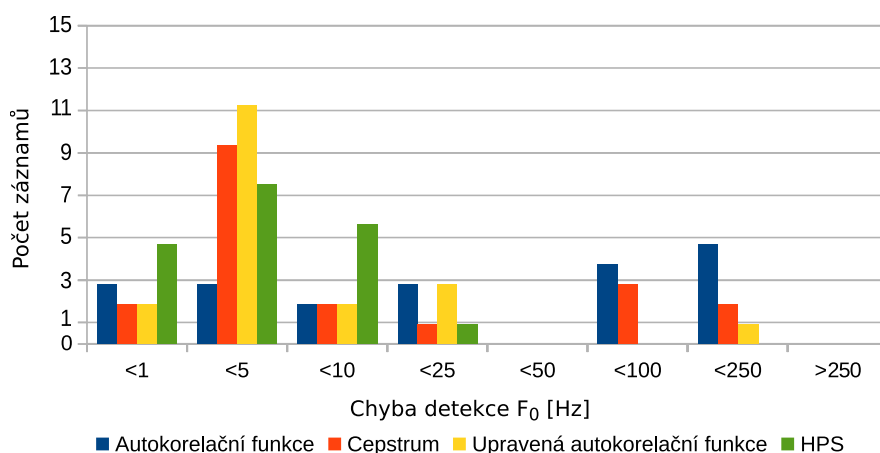
Do chyby úrovně 5 Hz je možné detekci základní frekvence považovat za správnou, protože počet vzorků záznamů je průměrně 8000 a podle vztahu 4.5 tak i referenční hodnota, která je určena podle spektra s rozlišením 5,38 Hz, může být určena s chybou až 2,69 Hz. Také při detekci kmitů je možné chybu

do úrovně 5 - 20 Hz, v závislosti na základní frekvenci, zanedbat. Algoritmus pro detekci počítá s frekvenčním kolísáním v průběhu fonace, hranice kmitů jsou hledány v rozsahu jedné čtvrtiny periody okolo předpokládaného konce periody, viz část 4.1.2.

Určení základní frekvence v EGG signálu bylo ověřeno vzhledem k dostupnosti dat pouze na 20 záznamech. Oproti zvukovému záznamu fonace EGG signál obsahuje velmi málo vysokých frekvencí, ale záznam obsahuje často velkou míru vysoko i nízkofrekvenčního šumu. Aby byl omezen vliv nízkofrekvenčního šumu, který zejména u autokorelačních metod znemožňuje detekci F_0 , je detekce prováděna na první derivaci EGG signálu.

Rozdíl F_0 oproti referenci [Hz]	Počet záznamů			
	ACF	UACF	Cepstrum	HPS
<1	3	2	2	5
<5	3	12	10	8
<10	2	2	2	6
<25	3	3	1	1
<50	0	0	0	0
<100	4	0	3	0
<250	5	1	2	0
>250	0	0	0	0
celkem	20	20	20	20

Tabulka 7.3: Četnosti chyb detekce F_0 na reálných záznamech EGG.



Obrázek 7.2: Histogram chyb detekce F_0 v EGG signálu.

7.2 Detekce kmitu

Úspěšnost detekce period kmitů v záznamu fonace je závislá na určení základní frekvence a závažnosti hlasové poruchy. Hlasová porucha může způsobit velké frekvenční kolísání a změnu tvaru periody, algoritmus detekce kmitu tak nemůže nalézt okrajové body periody podle zadaného pravidla. Detekce period tak může selhat nejen v případě zcela nepravidelné fonace, pro kterou již analýza odlišnosti kmitů nemá význam, ale také v případě ojedinělé nestandardní periody.

Zda algoritmus detekce našel všechny periody kmitů lze hodnotit porovnáním délky záznamu a součtu délek period. V ideálním případě by součet délek period měl být roven délce záznamu, ale ve většině případů počátek a konec záznamu není zarovnán k počátku, respektive konci periody a součet délek period bude vždy kratší než záznam. Algoritmus hledá počátek prvního kmitu od začátku záznamu v rozsahu jedné periody vypočtené ze základní frekvence signálu. Počet vzorků před počátkem první nalezené periody může dosáhnout téměř délky jedné periody. Obdobně pokud za poslední nalezenou periodou v záznamu je méně vzorků, než je délka jedné periody, je algoritmus ukončen.

Pro hodnocení úspěšnosti je nutné zvolit práh $C \in \langle 0; 2 \rangle$, kterým budou vykompenzovány neúplné periody na okrajích záznamu. Detekce period je označena jako úspěšná, pokud před počátkem první periody a za koncem poslední periody je méně než $C \cdot L_s$ vzorků, neboli rozdíl délky záznamu R a součtu délek period je menší než $C \cdot L_s$. Hodnota L_s odpovídá délce standardní periody, nejpresnější výsledky jsou dosaženy s prahem C v rozsahu $\langle 1, 6; 2 \rangle$.

$$R \in A, L_R - \sum_{k=1}^N L_k < C \cdot L_s \quad (7.1)$$

Ve výše uvedeném vztahu 7.1 označuje A množinu záznamů s úspěšně detekovanými kmity a L_R celkový počet vzorků záznamu.

Pro hodnocení úspěšnosti detekce je dále nutné ověřit, zda byly nalezeny periody kmitu odpovídající délce periody vypočtené podle vztahu 4.2 z referenční hodnoty základní frekvence. Kvůli frekvenčnímu kolísání v průběhu fonace nelze porovnat délku každé periody, pro ověření správnosti detekce bude hodnoceno, zda délka referenční periody leží v intervalu minimální a maximální nalezené délky periody, podle vztahu 7.2.

$$R \in B, L_{ref} \in \langle L_{min}, L_{max} \rangle \quad (7.2)$$

Ve výše uvedeném vztahu 7.2 označuje B množinu záznamů, kde délka referenční periody leží v intervalu minimální a maximální nalezené délky periody záznamu.

	Počet záznamů	
Nalezeny všechny kmity v záznamu ($C = 1$)	112	60,54%
Nalezeny všechny kmity v záznamu ($C = 1,5$)	160	86,49%
Nalezeny všechny kmity v záznamu ($C = 1,75$)	172	92,97%
Nalezeny všechny kmity v záznamu ($C = 2$)	175	94,59%
Nalezena správná délka periody	172	92,97%
$A_{C=2} \cap B$	164	88,65%
Celkový počet	185	100,00%

Tabulka 7.4: Úspěšnost detekce period kmitu v audiozáznamu.

Z uvedených hodnot vyplývá, že algoritmus dokáže správně detekovat periodu kmitu i pokud byla základní frekvence určena s chybou do úrovně 50 Hz. V naprosté většině záznamů algoritmus nalezne všechny periody kmitu hlasivek, zbývajících 5,5% představuje záznamy s velmi výraznou poruchou hlasu, případně nekvalitní záznamy. Všechny periody správné délky byly nalezeny v 88,65% záznamů.

	Počet záznamů	
Nalezeny všechny kmity v záznamu ($C = 1$)	15	75,00%
Nalezeny všechny kmity v záznamu ($C = 1,5$)	19	95,00%
Nalezeny všechny kmity v záznamu ($C = 1,75$)	20	100,00%
Nalezena správná délka periody	20	100,00%
$A_{C=2} \cap B$	20	100,00%
Celkový počet	20	100,00%

Tabulka 7.5: Úspěšnost detekce period kmitu v EGG záznamu.

Na dostupném omezeném vzorku EGG záznamů byla úspěšnost detekce period kmitů 100%. EGG signál obsahuje oproti zvukovému signálu jen velmi úzké spektrum frekvencí a působením hlasové poruchy se mění jen minimálně, detekce kmitů je tak výrazně přesnější.

7.3 Metoda SCORE

Úspěšnost klasifikace kvality závěru glottis metodou SCORE byla určena porovnáním známky přidělené záznamu fonace expertem ORL a známky určené metodou SCORE. Expert při hodnocení kvality závěru hlasivek určuje hodnocení zejména podle následujících parametrů určených MDVA softwarem: index měkké fonace, poměr šumu k harmonickým a frekvenční kolísání fonace. Výsledné hodnocení vychází také z vyšetření HSV a osobní zkušenosti experta [12].

Přesnost klasifikace je hodnocena na 200 záznamech fonace z HSV vyšetření. Tyto záznamy jsou specifické fonací vokálu „i“, která se kvůli umístění kamery v ústní dutině a omezení pohybu jazyka projevuje jako vokál „e“. Problematika záznamu byla blíže popsána v části 3.3. Data byla klasifikována třemi verzemi funkce SCORE, verze $SCORE_{25}$ je odvozená pro fonace vokálu „a“, $SCORE_{HSV20}$ s váhovými koeficienty odvozenými pro záznamy z vyšetření HSV a $SCORE_{HSV10}$ se zjednodušeným lineárním modelem také odvozena pro záznamy z vyšetření HSV.

Výsledek klasifikace je uveden ve formě porovnání s hodnocením experta. Pro každý záznam je určena chyba klasifikace ERR jako rozdíl výstupu metody SCORE $CLASS_{SCORE}$ a hodnocení experta $CLASS_{orl}$ podle vztahu 7.3.

$$ERR = CLASS_{SCORE} - CLASS_{orl} \quad (7.3)$$

Interpretace hodnoty ERR je následující:

$ERR > 0$... Metoda SCORE hodnotila horší známkou než expert - planý poplach.

$ERR < 0$... Metoda SCORE hodnotila lepší známkou než expert - přeměškaná diagnóza.

Hodnocení kvality závěru funkcí SCORE vychází z tvaru průběhu jedné periody kmitu hlasivek. Lze proto předpokládat, že hodnocení funkcí $SCORE_{25}$, jejíž váhové koeficienty jsou odvozeny pro fonace vokálu „a“, bude pro záznamy z HSV vyšetření nepřesné.

ERR	Počet záznamů	
-5	0	0,00%
-4	0	0,00%
-3	1	0,50%
-2	23	11,50%
-1	75	37,50%
0	64	32,00%
1	33	16,50%
2	4	2,00%
3	0	0,00%
4	0	0,00%
5	0	0,00%
celkem	200	100,00%

Tabulka 7.6: Chyba odhadu kvality závěru glottis modelem $SCORE_{25}$

Hodnocení závěru hlasivek shodné s hodnocením experta bylo určeno pro jednu třetinu záznamů. Více než polovina přidělených hodnocení se lišila o jednu úroveň od hodnocení experta. Hodnocení rozdílné o více než jednu úroveň bylo určeno pouze v minimálním množství záznamů. U modelů odvozených pro záznamy z HSV vyšetření lze očekávat vyšší přesnost hodnocení.

ERR	Počet záznamů	
-5	0	0,00%
-4	0	0,00%
-3	0	0,00%
-2	6	3,00%
-1	51	25,50%
0	88	44,00%
1	48	24,00%
2	7	3,50%
3	0	0,00%
4	0	0,00%
5	0	0,00%
celkem	200	100,00%

Tabulka 7.7: Chyba odhadu kvality závěru glottis modelem $SCORE_{HSV\ 20}$

Logaritmický model pro záznamy z HSV vyšetření na testovacích datech odhadl hodnocení shodné s expertem ORL v 44% záznamů. Hodnocení lišící

ERR	Počet záznamů	
-5	0	0,00%
-4	0	0,00%
-3	0	0,00%
-2	7	3,50%
-1	45	22,50%
0	97	48,50%
1	43	21,50%
2	8	4,00%
3	0	0,00%
4	0	0,00%
5	0	0,00%
celkem	200	100,00%

Tabulka 7.8: Chyba odhadu kvality závěru glottis modelem $SCORE_{HSV\ 10}$

se o jednu úroveň tato verze funkce určila v 49,5%. Hodnocení rozdílné o více než jednu úroveň bylo opět určeno pouze v minimálním množství záznamů.

Lineární model $SCORE_{HSV\ 10}$ oproti očekávání odhaduje na testovaných datech kvalitu závěru o přibližně 4% přesněji než logaritmický model $SCORE_{HSV\ 20}$.

Funkce SCORE s koeficienty váhové funkce odvozenými pro HSV záznamy fonace vokálu „e“ jsou na testovaných datech významně přesnější než původní verze funkce SCORE. Oba modely odvozené pro záznamy z HSV vyšetření odhadují kvalitu závěru s dostatečnou přesností, hodnocení s maximální chybou jednoho stupně oproti hodnocení experta je určeno v 93,5% testovaných záznamů. Chyba je u obou metod rovnoměrně rozložena mezi promeškané diagnózy a plané poplachy.

7.4 Kontaktní kvocient hlasivek

Metody výpočtu kontaktního kvocientu hlasivek jsou porovnány s kontaktním kvocientem určeným ze záznamu HSV vyšetření. Použitá HSV kamera snímá rychlostí 4000 snímků za vteřinu, časové rozlišení videozáznamu je tak více než desetinásobně nižší než časové rozlišení zvukového záznamu, který je vzorkován frekvencí 44 100 Hz. Hodnota kontaktního kvocientu určeného z videozáznamu tak není přesná, ale pouze orientační hodnota.

Hodnota CQ_{HSV} byla určena sledováním videozáznamu kmitání hlasivek, okamžik uzavření odpovídá snímku, ve kterém došlo k maximálnímu kontaktu hlasivek, okamžik otevření odpovídá snímku, ve kterém došlo k úplnému oddělení hlasivek. Z testované sady byly vyřazeny záznamy, kde se hlasivky v průběhu celého záznamu nedotýkají.

Průměrné hodnoty kontaktního kvocientu pro patnáct testovaných záznamů jsou uvedeny v tabulce 7.9. Metoda prahu pro celý záznam a metoda prahu pro jednotlivé periody mají práh v úrovni 30% amplitudy od minima do maxima.

Číslo záznamu	CQ_{30}	CQ_{T30}	DECOM	CQ_H	CQ_{HSV}
1	0,999	0,499	0,382	0,377	0,500
2	0,866	0,730	0,582	0,588	0,690
3	0,486	0,433	0,215	0,252	0,478
4	0,718	0,687	0,134	0,025	0,580
5	0,373	0,351	0,293	0,254	0,375
6	0,652	0,631	0,494	0,634	0,600
7	0,587	0,601	0,478	0,583	0,580
8	0,598	0,492	0,464	0,408	0,520
9	0,615	0,611	0,497	0,257	0,610
10	0,605	0,597	0,535	0,609	0,600
11	0,746	0,704	0,468	0,022	0,777
12	0,517	0,549	0,450	0,423	0,550
13	0,676	0,482	0,506	0,464	0,550
14	0,783	0,590	0,461	0,445	0,400

Tabulka 7.9: Průměrné hodnoty kontaktního kvocientu záznamů.

Na testovaných datech kombinovaná metoda a metoda DECOM ve většině případů určují hodnotu kontaktního kvocientu výrazně nižší než je hodnota CQ_{HSV} . Maximální rozdíl v případě kombinované metody tvořil až

0,755. Obě metody jsou založené na derivaci EGG signálu a jsou tak negativně ovlivněny vysokofrekvenčním šumem, který se vyskytuje ve většině záznamů. Prahové metody určují hodnotu CQ blíže k referenční hodnotě CQ_{HSV} , ale základní metoda s globálním prahem pro celý záznam selhává při výskytu nízkofrekvenčního šumu, což je nejvíce patrné v záznamu číslo 1. Nejspolehlivější a zároveň nejpřesnější výsledky na testovaných datech podává metoda prahu po jednotlivých periodách.

Nízkofrekvenční šum v EGG signálu lze redukovat horní propustí, testovaná data byla filtrována horní propustí typu IIR 2. řádu s mezní frekvencí 45 Hz. Hodnoty CQ určené testovanými metodami jsou uvedeny v tabulce 7.9.

Číslo záznamu	CQ_{30}	CQ_{T30}	DECOM	CQ_H	CQ_{HSV}
1	0,999	0,493	0,360	0,362	0,500
2	0,894	0,725	0,570	0,524	0,690
3	0,450	0,448	0,213	0,227	0,478
4	0,892	0,679	0,125	0,025	0,580
5	0,718	0,474	0,292	0,230	0,375
6	0,626	0,609	0,461	0,533	0,600
7	0,499	0,585	0,462	0,494	0,580
8	0,522	0,491	0,434	0,354	0,520
9	0,606	0,598	0,476	0,217	0,610
10	0,572	0,574	0,415	0,555	0,600
11	0,640	0,581	0,641	0,332	0,777
12	0,585	0,530	0,405	0,387	0,550
13	0,640	0,544	0,464	0,409	0,550
14	0,873	0,573	0,453	0,414	0,400

Tabulka 7.10: Průměrné hodnoty kontaktního kvocientu záznamů filtrovaného horní propustí.

