T A
Č R

# VKG 3.0 Software

Příloha k závěrečné zprávě

Číslo projektu: TH04010422

Název projektu: **VKG 3.0**

Předkládá: B. Zitová

**Název organizace**: UTIA AV ČR v.v.i.

**Hlavní řešitel**:

Ústav teorie informace a automatizace AV ČR, v.v.i.

**Spoluřešitelé**:

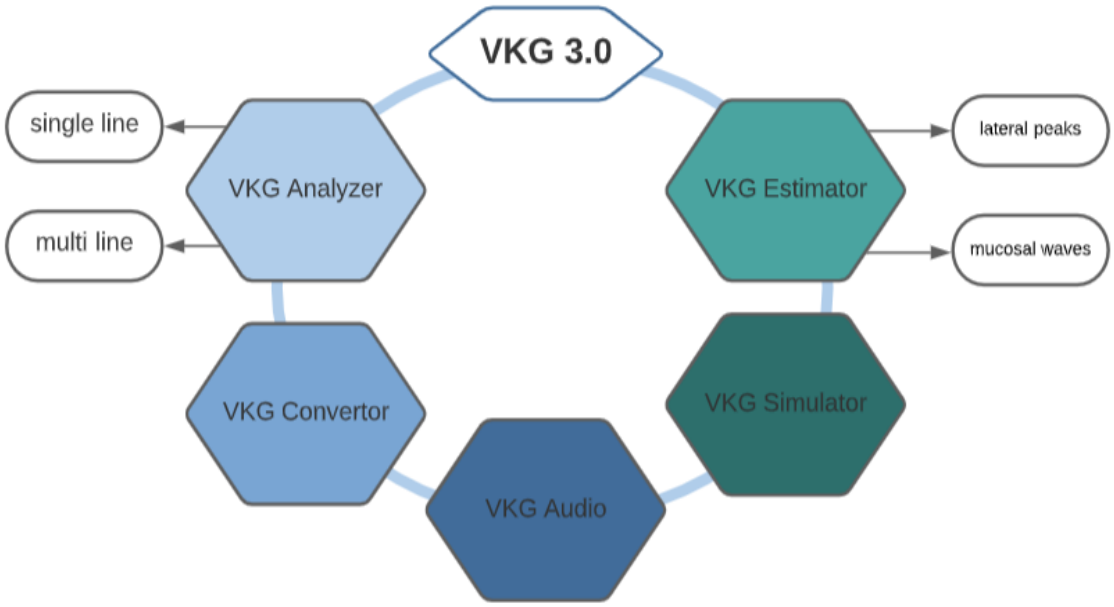
Medical Healthcom spol.s r.o.

ATE system s.r.o.

## 1. VKG 3.0 software

V roce 2021 byl dokončen vývoj software v rámci projektu VKG3.0. V následujících sekcích představíme funkcionalitu jednotlivých modulů. Modulové řešení je na http://zoi.utia.cas.cz/VKG3_0. Výsledná implementace se sestává z 5 navzájem se doplňujících se modulů, které umožňují komplexní analýzu a pořízených videokymografických dat:

- **VKG Analyzer** - analýza videokymogramů na jednom či více řádků s vizualizací výsledků
- **VKG Audio** - funkcionalita svázaná s analýzou audio informace
- **VKG Simulator** – simulace videokymogramů
- **VKG Convertor** - konverze dat mezi jednotlivým typy videokymografických kamer
- **VKG Estimator** - automatická detekce ostrosti laterálního vrcholu a odhadu délky mukosální vlny. Tento modul považujeme za nejdůležitější výsledek projektu VKG 3.0.

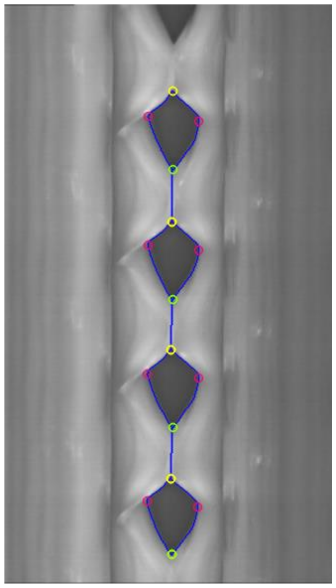


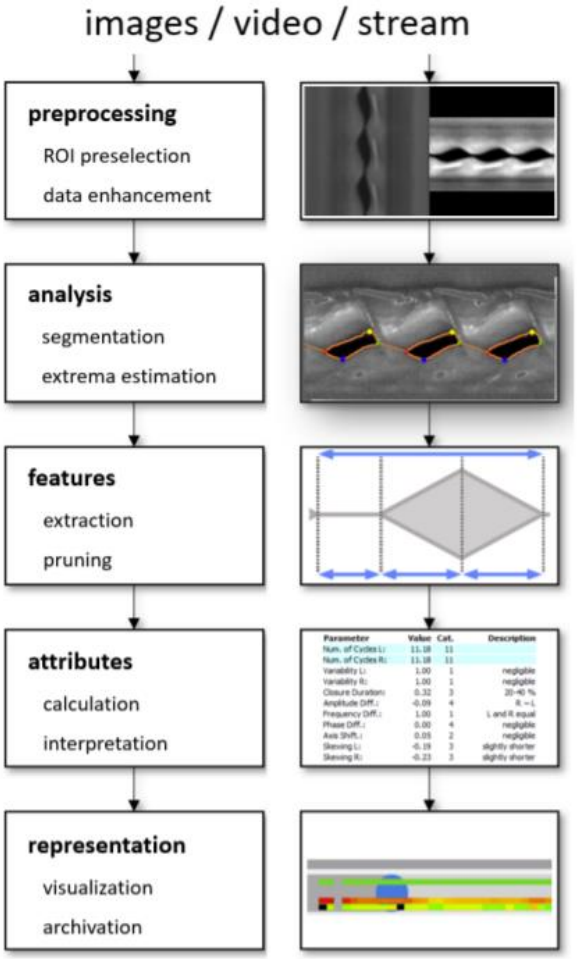

*Obr.1. VKG 3.0 software s příslušnými moduly – VKG Analyzer - analýza videokymogramů na jednom či více řádků s vizualizací výsledků. VKG Audio - funkcionalita svázaná s analýzou audio informace. VKG Simulator – simulace videokymogramů. VKG Convertor - konverze dat mezi jednotlivým typy videokymografických kamer a VKG Estimator - automatick detekce ostrosti laterálního vrcholu a odhadu délky mukosální vlny.*

# VKG 3.0 Analyzer

Tento modul nabízí detekci význačných struktur v pořízených videokymogramech a vyhodnocuje návazné příznaky používané pro vyhodnocování kymogramů. O jaké příznaky se jedná a jakým způsobem jsou detekované je blíže popsáno v našem článku Videokymogram Analyzer Tool: Human-computer comparison [1].