Filtrace horní propustí výstup metod na testovaných datech nezlepšila, naopak v některých případech došlo k mírnému zhoršení přesnosti. Parametry filtru byly nastaveny shodně pro všechny záznamy, pro optimální výsledky by bylo nutné pro každý záznam zvolit vhodné hodnoty filtru. Nejpřesnější výsledky na filtrovaných datech podala opět metoda prahu po jednotlivých periodách.

Porovnáním metod se zabýval ve své práci Christian Herbst [6]. K porovnání využil záznamů EGG a videokymogramů fonace zkušených amatérských

zpěváků bez hlasové poruchy. Každý záznam obsahoval fonaci vokálu „a“ přecházející v rozmezí jedné oktávy od F#2 do F#3. Kontaktní kvocient určený z videokymogramu CQ_{KYM} byl stanoven jako referenční. Nejvíce referenci odpovídaly prahové metody, konkrétně s prahem 20%, derivační metody určily CQ_{EGG} vždy nižší, rozdíl oproti CQ_{KYM} tvořil až 0,3. Výsledky tak odpovídají hodnotám určeným v této práci, zejména jsou-li brány v potaz rozdílné kvality testovaných záznamů.

7.5 Detekce nestandardních kmitů

Úspěšnost metod detekce nestandardních kmitů fonace nelze porovnat podle jediného parametru. Nestandardní perioda může vzniknout z mnoha příčin a jednotlivé metody mohou být na odlišnosti daného původu různě citlivé. Hodnocení je nutné provádět pro každý záznam individuálně a hodnotit výstup metod v závislosti na diagnóze a kvalitě záznamu.

Pro hodnocení a popis vlastností metod bylo vybráno šest reprezentativních záznamů, čtyři záznamy obsahují fonaci pacientů s diagnostikovanou poruchou hlasu a dva záznamy pacientů bez nálezu. Záznamy nemocných pacientů byly vybrány tak, aby obsahovaly převážně pravidelnou fonaci s několika periodami nestandardního tvaru.

Postup hodnocení

V každém záznamu jsou nejprve automaticky detekovány jednotlivé periody. V segmentovaném záznamu jsou označeny subjektivně vizuálně nejodlišnější periody fonace. Následně jsou určeny standardní periody následujícími metodami, které jsou popsány v části 5:

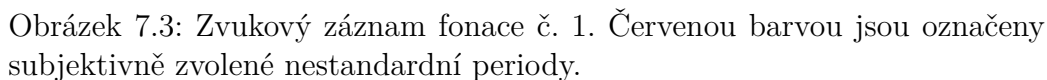
- Průměrná hodnota vzorků
- Medián vzorků
- Minimální eukleidovská vzdálenost mediánu amplitudového spektra
- Skóre četnosti amplitudového spektra

Pro každý standardní kmit je vypočítána odlišnost period podle níže uvedených metod detekce nestandardních kmitů, podrobně popsaných v části 6:

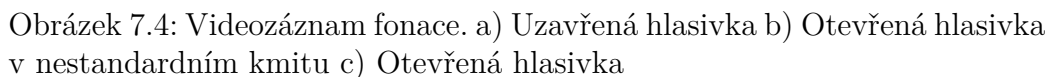
- Průměrná odchylka vzorků periody
- Vzdálenost vzorků periody
- Skóre vzorků mimo pásmo σ
- Eukleidovská vzdálenost amplitudového spektra
- Vzdálenost Fourierových deskriptorů

Úspěšnost metod je poté hodnocena porovnáním označených kmitů a kmitů s nejvyšší hodnotou odlišnosti. Dále je nutné hodnotit zda pravidelné kmity, které by měl reprezentovat standardní kmit fonace, mají minimální hodnotu odlišnosti.

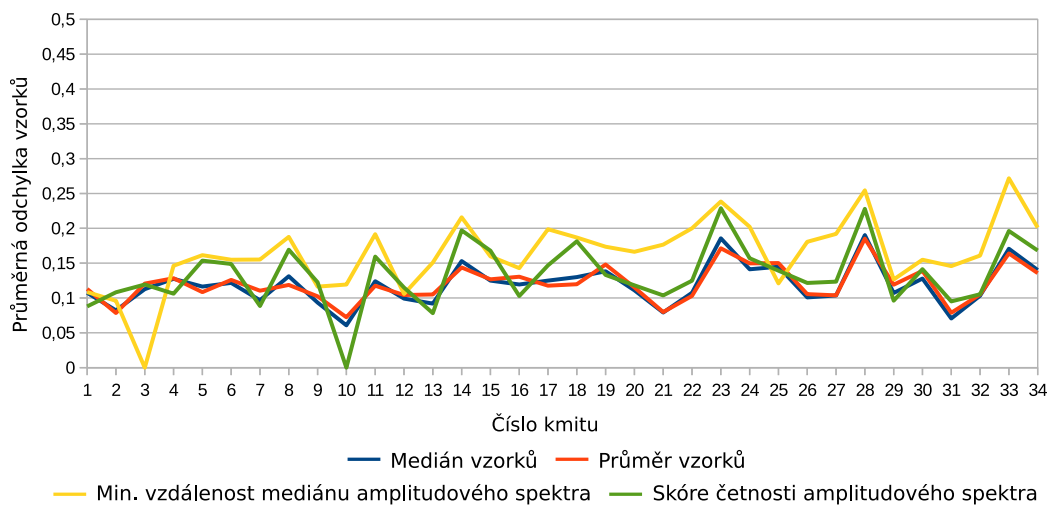
První záznam představuje fonaci pacienta s diagnózou bilaterální Reinkeho edém. Nestandardní kmity jsou náhodně rozloženy v průběhu fonace. Odlišnost kmitů je patrná, ale nevýrazná, periody se odlišují zejména v amplitudách frekvenčních složek. Průběh fonace je uveden na obrázku 7.3.



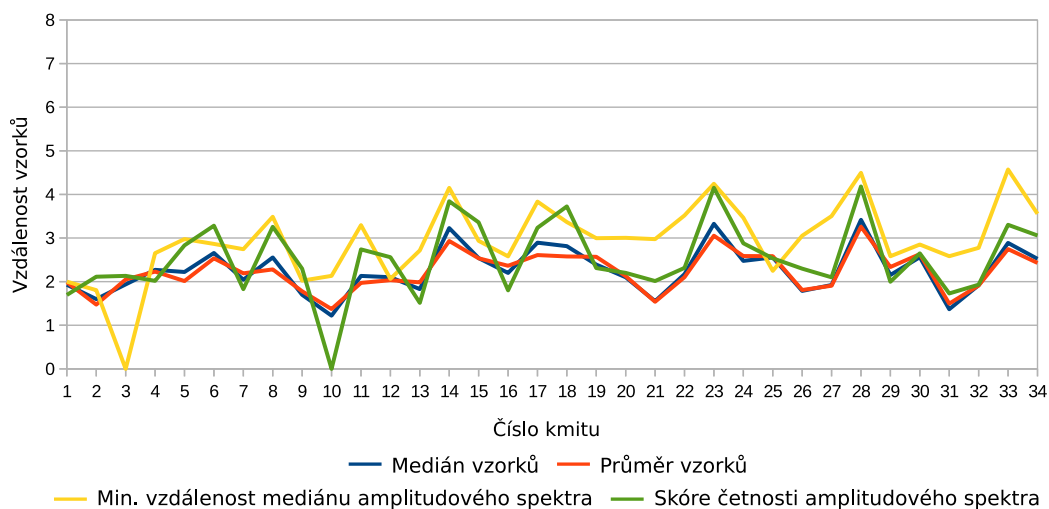
Ve videozáznamu fonace je patrný nesymetrický pohyb hlasivek, v označených odlišných kmitech dochází ve fázi otevření na hraně hlasivek k dotyku. Průběh pohybu hlasivek je uveden na obrázku 7.4.



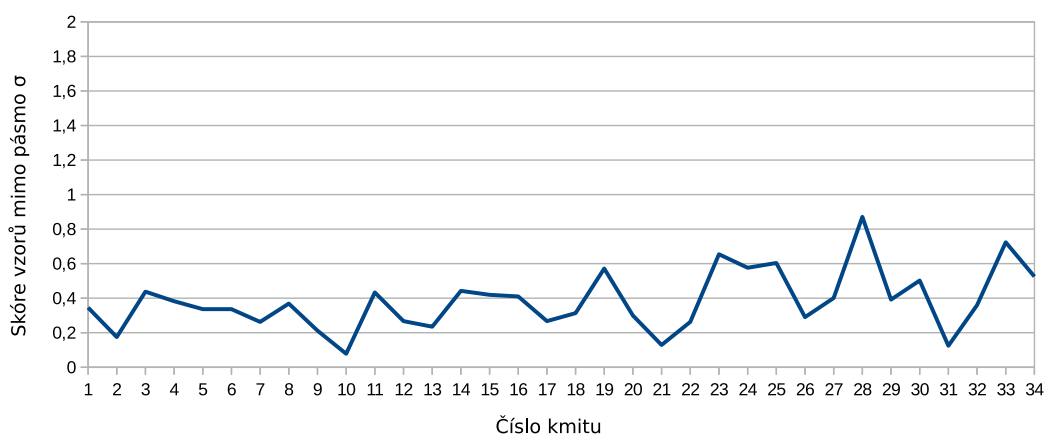
Detekce v časové oblasti



Obrázek 7.5: Průměrné odchylky vzorků period záznamu č.1.

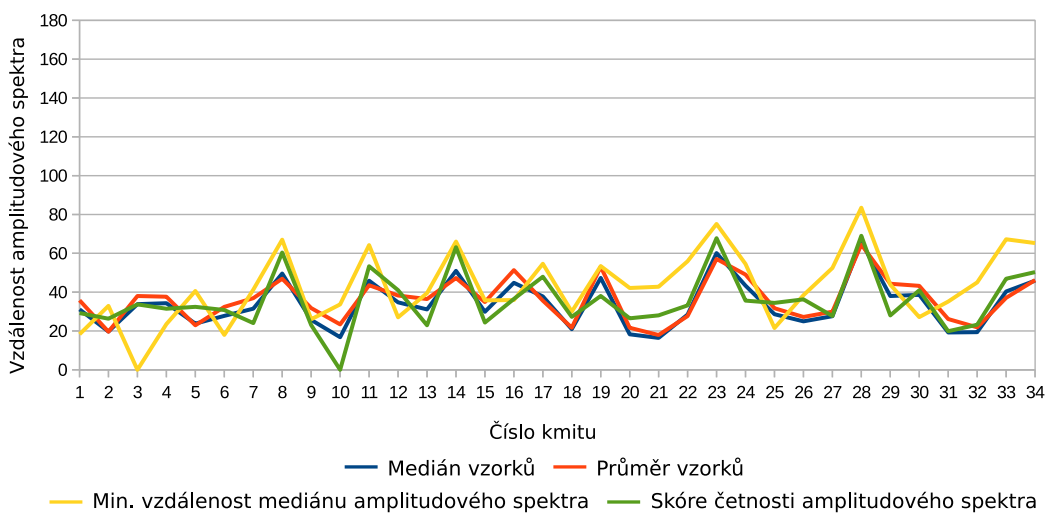


Obrázek 7.6: Vzdálenosti vzorků period záznamu č.1.

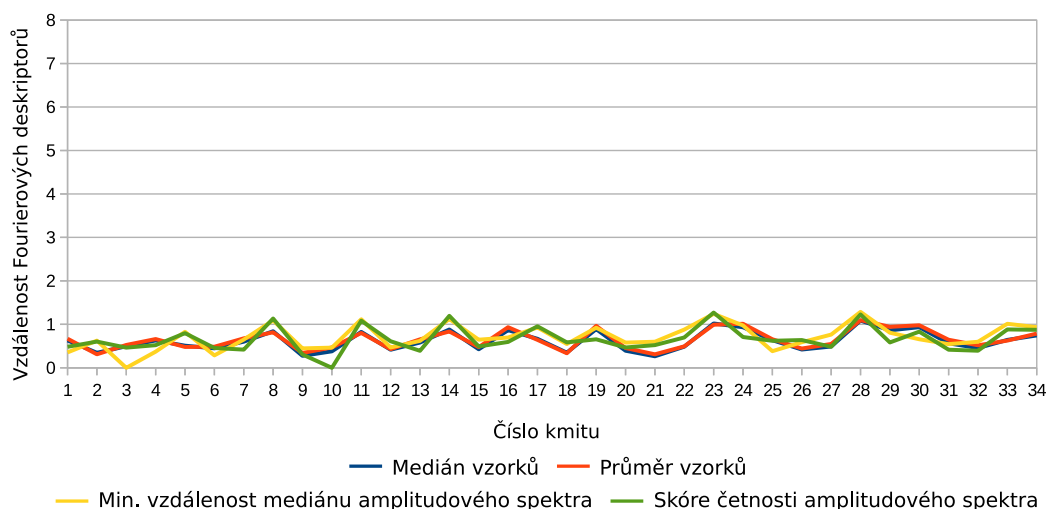


Obrázek 7.7: Skóre vzorků mimo pásmo σ jednotlivých period záznamu č.1.

Detekce ve frekvenční oblasti



Obrázek 7.8: Eukleidovská vzdálenost amplitudového spektra period záznamu č.1.



Obrázek 7.9: Vzdálenost Fourierových deskriptorů period záznamu č.1.

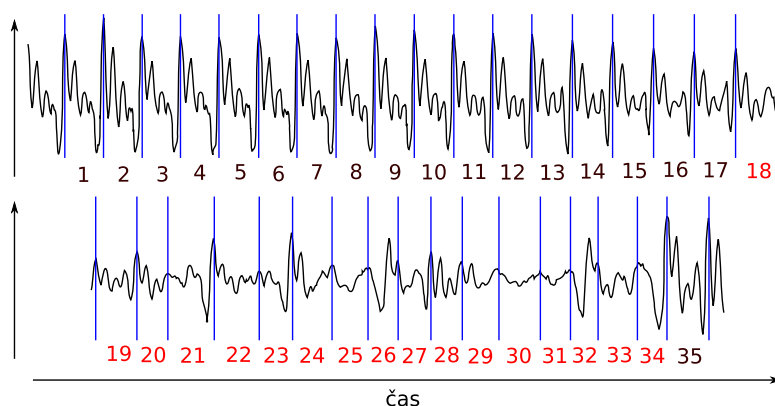
Všechny subjektivně zvolené nestandardní kmity odpovídají vrcholům v grafech odlišnosti pro všechny metody. V případě metod *Průměrná odchylka vzorků periody* a *Vzdálenosti vzorků period* je odlišnost kmitů výraznější při použití metod určení standardního kmitu z frekvenční oblasti, zejména *Skóre četnosti amplitudového spektra*. V případě metody *Skóre vzorků mimo pásmo σ* je hodnota odlišnosti označených kmitů často jen minimálně vyšší než hodnota odlišnosti okolních kmitů, například kmitů č. 8, č.14 a č. 23 v grafu 7.7.

Metody *Eukleidovská vzdálenost amplitudového spektra* a *Vzdálenost Fourierových deskriptorů* produkují pro tento záznam velmi podobné výsledky. V porovnání s metodami z časové oblasti jsou označené kmity hodnoceny jako výrazně odlišnější oproti sousedním kmitům.

Lepší výsledky ve frekvenční oblasti odpovídají vlastnostem nestandardních kmitů, které se odlišují převážně složením amplitudového spektra.

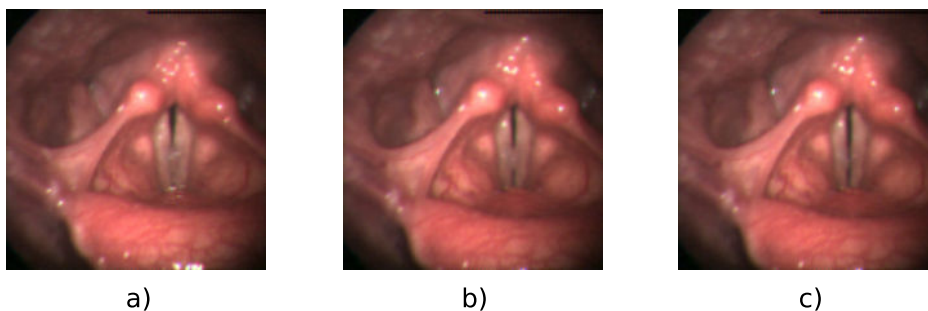
7.5.2 Záznam č. 2

Záznam číslo dva obsahuje z počátku pravidelnou fonaci, která v druhé části záznamu zcela selhává a periody získávají téměř náhodný tvar. Stanovená diagnóza pacienta jsou hlasivkové uzlíky, které způsobují nedomykavost hlasivek. Nestandardních kmitů je téměř polovina z celkového počtu a záznam byl zvolen pro demonstraci jejich vlivu na tvar standardní periody.



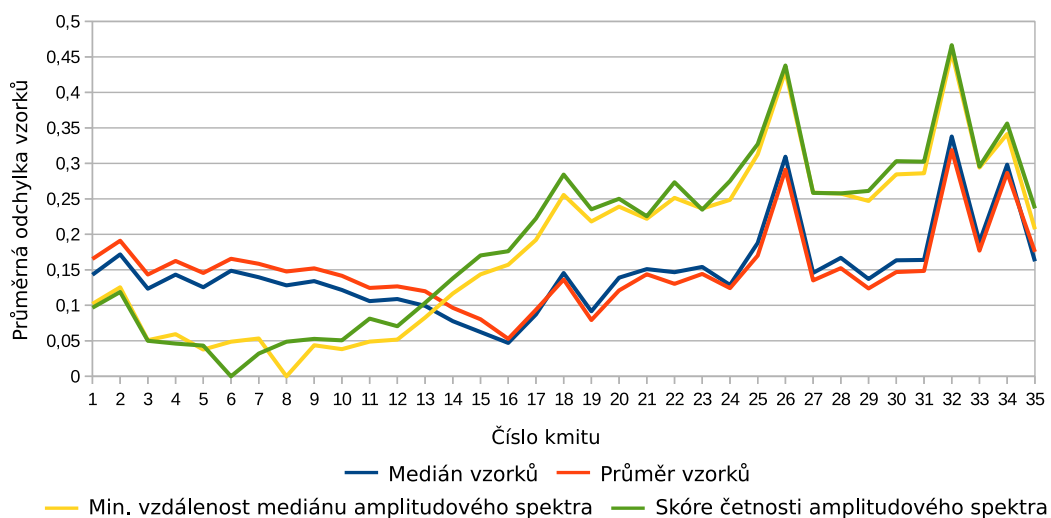
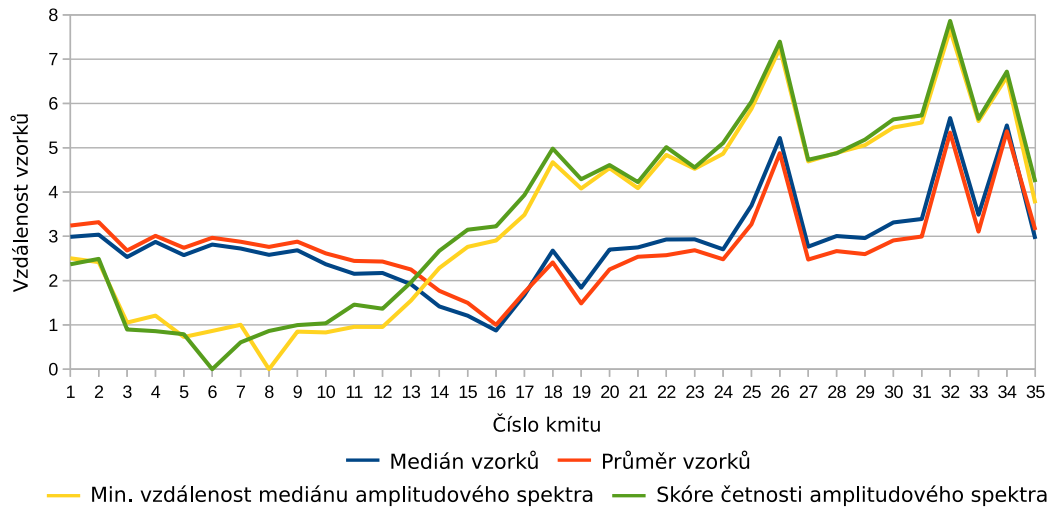
Obrázek 7.10: Zvukový záznam fonace č. 2. Červenou barvou jsou označeny subjektivně zvolené nestandardní periody.

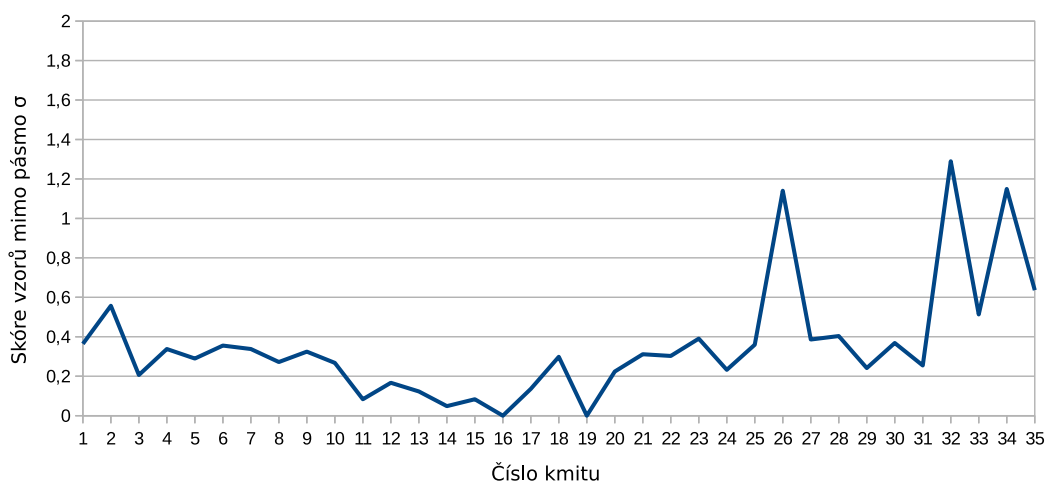
Změna tvaru period fonace je způsobena nedomykavostí hlasivek, která je dobře patrná z videozáznamu fonace. V pravidelných kmitech dochází k uzavření hlasivek od uzlíku k přední komisuře, v nepravidelné části fonace dochází k dotyku pouze v oblasti uzlíků. Stav hlasivek ve fázi uzavření je na obrázku 7.11.



Obrázek 7.11: Videozáznam fonace. a) Uzavřená hlasivka b) Uzavřená hlasivka v nestandardním kmitu c) Otevřená hlasivka

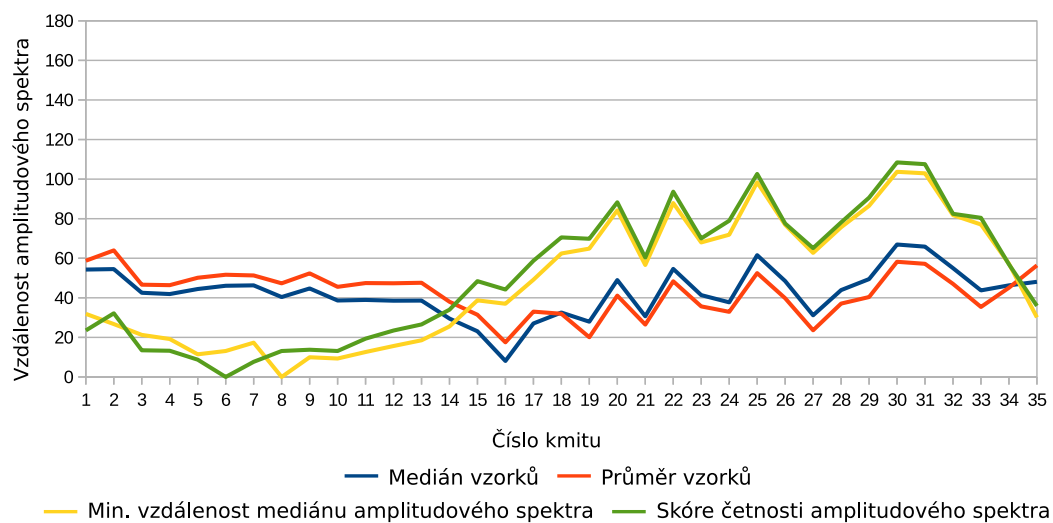
Detekce v časové oblasti

Obrázek 7.12: *Průměrné odchylky vzorků period záznamu č.2.*Obrázek 7.13: *Vzdálenosti vzorků period záznamu č.2.*

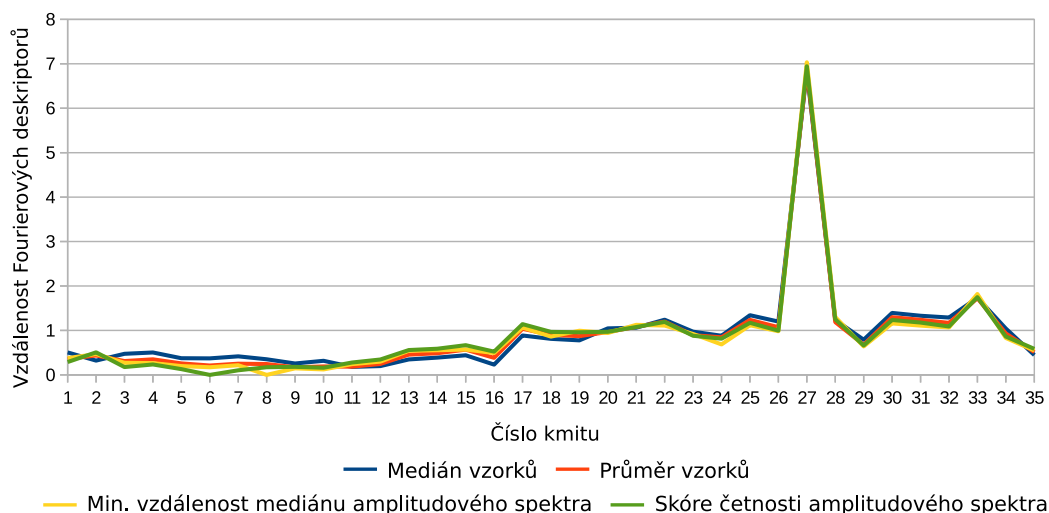


Obrázek 7.14: Skóre vzorů mimo pásmo σ jednotlivých period záznamu č.2.

Detekce ve frekvenční oblasti



Obrázek 7.15: Eukleidovská vzdálenost amplitudového spektra period záznamu č.2.



Obrázek 7.16: Vzdálenost Fourierových deskriptorů period záznamu č.2.

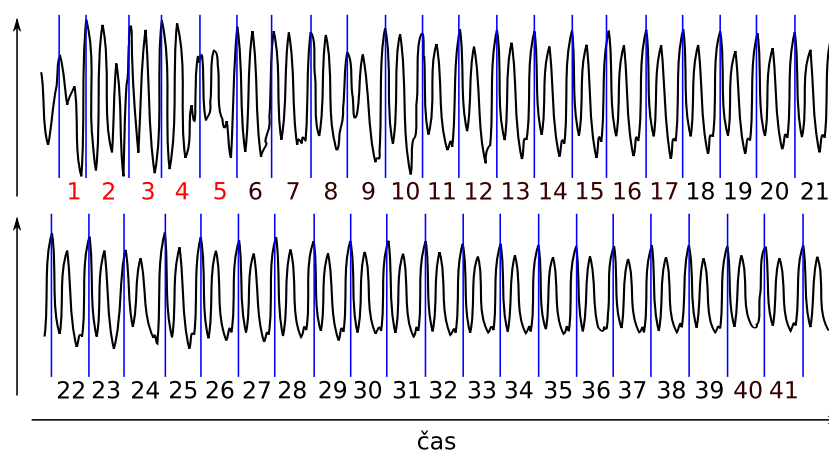
Odlišnost nestandardních kmitů v záznamu č.2 je velmi výrazná, ale kvůli vysokému počtu nepravidelných period se projeví také ve standardní periodě určené metodami *Průměr vzorků* a *Medián vzorků*. Odlišnost vzhledem k těmto standardním periodám je výrazná i pro pravidelné kmity v první části záznamu.

Metoda detekce *Skóre vzorků mimo pásmo σ* je vzhledem k velkému počtu nestandardních period také nepřesná, z celkového počtu kmitů jsou pouze 3 ohodnoceny jako výrazně odlišné.

Pro tento záznam jsou nejvhodnější metody volby standardního kmitu *Minimální eukleidovská vzdálenost mediánu amplitudového spektra* a *Skóre četnosti amplitudového spektra*. Všechny metody detekce s těmito standardními metodami správně hodnotí pravidelné kmity v první části záznamu jako minimálně odlišné.

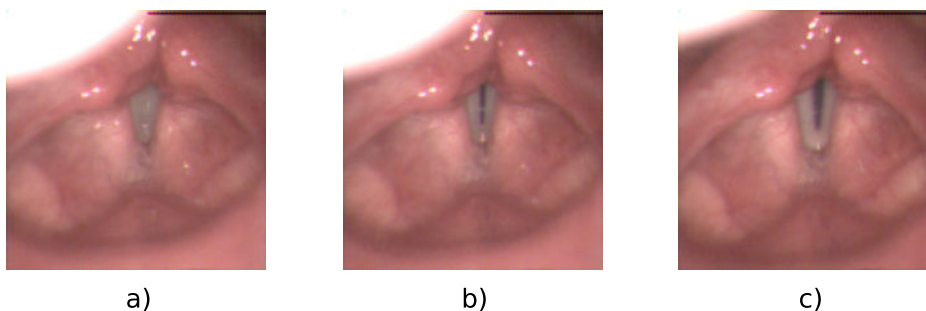
7.5.3 Záznam č. 3

Tento záznam obsahuje z převážné většiny pravidelnou fonaci, pouze prvních pět kmitů se výrazněji odlišuje. Záznam zachycuje fonaci hlasivek bez nálezu a odlišnost je dle analýzy videozáznamu způsobena přítomností hleny na počátku fonace.



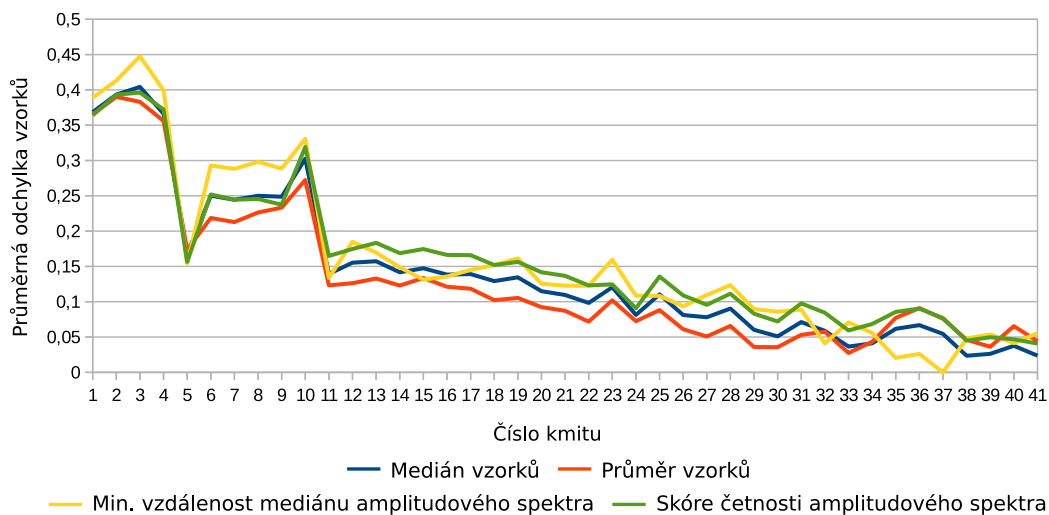
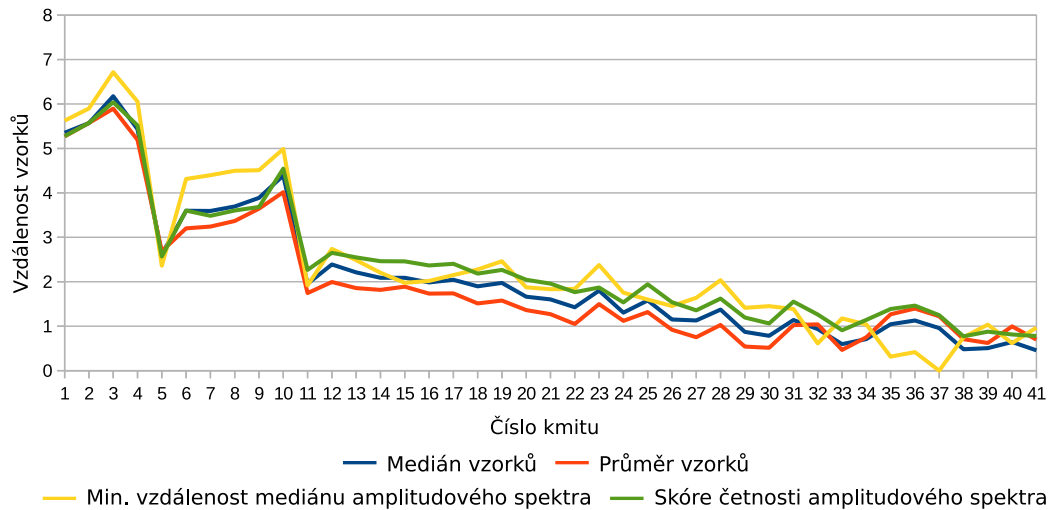
Obrázek 7.17: Zvukový záznam fonace č. 3. Červenou barvou jsou označeny ručně zvolené nestandardní periody.