| advanced features | notation and definition |
|---|---|
| generalized opening points | $\tilde{O}_i^j = \{O_i, M_i^j\}$ |
| generalized closing points | $\tilde{C}_i^j = \{C_i, M_i^j\}$ |
| opening phase duration | $t_i^{oj} = A_i^j(y) - \tilde{O}_i^j(y)$ |
| closing phase duration | $t_i^{cj} = \tilde{C}_i^j(y) - A_i^j(y)$ |
| open phase duration | $T_i^{oj} = t_i^{oj} + t_i^{cj} = \tilde{C}_i^j(y) - \tilde{O}_i^j(y)$ |
| closed phase duration | $T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{C}_i^j(y)$ |
| vibration cycle duration | $T_i^j = T_i^{oj} + T_i^{cj} = t_i^{oj} + t_i^{cj} + T_i^{cj} = \tilde{O}_{i+1}^j(y) - \tilde{O}_i^j(y)$ |
| vocal fold amplitudes | $a_i^j = \mathrm{mean}(\lvert A_i^j(x) - \tilde{O}_i^j(x)\rvert, \lvert A_i^j(x) - \tilde{C}_i^j(x)\rvert)$ |
| glottal amplitudes | $a_i = A_i^L(x) - A_i^R(x)$ |

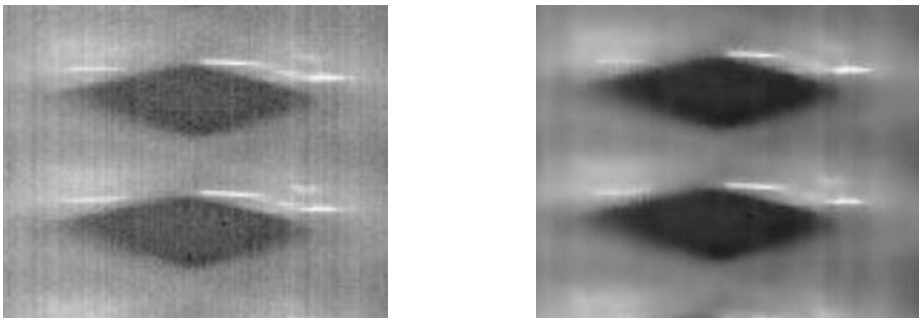





*Obr.2. Ukázka příznaků a detekovaných význačných bodů na videokymogramu (vlevo) a vývojový diagram základních funkcí (vpravo)*

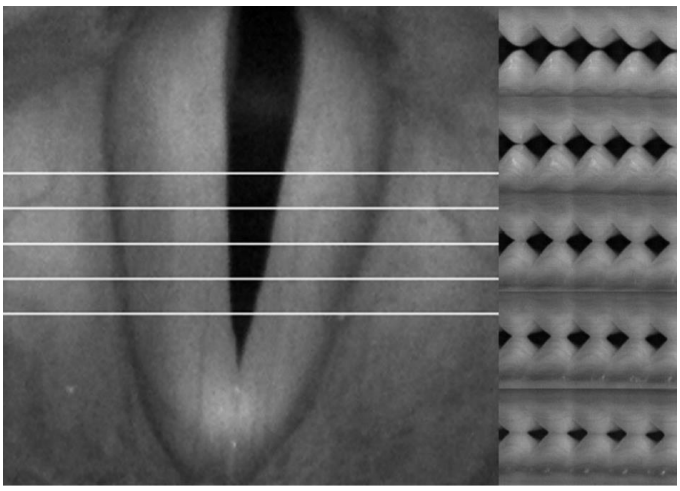
V rámci projektu VKG 3.0 došlo k rozšíření software o nástroje pro analýzu dat z více řádků najednou a hlavně pro jejich efektivní vizualizaci. Nová verze programu je schopna and rámec detekce význačných bodů a odhadu parametrů (viz. článek a *Obr.2*)

- Funkce **preprocessingu** pro vylepšení kvality dat, neboť víceřádkové kamery mají nižší kvalitu pořizovaných dat z důvodu rychlosti snímání



*Obr.3. Ukázka předzpracování VKG dat (vlevo originál, vpravo po vyhlazení)*

- Funkce zobrazování vstupu z víceřádkové kamery

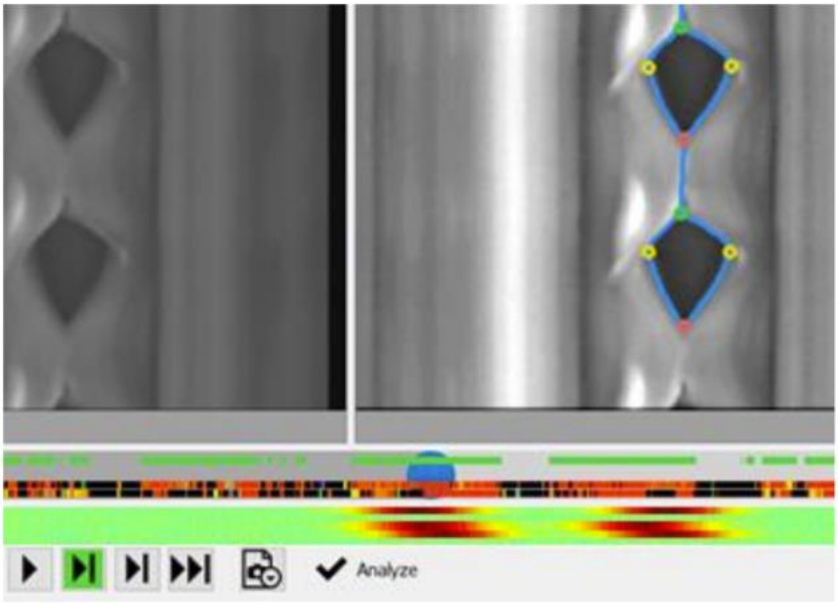


*Obr.4. Ukázka předzpracování VKG dat (vlevo originál, vpravo po vyhlazení)*

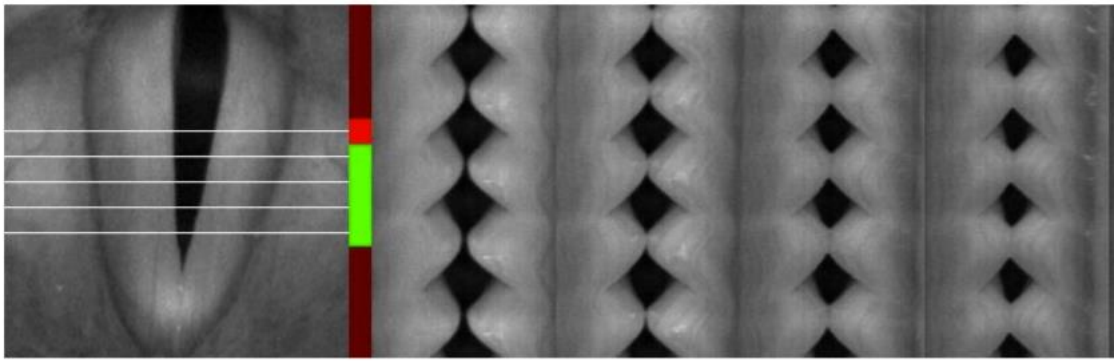
- Funkce zobrazení napočítaných parametrů pro víceřádkovou kameru (zpáva z 2021)

Napočítané příznaky můžeme rozdělit do několika skupin. **První skupinou** jsou ty, které se zpravidla ani u vážnějších problémů podél hlasivky nemění, např. počet cyklů. Tyto příznaky není třeba zobrazovat. **Další kategorie** příznaků mění hodnotu, ale zajímá nás v nějakém smyslu nejhorší hodnota. Tyto příznaky se dají zobrazit jako u jednořádkové kamery jedním pruhem, který mění barevný odstín podle hodnoty. **Poslední skupinou** jsou příznaky, které se podél hlasivky mění nějakým předvídatelným způsobem, např. relativní délku uzávěru hlasivky

(duration of closure) a související poměr mezi fází zavírání a otvírání. Tady používáme hodnotu v místě maximální amplitudy. Doplňkově je možné zobrazit porušení očekávaného průběhu hodnot podél hlasivky (např. když délka uzávěru neroste směrem ke krajům). Obdélník s hodnotami příznaků ze single módu na aktuálním řádku jsme doplnili o další obdélník, kde je stejným způsobem zobrazena jedna hodnota pro každý příznak. Případně pro komplexnější příznaky je možno přepnout do módu vertikálního, kdy je možno hodnoty také prostorově interpretovat vzhledem k tvaru hlasivky.