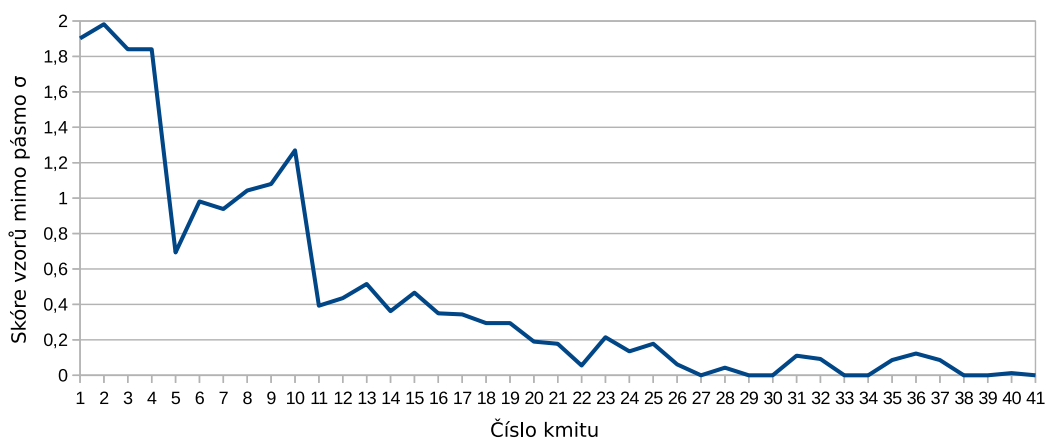
Na videozáznamu fonace je v prvních několika kmitech ve střední části glottis viditelné množství hleny. V následujících kmitech již dochází k čistému otevření a velmi dobrému uzavěru hlasivek, viz obrázek 7.18



Obrázek 7.18: Videozáznam fonace. a) Uzavřená hlasivka b) Otevřená hlasivka v nestandardním kmitu c) Otevřená hlasivka

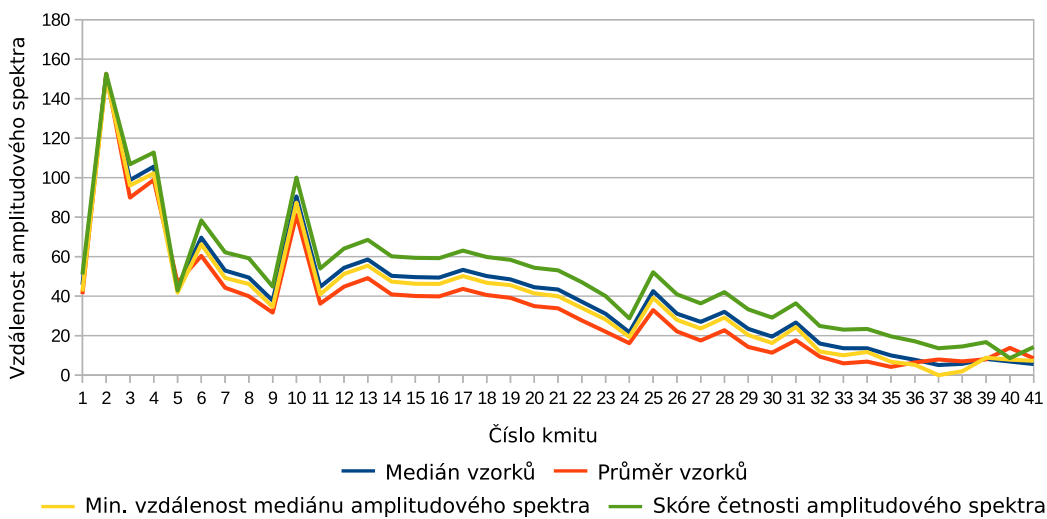
Detekce v časové oblasti

Obrázek 7.19: *Průměrné odchylky vzorků period záznamu č.3.*Obrázek 7.20: *Vzdálenosti vzorků period záznamu č.3.*

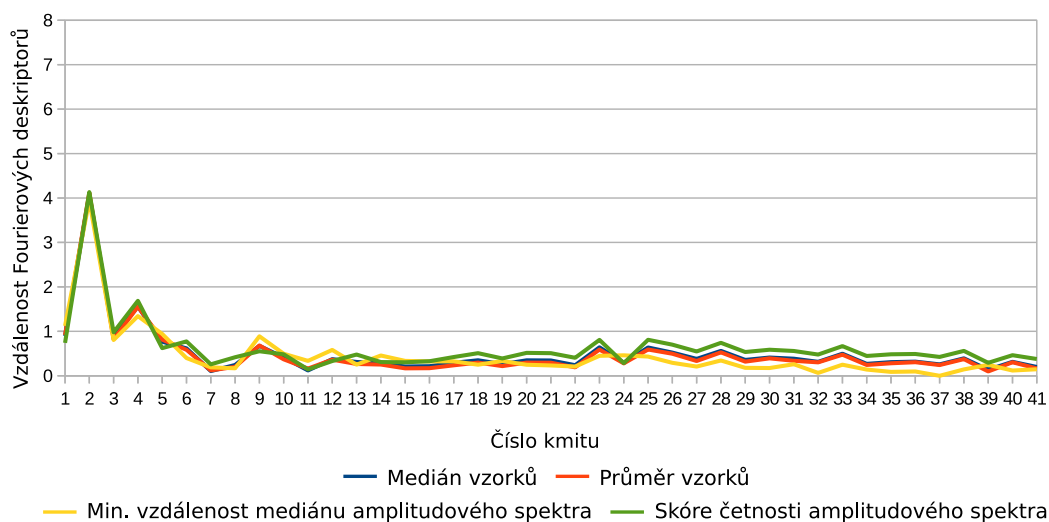


Obrázek 7.21: Skóre vzorků mimo pásmo σ jednotlivých period záznamu č.3.

Detekce ve frekvenční oblasti



Obrázek 7.22: Eukleidovská vzdálenost amplitudového spektra period záznamu č.3.



Obrázek 7.23: *Vzdálenost Fourierových deskriptorů* period záznamu č.3.

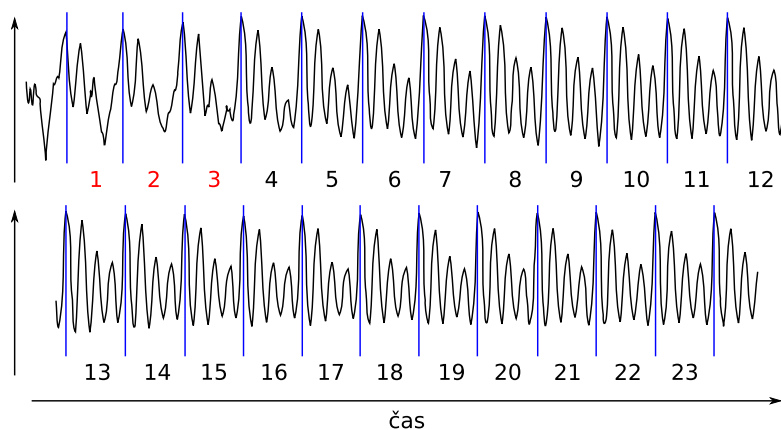
V tomto záznamu jsou hodnoty odlišnosti pro všechny standardní periody velmi podobné, výstupem všech metod určení standardních kmitů tedy byl téměř identický tvar periody.

Také výsledky metod detekce *Průměr vzorků* a *Medián vzorků* a *Skóre vzorků mimo pásmo σ* jsou podobné. Zvláštností je nižší odlišnost 5. kmitu, který se zdá tvarově zcela rozdílný, ale podrobnou analýzou a porovnáním se standardní periodou bylo potvrzeno, že se tvarově blíží pravidelným kmitům fonace. Metody určily významnou odlišnost také u 6. až 10. kmitu.

Hodnoty odlišnost určené metodami *Eukleidovská vzdálenost amplitudového spektra* a *Vzdálenost Fourierových deskriptorů* přesněji odpovídá označeným kmitům. Pro jednotlivé standardní periody jsou výsledky opět téměř shodné. V případě metody vzdálenost Fourierových deskriptorů se projevuje nezávislost na amplitudě signálu u 10. kmitu, který je ostatními metodami hodnocen jako výrazně odlišný.

7.5.4 Záznam č. 4

Záznam číslo čtyři obsahuje fonaci pacienta s diagnózou polyp hlasivky. Tvar period je pravidelný, pouze v prvních třech periodách je patrné odlišné složení amplitudového spektra.



Obrázek 7.24: Zvukový záznam fonace č. 4. Červenou barvou jsou označeny ručně zvolené nestandardní periody.

Polyp na pravé hlasivce zasahuje téměř polovinu hrany hlasivky. Nestandardní tvar prvních period fonace způsobuje odlišný pohyb zbytnělé tkáně, který se poté sjednotí s pohybem celé hlasivky.



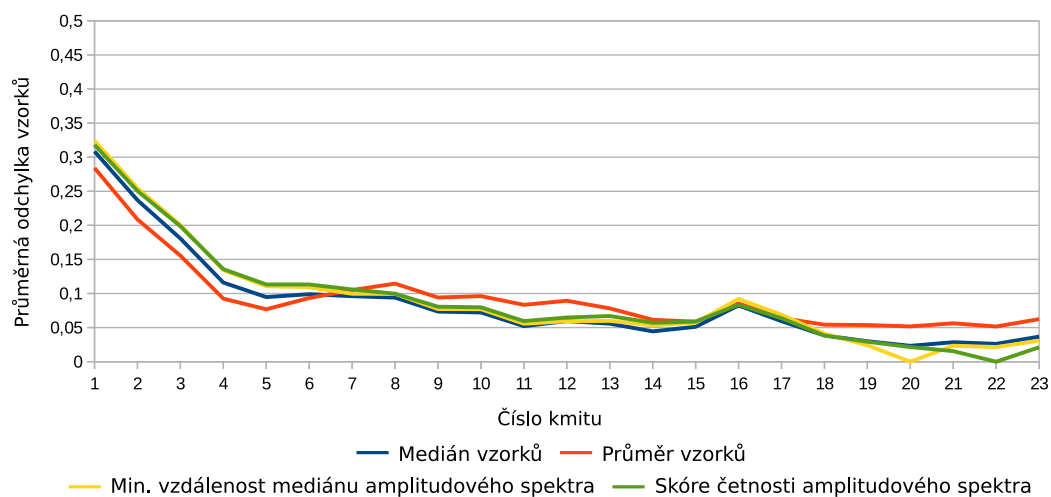
a)



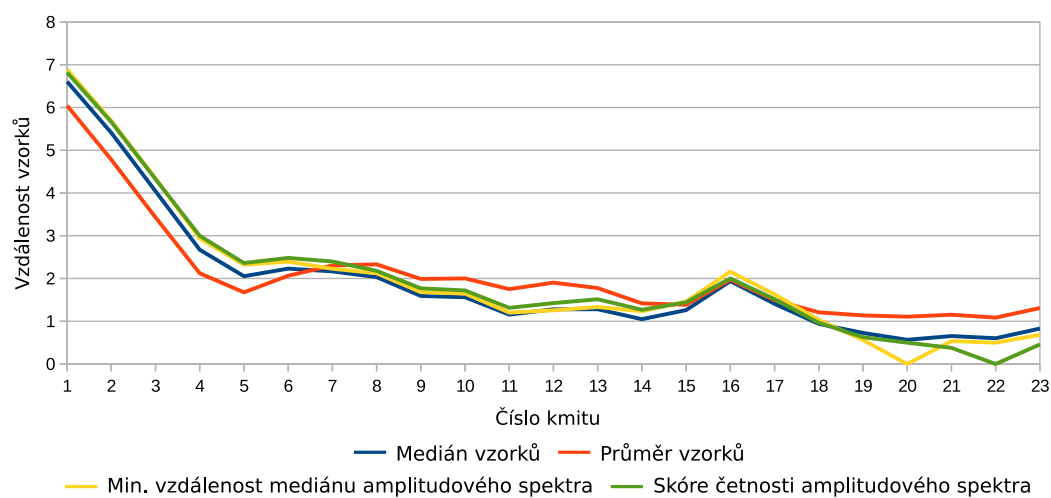
b)

Obrázek 7.25: Videozáznam fonace. a) Uzavřená hlasivka b) Otevřená hlasivka

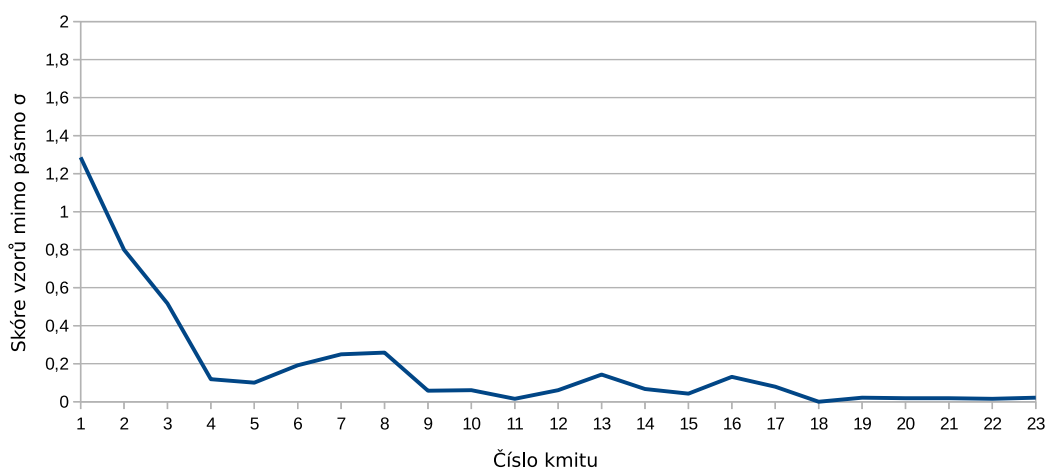
Detekce v časové oblasti



Obrázek 7.26: Průměrné odchylky vzorků period záznamu č. 4.

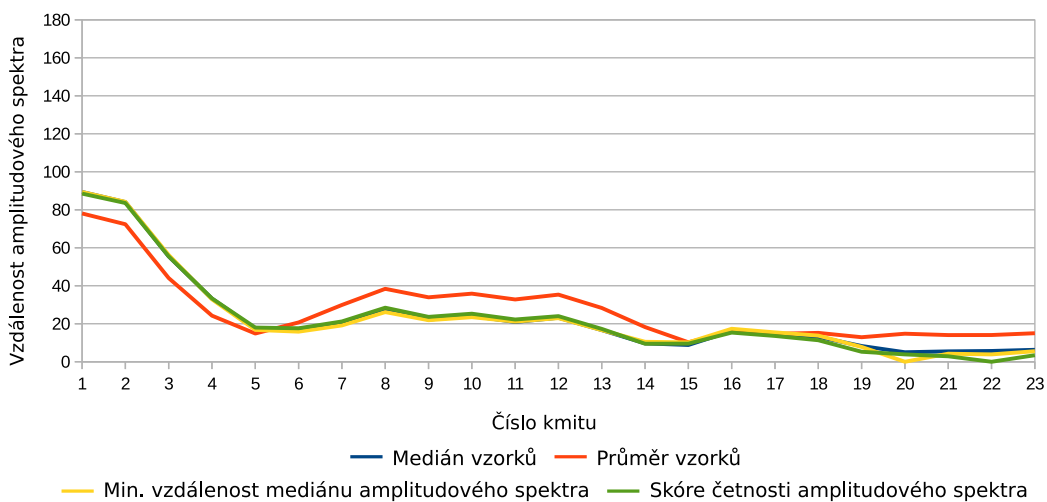


Obrázek 7.27: Vzdálenosti vzorků period záznamu č.4.