*Obr.5. Multiline mód zobrazování příznaků*

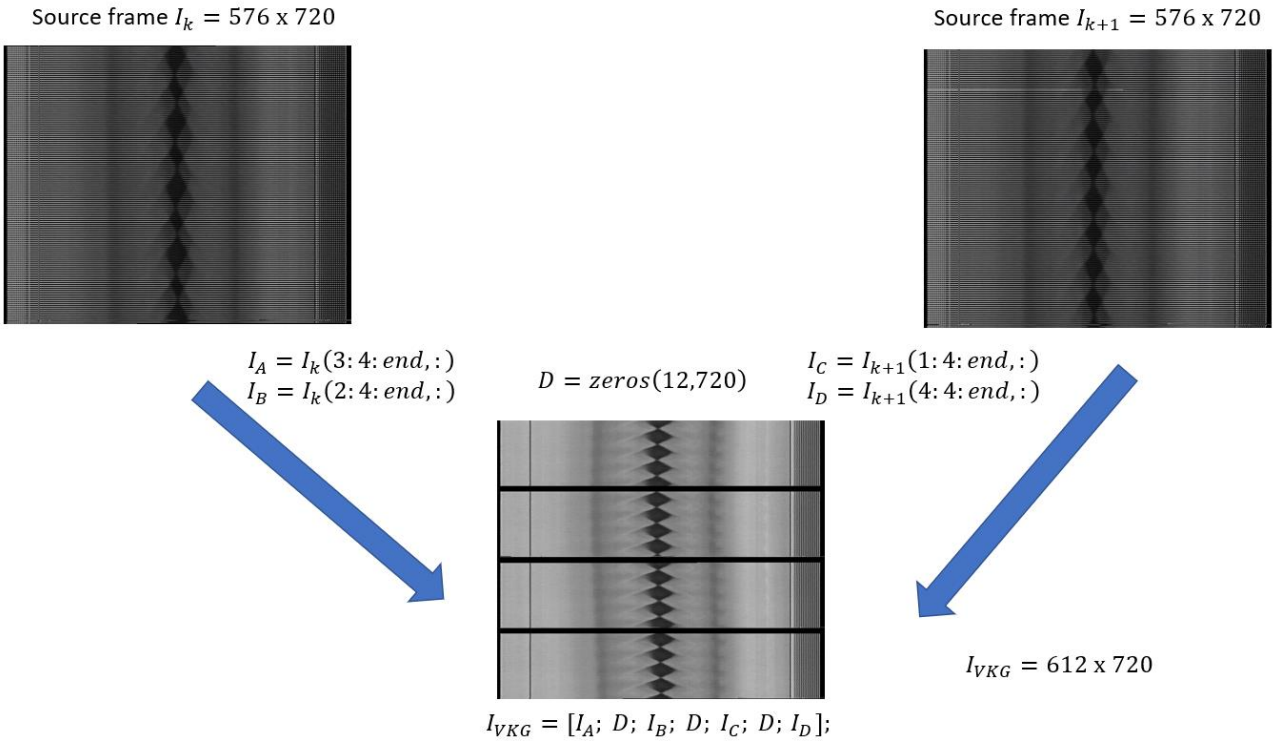


*Obr.6. Vertikální mód zobrazení. Umístění na vertikále odpovídá prostorově jednotlivým liniím snímání*

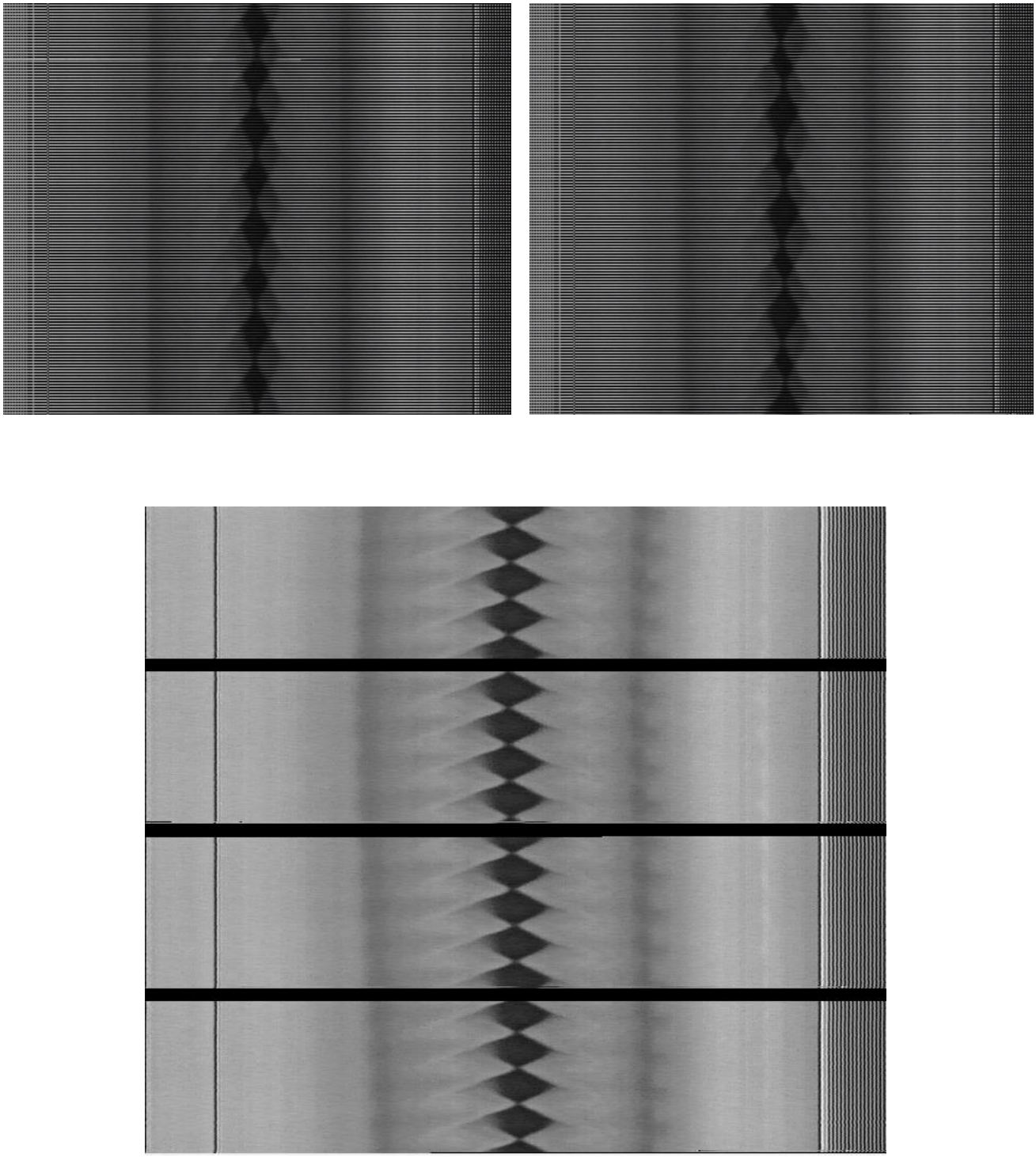
# VKG 3.0 Convertor

Tento modul umožňuje konvertovat VKG data pořízená prvním modelem VKG kamery do dnešního formátu. Byť se může zdát, že tento krok nepřinese nic nového, z podhledu metod založených na neuronových sítích každá datová sada, která je k dispozici, je cenná. Navíc je velkým přínosem mít možnost rozšířit sadu detekovaných diagnóz.

Původní snímání probíhalo v prokládaném režimu, a tedy je potřeba skládat dva snímky, aby vznikl jeden dnešní kvality. Postup je uveden na Obr.7. Na obrázku Obr. 8 pak je ukázka starých dat a vzniklého datového obrázku v současném formátu. Vše je možno spouštět dávkově.




*Obr.7. Systém konverze dat*



*Obr.8. Ukázka dvou vstupních souborů a vzniklého VKG snímku v současném formátu*



Evropská 1692/37, 160 00 Praha 6
+420 234 611 111
info@tacr.cz, www.tacr.cz

## VKG 3.0 Audio

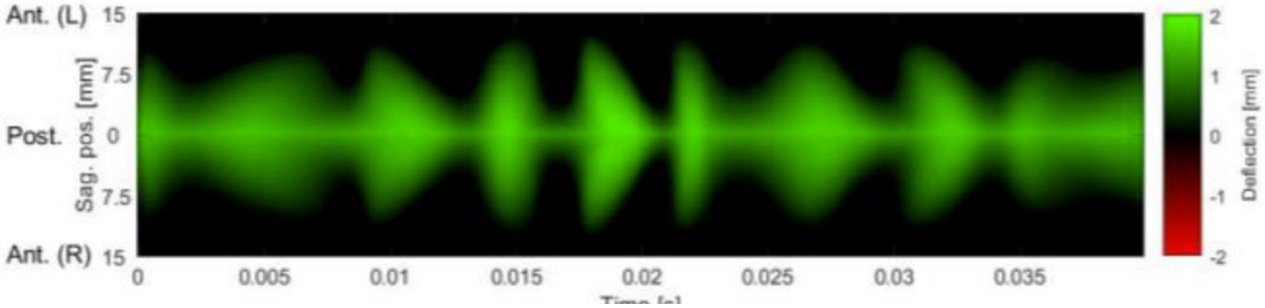
Třetí modul VKG 3.0 software implementuje funkce pro analýzu pořízených dat, které byly navrženy v roce 2020. Podrobnější popis významu parametrů je v článku členů týmu [2], vstupní parametry a jejich význam je ve zprávě z roku 2020. Parametry, které byly do modulu zahrnuty jsou následující

- sound pressure level SPL
- fundamental frequency $f_o$
- pitch strength

## VKG 3.0 Simulator

Dr. Švec vytvořil ve spolupráci s indickými (dr. S. Pravin Kumar) a rakouskými kolegy (Dr. P. Aichinger) a svým doktorandem na UP v Olomouci (H. Lehoux) simulátor víceřádkových videokymografických a vysokorychlostních videolaryngoskopických záznamů, který je založen na kinematickém modelu kmitání hlasivek.

Tento simulátor využívá matematicky přesně definovaný tvar hlasivek (*délku, tloušťku, zakřivení povrchu, konvergenci mediálního povrchu hlasivek a vzdálenost hlasivek od sebe*) a přesně definované kinematické parametry kmitání hlasivek (*frekvenci, amplitudu spodního a horního okraje hlasivek, rychlost slizniční vlny a dosah šíření slizniční vlny na horním povrchu hlasivek*). Změna jednotlivých parametrů modelu vede ke změně kymografického vzorce kmitání hlasivek. Navíc lze parametry modelu hlasivek měnit podél délky glottis a tak dosáhnout různého vibračního vzorce v různých místech podél délky hlasivek.



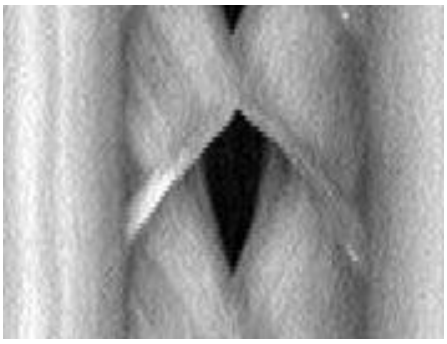

*Obr.9. Ukázka synteticky nagenerovaného videokymogramu*

Z velkého množství řádkových kymogramů simulátor vytváří nasimulovaný vysokorychlostní videolaryngoskopický záznam kmitání celých hlasivek. Simulátor umožňuje lépe pochopit jak jednotlivé rysy kmitání hlasivek viditelné v kymografických a laryngoskopických videích souvisí s kinematickými parametry kmitání hlasivek, což ve výsledku pomůže opodstatnit doagnostický význam jednotlivých parametr a jejich hodnot.

Kromě zdravých stabilních kmitů lze i simulovat různé fyziologické a patologické variace, jako např. fonaci se zipovitým otevíráním glottis, tlačenou fonaci, hlasový začátek, konstantní a proměnlivé pravo-levé a předo-zadní rozdíly kmitů, atd. Simulátor je také vhodným nástrojem pro validaci algoritmů obrazové analýzy kmitů hlasivek. Informace o simulátoru byly sepsány do dvou článků. Informace o simulátoru byly sepsány do dvou článků - [3] byl publikován ve sborníku konference **MAVEBA 2021** v italské Florencii a [4] byl zaslán k publikaci do impaktovaného časopisu **Journal of Speech, Hearing and Language Research.**

## VKG 3.0 Estimator

Posledním modulem je **Estimator**, který je zaměřen na odhad ostrosti laterálních vrcholů a délku mukosální vlny. Odhad těchto parametrů a jejich důležitost pro schopnost stanovení správné diagnózy je založena na poddajnosti sliznice.