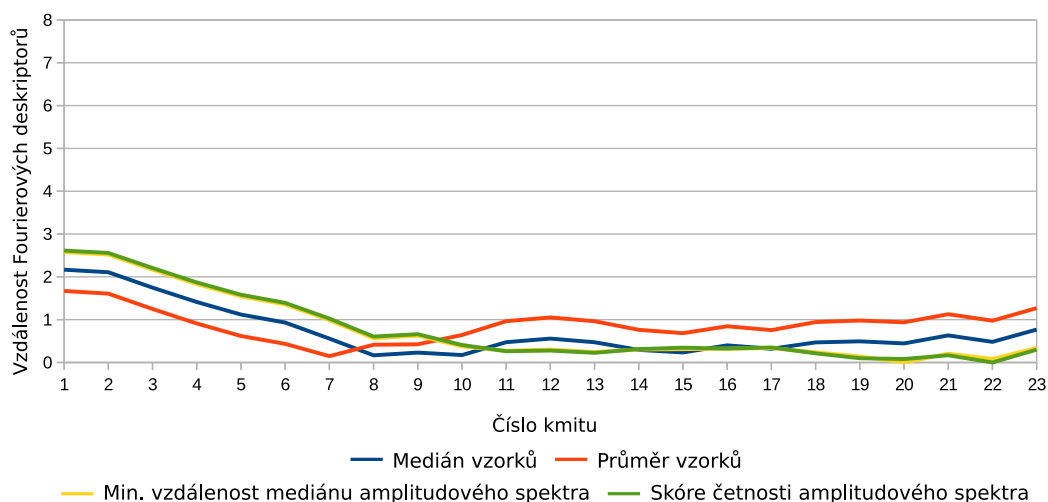


Obrázek 7.28: Skóre vzorků mimo pásmo σ jednotlivých period záznamu č.4.

Detekce ve frekvenční oblasti



Obrázek 7.29: Eukleidovská vzdálenost amplitudového spektra period záznamu č.4.



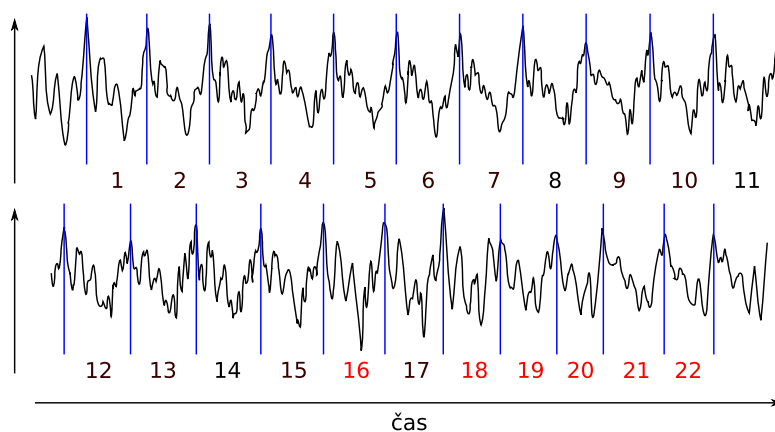
Obrázek 7.30: Vzdálenost Fourierových deskriptorů period záznamu č.4.

Všechny metody detekce v časové oblasti poskytují pro tento záznam velmi podobné výsledky, stejně tak všechny metody určení standardního kmitu. První tři kmity fonace jsou hodnoceny jako nejodlišnější všemi metodami.

Metoda *Eukleidovská vzdálenost amplitudového spektra* podává shodné výsledky pro všechny metody určení standardního kmitu s výjimkou metody průměr vzorků, ale odchylka od ostatních metod je minimální. Výsledky metody *Vzdálenost Fourierových deskriptorů* se od ostatních metod liší, zejména při použití standardních period metod *Průměr vzorků* a *Medián vzorků*.

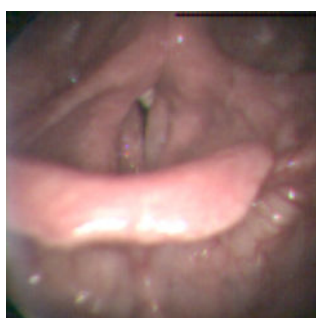
7.5.5 Záznam č. 5

Záznam číslo pět představuje fonaci s nepravidelným tvarem period, z nichž označené periody jsou subjektivně odlišné nejvíce. Hlasová porucha je v tomto případě velmi výrazná, stanovená diagnóza je karcinom hlasivek, v době vyšetření je diagnóza již zcela zjevná.



Obrázek 7.31: Zvukový záznam fonace č. 5. Červenou barvou jsou označeny ručně zvolené nestandardní periody.

Pravá hlasivka je zasažena karcinomem a podle videozáznamu fonace je zcela nehybná, poloha hlasivek v otevřené a uzavřené fázi je uvedena na obrázku 7.32.



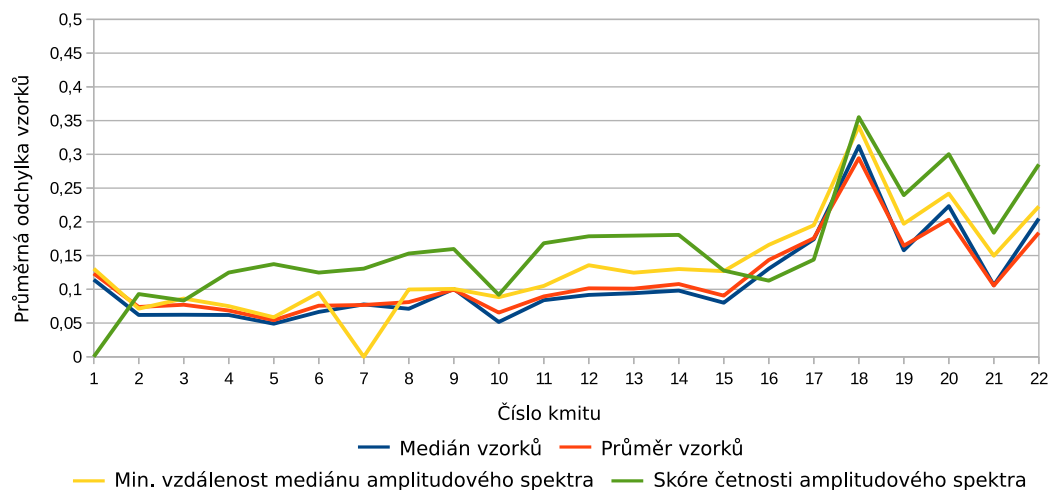
a)



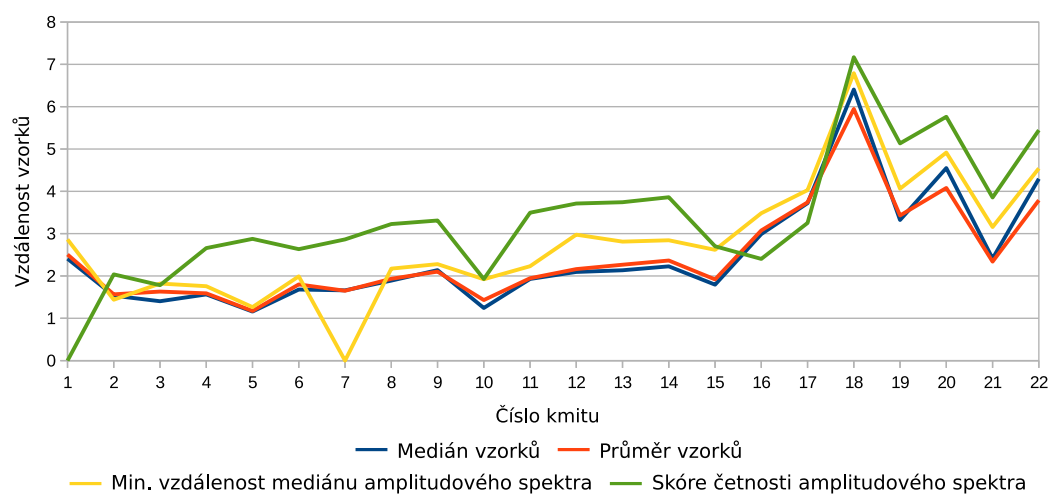
b)

Obrázek 7.32: Videozáznam fonace. a) Uzavřená hlasivka b) Otevřená hlasivka

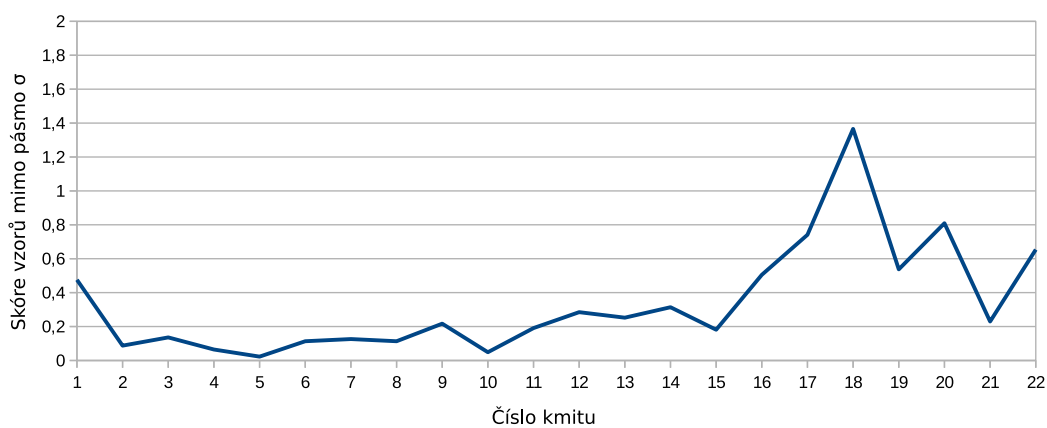
Detekce v časové oblasti



Obrázek 7.33: *Průměrné odchylky vzorků period záznamu č.5.*

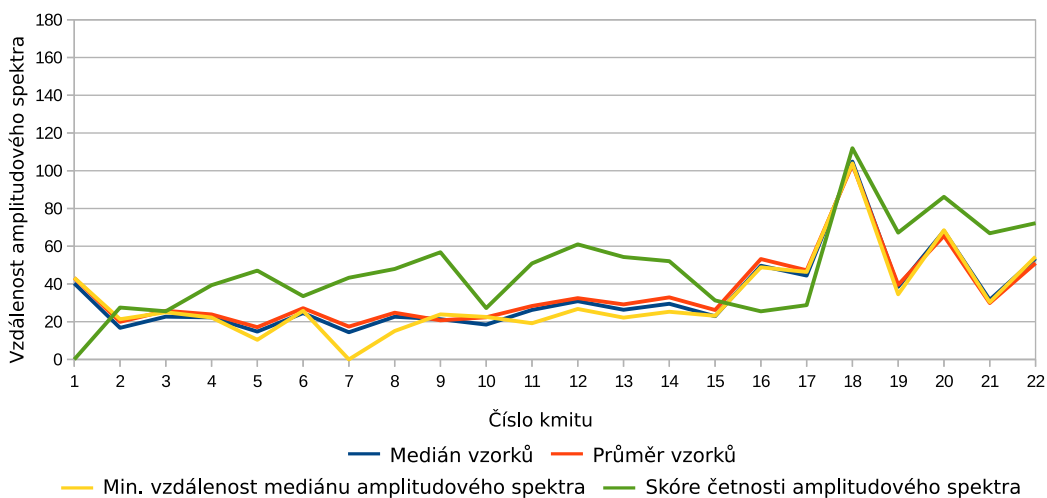


Obrázek 7.34: *Vzdálenosti vzorků period záznamu č. 5.*

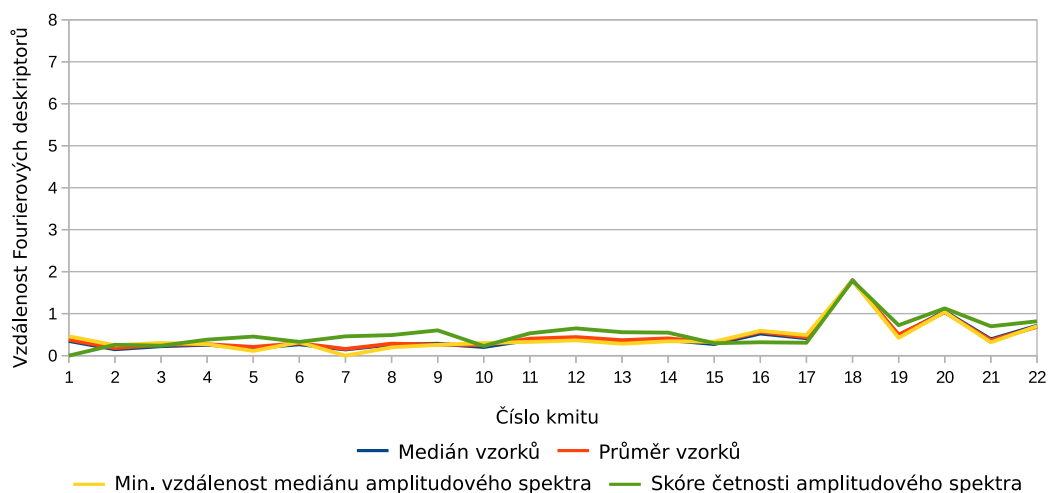


Obrázek 7.35: Skóre vzorků mimo pásmo σ jednotlivých period záznamu č.5.

Detekce ve frekvenční oblasti



Obrázek 7.36: Eukleidovská vzdálenost amplitudového spektra period záznamu č.5.



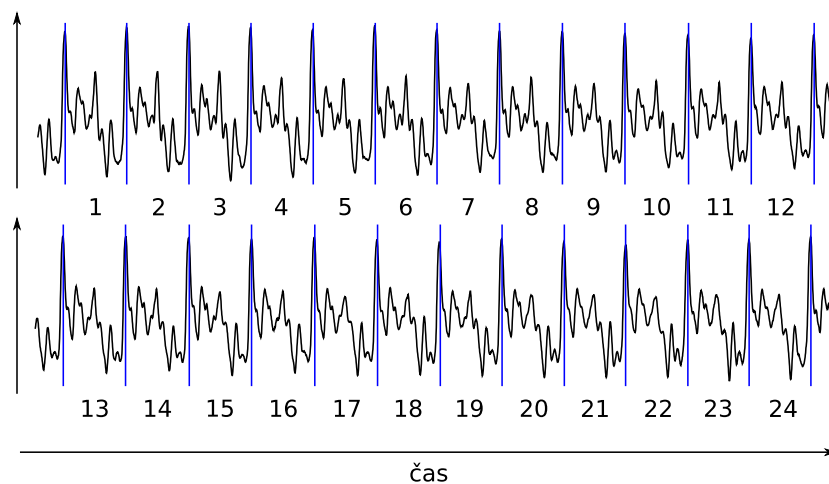
Obrázek 7.37: Vzdálenost Fourierových deskriptorů period záznamu č.5.

V časové oblasti jsou pro všechny metody detekce jako nejodlišnější označeny poslední kmity fonace a výsledky tak odpovídají subjektivně zvoleným periodám. Metoda určení standardního kmitu *Skóre četnosti amplitudového spektra* určila oproti ostatním metodám odlišný tvar periody. V případě takto nepravidelné fonace však nelze určit vhodnost zvoleného tvaru.

Výsledky metod pro detekci nestandardních kmitů ve frekvenční oblasti odpovídají výsledkům metod detekce v časové oblasti.

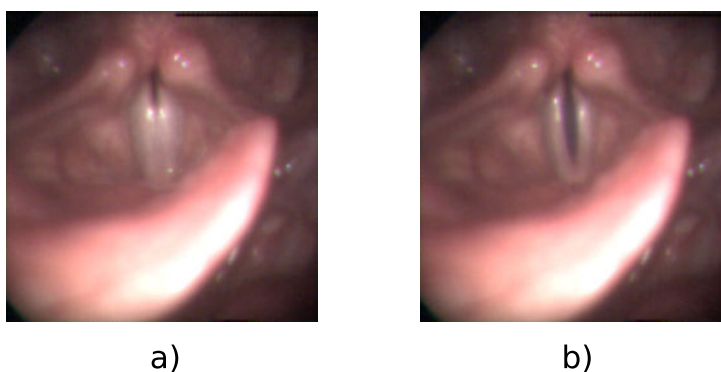
7.5.6 Záznam č. 6

Záznam č. 6 obsahuje fonaci zdravého pacienta, celá fonace je pravidelná a nelze vybrat odlišný kmit.