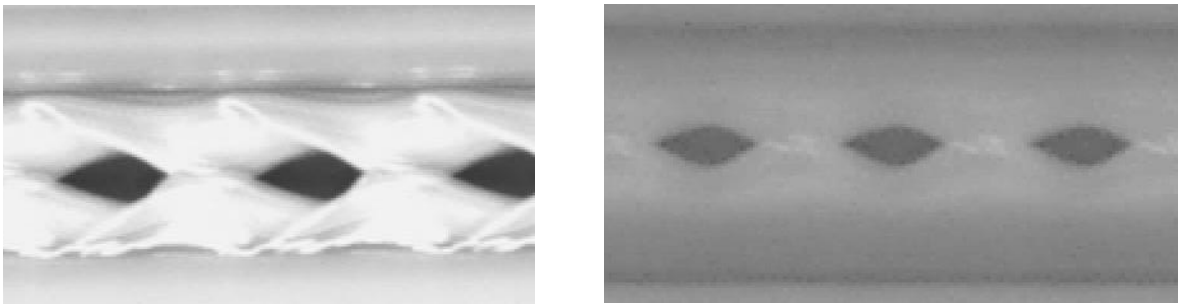
*Obr. 10: Slizniční vlna na mediálním povrchu hlasivek způsobuje fázový rozdíl mezi pohybem spodního a horního okraje hlasivek. Ve VKG snímcích se tento fázový rozdíl projevuje jako dvojitá kontura během uzavírací fáze glottis a jako ostré laterální vrcholy kmitů. Pokud je slizniční vlna omezena a fázové rozdíly mezi horním a spodním okrajem hlasivek jsou malé, ve VKG snímcích se to projeví jako zaoblení laterálních vrcholů kmitů.*

Ta je totiž důležitým faktorem pro zdravou fonaci. Poddajnost je tradičně laryngoskopicky sledována skrze přítomnost slizničních vln na hlasivkách [1, 2]. Nepřítomnost slizničních vln na hlasivkách nebo jejich omezení je tradičně brána jako patologický příznak související se zvýšenou tuhnutí sliznice hlasivek.

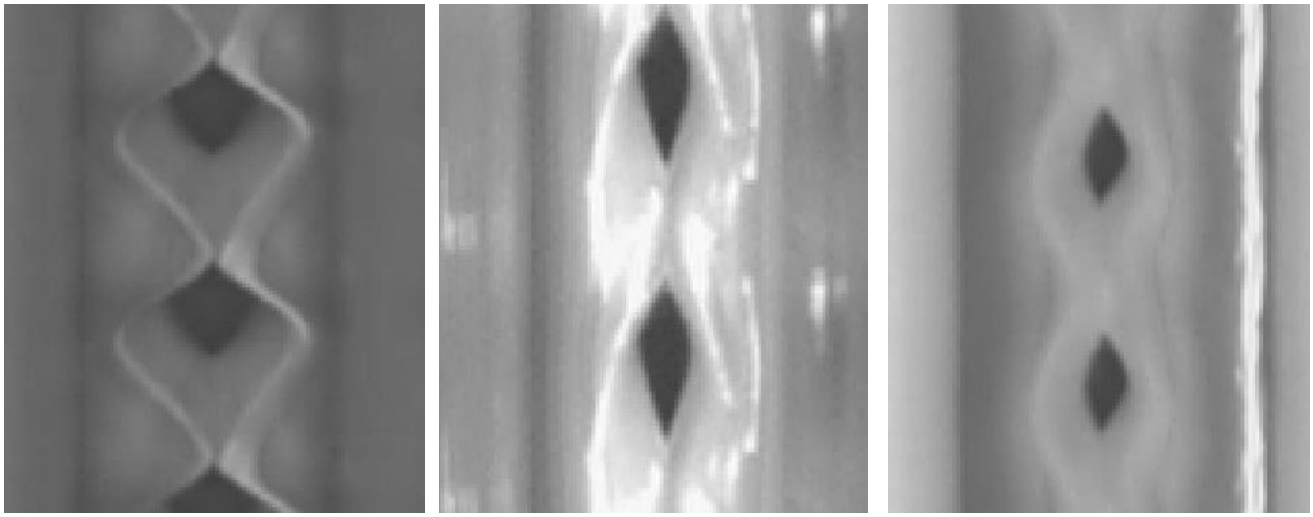
Základní frekvence zvuku hlasu je tvořena kmitavým pohybem sliznice hlasivek. Při tvoření hlasu se hlasivky přiblíží k sobě do tzv. fonačního postavení a zároveň proud vydechovaného vzduchu posouvá sliznici hlasivek zespoda vzhůru (kraniálně) a do strany (laterálně). Posun sliznice hlasivek proudem vydechovaného vzduchu je podstatou kmitavého pohybu sliznice hlasivek a zejména podstatou vytvoření základní frekvence zvuku hlasu. Kvalita hlasu je tedy přímo závislá na kvalitě (zdraví) sliznice hlasivek.

Pro včasnou diagnostiku onemocnění sliznice hlasivek je mimořádně důležité využití co nejkvalitnějších zobrazovacích technik, které mají schopnost odhalit anatomickou či funkční poruchu tkání v raném stadiu. Videokymografie má schopnost odhalit poruchu pohybu sliznice hlasivek. Podstatou poruchy pohybu sliznice je tzv. zvýšená tuhost slizniční tkáně. Tato zvýšená tuhost slizniční tkáně je dána typem infiltrace sliznice zánětlivými, nádorovými či dalšími patologickými procesy. Studiem vibračního vzorce zachyceného videokymografickým vyšetřením jsme schopni odhalit tyto slizniční patologie. Zásadní je absolutní znalost vibračního VKG vzorce zdravé slizniční tkáně. Jako hlavní parametr patologického zvýšení tuhosti sliznice hlasivek ve VKG obraze je porucha vibračního vzorce v oblasti tzv. laterálních vrcholů a porucha posunu slizniční vlny směrem laterálním.

Soustředili jsme se na studium videokymografických záznamů, kde ve vibračním vzorci je viditelná porucha laterálního okraje hlasivkových kmitů a dochází k většímu či menšímu omezení posunu slizniční vlny laterálním směrem. Tyto patologické nálezy jsme verifikovali medicínskými diagnostickými postupy k ověření správnosti diagnózy infiltrativního procesu sliznice hlasivek.



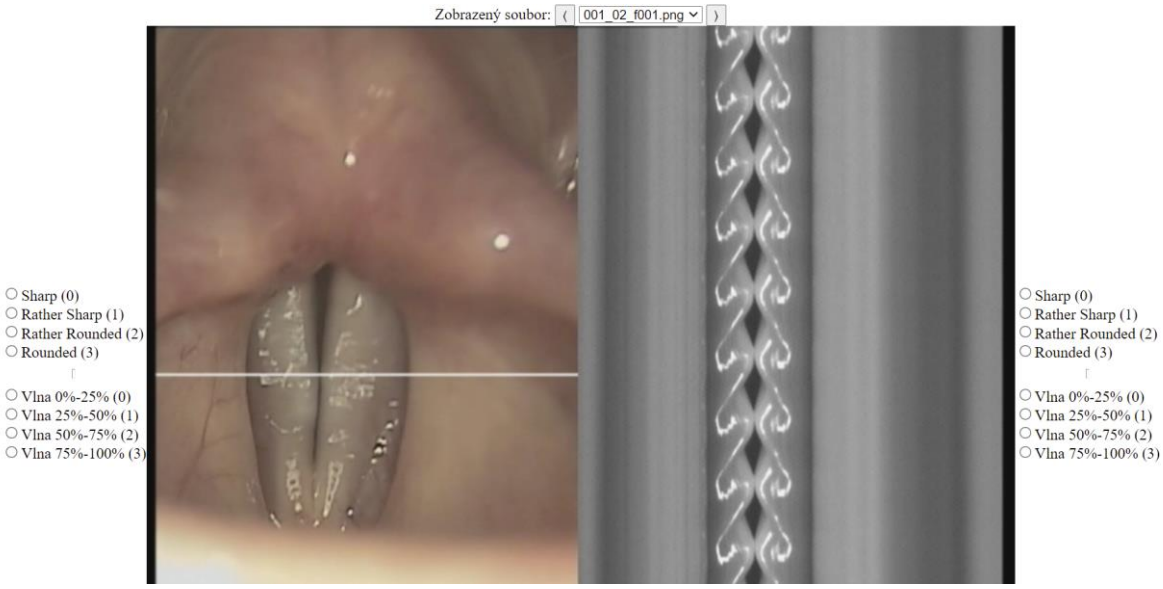
*Obr 11: Posun vlny v celém rozsahu (vlevo) a chybějící (vpravo)*