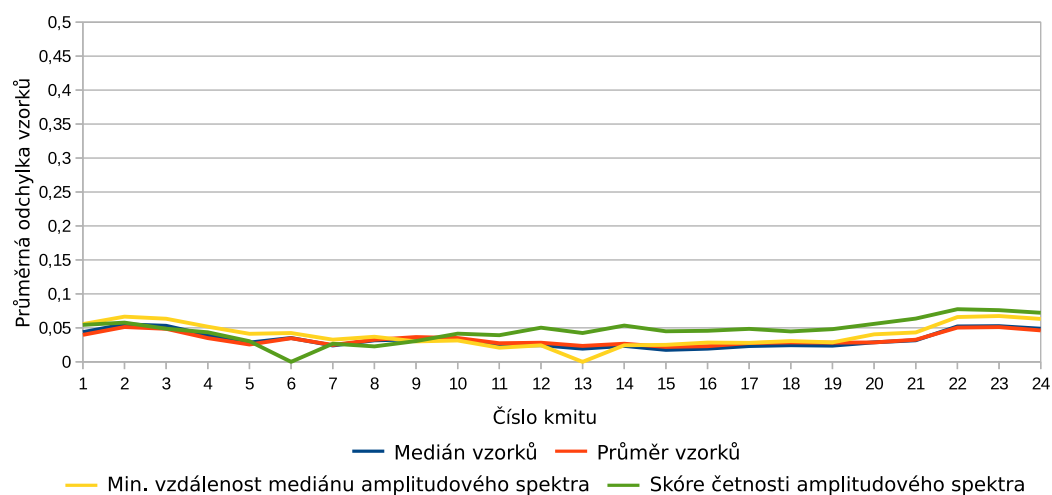
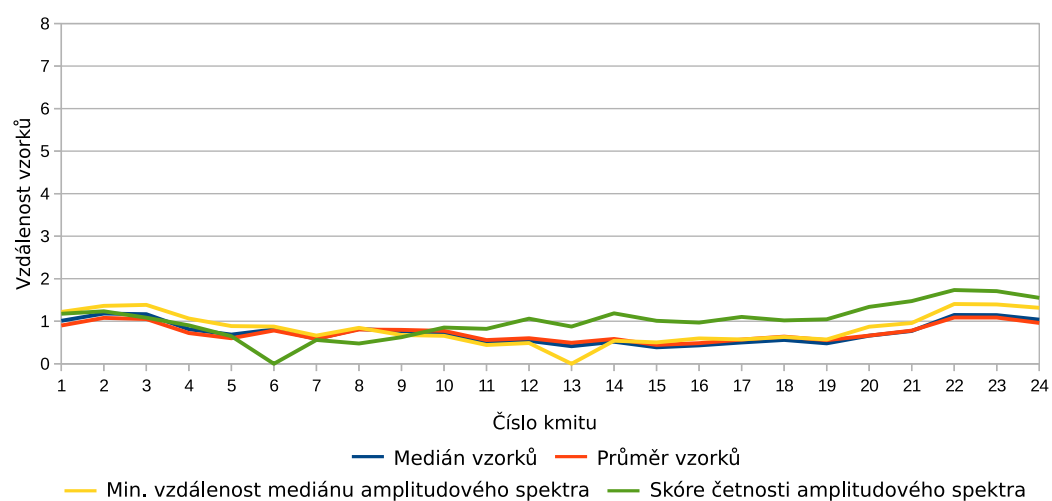
Obrázek 7.38: Zvukový záznam fonace č. 6. Červenou barvou jsou označeny ručně zvolené nestandardní periody.

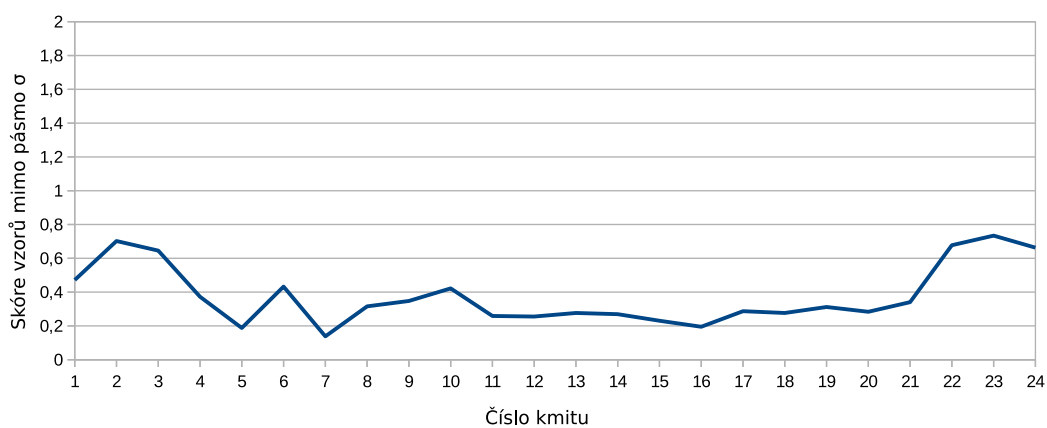
Hlasivky pravidelně kmitají i podle videozáznamu fonace. Hrana hlasivek je bez patologických nálezů a hlasivky ve fázi uzavření doléhají téměř po celé délce.



Obrázek 7.39: Videozáznam fonace. a) Uzavřená hlasivka b) Otevřená hlasivka

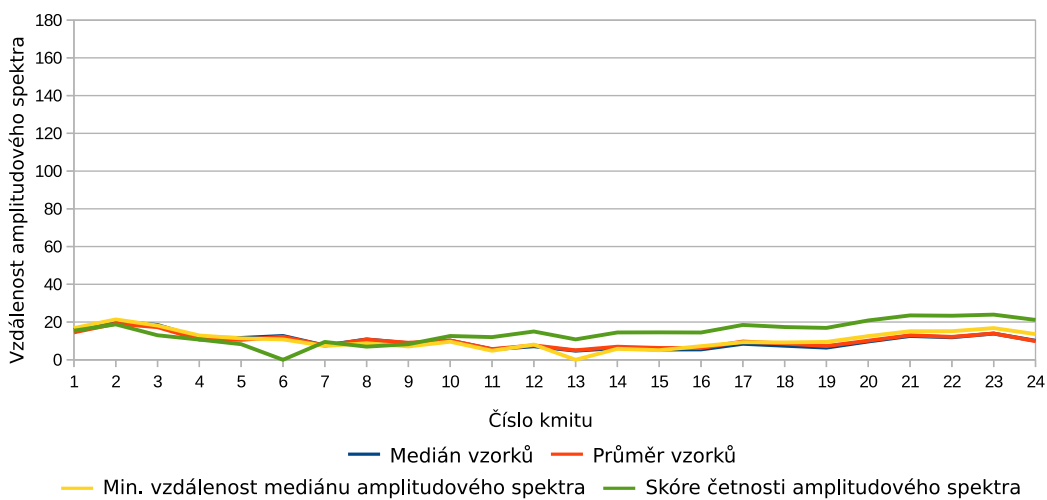
Detekce v časové oblasti

Obrázek 7.40: *Průměrné odchylky vzorků period záznamu č.6.*Obrázek 7.41: *Vzdálenosti vzorků period záznamu č.6.*

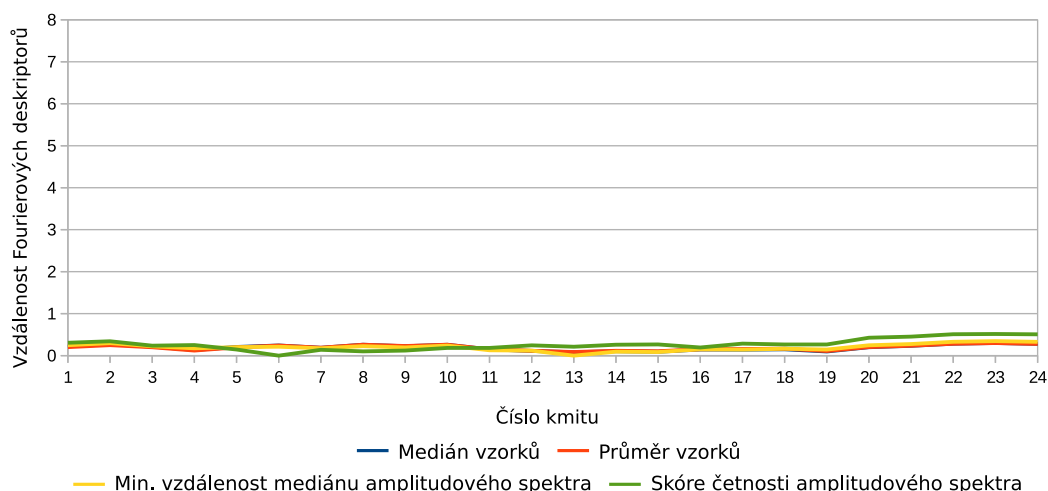


Obrázek 7.42: Skóre vzorků mimo pásmo σ jednotlivých period záznamu č.6.

Detekce ve frekvenční oblasti



Obrázek 7.43: Eukleidovská vzdálenost amplitudového spektra period záznamu č.6.



Obrázek 7.44: Vzdálenost Fourierových deskriptorů period záznamu č.6.

Metody *Průměrná odchylka vzorků periody* a *Vzdálenost vzorků periody* ohodnotily všechny kmity minimální hodnotou odlišnosti. Metoda *Skóre vzorků mimo pásmo σ* je založena na směrodatné odchylce vzorků period, pro takto pravidelnou fonaci je pásmo sigma velmi úzké a hodnocení odlišnosti je tak pouze relativní k ostatním kmitům fonace.

Také hodnocení metod detekce ve frekvenční oblasti je pro všechny kmity téměř shodné a míra odlišnosti je minimální.

7.5.7 Diskuze výsledků metod detekce

Podle výsledků pro jednotlivé záznamy navrhované metody ve všech případech ohodnotily nestandardní kmitu maximálními hodnotami odlišnosti v rámci záznamu. Na testovaných datech se neprojevila závislost výstupu jednotlivých metod detekce na typu hlasové poruchy.

Výsledky pro jednotlivé metody detekce byly téměř ve všech případech shodné, výjimku tvoří převážně metoda *Vzdálenost Fourierových deskriptorů*. Tato metoda je oproti ostatním nezávislá na amplitudě kmitu a odlišnost je tak hodnocena podle poměru hodnot amplitudového spektra.

Pokud je fonace z velké části nepravidelná, je problematické určit reprezentanta standardních kmitů, nejvýrazněji je tento problém patrný u zá-

znamu č. 2 a v menší míře u záznamu č. 5. Pro testované záznamy se jako vhodnější ukázaly metody určení standardní periody ve frekvenční oblasti, které určí jako standardní kmit vhodného reprezentanta z period fonace. Odhad standardního kmitu podle *Průměru vzorků* nebo *Mediánu vzorků* dobře reprezentuje periody fonace pokud existuje jen malé množství nestandardních kmitů. Pokud je fonace zcela nepravidelná, je odhad standardní periody velmi nepřesný.

Na testovaných datech se prokázala schopnost navrhovaných metod detekovat odlišné kmity ve fonaci s alespoň částečně pravidelnými kmity, metody jsou tak vhodné pro detekci náhodně se vyskytujících změn v kinematice hlasivek, které mohou být způsobeny počínajícím růstem útvaru různého původu na hlasivce, ale také například zahleněním.

7.6 Další možné rozšíření práce

Navrhované metody a jejich implementace v realizované aplikaci umožňují zpracovat záznamy vyšetření a určit časové úseky, kde dochází k nestandardnímu kmitání hlasivek podle zvoleného kritéria. Možným rozšířením práce by mohlo být zpracování rozsáhlejšího korpusu dat, na kterém by bylo možné s dostatečnou mírou spolehlivosti určit závislost hlasové poruchy na výstupu jednotlivých metod. Dále by bylo možné navrhnout kritériální funkci, která by podle výsledných hodnot odlišnosti klasifikovala standardní a nestandardní kmity.

Práci by bylo dále možné rozšířit o nové metody detekce nestandardních kmitů a metody určení standardního kmitu, které by vycházely z jiných atributů kmitů a mohly by tak lépe detekovat některé odchylky v záznamu.

Porovnáním výsledků navrhovaných metod s dalšími metodami vyšetření, například s parametry kmitů určenými analýzou vysokorychlostního záznamu kmitání hlasivek, viz [4], by bylo možné přesněji určit původ nestandardního pohybu hlasivek.

8 Závěr

Cílem této práce bylo navrhnout a implementovat systém vyhodnocení parametrů jednoho kmitu hlasivek, podle kterých by bylo možné určit vlastnosti fonace z hlediska pravidelnosti pohybu hlasivek a určit časové okamžiky, ve kterých dochází k nestandardnímu pohybu hlasivek. Tento systém by měl být používán jako součást vyšetření pro včasnou diagnózu hlasových poruch.

V rámci práce byl navržen postup určení jednotlivých period signálu založený na detekci základní frekvence signálu. V další části práce byly navrženy postupy určení referenčního kmitu hlasivek a metody hodnocení odlišnosti period zvukového signálu od periody odpovídající referenčnímu kmitu. Navrhované metody a jejich implementace byly ověřeny na dostupných datech, poskytnutých ORL klinikou FN Plzeň. Výsledky metody na testovaných datech prokázaly funkčnost navrhovaných postupů.

Realizovaný program umožňuje zpracovat audio, video a EGG záznamy fonace, detekovat jednotlivé kmity a vypočítat jejich parametry. Volbou referenčního kmitu lze hodnotit odlišnost kmitů a určit části záznamu, kde dochází k změně kinematiky hlasivek. Program je vhodně navržen, aby bylo možné doplnit další metody hodnocení odlišnosti kmitů, volby referenčního kmitu nebo další doplňující funkce. Aplikace dále umožňuje podrobně analyzovat videozáznam vysokorychlostní kamery s využitím hodnot odlišnosti kmitů a spravovat záznamy vyšetření ve vlastní databázi s možností pasivního přístupu k databázi systému HSV, viz tištěná příloha C.

Literatura

- [1] Understanding Voice Production [online], 2015. Dostupné z: <http://voicefoundation.org/health-science/voice-disorders/anatomy-physiology-of-voice-production/understanding-voice-production/>.
- [2] CUADRA, P. – MASTER, A. – SAPP, C. Efficient Pitch Detection Techniques for Interactive Music. In *Proceedings of the International Computer Music Conference*, 2001.
- [3] DLOUHÁ, O. – ČERNÝ, L. *Foniatricie*. Praha : Karolinum, 2012. ISBN 978-80-246-2048-0.
- [4] ETTLER, T. Analýza vysokorychlostního záznamu kmitání hlasivek, 2012. Diplomová práce, Západočeská univerzita v Plzni.
- [5] GRAY, H. *Anatomy of the human body*. Philadelphia : Lea & Febiger, 1918. ISBN 1-58734-102-6.
- [6] HERBST, C. Evaluation of Various Methods to Calculate the EGG Contact Quotient, 2004. Diploma Thesis, Kungliga Tekniska Högskolan.
- [7] KASTNER, M. J. Videokymografie a digitální kymografie, kymografické nálezy u pacientů po thyreoidektomii. *Endoskopie*. ISSN 1211-1074.
- [8] *Multi-Dimensional Voice Program, Model 5105 [online]*. KayPENTAX, 2015. Dostupné z: [http://www.kayelemetrics.com/index.php?option=com_product&view=product&Itemid=3&controller=product&cid\[\]=56&task=pro_details](http://www.kayelemetrics.com/index.php?option=com_product&view=product&Itemid=3&controller=product&cid[]=56&task=pro_details).
- [9] KUČERA, M. M. – FRÍČ, R. M. – HALÍŘ, M. M. *Praktický kurz hlasové rehabilitace a reedukace*. 2010. ISBN 978-8025465929.
- [10] MARKUZZIOVÁ, H. Popisy hranice objektů a jejich využití pro klasifikaci, 2008. Diplomová práce, Západočeská univerzita v Plzni.

-
- [11] NOVÝ, P. – VÁVRA, F. – KOTLÍKOVÁ, M. Voice Range Profile Examination Method and its Applications. In *Summer School DATASTAT 03*, 2003. ISBN 80-210-3564-1.
- [12] PEŠTA, J. et al. Hodnocení kvality závěru glottis. *Otorinolaryngologie a foniatrie*. 2010. ISSN 1210-7867.
- [13] PROF. MUDR. ALEXEJ NOVÁK, D. *Foniatrie a pedaudiologie II*. Praha : UNITISK, 1996.
- [14] PROF. MUDR. STANISLAV TICHÝ, D. a. k. *Otorinolaryngologie, Učebnice pro lékařské fakulty*. Avicenum, 1983. ISBN ČSR čj. 9.182/82-82-31.
- [15] BERG, J. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*. 1958. ISSN 1092-4388.
- [16] ŠLAPÁK, I. – JANEČEK, D. – LAVIČKA, L. *Základy otorinolaryngologie a foniatrie pro studenty speciální pedagogiky*. 2009.
- [17] ŠVEC, J. G. Základní metody pro vyšetření a analýzu hlasu, 2014.
- [18] ŠVEC, J. G. – ŠRAM, F. – SCHUTTE, H. Videokymografie: nová vysokofrekvenční metoda vyšetřování kmitů hlasivek. *Otorinolaryngologie a foniatrie*. 1999. ISSN 1210-7867.

A Adresářová struktura DVD

VFO Analyzer - Aplikace pro analýzu záznamu fonace realizovaná v rámci této práce.

BLD dekodér - Doplnující aplikace, umožňující prohlížet a převádět BLD soubory do formátu AVI nebo série snímků PNG.

Podrobné výsledky - Soubory s podrobnými výsledky postupů uvedených v této práci.

detekce f0 - Porovnání výstupu metod detekce F_0 .

detekce period - Parametry detekovaných period testovaných záznamů.

lpc-analýza - Příklad výstupu LPC analýzy pro určení polohy formantů a fonovaného vokálu.

score - Výpočet hodnocení kvality závěru glottis metodou SCORE.

kazuistika 1 - 6 - Výstup metod detekce nestandardních kmitů pro záznamy uvedené v diplomové práci.

Ukázkové soubory vyšetření - Záznamy vyšetření pro demonstraci funkcí realizované aplikace.

diplomová práce.pdf - Text diplomové práce ve formátu PDF.

B Uživatelský manuál

Implementace

Aplikace vytvořená v rámci této práce je implementována v jazyce C# a využívá knihovny .NET verze 4.0. Databáze pacientů je založena na Microsoft SQL Server Compact. Podporované operační systémy jsou Microsoft Windows XP a novější.

Hlavní okno programu

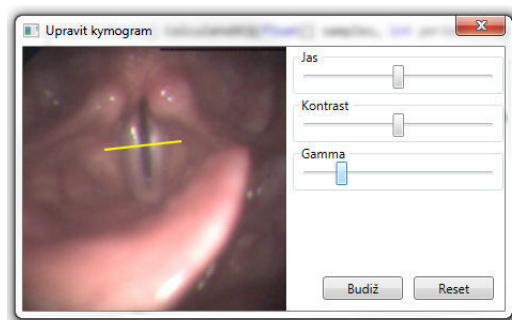
Hlavní okno aplikace tvoří zobrazení zvukového signálu, signálu EGG, derivace EGG signálu a videokymogram. Kompletní vyšetření lze otevřít volbou *Otevřít vyšetření* v nabídce *Soubor*, případně otevřením vyšetření z databáze. Aplikace umožňuje také otevřít pouze EGG nebo pouze zvukový záznam odpovídajícími volbami v nabídce *Soubor*.

Po načtení zvoleného vyšetření jsou vykresleny signály a rozdělení signálů na periody, viz obrázek B.1. Části záznamů je možné označit značkami, levým tlačítkem myši červenou a pravým modrou značkou. Pohybem myši nad signálem je vykreslen kurzor s číslem vzorku, nad kterým se kurzor nachází.



Obrázek B.1: Hlavní okno aplikace s vyšetřením.

Zobrazení signálů lze upravit možnostmi v panelu nástrojů. Tlačítko *Upravit kymogram* zobrazí okno s možností zvolit řez ve snímku videa, z kterého bude složen videokymogram. Dále je možné provést jednoduché jasové úpravy pro zvýraznění hlasivkové štěrbin v kymogramu. Okno úprav kymogramu je uvedeno na obrázku B.2.

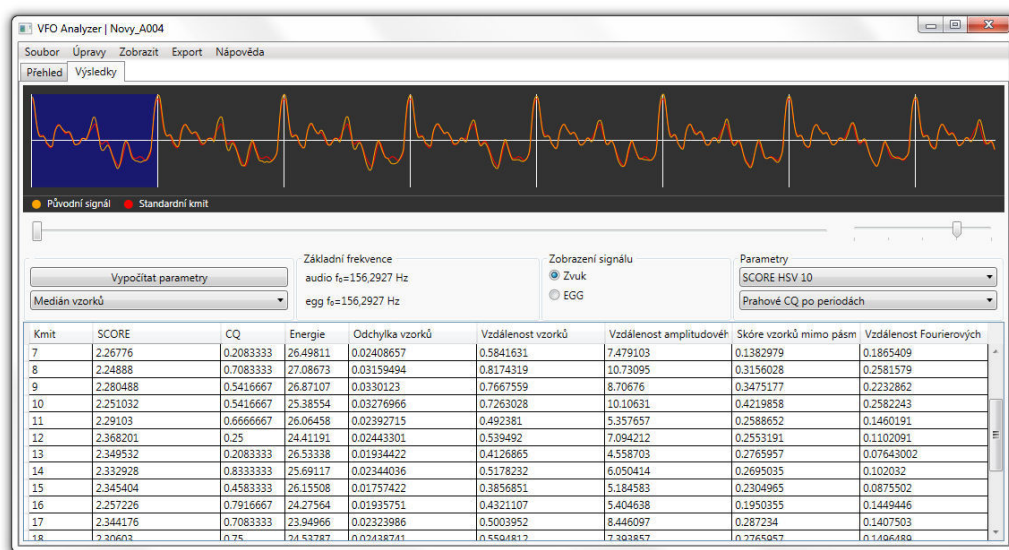


Obrázek B.2: Okno úprav kymogramu.

Tlačítka v panelu nástrojů je dále možné invertovat polaritu audio signálu a posunout audio signál vzhledem k EGG, aby bylo odstraněno zpoždění způsobené vzdáleností mikrofону. Zpoždění zvukového signálu lze odstranit automaticky, nebo zadáním vzdálenosti mikrofónu od hlasivek v milimetrech.

V nabídce úpravy je možné ručně nastavit délku periody audio a EGG signálu v případě selhání automatické detekce základní frekvence. Pokud selže automatická detekce period kmitů, je možné zvolit počátek první periody, z které bude algoritmus automatické detekce vycházet. V menu *Úpravy* je dále možné filtrovat audio a EGG signál pásmovou propustí.

Po úpravě signálů lze určit parametry jednotlivých kmitů v panelu *Výsledky*. Panel výsledky je tvořen třemi částmi, zobrazením signálu, panelem voleb a tabulkou parametrů. Panel *Výsledky* je uveden na obrázku B.3. Zobrazení signálu umožňuje porovnat periody signálu se standardní periodou určenou podle zvolené metody.



Obrázek B.3: Panel s parametry jednotlivých kmitů.

V panelu nástrojů lze zvolit metody výpočtu vybraných parametrů a v tabulce parametrů tyto hodnoty porovnat a určit oblast signálu, kde došlo ke změně. Orientaci v periodách signálu usnadňuje zvýraznění periody odpovídající vybrané řádce tabulky parametrů. Odlišné periody lze podrobně analyzovat zobrazením přehrávače HSV videozáznamu, viz B.4. Přehrávač umožňuje procházet celý záznam, nebo sledovat smyčku pro jednotlivé kmitů. Záznam je možné krokovat po snímcích, nebo automaticky přehrávat. Kmitů jejichž odlišnost překračuje pásmo směrodatné odchylky sigma pro danou metodu jsou barevně zvýrazněny.

Databáze pacientů

Další částí aplikace je databáze pacientů a vyšetření. Databázi pacientů lze zobrazit volbou *Otevřít databázi* v menu *Soubor* v hlavním okně programu. Záznamy lze do databáze automaticky importovat zvolením adresáře se soubory z HSV vyšetření, aplikace automaticky zpracuje informace o pacientech a jednotlivých vyšetřeních. Další možností je ruční vyplnění karty pacienta. Databáze pacientů je uvedena na obrázku B.5 a karta pacienta na obrázku B.6. Dvojitým poklepáním na záznam vyšetření v kartě pacienta je vyšetření načteno do hlavního okna aplikace.



Obrázek B.4: Okno prohlížeče videozáznamu.

The screenshot shows a window titled "Databáze pacientů". It contains a table with four columns: "Jméno", "Příjmení", "Datum narození", and "Počet záznamů". The table lists nine patients, each with a first letter of the first name (a-i) and a fixed birth date of 1/1/1800 12:00:00 AM, and a count of 1 record. Below the table is a large empty rectangular area.

Jméno	Příjmení	Datum narození	Počet záznamů
	a	1/1/1800 12:00:00 AM	1
	b	1/1/1800 12:00:00 AM	1
	c	1/1/1800 12:00:00 AM	1
	d	1/1/1800 12:00:00 AM	1
	e	1/1/1800 12:00:00 AM	1
	f	1/1/1800 12:00:00 AM	1
	g	1/1/1800 12:00:00 AM	1
	h	1/1/1800 12:00:00 AM	1
	i	1/1/1800 12:00:00 AM	1

Obrázek B.5: Databáze pacientů

Osobní údaje

Osobní údaje

Jméno: Lukas Datum narození: 23.10.1990

Příjmení: Kroupa Datum registrace: 13.2.2015

Pohlaví: ☒ Muž ☐ Žena Rodné číslo: 0

Kód pojišťovny: 0

Poznámky

Vyšetření

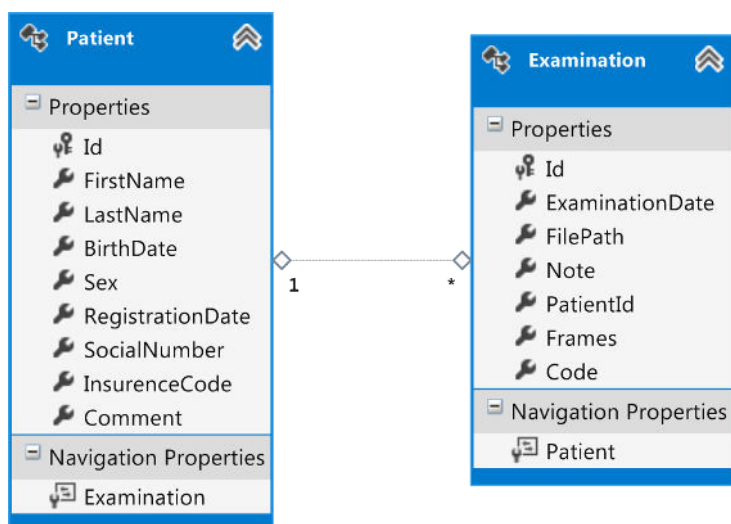
Datum záznamu	Kód	Poznámka	Počet snímků	Cesta
3/18/2015 1:36:04 PM	Kroupa_A004		642	D:\diplomka_data\Kroupa_

Přidat vyšetření Odstranit vyšetření Budiž Zrušit

Obrázek B.6: Okno karty pacienta.

C Struktura databáze pacientů a vyšetření

Součástí aplikace je databáze pro správu pacientů a vyšetření. Struktura databáze odpovídá adresářové struktuře záznamů vyšetření HSV systému a tuto strukturu lze pro zjednodušení přístupu k záznamům vyšetření do databáze importovat. Do souborů vyšetření HSV systému nejsou aplikací ukládána žádná data, přístup aplikace je pouze pro čtení.



Obrázek C.1: Relační databázový model aplikace.

Pro každého pacienta je kromě jeho osobních údajů uložen krátký textový popis diagnózy nebo předchozí léčby. Záznamy vyšetření obsahují datum vyšetření, textovou poznámku a cestu k souboru se záznamem.

D Dekódování souboru BLD

Tyto videozáznamy jsou ukládány softwarem dodávaným ke kameře do souborů s příponou *bld*. Tento formát souborů je proprietární formát výrobce softwaru a není standardizovaný. Výrobce softwaru neposkytuje dokumentaci, či zdrojové kódy dekodéru nebo kodéru tohoto formátu, jediná možnost jak tento soubor dekodovat, zobrazit nebo převést do standardizovaného formátu je pomocí dodávaného softwaru.

Tento postup je ale pro potřeby vyvíjeného softwaru nepoužitelný kvůli časové náročnosti procesu, velkému počtu potřebných videozáznamů a omezenému přístupu k videokameře a obslužnému softwaru. Vyvíjené aplikace musí pracovat přímo s *bld* soubory a je proto nutné popsat strukturu souboru a implementovat dekodér, který bude poskytovat výstup srovnatelný s dekodérem dodávaným k vysokorychlostní kameře.

Z důvodu nedostupnosti dokumentace formátu *bld* je pro implementaci dekodéru nutné provést zpětnou analýzu. Zpětná analýza, nebo též *reverzní inženýrství* je definováno jako *proces analýzy předmětného systému s cílem identifikovat komponenty systému a jejich vzájemné vazby a/nebo vytvořit reprezentaci systému v jiné formě nebo na vyšší úrovni abstrakce*¹. Tento postup je často využíván nejen v oblasti výpočetní techniky. Hlavní cíle zpětné analýzy jsou kromě proprietárních formátů také ovladače hardwaru a síťové protokoly. Výrobci těchto proprietárních systémů často znemožňují, či ztěžují zpětnou analýzu šifrováním, obfuskací kódu a dalšími prostředky.

Z právního hlediska je reverzní inženýrství legální s výjimkou, kdy je nutné odstranit, či jinak obejít ochranné prvky zkoumaného systému, nebo v případě, kdy je zpětná analýza zakázána v rámci smluvních podmínek.

¹Údržba softwaru, Jiří Sochor, FI MU, <http://webserver.ics.muni.cz/zpravodaj/articles/61.html>

D.1 Analýza souboru bld

Každý *bld* soubor vytvořený proprietárním softwarem výrobce je doplněn textovým souborem s příponou *par*. Tento soubor obsahuje popis některých vlastností *bld* formátu a obsahu souboru, například popis barevné hloubky, rozměr snímku, snímkovací frekvenci a celkový počet snímků v souboru. Příklad *par* souboru je uveden níže:

```
CAMERATYPE: HRES_ENDOCAM_COLOR
DIGITALGAIN: 1
COLORSPACE: RGB24
RESOLUTION: 256x256
REDUCTION: RT_CHECKERED
FIRSTFRAME: 6452
FRAMES: 504
FRAMES/S: 4000
DATE: 2012-Apr-03 11:41:45
FIRSTSAMPLE: Blue
RGB_MATRIX: MAT3x3 256,0,0,0,256,0,0,0,256
```

Tyto informace odpovídají obsahu hlavičky běžně používaných obrazových a video formátů, prohlížečem binárních souborů bylo ověřeno, že formát *bld* hlavičku souboru neobsahuje.

Informace ze souboru *par* poskytují základ pro analýzu souboru *bld*. Podle uvedené barevné hloubky, rozměrů snímku a počtu snímků je možné vypočítat počet bajtů jednoho snímku za předpokladu, že počet bajtů snímků je konstantní. Barevný prostor je ve formátu **RGB24**, tedy každá barevná složka je vyjádřena osmi bity. Podle uvedených informací jsou předpokládány následující vztahy pro počet bajtů jednoho snímku:

$$RESOLUTION \cdot COLORSPACE = \frac{velikost\ souboru}{FRAMES}$$

Pokud dosadíme hodnoty z vzorového *par* souboru a doplníme, že počet bajtů odpovídajícího *bld* souboru je 16515072, získáme následující hodnoty:

$$256 \cdot 256 \cdot 3 = \frac{16515072}{504}$$

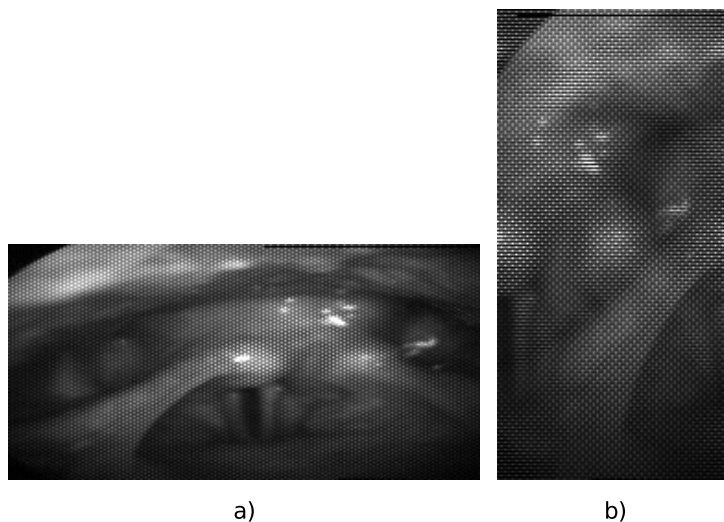
$$196608 \neq 32768$$

Získané hodnoty nejsou v rovnosti, jednotlivé snímky tedy nejsou uloženy jako obraz s rozlišením 256×256 s barevnou hloubkou 24 bitů.