*Obr 12: Posun vlny ostré, středně ostré, oblé*

Analyzovali jsme pří činy p oruchy kmitavého pohybu hlasivek zachycené ve VKG obraze. Srovnávali je s klinickými a laboratorními nálezy. Nejčastějšími příčinami zvýšené tuhosti sliznice hlasivek byla zánětlivá infiltrace slizniční tkáně: akutní a chronické zánětlivé procesy, v menší míře nádorová infiltrace slizniční tkáně: nádorové procesy v počínajícím stadiu. Schopnost diagnostiky iniciální fáze nádorové infiltrace sliznice je největším přínosem vyšetřovací metody VKG. Charakteristickým klinickým markerem pro zvýšenou tuhost tkáně (infiltraci sliznice) je omezení posunu slizniční vlny při fonaci a zaoblení laterálních vrcholů kmitů. Tato ostros-oblost je důsledkem fázového rozdílu mezi pohybem spodního a horního okraje hlasivek, více ve zprávě z roku 2020. Čím je sliznice vlivem patologie tužší, tím jsou fázové rozdíly mezi pohyby spodního a horního okraje hlasivek menší a tím je laterální vrchol kmitů oblejší. Ve studii významnosti videokymografie pro klinickou praxi se ostrost/zaoblenost laterálního vrcholu kmitů hlasivek ukázala být nejvýznamnějším klinickým parametrem pro funkční diagnostiku poruch hlasu [7].

Pro úlohu automatické detekce jsme použili postupy konvolučních neuronových sítí (CNN). Tyto postupy vycházejí z použití předtrénované architektury CNN, s následným dotrénovaním postaveném na množině trénovacích dat, pocházejících z dané aplikační oblasti, v našem případě videokymogramů. Jedná se o klasifikační úlohu, kdy je potřeba laterální vrcholy videokymogramu zařadit do jedné ze čtyř tříd ostrosti laterálního vrcholu [***zakulacený, spíše zakulacený, spíše ostrý, ostrý***]. Robustnost přístupů, založených na CNN, jsou silně závislé na kvalitě použité trénovací množiny, proto jsme se věnovali sběru dat a následné anotaci pomocí navrženého webového rozhraní pro zadávání hodnot, viz.Obr. 12.



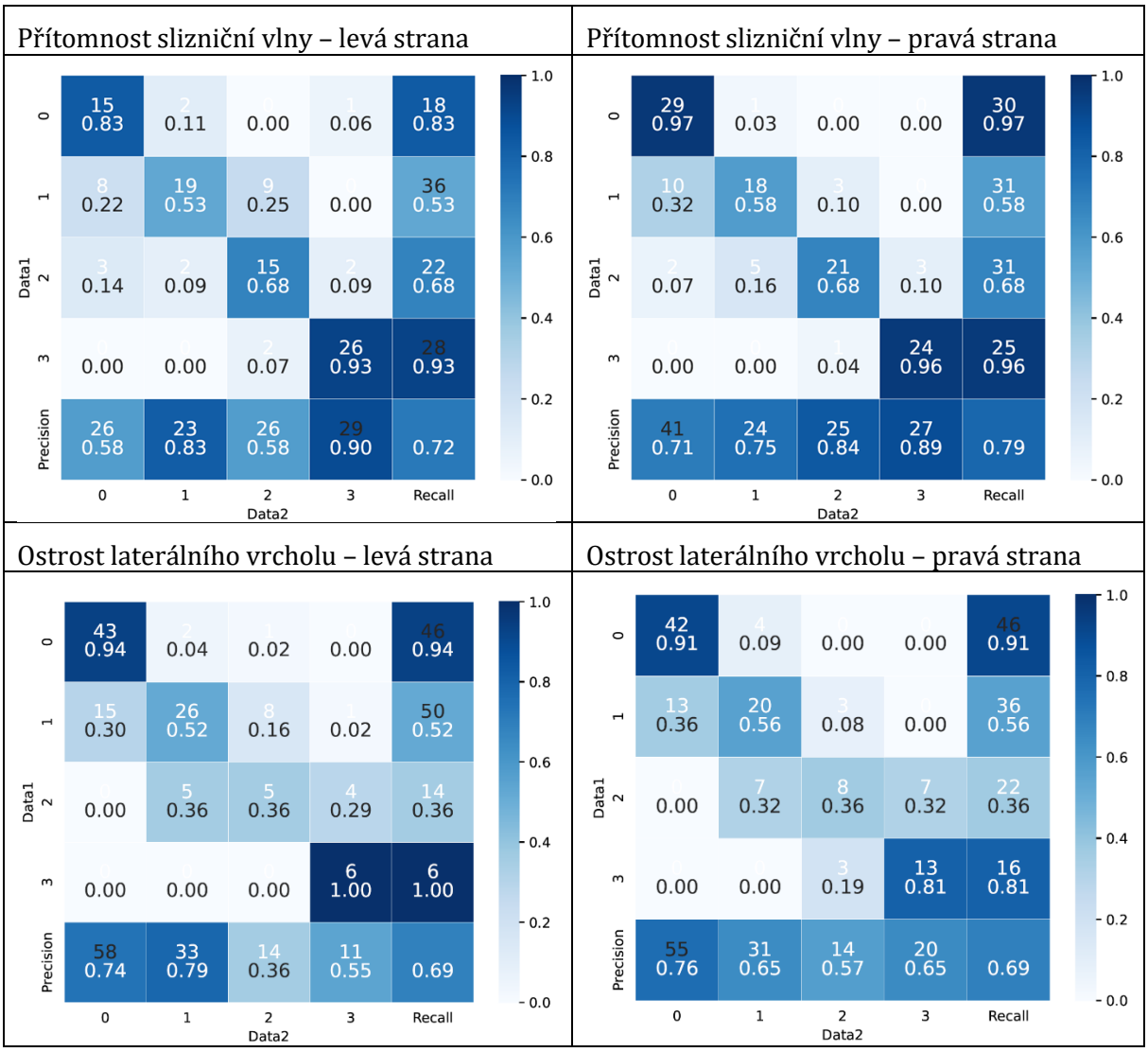

*Obr 13: Anotační nástroj pro oblost-ostrost a délku vlny.*

V roce 2021 byla sada z roku 2020 rozšířena a oanotována. Datová sada byla podrobena kontrole konzistence, protože je důležité, aby se samotné odborníci shodli na chápání jednotlivých jevů.