Hodnota podílu počtu snímků a velikosti souboru je celé číslo a zároveň mocnina dvou, což naznačuje, že jednotlivé snímky jsou uloženy vždy v 32 768 bajtech. Skutečná velikost souboru je menší než očekávaná, jednotlivé snímky jsou zřejmě komprimovány a pro správné vykreslení snímku bude nutné určit význam jednotlivých bajtů.

D.2 Dekódování souboru bld

Analýzou souboru byla zjištěna velikost snímku 32 768 bajtů, faktorizací na součin dvou čísel, tedy horizontální a vertikální rozměr snímku, je toto číslo možné vyjádřit jako 256×128 , případně 128×256 . Vykreslením prvních 32 768 bajtů ve formě rastrového obrazu ve stupních šedi s barevnou hloubkou 8 bitů získáme výstup uvedený na obrázku D.1.



a)

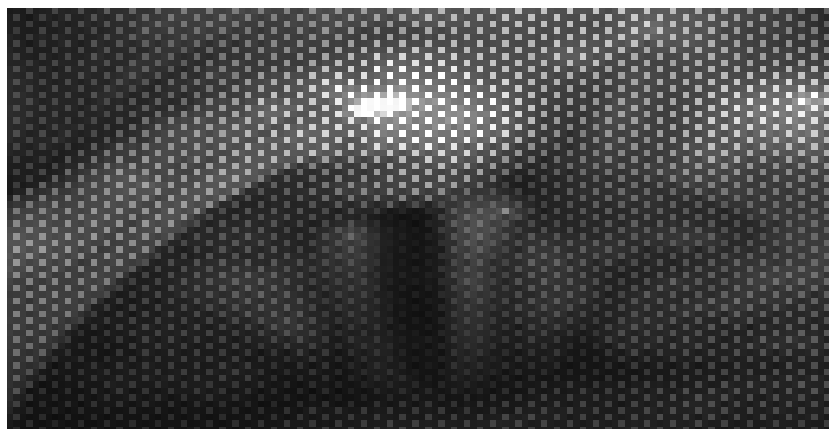
b)

Obrázek D.1: Zobrazení prvních 32 768 bajtů jako rastrový obraz s rozměry:

a) 256×128

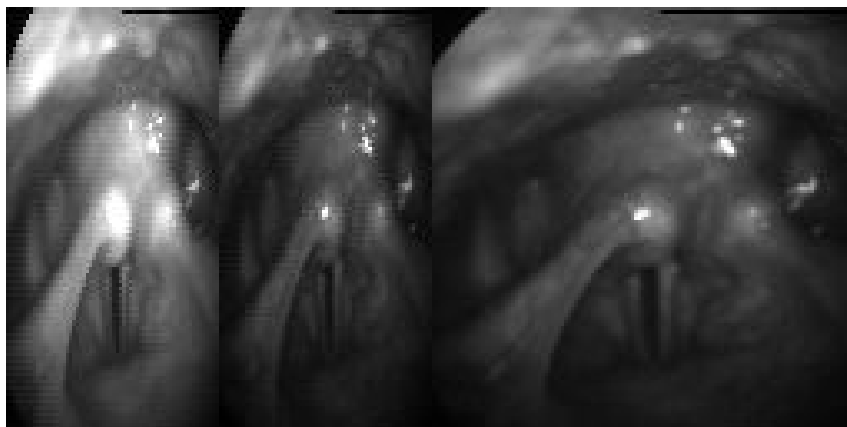
b) 128×256

Reprezentace dat v podobě obrazu s rozměry 256×128 obsahuje hledanou scénu. Při bližším pohledu je patrné, že obrazové body tvoří opakující se vzor (viz obrázek D.2).



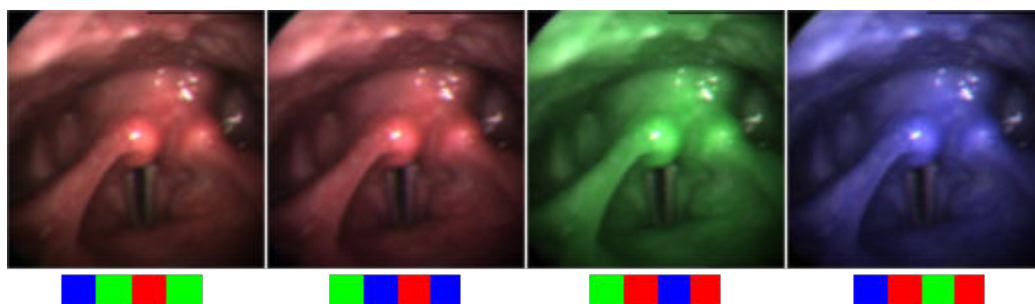
Obrázek D.2: Detailní pohled na opakující se vzor ve snímku.

Rozdělením obrazových bodů na liché a sudé a poté rozdělením první skupiny opět na liché a sudé získáme tři skupiny bodů se shodným jasnem scény. Zobrazení bodů z jednotlivých skupin pohromadě ve formě rastrového obrazu je uvedeno na obrázku D.3. Posun o jeden bod v každém sudém řádku v prvních dvou částech snímku je způsoben střídavým uložením bodů z těchto skupin v surových datech.



Obrázek D.3: Zobrazení bodů obrazu podle barevných kanálů.

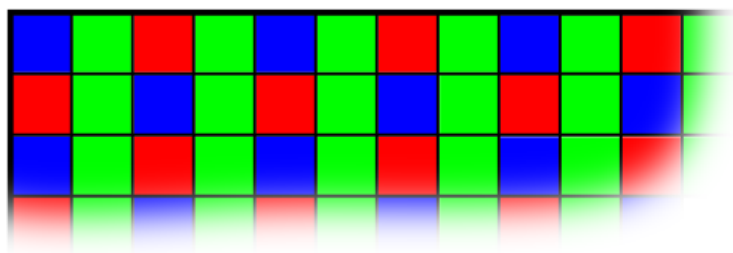
Podle jasů jednotlivých částí obrázku D.3 lze odvodit, že jednotlivé části reprezentují barevné kanály. První část má nejvyšší jas a zřejmě obsahuje hodnoty pro červený kanál, jelikož výsledný obraz složený ze všech kanálů se očekává zbarvený do červena. Pořadí barevných složek lze ověřit vykreslením snímku v 24 bitové barevné hloubce s použitím jednotlivých částí jako barevných kanálů. Dvě menší části byly pro vykreslení horizontálně roztaženy na dvojnásobnou velikost.



Obrázek D.4: Vykreslení snímku s různou reprezentací barevných kanálů.

Správný význam bajtů ve snímku zřejmě odpovídá prvnímu či druhému případu uvedenému na obrázku D.4. Modrá a zelená složka má téměř shodný jas a správné pořadí nemusí být zřejmé, ale podle informace ze souboru *par* reprezentuje první bajt snímku modrý kanál.

Význam jednotlivých bajtů snímku tedy odpovídá schématu na obrázku D.5. Každý druhý bajt odpovídá zelenému kanálu, zbývající bajty reprezentují střídavě modrý a červený kanál.



Obrázek D.5: Význam bajtů ve snímku při zobrazení 256×128 .

D.2.1 Rekonstrukce snímku v rozlišení 256×256

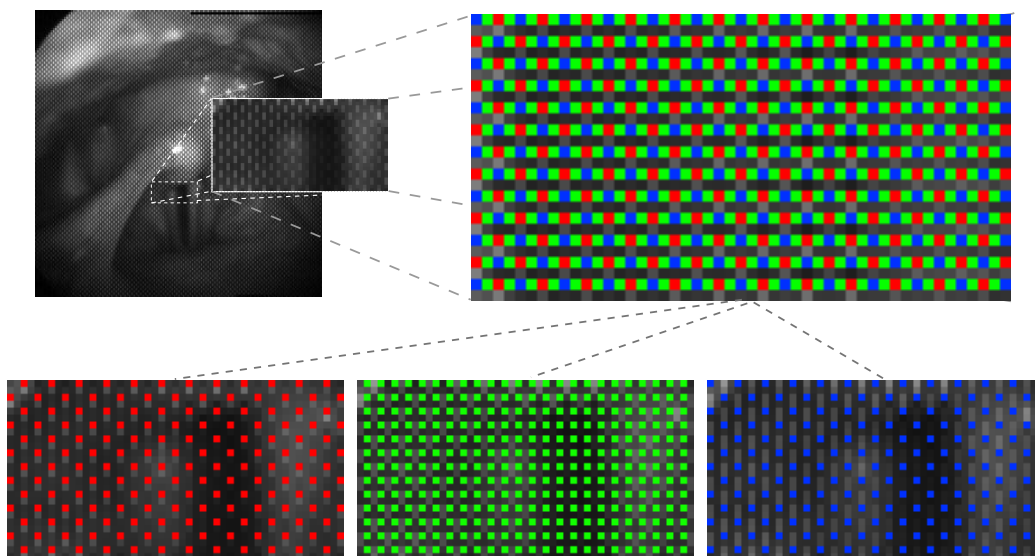
Postupem uvedeným výše je možné dekodovat barevné snímky v rozlišení 128×128 bodů, kde ovšem červená a modrá barevná složka je horizontálně podvzorkována 2:1. Podle popisu v *par* souboru a dle výstupu z proprietárního softwaru jsou rozměry dekódovaného snímku 256×256 .

Pro zvětšení snímku na dvojnásobek nelze použít prostou bikubickou interpolaci, kvůli posunutí modrého a červeného kanálu v sudých řádcích. Při interpolaci je tato vada znásobena, viz obrázek D.6.



Obrázek D.6: Snímek převzorkovaný na rozlišení 256×256 .

Pro nalezení vhodné metody, jak škálovat červenou a modrou barevnou složku v rozlišení 64×128 a zelenou barevnou složku v rozlišení 128×128 na obrázek v rozlišení 256×256 je vhodné uvést, jak jednotlivé bajty ze surových dat odpovídají výslednému snímku. Na obrázku D.7 jsou v detailu znázorněny bajty reprezentující jas bodů barevných složek ve vztahu k výslednému snímku. Z detailního zobrazení je zřejmé, pro kolik okolních bodů musí být hodnota jednoho bodu barevné složky interpolována.



Obrázek D.7: Zobrazení bajtů a jejich barevné reprezentace na výsledný snímek v rozlišení 256×256 .

Interpolace maskou

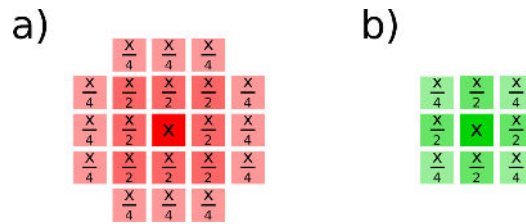
Tato metoda interpolace vypočítává hodnoty jasu bodů v okolí pomocí masky. Rozměry masky jsou pro jednotlivé barevné kanály různé.

Interpolační maska modré a červené složky musí sahát do vzdálenosti 2 obrazových bodů v přímém okolí a v nepřímém okolí do vzdálenosti jednoho bodu. Koeficienty hodnoty jasu pro body masky v osmiokolí jsou 0,5 a 0,25 pro body ve vzdálenosti 1. Zelená složka má dvojnásobné rozlišení a její interpolační maska bude ve tvaru osmiokolí. Koeficient hodnoty jasu pro body v přímém okolí je 0,5 a pro body v nepřímém okolí 0,25.

Tvar a koeficienty jasu obou masek jsou zvoleny tak, aby se při interpolaci okrajové body masky překrývaly a součet koeficientů v každém bodě byl roven 1. Masky jsou uvedeny na obrázku D.8.

Kombinovaná interpolace

Kombinovaná metoda využívá minimální masky ve tvaru čtyřkolí pro interpolaci modré a červené složky na rozměr 128×128 . Všechny barevné složky



Obrázek D.8: Interpolační maska a) červeného a modrého kanálu b) zeleného kanálu.

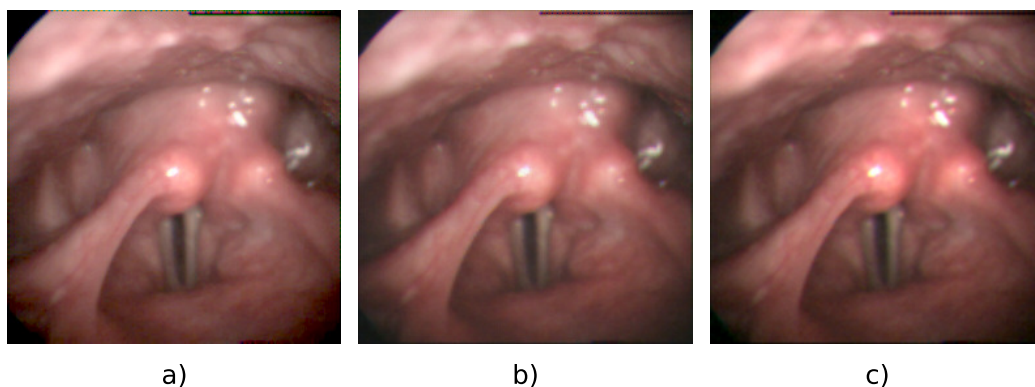
jsou poté bikubickou interpolací škálovány na výsledný snímek v rozlišení 256×256 . Bikubická interpolace při výpočtu vychází z 4×4 bodů oproti 2×2 bodům bilineární interpolace a produkuje jemnější obraz bez výrazných artefaktů.

D.3 Hodnocení výstupu

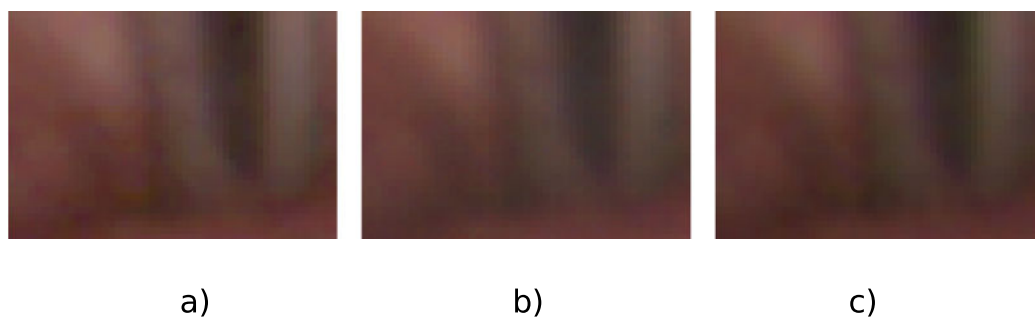
Oba navrhované postupy dekodování produkují použitelný výstup. Výstup získaný interpolací maskou ale obsahuje interpolační artefakty o velikosti masky, které jsou dále zvýrazněny jasovými úpravami.

Výstup kombinované metody netrpí žádnými výraznými nedostatky v porovnání s výstupem softwaru výrobce. Výstup obou metod je spolu s výstupem proprietárního softwaru dodávaného výrobcem uveden na obrázku D.9 a dále v detailu na obr. D.10.

Obě metody stejně jako metoda použitá v softwaru výrobce kamery jsou interpolovány ze surových dat v souboru *bld*. Porovnání úspěšnosti dekodování je tak pouze relativní, výstup nelze porovnat s nekomprimovanými snímky v plném rozlišení. Výstup by bylo dále možné upravit, například doostřit použitím masky ostroty, ale takto získaný obraz je vhodný pouze pro vizuální kontrolu. Podobné úpravy pouze mění až znehodnocují zaznamenanou informaci a pro další zpracování jsou nevhodné.



Obrázek D.9: a) Proprietární software b) Interpolace maskou c) Kombinovaná interpolace



Obrázek D.10: a) Proprietární software b) Interpolace maskou c) Kombinovaná interpolace

Analýzou souboru *bld* byla zjištěna struktura souboru, formát jednoho snímku a interpretace jednotlivých bajtů. Ze získaných informací vyplývá, že *bld* soubor je ztrátově komprimovaný, ale jedná se spíše o záznam surových dat ze snímače. Data odpovídají snímkům v rozlišení 128×128 bodů, tedy polovině rozlišení udávaného v dokumentaci dodávané výrobcem.

Pro získání snímků v uváděném rozlišení byly navrženy metody interpolace, které poskytují výstup srovnatelný s dodávaným softwarem. Výstup je použitelný pro účely uvedené v zadání práce.