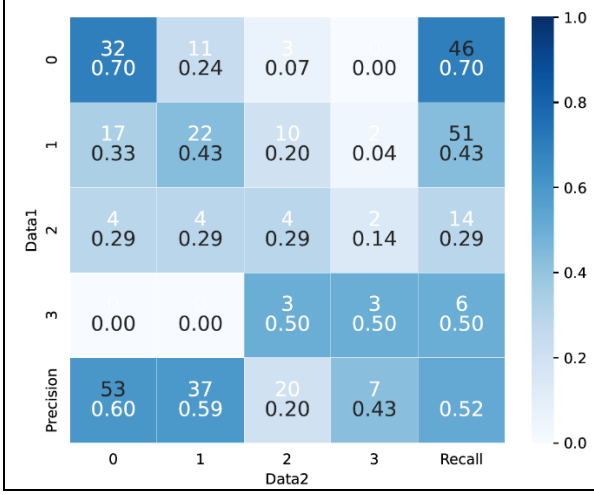
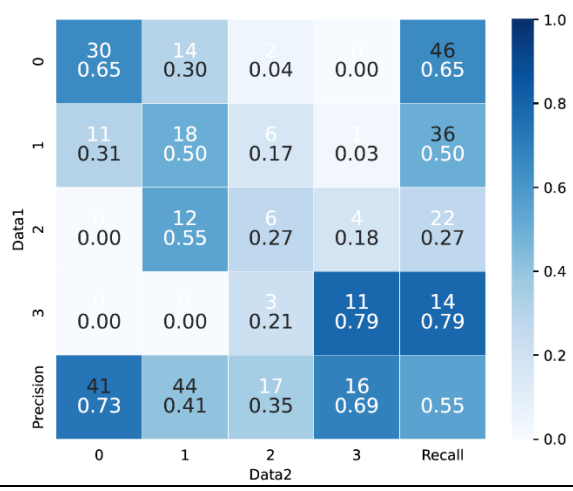
Pro analýzu jsme použili dataset 319 kymogramů, ohodnocených dvěma experty a doktorandem provádějící výzkum v oblasti videokymografie pod vedením jednoho z expertů. Hodnotitelé uváděli hodnoty dvou příznaků na škále 0, 1, 2 nebo 3. Pro případ přítomnosti slizniční vlny odpovídala hodnota 0 nepřítomné vlně, 3 naopak dobře viditelné. Pro případ ostrosti laterálního vrcholu kymogramu odpovídala hodnota 0 ostrému vrcholu, naopak 3 vrcholu dokonale neostrému.

Nejdříve zde uvádíme kontingenční tabulku Tab.1 charakterizující konzistenci hodnocení jednoho z expertů. Za tímto účelem jsme nechali experta ohodnotit jednotlivé kymogramy v datové sadě, a to dvakrát s odstupem několika dní. Při druhém hodnocení bylo pořadí kymogramů náhodně zpřeházeno. Pro zobrazení výsledků jsme použili způsob známý ve strojovém učení jako konfuzní matice. Díky tomu budeme schopni přímo porovnat konzistenci hodnocení s výsledky strojového učení. První čtyři sloupce a řádky udávají počty jednotlivých kombinací možných hodnot (0-3) v rámci prvního a druhého hodnocení, které bylo provedeno v jiném náhodném pořadí předkládaných kymogramů. Poslední řádek a sloupec udává agregované hodnoty pro jednotlivé hodnoty v rámci prvního, respektive druhého hodnocení, viz popis výsledků strojového učení. V pravém dolní rohu vidíme celkovou úspěšnost opakování přesně stejné hodnoty. Vidíme, že expert dosahuje poměrně dobré shody obou hodnocení a pokud uvede jinou hodnotu, zpravidla zvolí hodnotu o jedničku menší či větší.

Porovnali jsem také hodnocení experta s hodnocením mladého odborníka (Tab. 2), který byl seznámen se způsobem, jakým hodnotit ostrost laterálního vrcholu na jiné datové sadě. Mladý odborník se účastnil pouze hodnocení příznakem ostrost laterálního vrcholu, a to pouze na polovině dat použité datové sady. Výsledek můžeme vidět na následujícím obrázku. Vidíme, že míra shody je již výrazně horší než vlastní konzistence experta a dosahuje hodnot 0.52 a 0.55 pro levou a pravou stranu hlasivky. Ve většině případů se ale hodnocení také neodchyluje více než o jedna.



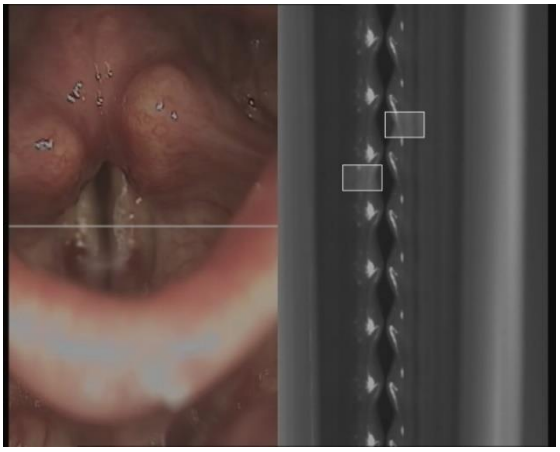
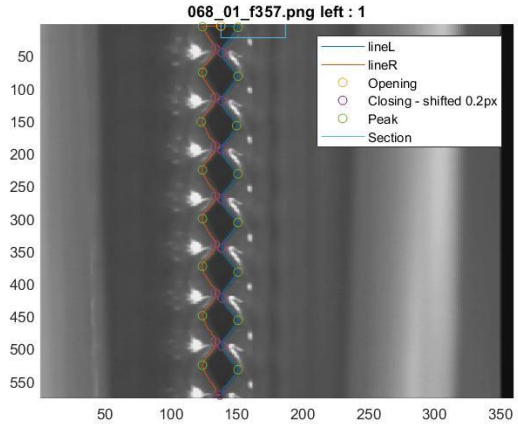

*Tabulka 1: Expert proti sobě, dvě hodnocení, náhodné, v jiném čase. Popis v textu*

| Srovnání expert – student – levá strana | Srovnání expert – student – pravá strana |
|---|---|
|  |  |



*Tabulka 2: Expert a* mladý odborník*. Popis v textu*

## Automatické odhadování ostrosti a přítomnosti slizniční vlny

| Kymogram s naznačenými výřezy pro rozpoznávání | Detekce kontur a významných bodů v kymogramu |
|---|---|
|  |  |



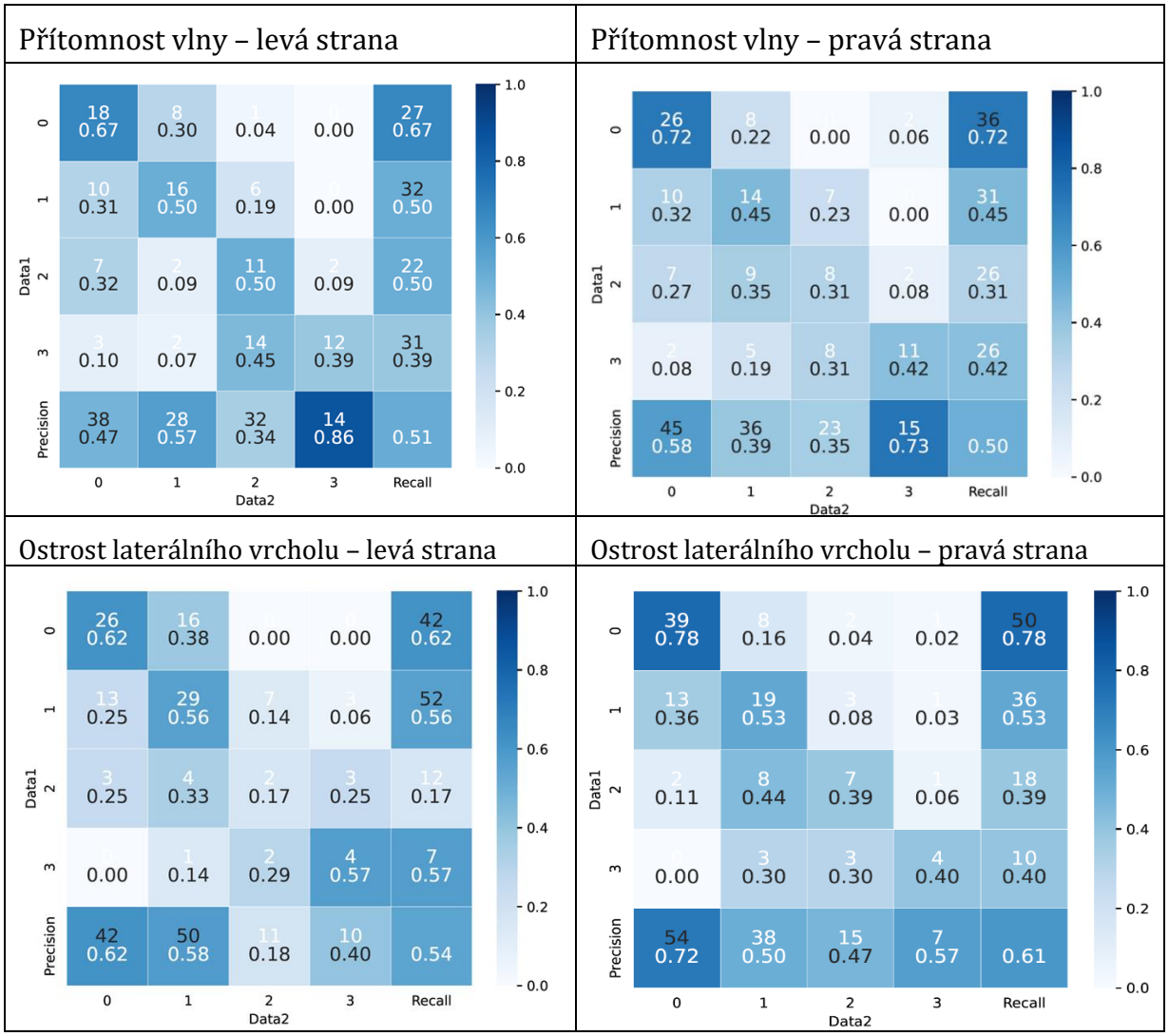
*Obr 14: Automatická detekce.*

V této části projektu jsme se využili metody konvolučních neuronových sítí pro automatický odhad ostrosti laterálního vrcholu a přítomnosti slizniční vlny. Nejlepší výsledky jsme docílili následujícím postupem:

1. Využili jsme toho, že videokymogramy typicky obsahují několik velmi podobných cyklů otevření a uzavření hlasivky. Použili jsme robustní postup na segmentaci hlasivky a detekci významný bodů (laterální vrcholy, otevření, uzavření) z Analyzéru, které nám udávají počátek a konec každého cyklu. V rámci každého cyklu jsme vyřízli obrázek lokálního okolí daného cyklu, který je pak vstupem neuronové sítí pro rozpoznávání. Tímto způsobem detekujeme příslušný příznak zvlášť pro každý cyklus a výslednou hodnotu pro celý kymogram dostaneme jako nejčastější hodnotu (modus) z hodnot na jednotlivých cyklech. Shoda výsledků na jednotlivých cyklech v rámci jednoho kymogramu nám navíc umožňuje stanovit spolehlivost odhadu.

2. Pro učení jsme použili dataset ohodnocený expertem. Dataset (319 kymogramů, 3695 výřezů) jsme rozdělili na trénovací a testovací množinu v poměru 3:2. Pro jsme použili konvoluční neuronovou síť MobileNetV2 [5], k optimalizace používáme algoritmus Adam [6]. Pro učení obou příznaků jsme použili stejný model sítě a stejný dataset. Levou a pravou půlku kymogramu jsme učili dohromady s tím, že levá strana byla nejdříve otočena kolem osy hlasivky do stejné polohy jako pravá.

3. Výsledky trénování na validační množině v podobě konfuzních matic můžeme vidět na následujících obrázcích, a to zvlášť pro levou a pravou stranu a oba hodnocené příznaky. Svislá osa vždy odpovídá správné hodnotě stanovené expertem (hodnoty 0 až 3), vodorovná osa pak hodnotám určených automatickým algoritmem strojového učení. Celé hodnoty odpovídají počtům případů, tj. kontingenční tabulce kombinací hodnot experta a algoritmu (Tab. 3). V tabulce jsou pod těmi hodnotami uvedeny tytéž hodnoty normované na součet jedna v rámci dané hodnoty stanovené expertem. Pravý sloupec udává hodnoty ... (recall), dolní řádek ... (precision) pro jednotlivé hodnoty. V pravém dolní rohu je pak celkové procento přesné shody (accuracy).




*Tabulka 3: Expert a SW. Popis v textu*

Navržené algoritmy strojového učení dosahují přesnosti 0.54/0.61) pro ostrost vpravo a vlevo, což převyšuje shodu úspěšnosti mladého odborníka (0.52/0.55). Přesto že datová sada se jevila jako dostatečně velká, nezahrnovala potřebný počet extrémních datových sad. Po doplnění těchto případů se přesnost ještě zvýší. Tento modul považujeme za hlavní výsledek projektu, protože umožní automatizovat a objektivizovat hodnocení důležitého parametru pro vyhodnocování stavu hlasivek.

T A
Č R

## Reference


[1] Zita A, Novozámský A, Zitová B, Šorel M, Herbst CT, Vydrová J, Švec JG (2021). Videokymogram Analyzer Tool: Human-computer comparison. (Submitted to Biomedical Signal Processing and Control on Dec 29, 2021).

[2] Kumar, S. P., & Svec, J. G. (2018). A simple method to obtain basic acoustic measures as video subtitles in laryngeal *videoendoscopic* recordings. Journal of Speech, Language, and Hearing Research, 61(9), 2196-2204

[3] Aichinger P, Kumar SP, Lehoux H, Švec JG (2021). Artificial high-speed videos of normal and dysphonic vocal fold vibration. In C. Manfredi (Ed.), *Models and Analysis of Vocal Emissions for Biomedical Applications - 12th International Workshop, MAVEBA 2021* (pp. 93-96). Firenze, Italy: Firenze University Press [ISBN: 978-837880799-5]. https://www.fupress.com/isbn/9788855184496

[4] Aichinger P, Kumar SP, Lehoux H, Švec JG: Artificial high-speed videos for studying the kinematics of normal and dysphonic vocal fold vibration (submitted Dec 22, 2021)

[5] SMobileNetV2L - www.tensorflow.org/api_docs/python/tf/keras/applications/MobileNetV2

[6] Optimizer https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam