

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta

**Automatické rozpoznávání rypošů na
základě jejich vokalizace**

Bakalářská práce

Jan Karásek

Školitel: Ing. Ladislav Ptáček, Ph.D.

České Budějovice 2021

Bibliografické údaje:

J. Karásek, 2021: Automatické rozpoznávání rypošů na základě jejich vokalizace [Automatic recognition of naked mole-rat based on their vocalization, Bc. Thesis, in Czech] – 44 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Abstrakt:

V této práci jsem se zabýval automatickým rozpoznáváním rypoše lysého na základě jeho vokalizace. Ve sklepení Přírodovědecké fakulty, kde se nachází chov rypoše lysého, jsem za dozoru Mgr. Emy Hrouzkové, Ph.D. nahrál 20 jedinců při třech mírách motivace. Poté jsem nahrávky zpracoval v programu Audacity, kde jsem odstranil úvodní identifikaci jedince a výrazné rušivé zvuky. Nahrávky jsem dále rozdělil po 20 sec. Nahrávky jsem z důvodů velké náročnosti a komplexnosti problematiky předal k experimentům Ing. Ladislavu Ptáčkoví, Ph.D. Výsledky z experimentů jsem zpracoval formou tabulek a grafů. Z výsledků jsme zjistili, že systém dokáže při správné kombinaci vstupních parametrů rozeznat individualitu jedince. Rozhodující stránkou pro vyhodnocení výsledků se ukázala být míra motivace použita pro GMM (hlavní model jedince).

Abstract:

In this thesis, I dealt with the automatic recognition of the naked mole-rat on the basis of its vocalization. In the basement of the Faculty of Science, where the naked mole-rat breeding is located, I recorded 20 individuals with three levels of motivation, under the supervision of Mgr. Emy Hrouzková, Ph.D. Then I processed the recordings in the Audacity program, where I removed the initial identification of the individual and significant disturbing sounds. Further, I divided the recordings after 20 sec. Due to the great demands and complexity of the issue, I handed over the recordings to the experiments of Ing. Ladislav Ptáček, Ph.D. I processed the results from the experiments in the form of tables and graphs. From the results, we found that the system can recognize the individuality of an individual with the right combination of input parameters. The degree of motivation used for the GMM (Gaussian Mixture Model) turned out to be a decisive aspect for the evaluation of the results.

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47 b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejich internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích,

dne

Podpis studenta

Poděkování

Touto cestou chci poděkovat vedoucímu této bakalářské práce panu Ing. Ladislavu Ptáčkovi, Ph.D. za cenné rady, připomínky a vstřícnost při vypracování této bakalářské práce. Dále děkuji paní Mgr. Emě Hrouzkové, Ph.D. za odborný dohled při nahrávání a přátelskou atmosféru. V neposlední řadě děkuji mé rodině a kolegům za podporu a schovívavost.

Jan Karásek

Obsah

1 Úvod	1
2 Teoretická část	2
2.1 Rypoš lysý	2
2.2 Rozpoznávání zvuků zvířat a lidské řeči, rozdíly	5
2.3 Technický úvod	9
2.3.1 Identifikace řečníka	9
2.3.2 Kepstrum	12
2.3.3 Hammingovo okno	13
2.3.4 Mel-kepstrální koeficienty, MFCC	13
2.3.5 GMM, Gaussian Mixture Model	14
2.3.6. Vokální trakt	15
2.3.7 Sběr dat	16
2.3.8 Dostupnost zdrojů a literatury	16
2.3.9 Rozpoznávání zvuků zvířat, specifikace úloh	17
3 Praktická část	18
3.1 Použité vybavení pro nahrávání	18
3.2 Výsledky nahrávání	20
3.3 Příprava nahrávek pro experiment	20
3.4 Koncepce experimentů	22
3.4.1 GMM (Gaussian Mixture Model)	23
3.4.2 UBM (Universal Background Model)	23
3.4.3 TRIAL	24
3.4.4 Vyhodnocení experimentu	24
3.5 Popis testovacích sad Experimentu I	26
3.6 Experiment I.: Zjištění optimálních parametrů	28
3.6.1 Minimální frekvence	28
3.6.2 Maximální frekvence	30
3.6.3 Počet filtrů	31
3.6.4 Typ filtru	33
3.6.5 DeltaDelta	34
3.6.6. Počet cepstrálních koeficientů	35
3.6.7 Počet Delta koeficientů s vypnutým a zapnutým DeltaDelta	36
3.7 Experiment II.: Verifikace jedinců s optimálními parametry	38
3.7.1 Experiment II. fáze 1	39

3.7.2 Experiment II. fáze 2.....	40
3.7.3 Experiment II. fáze 3.....	41
4 Závěr.....	43
Seznam literatury.....	45
Přílohy.....	47

1 Úvod

Cílem této práce je zjistit, zda je možné automaticky pomocí software rozpoznat informace o jedinci z obsahu jeho zvuků (odborně vokalizace), které vydává rypoš lysý. Pro tuto práci byly využity nahrávky pořízené na Přírodovědecké fakultě Jihočeské univerzity, kde se nachází chov rypoše lysého v četnosti 130 jedinců žijících v několika koloniích.

Práce obsahuje teoretickou a praktickou část. V teoretické části bude čtenář seznámen s rypošem lysým, s prostředím ze kterého pochází a s jeho způsobem života ve volné přírodě. Dále bude v teoretické části vysvětlen základ metody automatického rozpoznávání zvuků zvířat. Praktická část se věnuje experimentům s těmito nahrávkami. Bude uveden postup při pořizování nahrávek, které byly použity jako vstup do automatického rozpoznávání rypoše lysého podle vokalizace. Hlavní částí práce jsou výsledky experimentů automatického rozpoznávání, uváděných ve formě tabulek a grafů.

Původním cílem práce bylo zjistit, zda na základě vokalizace rypošů dovedeme rozeznat jedince (jejich individualitu), pohlaví, velikost (hmotnost) a motivaci u každého jedince. Pojem motivace je zoology využíván pro vyjádření míry stresu, které byl rypoš vystaven při pořizování nahrávky a souvisí též se společenským statutem zvířete v kolonii. Tuto míru lze měnit tím, že jedince hladíme nebo pošťuchujeme propiskou či dokonce taháme za ocas (tedy jej stresujeme).

Jak bude vysvětleno později, oproti původnímu očekávání byly z rozhodnutí vedoucího mé práce původní cíle zúženy na jedinou úlohu a to rozpoznání individuality, tedy zda dokážeme rozpoznat jedince podle jejich vokalizace.

Ukázalo se, že rozsah experimentů jen pro tuto úlohu byl tak náročný, že nebylo reálné se věnovat i dalším cílům. Zvolen byl proto, že se jedná o klíčový experiment. Další úlohy z něj vycházejí a bude proto možné v budoucnu na ně navázat.

2 Teoretická část

2.1 Rypoš lysý

Rypoš lysý (*Heterocephalus glaber*) je hlodavec, který je jediným zástupcem čeledi *Heterocephalidae*. Vyskytuje se v aridních oblastech afrického rohu, tedy oblastech se suchým podnebím a velkými výkyvy teplot, konkrétně Djibouti, jihovýchodní Etiopie, Somálsko, severovýchodní Keňa. Obývají polopouště 300-1500 m.n.m., kde je suché a horké klima [1].

Rypoši lysí jsou stejně jako ostatní rypoši striktně podzemní hlodavci. Celý život tráví ve svých podzemních systémech včetně páření, odpočinku a hledání potravy. Jejich podzemní systémy zahrnují hnízdo, několik toalet a zásobáren a síť chodeb vedoucích k potravě o celkové délce až 3 km. Nad zem se vydávají jen výjimečně mladí jedinci snažící si najít partnera [1].

Z hlediska smyslů je podzemní prostředí velmi odlišné od nadzemního. Je tam tma a vzduch proudí jen v omezené míře, proto je zrak u většiny striktně podzemních druhů slabý a pachové stopy se mohou šířit jen na krátké vzdálenosti [9]. Pro rypoše jsou důležité smysly hmat a sluch [2]. Hmat může sloužit jen na krátkou vzdálenost a na střední a delší vzdálenosti slouží sluch. Co pro mou práci je důležité, je fakt, že šíření zvukových signálů je omezeno specifickým akustickým prostředím podzemních tunelů. Nejlépe se šíří zvuky o nízkých frekvencích (kolem 440 Hz). Vysokofrekvenční zvuky jsou odraženy a pohlcovány stěnami tunelů a mohou se šířit jen na krátké vzdálenosti a navíc vytvářejí vysokou hladinu šumu, díky níž je sluch podzemních hlodavců velice málo citlivý [10]. Vokalizace je přizpůsobena těmto specifickým podmínkám, signály mají vzhledem k velikosti těla nízké frekvence. Problém je i s vyjadřováním motivace (tedy míře stresu). Obecná tzv. Mortonova motivačně-strukturní pravidla o vokalizaci obratlovců popisuje, že signály vydávané v agresivních situacích bývají netonální a nízké (zvíře se snaží vypadat větší), zatímco hlasy kontaktní nebo vyděšené jsou spíše vysoké a tonální [11]. Podzemní druhy mají k dispozici jen úzké spektrum frekvencí, které se dobře šíří, proto motivaci vyjadřují spíš mírou tonality než změnou frekvence [12].

Většina rypošů lysých zůstává ve své rodné kolonii, která obsahuje 40-80 jedinců, ale byli nalezeny i kolonie s 290 jedinci. V kolonii se o množení stará vždy jen jedna samice a jeden nebo dva samci, ostatní pomáhají s provozem společenství. Nejmladší a nejmenší se starají o mláďata a uklízejí hnízdo a okolí. Jak rostou tak se postupně přesouvají ke kopání,

hledání potravy a největší jedinci se starají o obranu teritoria. Hlavní samice udržuje ostatní v předpubertálním stádiu vývoje pomocí feromonů. Proto existují jen malé rozdíly mezi pohlavími. Samice může mít až tři vrhy ročně, každý v průměru 12 mlád'at [1].

Ostatní druhy rypošů žijí buď osamoceně, nebo ve skupinách po 8 až 12 jedincích. Jedinci, kteří se nemnoží, jsou pohlavně dospělí, samice nemají ovulaci díky agresivnímu chování hlavní samice, ale po opuštění rodiny jsou schopné se okamžitě začít množit. Mají jen jeden až dva vrhy ročně po 1–4 mlád'atech [1].

Složitost a bohatost vokálního repertoáru u většiny druhů zvířat odpovídá sociální složitosti skupiny. To ale nemusí odpovídat prostému počtu členů skupiny. Závisí to také na složitosti jejich interakcí, tím např. na stabilitě skupiny, hustotě, hierarchii a na počtu sociálních rolí. Počet sociálních rolí úzce souvisí se schopností individuálního rozpoznávání. Pokud se jedinci dokáží rozpoznávat individuálně, mají více sociálních rolí než jedinci, kteří v rámci skupiny dokáží určit jen příslušnost k určité podskupině (množící se jedinci, jedinci s potlačenou reprodukcí) [13].

U rypošů bohatost vokálního repertoáru neodpovídá prostému počtu členů skupiny. Rypoš lysý s největšími skupinami má méně druhů vokalizací než rypoš obří s poměrně malými rodinami (*Tabulka 1*).

	Heterocephalus	Fukomys mechowii	Fukomys darlingi
	rozsah [kHz]	rozsah [kHz]	rozsah [kHz]
Contact calls	Soft chirp (2,5 - 4,5)	Twitter - 3,65 (3,0 - 9,7)	Twitter - 3,5 (1,5 - 3,7)
	Toilet call (1,5 - 4,0)	Twitter-like - 1,91 (0,4 - 8,7)	Cheep2 - 5,13 (3,5 - 6,4)
		Long twitter - 1,57 (0,3 - 11,0)	Cheep1 - 4,20 (1,9 - 4,7)
		Squeak - 1,28 (0,1 - 11,0)	
		Gabbling - 0,96 (0,3 - 4,3)	
	Grunt - 0,25 (0,3 - 12,1)		
Počet	2	6	3
Distress calls	Scream (2,0 - 4,0)	Alert - 4,45 (0,3 - 22,0)	Alert - 5,21 (3,7 - 5,5)
			Squeal - 3,80 (2,0 - 5,3)
			Harsh call - 3,77 (1,5 - 10,4)
Počet	1	1	3
Aggressive calls	Hiss (0,4 - 1,3)	Hiss - 2,58 (0,9 - 4,4)	Snort - 0,48 (0,1 - 6,6)
	Grunt (0,2 - 20,0)	Snorting - 2,31 (0,3 - 20,5)	
	Upsweep trill (1,0 - 9,0)	Scream - 2,23 (0,3 - 14,4)	
	Loud chirp (1,0 - 6,0)	High trill - 0,83 (0,4 - 12,5)	Squeak - 4,57 (3,3 - 5,7)
		Swing trill - 0,51 (0,3 - 0,8)	
Počet	4	5	3
Mating calls	V-trill (1,5 - 6,0)	Harsh - 1,48 (0,3 - 9,0)	Cluck - 1,50 (0,7 - 1,9)
		Cluck - 0,43 (0,3 - 1,4)	Shriek - 0,98 (0,5 - 1,4)
		Shriek - 0,68 (0,3 - 8,8)	
Počet	1	3	2
Alarm calls	Tap (0,2 - 1,2)	Trill - 0,94 (0,4 - 5,8)	
	Sneeze (5,0 - 16,0)		
	Low-pitched chirp (1,0 - 1,0)		
Počet	3	1	0
Mechanical sounds	Tooth grinding (1,0 - 1,0)	Teeth grinding - 3,95 (0,3 - 1,9)	Teeth grinding - 1,34 (0,4 - 1,9)
		Seismic - 0,5 (0,4 - 1,9)	
Počet	1	2	1
Celkový počet vokalizací	12	18	12
Reference	Pepper et al. 1991	Bednářová et al. 2013	Dvořáková et al. 2016
Prům. hmotnost (samci/samice) g	40/38	380 / 260	65 / 63
Prům. vel. skupiny	78,3 ± 55,3	9,9 ± 2,49	7,4 ± 1,7

Tabulka 1: Bohatost vokálního repertoáru rozděleného do jednotlivých kategorií na základě behaviorálního kontextu. Modře zbarvená políčka vyjadřují silně nelineární zvuky, oranžová barva znázorňuje přechodné typy zvuků a nezvýrazněná políčka představují lineární tonální zvuky. Údaje o průměrných hmotnostech jednotlivých druhů jsou přejaté z chovů Přírodovědecké fakulty Jihočeské univerzity v Českých Budějovicích.

Tato skutečnost by mohla být u rypošů způsobena rozdílnou schopností individuálního rozpoznávání. Pro individuální rozpoznávání a pro rozpoznávání členů skupiny může sloužit pach a vokalizace [14]. U rypošů lysích je známá jejich schopnost poznat členy svojí rodiny podle pachu samice. Stejně tak mají společný pach kolonie i rypoši obří [15]. Vokální individuální rozpoznávání bylo zatím zkoumáno jen u rypoše darlingova (*Fukomys darlingi*) na základě jejich signálů vydávaných při námluvách. V tomto případě se signály individuálně lišily a i rypoši je byli schopní rozeznat. Navíc dokázali poznat z vokalizace neznámých jedinců jejich reprodukční status [16].

Rypoši jsou velmi vhodná zvířata pro výzkum vlivu kolektivní celistvosti na složitost repertoáru. Žijí ve stejných podmínkách a liší se právě jen sociálním systémem. Zahrnují samotářské druhy, i druhy žijící společně ve velkých rodinách po mnoho generací, kde jsou jedinci rozděleni do pracovních kast [1]. Z nich je právě rypoš lysý nejunikátnější velikostí skupin, přesahem generací i skladbou vokálního repertoáru. Při setkání dvou jedinců produkují rypoši lysý unikátní kontaktní signál, ve kterém se střídají. Bylo zjištěno, že nese informaci o velikosti a tím i o společenském postavení jedince [17]. Při střetu dvou jedinců začnou produkovat krátký syčivý zvuk. Pokud důvod k agresí nepřestává, začne se objevovat krátký ostrý hrdelní zvuk, a poté většinou následuje fyzický útok v podobě kousnutí [17].

2.2 Rozpoznávání zvuků zvířat a lidské řeči, rozdíly

Lidská řeč i zvukové projevy zvířat slouží k mnoha účelům: Komunikace, hájení teritoria, varování před nebezpečím, namlouvání, vyjadřování emocí (strach, spokojenost) a další. Mnohé z nich jsou lidem i zvířatům společné, jiné jsou specifické pro daný druh. Ale protože člověk zvířatům nerozumí a je otázkou, zda si plně rozumí rozdílné zvířecí druhy navzájem, je logický závěr, že nikdy nebudeme schopni přesně interpretovat významy zvířecích zvuků. Přesto jsme často schopni porozumět alespoň některým z nich.

Takové znalosti byly získány díky pozorování zoologů (ptáci, opice, sloni) či trvalému soužití se zvířetem (pes, kůň, delfin). Z pohledu automatického rozpoznávání zvuků zvířat je důležité určit odlišnosti, které budou významné při rozpoznávání zvuků zvířat v porovnání s lidskou řečí. Jelikož rozpoznávání lidské řeči a identifikace mluvčího jsou podrobně zkoumány řadu let, je žádoucí aplikovat tyto vyvinuté metody s přihlédnutím ke zmíněným odlišnostem.

Protože tato práce se věnuje automatickému rozpoznávání vokalizace rypoše lysého, uvádíme hlavní rozdíly mezi zvukovým projevem člověka a zvířat se zaměřením k těmto hlodavcům:

- Omezený repertoár vydávaných zvuků.

Řádově se jedná o jednotky zvuků.

- Spektrální a časové charakteristiky.

Základní frekvenční pásmo lidské řeči leží zhruba v pásmu 300Hz-3,5 kHz, hladina akustického tlaku se pohybuje kolem 50 dB. Standardní řečová dynamika je pro češtinu asi 30 dB.

Projevy zvířat jsou buď jednoduché zvuky (skřek, hvizd), nebo vázané projevy s určitou strukturou (zpěv ptáků, velryb). Spektrální pásmo zvuků zvířat se pohybuje od desítek Hz (sloni) až po desítky kHz (pěvci).

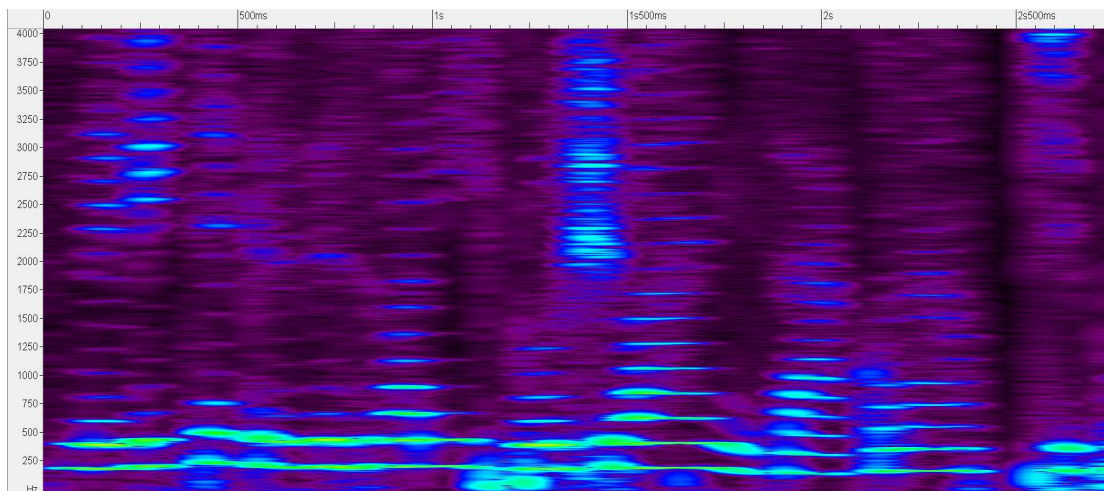
Konkrétně u rypoše lysého leží zvukový projev v pásmu 200 Hz až 20 kHz, dynamika projevu je asi 20 dB.

- Harmonická struktura.

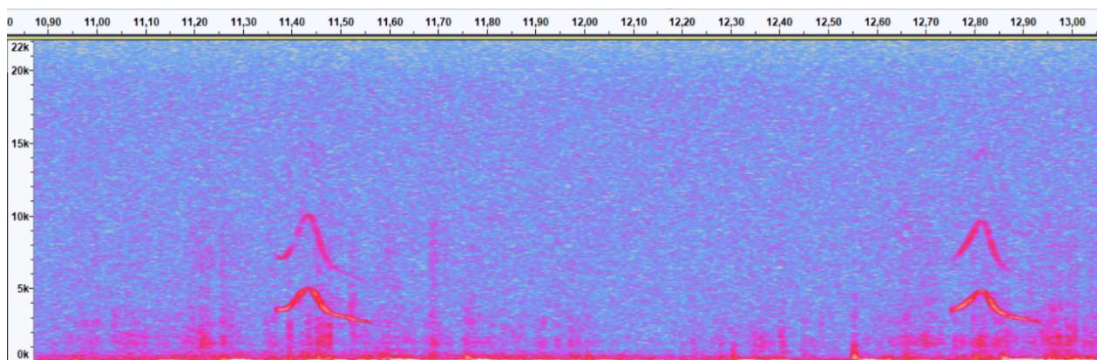
Lidská řeč je bohatá na množství obsažených harmonických. Ze spektrogramu (*Obrázek 1*) je patrné, že řeč obsahuje velké množství složek zabírajících úzká frekvenční pásma.

Na rozdíl od člověka vydávají zvířata většinou zvuky s menším množstvím harmonických. Některá zvířata mohou produkovat čistě sinusoidální nebo čistě šumový charakter (pěvci, drobní hlodavci).

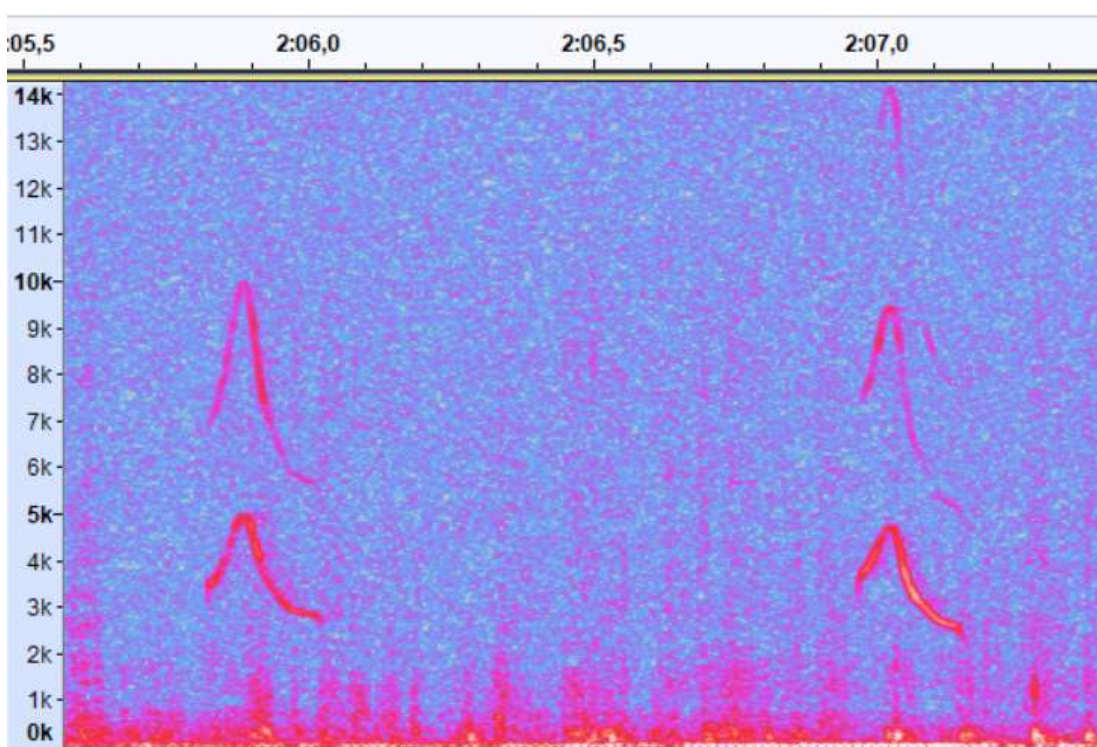
Na spektrogramu rypoše lysého (*Obrázek 2* a *Obrázek 3*) je patrné, že jeho projev sestává z jedné složky, která zabírá široké frekvenční pásmo a téměř neobsahuje vyšší harmonické.



Obrázek 1: Lidská řeč. Studiová kvalita, žena, český jazyk, čteno bez emocí (obchodní sdělení). Zobrazeno pásmo 0 – 4 kHz.



Obrázek 2: Rypoš lysý.
Na ose X čas. Na ose Y frekvence 0 – 22 kHz.



Obrázek 3: Detail vokalizace rypoše lysého, jedna slabika.
Na ose X čas. Na ose Y frekvence 0 – 14 kHz.

Pro přehlednost jsou hlavní rozdíly rypoše lysého v porovnání s lidskou řečí sestaveny do tabulky (*Chyba! Nenalezen zdroj odkazů.*).

		Lidská řeč	Rypoš lysý
Šířka frekvenčního pásma		3 kHz <i>100 Hz ÷ 3 kHz</i>	4 kHz <i>200 Hz ÷ 20 kHz</i>
Průměrné trvání jednoho slova resp. jedné slabiky		1 sec	200 ms
Harmonické složky	Počet složek	2 ÷ 20	1 ÷ 3
	Délka, po kterou lze složku považovat za statickou	150 ÷ 300 ms	10 ÷ 30 ms
	Šířka frekvenčního pásma, přes kterou leží jedna složka	20 Hz ÷ 130 Hz	150 Hz ÷ 2,5 kHz

Tabulka 2: Porovnání lidské řeči a vokalizace rypoše lysého

- Tvorba zvuku

Způsob tvorby zvuku (vokální trakt) je u mnoha druhů zvířat velmi podobný lidskému: opice, někteří pěvci, kytovci. Někteří ptáci mohou navíc zpívat dvojhlasně (skřivan, slavík, drozd).

U některých druhů je tvorba zvuku zcela odlišná a děje se bez využití hlasového traktu. Například kmitáním křídel (komár), užitím zvláštní membrány (cikáda), třením křídel o sebe (cvrček), u hlodavců tloučením hlavou o stěnu nory (hlodoun) ad.

- Přenášená informace.

Je nemožné či přinejmenším značně obtížné sestavit pro zvířata něco jako slovník, lexikon či řečový korpus. Člověk zvířatům „nerozumí“. U některých druhů to lze provést, i když s obtížemi a se zřejmými výhradami k nedokonalosti takových interpretací: neúplnost, mnohoznačnost, kdy má jeden zvuk více významů, atd. Např. pes [2.2] rozpoznává 5 různých druhů štěknutí podle „významu“ (radost, varování, smutek, ...).

Lidská řeč je unikátní množstvím přenášené informace. Zásadní rozdíly lze spatřovat ve výkonnosti mozku. S tím souvisí v přírodě jedinečné schopnosti tvorby zvuků a samozřejmě i sluchové schopnosti člověka včetně zpracování předané informace.

U zvířat, kde je vokální trakt uspořádán podobně jako lidský, lze teoreticky usuzovat, že s lépe vybaveným mozkiem by vydávané zvuky těchto druhů mohly být více podobné člověku. Proti tomu se staví fakt, že někteří pěvci, kteří dokáží napodobit lidskou řeč (andulka, korela, špaček, loskuták), zjevně nemají výkonnější mozek než jiné druhy. Je jisté, že schopnost napodobit lidskou řeč souvisí s rozvinutější částí mozku podílející se na ovládnání hlasového traktu. Zároveň ale špaček nebo korela nevnímají obsah vyslovovaných slov, pouze (z jejich pohledu) vydávají „zajímavé zvuky“ díky vyvinuté hudební paměti (otisku lidského slova). Tento mechanismus přesně neznáme.

2.3 Technický úvod

V souladu s náplní práce budou při automatickém rozpoznávání rypošů využívány stejné metody jako u úlohy automatického rozpoznávání řečníka (Speaker recognition), konkrétně úlohy identifikace mluvčího (Speaker identification). Rozdíly mezi oběma typy signálů, ze kterých vyplývá odlišné nastavení parametrizačních koeficientů, jsou uvedeny v kapitole 2.2.

Jako hlavní dva rozdíly uvedeme:

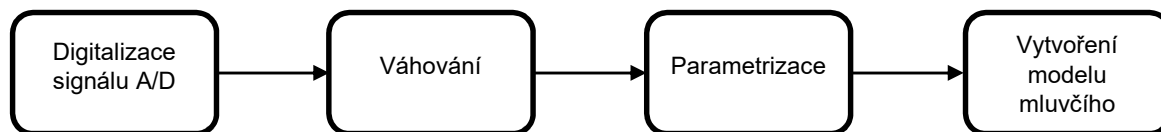
- Rozdílnou časovou základnu, po kterou je možné považovat signál za stacionární
Má vliv na délku chovacího okna a na šířku překrytí váhovacích oken
- Rozdílné frekvenční pásmo
Má vliv na správné nastavení zpracovávaného pásma, včetně vhodně zvolené vzorkovací frekvence

2.3.1 Identifikace řečníka

Hlavním cílem metody je nalezení modelu vokalizace jedince (řečníka), který bude vykazovat nejlepší shodu při následné verifikaci. Metoda musí být nezávislá na obsahu. Tedy nezáleží na tom, jaké části vokalizace jsou užívány. To konkrétně u zvířat přináší zásadní výhody, viz. diskuse v kapitole 2.2

Model je závislý na parametrech hlasového traktu zvířete. Bylo pozorováno, že vokalizace rypoše a jeho hlas se v průběhu času (měsíců, let) mění. Nebylo však dosud nikde zkoumáno, zda mají tyto změny vliv na jeho model.

Podle [3] lze postup identifikace řečníka popsat následujícím schématem:



- Digitalizace signálu A/D
Pro potřeby naší úlohy je vzorkovací frekvence zvolena $f_{vz} = 22.050$ Hz, jelikož frekvenční rozsah rysoše lysého se pohybuje v pásmu $0,2 \div 20$ kHz.
- Váhování
Pro získávání parametrů je nutné rozdělit digitalizovaný signál v čase na takové části, po které bude signál považován za statický. Pro lidskou řeč se jeví jako vhodná délka okna 20 ms s překryvem zpravidla 10 ms. Výběr těchto částí se provádí váhováním.
- Pro účely této práce je zvoleno váhování Hammingovým oknem. Pro vokalizaci rysoše musí být vhodné hodnoty teprve určeny. Na základě prvních pokusů jsou stanoveny na 10 ms s překryvem 5 ms.
- Parametrizace
Ze signálu získáme váhováním statické parametry x_i . Ty jsou zvoleny tak, aby vykazovaly nízkou variabilitu pro jednotlivé jedince i za odlišných podmínek (rozdílné emoce, obsah projevu). Naopak by měly vykazovat signifikantní rozdíly pro různé jedince. Měly by být též odolné proti šumu a zkreslení signálu [3]. Pro každý vzorek tak vznikne vektor příznaků.

$$X = \{x_1, x_2, \dots, x_N\}$$

Vektor zahrnuje kepstrální koeficienty a další vybrané parametry. Z každého okna získáme přibližně 40 těchto parametrů. Z těchto koeficientů vznikne Matice C, přičemž na ose X je počet oken (čas) a na ose Y kepstrální koeficienty. Pro vokalizaci rysoše budou vhodné parametry určeny na základě další práce. Během prvních pokusů jsme pracovali s 20 kepstrálními koeficienty plus nultý, který je úměrný logaritmu energie signálu.

- Vytvoření modelu mluvčího.
Pro vytvoření modelu mluvčího jsou nejčastěji využívány principy [3] Vektorové kvantování (VQ) nebo Gaussian Mixture Model (GMM). V další práci jsou s ohledem

na dobré výsledky využívány směsi gaussovských rozložení. K tomu bude využíván programový balík v programovacím jazyce Matlab, vytvořený mým školitelem. Ten umožňuje automatizované spouštění experimentů a jejich hromadné vyhodnocování. Programový balík obsahuje:

- Správce experimentu.
Proces experimentu je plně automatizovaný. Uživatel může spustit libovolný počet experimentů najednou. Parametry se nastavují jednotlivě pro každý experiment a je implementován v Matlabu.
- Správce experimentálních sad.
Správce třídí nahrávky pro GMM, UBM a TST (TRIAL). Generuje také seznam TST, což jsou dvojice porovnávaných nahrávek. Seznamy jsou seskupeny do takzvaných testovacích sad. Proces je plně automatizovaný s pokročilými funkcemi, tj. sekvenčním nebo náhodným tříděním souborů. Vše je implementováno v Matlabu.
- Extrakce parametrů.
Extrakční algoritmy jsou implementovány v Matlabu i v nástroji SV (Speaker Verification).
- Podpůrné moduly.
Sem patří moduly VAD (detektor vokalizace, Voice Activity Detector), BAF (filtr přizpůsobený ptačí vokalizaci, Bird Audiogram Filtr) a modul pro slučování. Všechny jsou implementovány v Matlabu, VAD je implementován v SV nástroji.
- Vytvoření modelu.
GMM / EM modul, modul přizpůsobení MAP a kalibrace algoritmů jsou implementovány v SV nástroji.
- Ověření.
Rozhodovací algoritmy jsou implementovány v nástroji SV.
- Vyhodnocení experimentu.
Získají se výsledné statistiky a výpočet EER. Hodnocení je automaticky zpracováno a propojeno s datovými soubory spolu s experimentálními parametry, to vše je implementováno v Matlabu.

2.3.2 Kepstrum

Řečový signál $s(t)$ je tvořen konvolucí signálu z generátoru buzení řeči $x(t)$ a impulzní charakteristikou řečového traktu $h(t)$ [3]. Pro diskrétní signál platí:

$$s(n) = x(n) * h(n) \quad (1)$$

Pro úlohu rozpoznání řečníka je cílem od sebe signály $x(n)$ a $h(n)$ oddělit a získat tak koeficienty popisující samotný hlasový trakt a buzení. Nesou totiž charakteristickou informaci o řečníkovi [3], viz. *Obrázek 5*. Jsou vzájemně málo korelovány, vykazují časovou stálost.

Platí

$$DFT\{s(n)\} = DFT\{x(n) * h(n)\} = DFT\{x(n)\} \cdot DFT\{h(n)\}, \quad (2)$$

Kde pro FT obrazy těchto signálů tedy platí

$$S(k) = X(k) \cdot H(k). \quad (3)$$

Pro dekonvoluci dvou signálů bude vynásobení signálů převedeno na jejich součet. K tomu je využit přirozený logaritmus spektra složeného signálu [3].

Protože

$$\ln[|S(k)|] = \ln[|X(k) \cdot H(k)|] \quad (4)$$

platí

$$\ln[|S(k)|] = \ln[|X(k)|] + \ln[|H(k)|]. \quad (5)$$

Jednotlivé koeficienty reálného kepstra tak lze určit jako

$$c(n) = \text{Re}\{DFT^{-1}[\ln|S(k)|]\} = c_x(n) + c_h(n) \quad (6)$$

Komplexní kepstrum se využívá též při analýze zvukových signálů pro detekci ozvěn a zpětnou rekonstrukci původního signálu.

2.3.3 Hammingovo okno

Pro váhování je využíváno Hammingovo okno, které omezuje prosakování ve spektru, dostatečně potlačuje postranní laloky a nemá příliš velkou šířku laloku hlavního [4].

Je definováno jako

$$w[n] = 0,54 - 0,46 \cos \frac{2\pi n}{N}, \quad \text{kde } 0 \leq n \leq (N - 1). \quad (7)$$

N ... udává počet vzorků na délku okna.

Při zpracování zpěvu ptáků je plánováno využití Hammingova okna, pokud experimenty neprokáží vhodnost jiného typu váhování.

2.3.4 Mel-kepstrální koeficienty, MFCC

Lidský sluch vnímá zvuk nerovnoměrně v závislosti na frekvenci. Proto je žádoucí upravit vstupní signál tak, aby se jeho rozložení přibližovalo vnímání lidským uchem. Tedy aby byly jednotlivé frekvence buď potlačeny nebo zvýrazněny.

Toho se pro lidský sluch dosahuje využitím tzv. Melovské banky filtrů, které převádí frekvence f na f_{mel} které tento jev zohledňují. Index *mel* značí *melodicko-frekvenční osu v melech*.

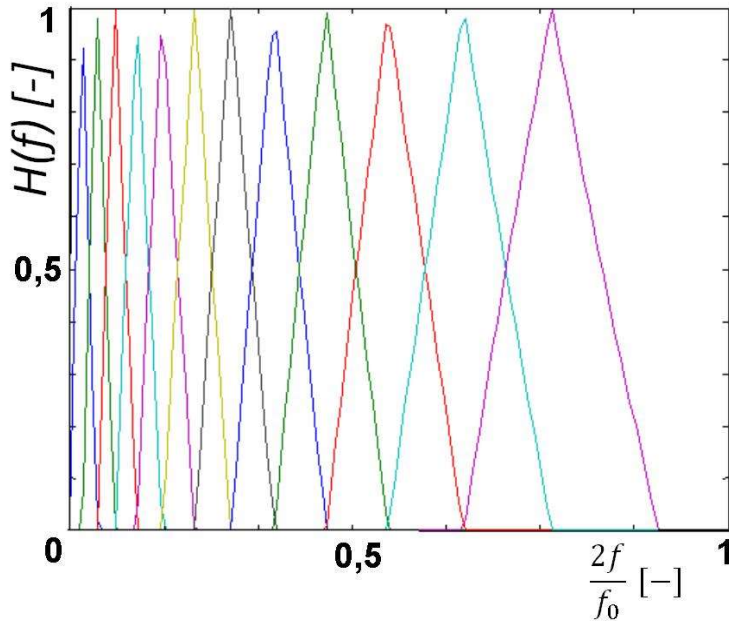
Podle definice [3] platí pro *mel-frekvence* převodní vztah

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (8)$$

a zpětně

$$f = 700 \left(10^{\frac{f_{mel}}{2595}} - 1 \right). \quad (9)$$

Základem výpočtu mel-kepstrální koeficientů (MFCC) je výstup z banky filtrů [3]. Ta obsahuje M pásem, který je standardně 22. Tyto filtry mají pro lineární frekvenční stupnici trojúhelníkový tvar s překrývajícími se pásmy o polovinu (Obrázek 4):



Obrázek 4:

Z pohledu vnímání rysoše lysého zatím není zřejmé, zda budou MFCC využity. Během prvních experimentů bylo melovské rozložení filtrů vypnuto.

Při experimentech s ptáky se používají adaptivní tvary filtrů, které byly nastaveny podle spektrální charakteristik jejich vokalizace BAF (Bird Adapted Filter) a jsou popsány v [8].

2.3.5 GMM, Gaussian Mixture Model

Každý segment řeči je aproximován lineární kombinací gaussovských rozložení. Jejich počet se liší podle požadované úrovně aproximace [4]. Pro popis každé gaussovské složky je užitá střední hodnota a rozptyl. Normální rozdělení je definováno jako

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (10)$$

Pokud má vektor příznaků \mathbf{X} právě N parametrů, bude pravděpodobnostní funkce definována jako lineární kombinace váhovacích koeficientů w_i a směsí gaussovských rozdělení p_i

$$p(\mathbf{x}|\lambda) = \sum_{n=1}^M w_n p_w(x), \quad (11)$$

kde \mathbf{M} je počet gaussovských rozdělení. Ty jsou definovány [3] jako

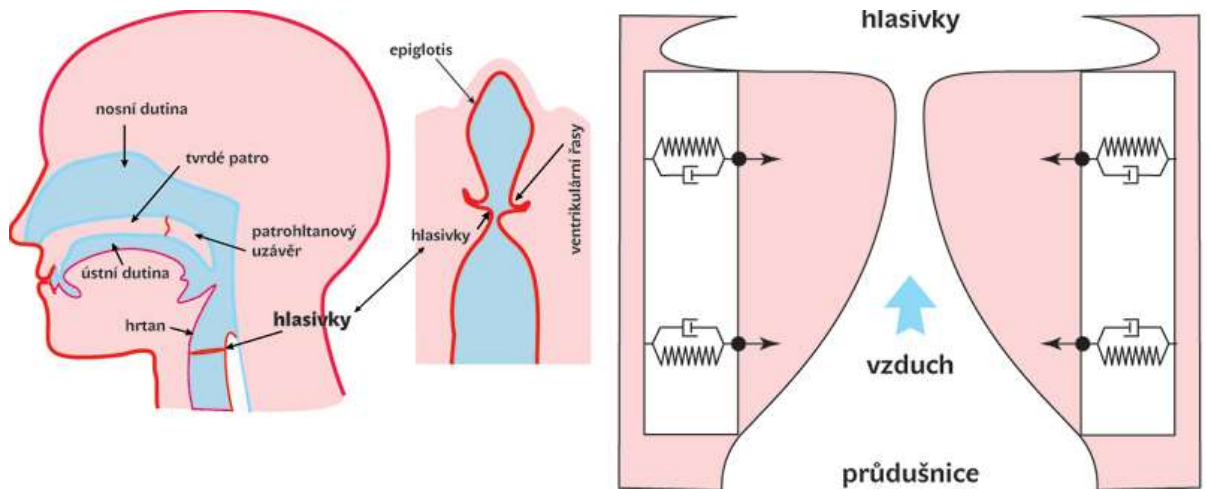
$$p_i(x; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (12)$$

Parametr λ je tedy určen kombinací

$$\lambda = \{w_i, \mu_i, \sigma_i^2\} \quad i = 1, \dots, M. \quad (13)$$

2.3.6. Vokální trakt

Způsob tvorby zvuku u rysořů je aproximován podobně, jako u lidského vokálního traktu, konvolucí budícího signálu a impulzní odezvy hlasového traktu (*Chyba! Nenalezen zdroj odkazů.*):



Obrázek 5: Průřez hlasovým traktem [5]

$$s(n) = x(n) * h(n), \quad (14)$$

$s(t)$ řečový signál

$x(t)$ buzení (*source signal*)

$h(t)$ impulzní charakteristika vokálního traktu (*vocal tract filter*)

Pro zpětné oddělení obou složek se využívá kepstrum, viz. kapitola 0. Spodní koeficienty reprezentují spektrální obálku, tedy vokální trakt, vyšší koeficienty pak buzení. Typicky je pro automatické rozpoznávání využito asi 20 cepstrálních koeficientů.

2.3.7 Sběr dat

Získávání nahrávek rypošů lysých proběhlo v budově B Přírodovědecké fakulty Jihočeské univerzity v Českých Budějovicích ve skladu biologického odpadu. Sběr dat byl situován do několika časových bloků. Nahrávky byly pořízeny při třech mírách motivace (stresu) a to malé, střední a vysoké míře motivace.

Při **malé** míře stresu byl rypoš ponechán, aby se sám projevil. Pokud se neprojevoval, byl motivován doteky na hřbetu a bocích, aby mu byla dána najevo má přítomnost a vokalizoval.

Při **střední** míře stresu bylo nahrávanému jedinci bráněno v cestě a byl mu lehce znepríjemňován život v podobě strkáním do hřbetu a boků gumovým koncem tužky.

Při **vysoké** míře stresu byl na počátku nahrávky jedinec držěn za ocas hlavou dolů a bylo na něj foukáno ústy (rypoši jsou extrémně citliví na průvan). Dále byl nahrávaný jedinec vrácen do nahrávacího boxu a byl mu znepríjemňován život gumovým koncem tužky.

Na počátku každé nahrávky bylo uvedeno ID nahrávaného jedince, dále pohlaví, hmotnost, datum pořízení nahrávky a míra motivace. Každá z nahrávek má délku přibližně 5 minut.

Každá z nahrávek trvající 5 minut byla po vyčištění sestříhána na 20 sec úseky, které byly využity pro experimenty.

Nahrávky jsem pořídil pod vedením dr. Hrouzkové z katedry zoologie.

Jedním ze zásadních problémů při získávání nahrávek byla individualita jedinců rypoše lysého. Někteří jedinci při nahrávání malé míry stresu nechtěli vokalizovat. Avšak když jsem zvýšil míru motivace na střední míru stresu, někteří nahrávaní jedinci přešli rovnou do vokalizace odpovídající vysoké míře stresu. Dalším problémem byl fakt, že některé jedince nebylo možné přimět k vokalizaci.

2.3.8 Dostupnost zdrojů a literatury

Rozpoznávání zvuků zvířat je poměrně novou oblastí, které se týká relativně malé množství zdrojů. Důvodem může být menší atraktivita tématiky a v zásadě i využitelnost oproti rozpoznávání lidského hlasu. Dalším důvodem může být obrovská paleta druhů a zvuků

v živočišné říši oproti „jedné“ lidské řeči. Tím hlavním důvodem ale bude velká náročnost technologií, které se k automatickému rozpoznávání využívají (matematický aparát, znalost programování, atd.). Pro jejich nasazení do praxe je zásadní propojení techniků, kteří tyto techniky ovládnou a zoologů, kteří naopak přinášejí nadměru zajímavé aplikace z oblasti zoologie. K takovému propojení inženýrů a zoologů dochází jenom na specializovaných pracovištích, jako jsou například právě přírodovědecké fakulty.

V současné době se všechny výzkumy věnují zpracování off-line, tedy provádí rozbor již dříve získaných nahrávek. Teoreticky bude možné provádět rozpoznávání vokalizace též on-line. Nezaznamenal jsem žádnou práci, která by se tomuto tématu věnovala.

2.3.9 Rozpoznávání zvuků zvířat, specifikace úloh

Zvířata na rozdíl od lidí nepřečtou přesně specifikovaný text. Nevydají na povel přesně specifikovaný zvuk. Zvířatům nerozumíme, nevíme „co říkají“.

Rozpoznávání zvířat můžeme rozdělit na dvě základní úlohy:

1. Rozpoznávání zvířecího druhu.
2. Rozpoznání konkrétního jedince.

Úlohu rozpoznání zvířecího druhu lze vzdáleně přirovnat k úloze typu *Language recognition* z oblasti automatického rozpoznávání lidské řeči. Mezi zvuky vydávanými zvířaty jsou obrovské rozdíly: Kytovci, savci i pěvci vydávají širokou škálu zvuků. Avšak nejen to, jejich hlasová ústrojí se často značně liší. To je zřejmě hlavní rozdíl mezi touto úlohou a *Language recognition*.

Podobně lze úlohu *Rozpoznávání konkrétního jedince* přirovnat k úloze *Speaker recognition*. Zde je totožnost obou úloh stejná, hlasové trakty zkoumané množiny (člověk, rypoš) je vždy stejná.

K oběma úlohám rozpoznávání druhu či jedince se přistupuje s využitím stejných technik, kterých je využíváno pro lidskou řeč. Je samozřejmě nezbytné přizpůsobení těchto postupů. To je dáno zejména odlišnými frekvenčními a časovými charakteristikami obou typů zvuků (*člověk x zvíře*).

3 Praktická část

Nahrávky jsem pořídil ve sklepení Jihočeské univerzity, kde se nachází chov rypošů lysých, konkrétně 130 jedinců. Jako nejvhodnější místo pro nahrávání s ohledem na nežádoucí znečištění nahrávek okolním hlukem se ukázala místnost určená pro sklad biologického odpadu. Nahrával jsem v několika etapách. Na začátku nahrávacího dne jsem s doktorkou Hrouzkovou odchytl 3-6 očipovaných jedinců. Při odchytu jsem dle jejich pokynů úmyslně nevybral hlavní samici a hlavního samce a to z důvodu, že by mohli mít vliv na parametry vokalizace v porovnání s ostatními členy skupiny. Podle identifikačního čísla odečteného bezkontaktní čtečkou z podkožního čipu jsme dohledali kartu každého jedince. Nádoby s odchycenými jedinci jsem umístil na chodbu před dveře skladu, aby nedošlo vokalizací odchycených jedinců ke znečištění nahrávky. Z nádoby jsem odebral vždy jednoho rypoše a čtečkou podkožních čipů jsem si ověřil identifikační číslo jedince. Poté jsem vybraného rypoše umístil do nahrávacího boxu a několik minut nechal aklimatizovat. Nahrával jsem při třech mírách motivace a to při malé, střední a vysoké míře stresu tak jak je uvedeno v podkapitole 2.3.7 Sběr dat.

3.1 Použité vybavení pro nahrávání

Digitální záznamník Marantz PMD661MKII viz *Obrázek 6*.

- Frekvenční rozsah 20 Hz – 20 kHz (± 1 dB)
- Vážený poměr S / N podle IEC-A:
MIC: 65dB
LINE: 85dB
- Celkové harmonické zkreslení při 0 VU:
MIC: 0,08%
LINE: 0,02%
- Dynamický rozsah 85dB



Obrázek 6: Marantz PMD661MKII [6]

Sluchátka Sennheiser HD 215 viz *Obrázek 7*

- Harmonické zkreslení (THD): $< 0,2 \%$
- Jmenovitá impedance: 32Ohm
- Frekvenční rozsah: 12Hz – 22kHz
- Akustický tlak (1 kHz, 1 Vrms): 112dB



Obrázek 7: Sennheiser HD215 [7]

Mikrofon Sennheiser MD 431 II viz *Obrázek 8*

- Frekvenční rozsah: 40Hz – 18kHz
- Typ snímače: Dynamický
- Citlivost: 2,0 mV/Pa \pm 3 dB
- Jmenovitá impedance: 250Ohm



Obrázek 8: Sennheiser MD431II [7]

3.2 Výsledky nahrávání

Výsledkem nahrávání bylo celkem 60 nahrávek od 20 jedinců ve formátu wav a to 10 samic a 10 samců. Každá nahrávka má délku přibližně 5 minut. Na počátku každé nahrávky je hlasem zaznamenána identifikace jedince: identifikační číslo, pohlaví, aktuální hmotnost a datum pořízení nahrávky. Identifikace jedince trvala přibližně 15 vteřin. Proto jsem se snažil nahrát každý záznam o délce alespoň 5 minut a 20 vteřin s přihlédnutím k tomu, že bude pravděpodobně nutné z nahrávky odstranit další rušivé zvuky, jako jsou například nechtěné údery do mikrofону, zazvonění mobilního telefonu atp.

3.3 Příprava nahrávek pro experiment

Pořízené nahrávky jsem označil písmeny od A do T a podle úvodní identifikace doplnil další informace do *Tabulky 6*. Nahrávky jsem zpracoval v programu Audacity, kde jsem odstranil úvodní identifikaci a případné rušivé zvuky, jako jsou například nechtěné údery do mikrofону. Drobné ruchy v pozadí jako jsou spláchnutí kanalizace, ruch na chodbě a můj dech jsem po konzultaci s doktorem Ptáčkem z nahrávky neodstraňoval a to z důvodu, že se snažíme naše experimenty koncipovat tak, aby byl nutný co nejmenší podíl postprodukční práce (stříhání nahrávek, úpravy, apod.). Využitý algoritmus detektoru vokalizace VAD (Voice Activity Detection) umožňuje s poměrně velkou přesností detekovat a odlišit vokalizaci zvířete od okolního ruchu. Minimalizace úprav považujeme za velkou výhodu našich experimentů v porovnání s běžnými postupy, kdy jsou před vložením nahrávek do rozpoznávacího procesu tyto upravovány a vyhlazovány, aby bylo dosaženo co nejlepšího výsledku. To ovšem znamená vysoké nároky na čas obsluhy a snižuje přínos automatického rozpoznávání – které by mělo být zejména o ušetření času a zautomatizování procesu.

Nahrávky jsem rozdělil po 20 vteřinových úsecích a uložil ve formátu názvu X_YZZ:

Symbol	X	YZZ
Formát	Znak	Číslo
Popis	Identifikace jedince A až T	Y..míra stresu 1 až 3 ZZ.. pořadí nahrávky 01 až 99

Například názvy A_101-A_115 jsou nahrávky od jedince A s nízkou mírou motivace, A_201-A_215 nahrávky od jedince A se střední mírou motivace a A_301-A_315 nahrávky od jedince A s vysokou mírou motivace. Ve stejném formátu jsem uložil nahrávky od každého jedince, celkem 951 nahrávek.

Pro účely zpracování v Matlabu musely být nahrávky přejmenovány do formátu X_YY_ZZZ:

Symbol	X	YY	ZZZ
Formát	Znak	Číslo	Číslo
Popis	Identifikace jedince A až T	Míra stresu: 01 až 03	Pořadí nahrávky: 001 až 999

Jde pouze o změněný formát pro kompatibilitu s již vytvořeným kódem, přenášená informace je totožná s předchozím pojmenováním.

Označení jedince	ID zvířete	Pohlaví	Hmotnost	Datum narození
A	8425	samec	53 g	12. 08. 2003
B	7343	samec	40,2 g	09. 11. 2015
C	8449	samice	35,6 g	03. 08. 2012
D	6412	samice	23,5 g	28. 07. 2019
E	5101	samec	34,8 g	nelze dohledat
F	8445	samice	40,9 g	12. 08. 2013
G	7303	samice	46 g	08. 02. 2016
H	3102	samice	47,7g	21. 08. 2017
I	8461	samec	42 g	30. 08. 2012
J	8451	samec	60,5 g	30. 08. 2012
K	7307	samec	41,9 g	nelze dohledat
L	4317	samice	24 g	27. 02. 2020
M	2378	samec	26 g	10. 12. 2019
N	Heinrich	samec	49,2 g	nelze dohledat
O	1491	samice	47,6 g	nelze dohledat
P	6704	samec	42,2 g	nelze dohledat
Q	9246	samice	41,6 g	nelze dohledat
R	7809	samec	49,3 g	nelze dohledat
S	2467	samice	23,6 g	27. 02. 2020
T	7309	samice	32,5 g	09. 11. 2015

Tabulka 3: Seznam nahaných jedinců

3.4 Koncepce experimentů

Využitý software je koncipován tak, že každý experiment je postavený na verifikaci. Klíčové charakteristiky jsou:

Verifikace, TRIAL

- V každém kroku experimentu porovnáváme jednu dvojici nahrávek, tedy verifikujeme. Výstupem je míra pravděpodobnosti, s jakou jsou si jedinci podobní. V ideálním případě dostaneme hodnoty:
 - výsledek 100%...jde s naprostou jistotou o dva totožné jedince
 - výsledek 0%...jde s naprostou jistotou o dva odlišné jedince.
- Každý experiment má tolik kroků, kolik je požadováno porovnat dvojic. V našich experimentech jsme pracovali průměrně s 6.000 dvojicemi na experiment, ale v některých případech bylo porovnávání více než 34.000 dvojic..
- Seznam těchto dvojic je zapsán v souboru TRIAL

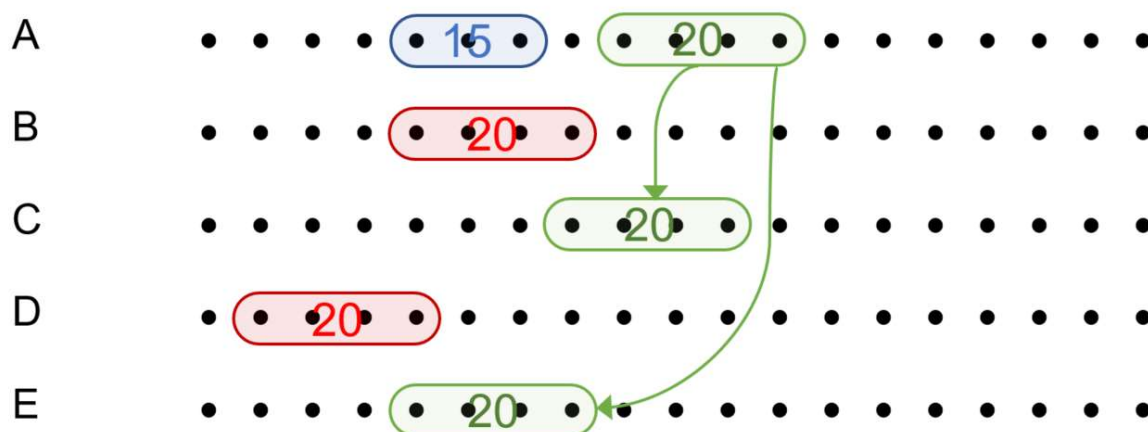
Target jedinec, GMM

- V každém experimentu je vždy jeden jedinec vybrán jako hlavní = *target*. V průběhu experimentu se systém naučí tohoto jedince rozpoznat od všech ostatních.
- Tyto nahrávky označujeme jako GMM podle názvu modelu *Gaussian Mixture Model*.
- Protože pokaždé porovnáváme dvojice nahrávek, je jedna nahrávka v TST dvojici **vždy** od target zvířete. Druhá nahrávka v TST dvojici je:
 - odlišné zvíře: pravděpodobnost jejich podobnosti očekáváme → 0%
 - totožné zvíře (target): pravděpodobnost jejich podobnosti očekáváme → 100%

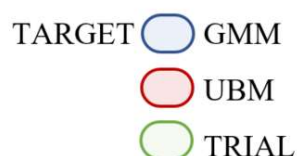
Okolí, UBM

- Systém se musí naučit, jak mohou znít jedinci odlišní od target jedince. Musí se natrénovat tak, aby dokázal rozpoznat zvuky z okolí.
- K tomu slouží nahrávky zařazené do skupiny UBM: Universal Background Model (univerzální model pozadí)

Obrázek 9 ukazuje příklad možného rozdělení nahrávek pěti jedinců (A až E) do jednotlivých množin.



VERIFIKACE



Obrázek 9: GMM, UBM, TRIAL. A-E znamená označení jedinců. Tečky symbolizují jednotlivé nahrávky.

3.4.1 GMM (Gaussian Mixture Model)

Systém se naučí rozpoznávat dle několika náhodně vybraných nahrávek (např. 15 nahrávek z 36) od jednoho jedince „A“, jednoho konkrétního rypoše „A“. Z těchto nahrávek vytvoříme sady dvojic způsobem „každá nahrávka s každou“ a vytvoříme podklad pro GMM. Tedy naučíme SW, jak vypadá rypoš „A“ a pro ten si systém vytvoří model GMM neboli TARGET.

3.4.2 UBM (Universal Background Model)

Dále musíme systém naučit, jak zní univerzální hlas rypoše, aby rypoš „A“ dokázal rozpoznat mezi ostatními hlasy rypošů. Podobně jako rozeznáme hlas blízké osoby od ostatních hlasů. Je to proto, že jsme od narození obklopeni hlasy lidí a umíme rozpoznat hlas například naší matky nebo sourozenců, protože máme srovnání právě s ostatními hlasy. To uděláme tak, že náhodným výběrem určíme např. 20 nahrávek od jedince „B“ a 20 od jedince „D“. Z těchto nahrávek vytvoříme dvojice tak, že zkombinujeme každou nahrávku s každou a vytvoříme podklad pro UBM. Nahrávky, které byly použity pro UBM nesmí být použity pro

TRIAL. Tím by došlo k ovlivnění výsledků (dostávali bychom lepší výsledky, jelikož by systém už nahrávky *znal* z předchozích iterací).

3.4.3 TRIAL

Použijeme např. 20 nahrávek od target jedince „A“. Zásadně vybíráme nahrávky, které nebyly použité pro trénování GMM. Ty využijeme pro porovnání s:

- např. 20 náhodně vybranými nahrávkami od jedince „C“
- např. 20 náhodně vybranými nahrávkami od jedince „E“
- např. s 10 náhodně vybranými nahrávkami od jedince „A“ (porovnáme samo se sebou)

Rypoši C a E jsou pro systém zcela neznámí jedinci. Takto vytvoříme seznamy pro TRIAL, kde postupně proběhne porovnání všech dvojic.

3.4.4 Vyhodnocení experimentu

Při vyhodnocení experimentu vycházíme z výsledků, které spočítá SW na základě vstupních parametrů a podobnosti dvojic nahrávek. Systém je naprogramovaný tak, že výsledky zapisuje přímo do Excelu. Je to přehlednější než práce s výstupní maticí z Matlabu. Ukázka výsledků experimentů jsou v *Tabulce 4*.

	A	B	C	D	E	F	G	H	I	J	K
1	param_01	param_30	param_31	param_32	param_33	param_34	param_35	param_36	param_37	param_38	param_39
2	NrExpermt	EER	TotalTrials	TotalError	TotalRate	TrueTrials	FR	TrueRate	FalseTrials	FA	FalseRate
3	1	33,5%	3900	781	80,0%	400	202	49,5%	3500	579	83,5%
4	2	25,0%	3600	608	83,1%	360	126	65,0%	3240	482	85,1%
5	3	19,3%	3900	499	87,2%	400	114	71,5%	3500	385	89,0%
6	4	19,4%	3600	356	90,1%	360	129	64,2%	3240	227	93,0%
7	5	40,0%	3900	400	89,7%	400	400	0,0%	3500	0	100,0%
8	6	34,2%	3600	752	79,1%	360	168	53,3%	3240	584	82,0%
9	7	21,3%	3900	696	82,2%	400	100	75,0%	3500	596	83,0%
10	8	23,1%	3600	454	87,4%	360	141	60,8%	3240	313	90,3%
11	9	17,8%	3900	597	84,7%	400	90	77,5%	3500	507	85,5%
12	10	28,3%	3600	360	90,0%	360	360	0,0%	3240	0	100,0%

Tabulka 4: Výsledky experimentu

- NrExpermt: Číslo experimentu
- EER: Chyba EER (Equal Error Rate)
- TotalTrials: Celkový počet dvojic
- TotalErrors: Celkový počet chybně vyhodnocených dvojic
- TotalRate: Celková úspěšnost
- TrueTrials: Počet dvojic od shodného jedince
- FR: Počet chybně odmítnutých dvojic od shodného jedince
- TrueRate: Úspěšnost při vyhodnocení dvojic od shodného jedince
- FalseTrials: Počet nahrávek od různých jedinců

- FA: Počet chybně přijatých dvojic od různých jedinců
- FalseRate: Úspěšnost při vyhodnocení dvojic od různých jedinců

Parametry používané pro vyhodnocení úspěšnosti jsou FA, FR a EER:

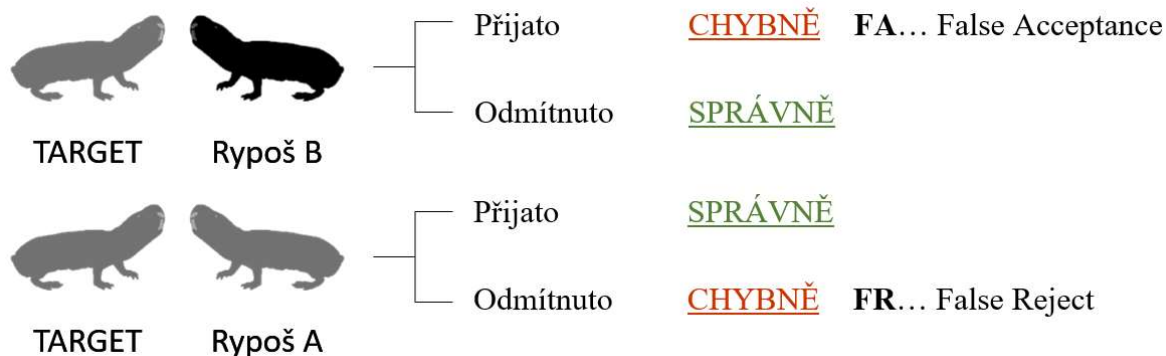
- EER...Equal Error Rate [%], kdy EER 0% znamená absolutně úspěšný experiment, EER 100% znamená zcela chybový experiment.
- Chyba FA (False Acceptance) [%]... . znamená chybně vyhodnocenou shodu dvojice

$$FA = \frac{\text{počet dvojic s rozdílnými jedinci, vyhodnocených, že se jedná o stejné jedince}}{\text{celkový počet dvojic s rozdílnými jedinci}}$$

- Chyba FR (False Reject)... znamená chybně odmítnutou shodnou dvojici.

$$FR = \frac{\text{počet dvojic se stejnými jedinci, vyhodnocených, že se jedná o rozdílné jedince}}{\text{celkový počet dvojic se stejnými jedinci}}$$

Pro názornost jsou chyby FA a FR znázorněny na *Obrázku 10*.



Obrázek 10: Chyba FA a FR

Nejdůležitějším parametrem pro vyhodnocení experimentů je chyba EER (Equal Error Rate), která se mění podle hodnoty FA a FR. To znamená, že není založena na obyčejném průměru chyb a počtu testovaných dvojic. Chyba EER nám zjednodušeně řečeno znázorňuje, kdy je počet chyb FA roven počtu chyb FR. Podobně jako medián použije chybu FA a FR a

vytvoří z nich výslednou chybu EER. Čím menší je hodnota EER, tím nižší je chybovost a vyšší úspěšnost experimentu.

V *Tabulce 4* například vidíme, že u experimentu číslo 1 je vysoká úspěšnost TotalRate 80,0% i FalseRate 83,5%, ale nízká úspěšnost TrueRate 49,5%. Zde vyšla chybovost EER 33,5%. Naopak v experimentu číslo 3 vyšla vysoká úspěšnost TotalRate 90,1%, FalseRate 89,0% a jen o něco málo horší TrueRate 71,5%. Chyba je EER je v tomto případě nízká 19,3%. Proto při vyhodnocení experimentů upřednostňujeme chybovost EER, která nám dává objektivnější náhled na dosažený výsledek.

3.5 Popis testovacích sad Experimentu I

Cílem Experimentu I. bylo určit nejvhodnější nastavení parametrů rozpoznávání.. Nebylo tedy důležité pracovat se všemi jedinci, ale udělat co nejrozsáhlejší sady nahrávek. Po úvaze bylo pro Experiment I. použito celkem 5 sad s různým hlavním modelem jedince (TARGET). Detailní rozložení nahrávek mezi jednotlivé sady ukazuje *tabulka 5 až tabulka 9*.

Sada 1: TARGET jedinec A (Tabulka 5)

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	1	0	10	11
B	66	50	1	0	0	28
C	67	45	1	21	0	0
D	68	49	1	0	0	0
E	69	48	1	0	0	20
F	70	47	1	25	0	0
G	71	49	1	0	0	0
H	72	47	1	0	0	23
I	73	48	1	21	0	0
J	74	47	1	20	0	0
K	75	48	1	28	0	0
L	76	47	1	0	0	29
M	77	47	1	0	0	0
N	78	49	1	29	0	0
O	79	48	1	0	0	24
P	80	46	1	0	0	0
Q	81	48	1	0	0	0
R	82	50	1	0	0	25
S	83	45	1	30	0	0
T	84	47	1	0	0	27

Tabulka 5: Sada 1

Sada 2: TARGET jedinec E (Tabulka 6)

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	1	0	0	27
B	66	50	1	0	0	28
C	67	45	1	30	0	0
D	68	49	1	0	0	0
E	69	48	1	0	12	13
F	70	47	1	23	0	0
G	71	49	1	0	0	25
H	72	47	1	0	0	0
I	73	48	1	29	0	0
J	74	47	1	25	0	0
K	75	48	1	25	0	0
L	76	47	1	0	0	20
M	77	47	1	0	0	0
N	78	49	1	29	0	0
O	79	48	1	23	0	0
P	80	46	1	0	0	0
Q	81	48	1	0	0	0
R	82	50	1	0	0	24
S	83	45	1	22	0	0
T	84	47	1	0	0	29

Tabulka 6: Sada 2

Sada 3: TARGET jedinec J (Tabulka 7)

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	1	21	0	0
B	66	50	1	0	0	0
C	67	45	1	0	0	0
D	68	49	1	24	0	0
E	69	48	1	0	0	0
F	70	47	1	0	0	22
G	71	49	1	0	0	0
H	72	47	1	0	0	0
I	73	48	1	27	0	0
J	74	47	1	0	10	11
K	75	48	1	0	0	0
L	76	47	1	0	0	23
M	77	47	1	20	0	0
N	78	49	1	23	0	0
O	79	48	1	0	0	21
P	80	46	1	26	0	0
Q	81	48	1	0	0	0
R	82	50	1	0	0	27
S	83	45	1	0	0	30
T	84	47	1	23	0	0

Tabulka 7: Sada 3

Sada 4: TARGET jedinec N (Tabulka 8)

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	1	26	0	0
B	66	50	1	24	0	0
C	67	45	1	0	0	22
D	68	49	1	0	0	26
E	69	48	1	0	0	28
F	70	47	1	25	0	0
G	71	49	1	0	0	24
H	72	47	1	0	0	22
I	73	48	1	24	0	0
J	74	47	1	27	0	0
K	75	48	1	0	0	20
L	76	47	1	0	0	0
M	77	47	1	0	0	0
N	78	49	1	0	12	14
O	79	48	1	29	0	0
P	80	46	1	30	0	0
Q	81	48	1	0	0	0
R	82	50	1	0	0	29
S	83	45	1	0	0	28
T	84	47	1	25	0	0

Tabulka 8: Sada 4

Sada 5: TARGET jedinec R (Tabulka 9)

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	1	0	0	24
B	66	50	1	25	0	0
C	67	45	1	23	0	0
D	68	49	1	0	0	23
E	69	48	1	0	0	30
F	70	47	1	25	0	0
G	71	49	1	0	0	30
H	72	47	1	0	0	0
I	73	48	1	0	0	0
J	74	47	1	0	0	26
K	75	48	1	0	0	0
L	76	47	1	26	0	0
M	77	47	1	0	0	0
N	78	49	1	24	0	0
O	79	48	1	0	0	25
P	80	46	1	21	0	0
Q	81	48	1	0	0	29
R	82	50	1	0	10	11
S	83	45	1	23	0	0
T	84	47	1	27	0	0

Tabulka 9: Sada 5

3.6 Experiment I.: Zjištění optimálních parametrů

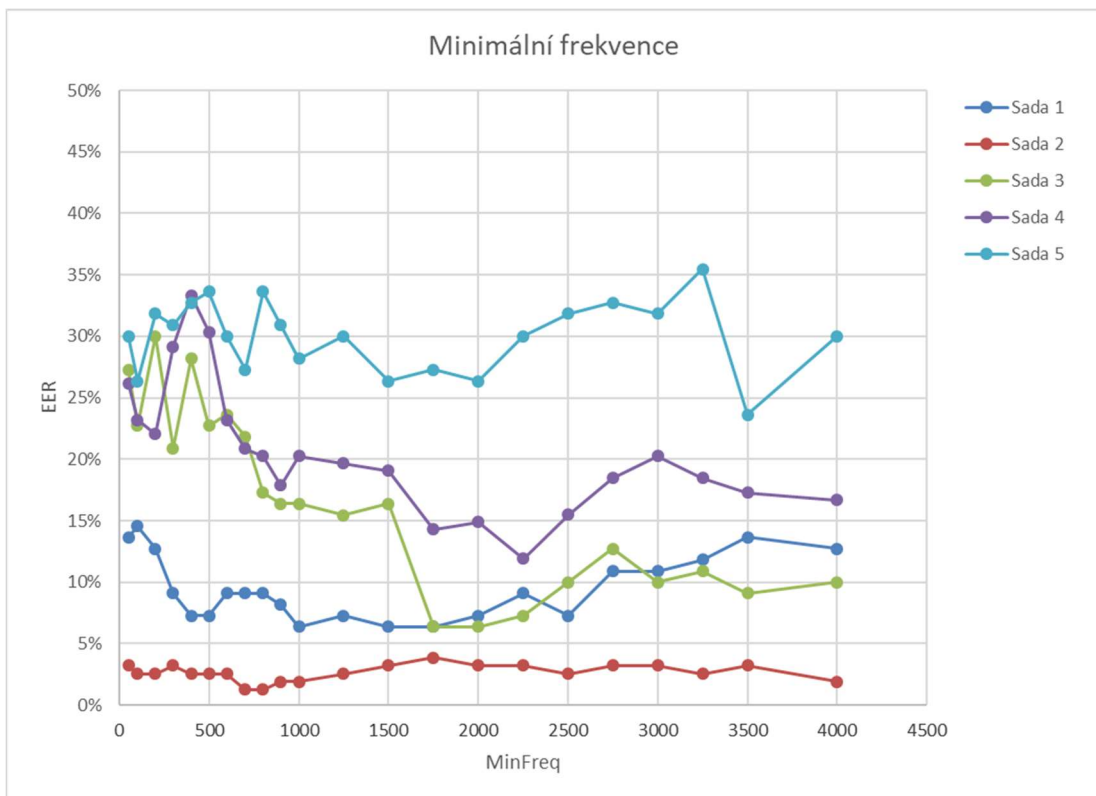
V této kapitole se stručně seznámíme se sledovanými parametry, které byly měněny v průběhu Experimentu I. za účelem nalezení jejich optimálního nastavení. Cílem bylo dosažení co nejnižší chyby EER. Tyto parametry budou dále využity pro Experiment II.

3.6.1 Minimální frekvence

Parametr MinFreq udává hodnotu, o kterou je frekvenční pásmo omezeno zleva, tedy hodnotu, o kterou je zúženo na nízkých frekvencích. Dosažené výsledky závislosti omezení chyby EER na MinFreq každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v *Tabulce 10* a *Grafu 1*.

MinFreq	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
50	14%	3%	27%	26%	30%	20%
100	15%	3%	23%	23%	26%	18%
200	13%	3%	30%	22%	32%	20%
300	9%	3%	21%	29%	31%	19%
400	7%	3%	28%	33%	33%	21%
500	7%	3%	23%	30%	34%	19%
600	9%	3%	24%	23%	30%	18%
700	9%	1%	22%	21%	27%	16%
800	9%	1%	17%	20%	34%	16%
900	8%	2%	16%	18%	31%	15%
1000	6%	2%	16%	20%	28%	15%
1250	7%	3%	15%	20%	30%	15%
1500	6%	3%	16%	19%	26%	14%
1750	6%	4%	6%	14%	27%	12%
2000	7%	3%	6%	15%	26%	12%
2250	9%	3%	7%	12%	30%	12%
2500	7%	3%	10%	15%	32%	13%
2750	11%	3%	13%	18%	33%	16%
3000	11%	3%	10%	20%	32%	15%
3250	12%	3%	11%	18%	35%	16%
3500	14%	3%	9%	17%	24%	13%
4000	13%	2%	10%	17%	30%	14%
MIN	6%	1%	6%	12%	24%	12%
MAX	15%	4%	30%	33%	35%	21%

Tabulka 10: Minimální frekvence EER



Graf 1: Minimální frekvence EER

Z výše uvedeného je patrné, že nejlepších výsledků, tedy nejnižšího EER, dosáhneme od hodnoty 1500Hz viz *Tabulka 11*.

parametr	doporučené nastavení (Hz)	dosažená minimální chyba EER
MinFreq	1750, 2000, 2250	12%
MinFreq	2500, 3500	13%
MinFreq	1500, 4000	14%
MinFreq	900, 1000, 1250, 3000	15%

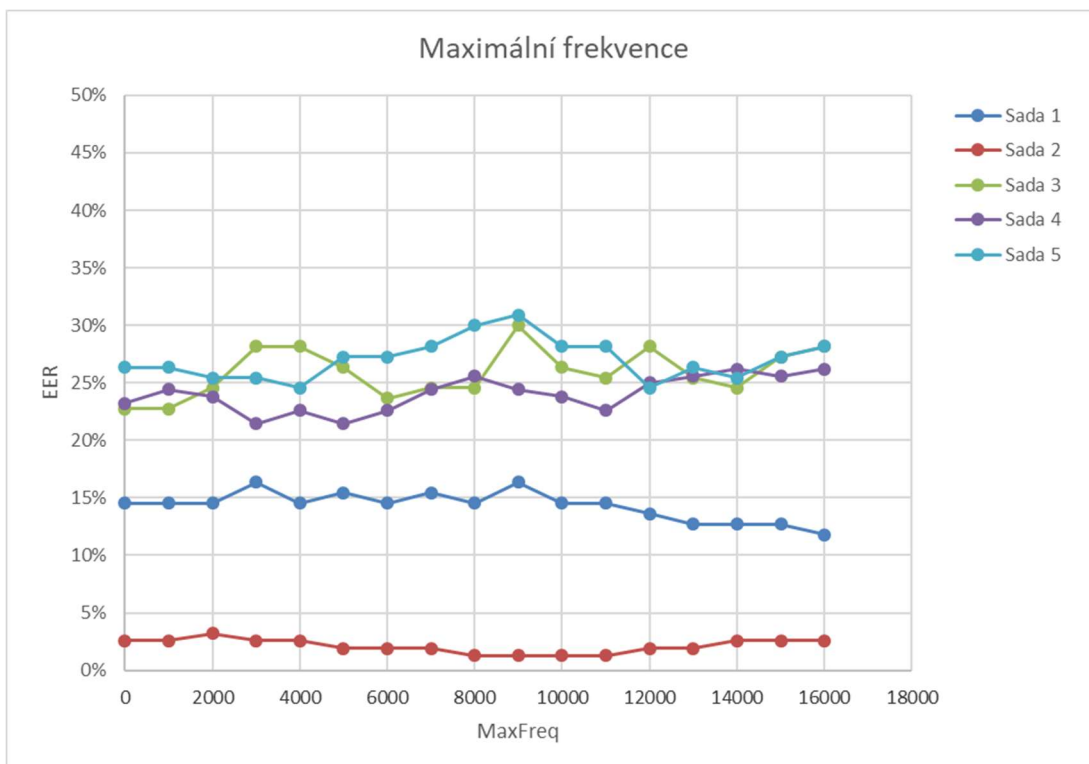
Tabulka 11: Doporučená minimální frekvence

3.6.2 Maximální frekvence

Parametr MaxFreq udává hodnotu, o kterou je frekvenční pásmo omezeno zprava, tedy o jakou hodnotu je sníženo od vzorkovací frekvence, v našem případě 22.500 Hz.. Dosažené výsledky závislosti chyby EER na omezení MaxFreq každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v *Tabulce 12* a *Grafu 2*.

MaxFreq	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
0	15%	3%	23%	23%	26%	18%
1000	15%	3%	23%	24%	26%	18%
2000	15%	3%	25%	24%	25%	18%
3000	16%	3%	28%	21%	25%	19%
4000	15%	3%	28%	23%	25%	18%
5000	15%	2%	26%	21%	27%	18%
6000	15%	2%	24%	23%	27%	18%
7000	15%	2%	25%	24%	28%	19%
8000	15%	1%	25%	26%	30%	19%
9000	16%	1%	30%	24%	31%	21%
10000	15%	1%	26%	24%	28%	19%
11000	15%	1%	25%	23%	28%	18%
12000	14%	2%	28%	25%	25%	19%
13000	13%	2%	25%	26%	26%	18%
14000	13%	3%	25%	26%	25%	18%
15000	13%	3%	27%	26%	27%	19%
16000	12%	3%	28%	26%	28%	19%
MIN	12%	1%	23%	21%	25%	18%
MAX	16%	3%	30%	26%	31%	21%

Tabulka 12: Maximální frekvence EER



Graf 2: Maximální frekvence EER

Z výše uvedeného vyplývá, že omezení maximální frekvence patrně nemá vliv na hodnotu EER. Chyba se pohybuje v rozmezí 18%-19%. Pouze na hodnotě omezení 9kHz je průměrná hodnota EER 21% viz *Tabulka 13*.

parametr	doporučené nastavení (Hz)	dosažená minimální chyba EER
MaxFreq	0, 1000, 2000, 4000, 5000, 6000, 11000, 13000	18%
MaxFreq	3000, 7000, 8000, 10000, 12000, 15000, 16000	19%
MaxFreq	9000	21%

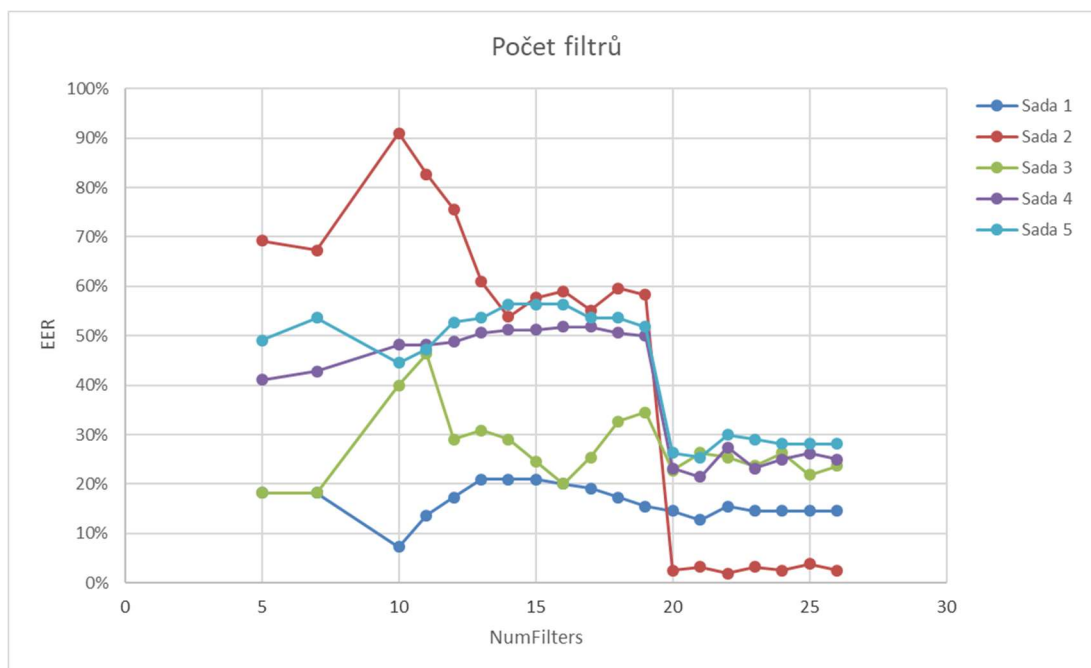
Tabulka 13: Doporučená maximální frekvence

3.6.3 Počet filtrů

Parametr NumFilters určuje počet trojúhelníkových filtrů, které dělí frekvenční pásmo, viz kapitola 2.3.4. Dosažené výsledky závislosti chyby EER na počtu NumFilters každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v *Tabulce 14* a *Grafu 3*.

NumFilters	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
5	18%	69%	18%	41%	49%	39%
7	18%	67%	18%	43%	54%	40%
10	7%	91%	40%	48%	45%	46%
11	14%	83%	46%	48%	47%	48%
12	17%	76%	29%	49%	53%	45%
13	21%	61%	31%	51%	54%	43%
14	21%	54%	29%	51%	56%	42%
15	21%	58%	25%	51%	56%	42%
16	20%	59%	20%	52%	56%	41%
17	19%	55%	25%	52%	54%	41%
18	17%	60%	33%	51%	54%	43%
19	15%	58%	35%	50%	52%	42%
20	15%	3%	23%	23%	26%	18%
21	13%	3%	26%	21%	25%	18%
22	15%	2%	25%	27%	30%	20%
23	15%	3%	24%	23%	29%	19%
24	15%	3%	26%	25%	28%	19%
25	15%	4%	22%	26%	28%	19%
26	15%	3%	24%	25%	28%	19%
MIN	7%	2%	18%	21%	25%	18%
MAX	21%	91%	46%	52%	56%	48%

Tabulka 14: Počet filtrů EER



Graf 3: Počet filtrů EER

Z výsledků můžeme vyčíst, že nejlepší EER je v rozmezí počtu filtrů 20-26, viz *Tabulka 15*

parametr	doporučené nastavení	dosažená minimální chyba EER
NumFilters	20,21,	18%
NumFilters	23-26	19%
NumFilters	22	20%

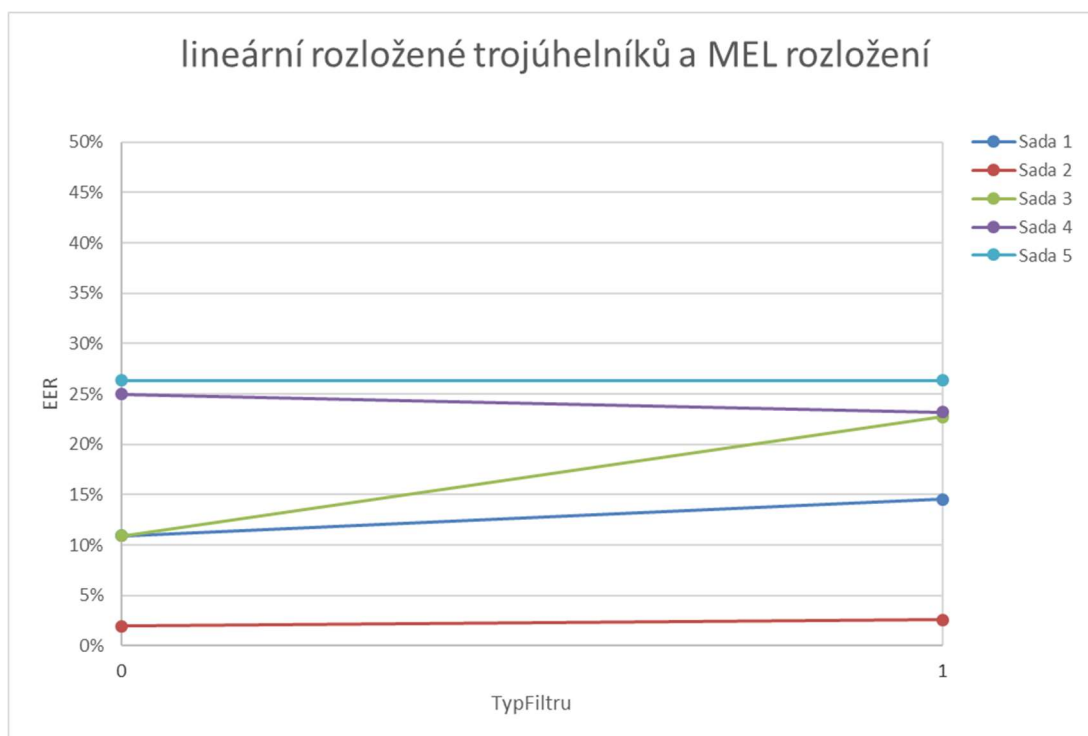
Tabulka 15: Doporučený počet filtrů

3.6.4 Typ filtru

Parametr TypFiltru nám udává filtr, který je použitý a to: 0- lineární filtr viz kapitola 2.3.5 a 1- MEL filtr viz kapitola 2.3.4. Dosažené výsledky závislosti chyby EER na typu filtru NumFilters každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v *Tabulce 16* a *Grafu 4*.

TypFiltru	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
0	11%	2%	11%	25%	26%	15%
1	15%	3%	23%	23%	26%	18%
MIN	11%	2%	11%	23%	26%	15%
MAX	15%	3%	23%	25%	26%	18%

Tabulka 16: Typ filtru EER



Graf 4: Typ filtru EER

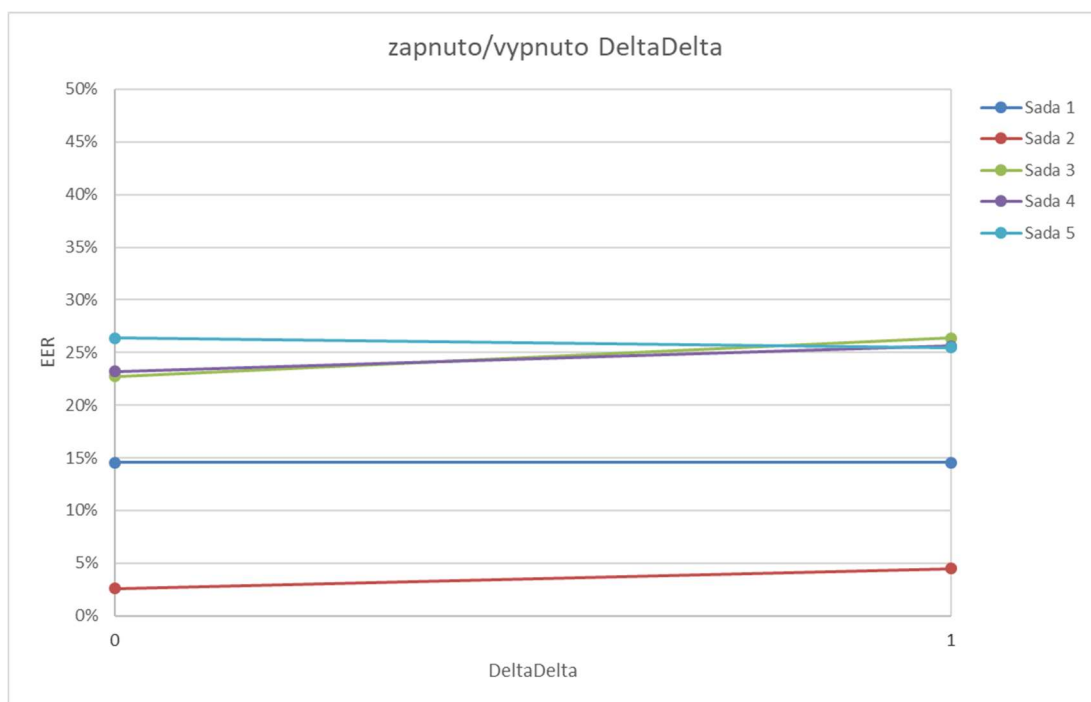
Parametr TypFiltru je nejlepší nastavit na hodnotu 0, tedy lineární filtr.

3.6.5 DeltaDelta

Parametr DeltaDelta udává stav vypnutých, nebo zapnutých DeltaDelta koeficientů. Dosažené výsledky závislosti chyby EER na stavu DeltaDelta každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v *Tabulce 17* a *Grafu 5*.

DeltaDelta	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
0	15%	3%	23%	23%	26%	18%
1	15%	4%	26%	26%	25%	19%
MIN	15%	3%	23%	23%	25%	18%
MAX	15%	4%	26%	26%	26%	19%

Tabulka 17: DeltaDelta EER



Graf 5: DeltaDelta EER

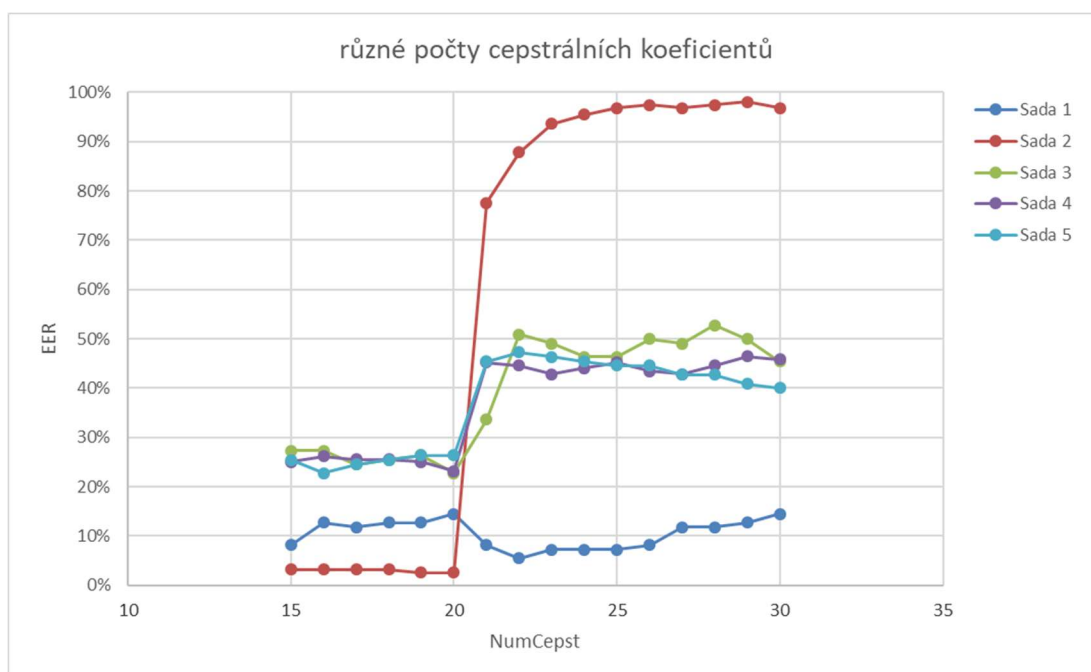
Z výsledků můžeme vyčíst, že zapnutí či vypnutí DeltaDelta filtrů nemá vliv na výsledné průměrné EER. Rozdíl je pouze nepatrný a to 18% a 19%.

3.6.6. Počet cepstrálních koeficientů

Parametr NumCepst udává počet cepstrálních koeficientů viz kapitola 2.3.2. Dosažené výsledky závislosti chyby EER na počtu NumCepst každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v Tabulce 18 a Grafu 6.

NumCepst	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
15	8%	3%	27%	25%	25%	18%
16	13%	3%	27%	26%	23%	18%
17	12%	3%	25%	26%	25%	18%
18	13%	3%	25%	26%	25%	18%
19	13%	3%	26%	25%	26%	19%
20	15%	3%	23%	23%	26%	18%
21	8%	78%	34%	45%	45%	42%
22	5%	88%	51%	45%	47%	47%
23	7%	94%	49%	43%	46%	48%
24	7%	96%	46%	44%	45%	48%
25	7%	97%	46%	45%	45%	48%
26	8%	97%	50%	43%	45%	49%
27	12%	97%	49%	43%	43%	49%
28	12%	97%	53%	45%	43%	50%
29	13%	98%	50%	46%	41%	50%
30	15%	97%	45%	46%	40%	49%
MIN	5%	3%	23%	23%	23%	18%
MAX	15%	98%	53%	46%	47%	50%

Tabulka 18: Počet cepstrálních koeficientů EER



Graf 6: Počet cepstrálních koeficientů EER

Dle dostupných dat můžeme nastavení cepstrálních koeficientů doporučit na rozmezí 15-20, viz *Tabulka 19*.

parametr	doporučené nastavení	dosažená minimální chyba EER
NumCepst	15, 16, 17, 18, 20	18%
NumCepst	19	19%

Tabulka 19: Doporučený počet cepstrálních koeficientů

3.6.7 Počet Delta koeficientů s vypnutým a zapnutým DeltaDelta

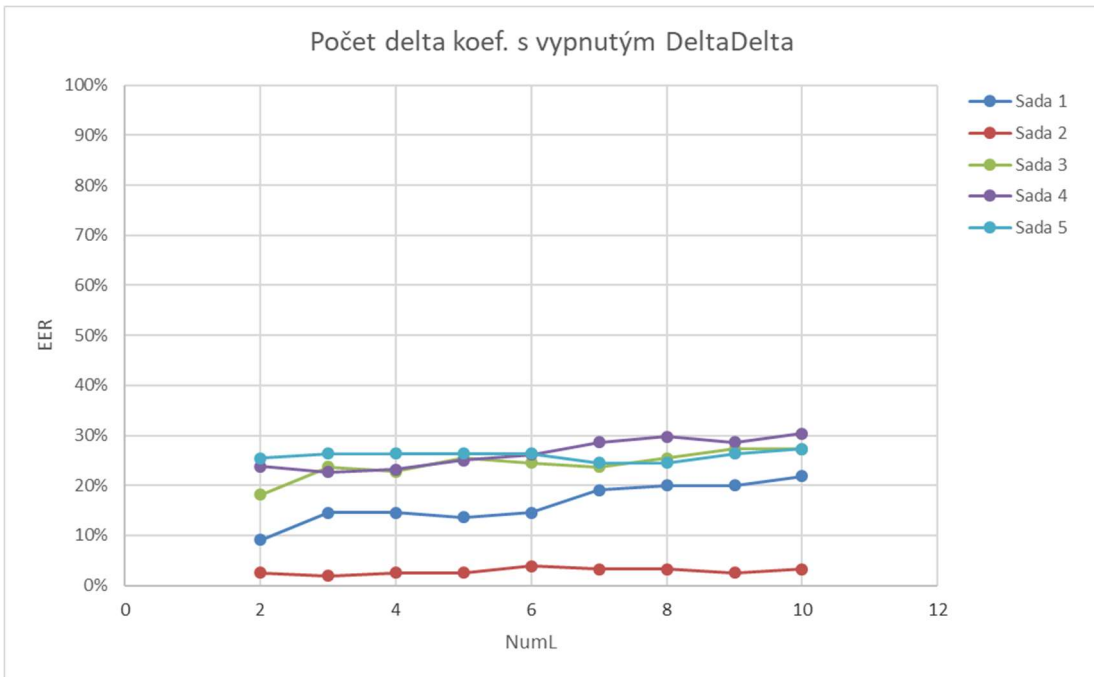
Parametr NumL udává počet Delta koeficientů navíc skombinovaný se zapnutým/vypnutým DeltaDelta. Dosažené výsledky závislosti chyby EER na NumL každé jednotlivé sady 1-5 a průměrné chyby EER, stejně tak jako minimální a maximální chyba EER každé sady jsou uvedeny v *Tabulce 20 a 21 a Grafech 7 a 8*.

NumL	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
2	9%	3%	18%	24%	25%	16%
3	15%	2%	24%	23%	26%	18%
4	15%	3%	23%	23%	26%	18%
5	14%	3%	25%	25%	26%	19%
6	15%	4%	25%	26%	26%	19%
7	19%	3%	24%	29%	25%	20%
8	20%	3%	25%	30%	25%	21%
9	20%	3%	27%	29%	26%	21%
10	22%	3%	27%	30%	27%	22%
MIN	9%	2%	18%	23%	25%	16%
MAX	22%	4%	27%	30%	27%	22%

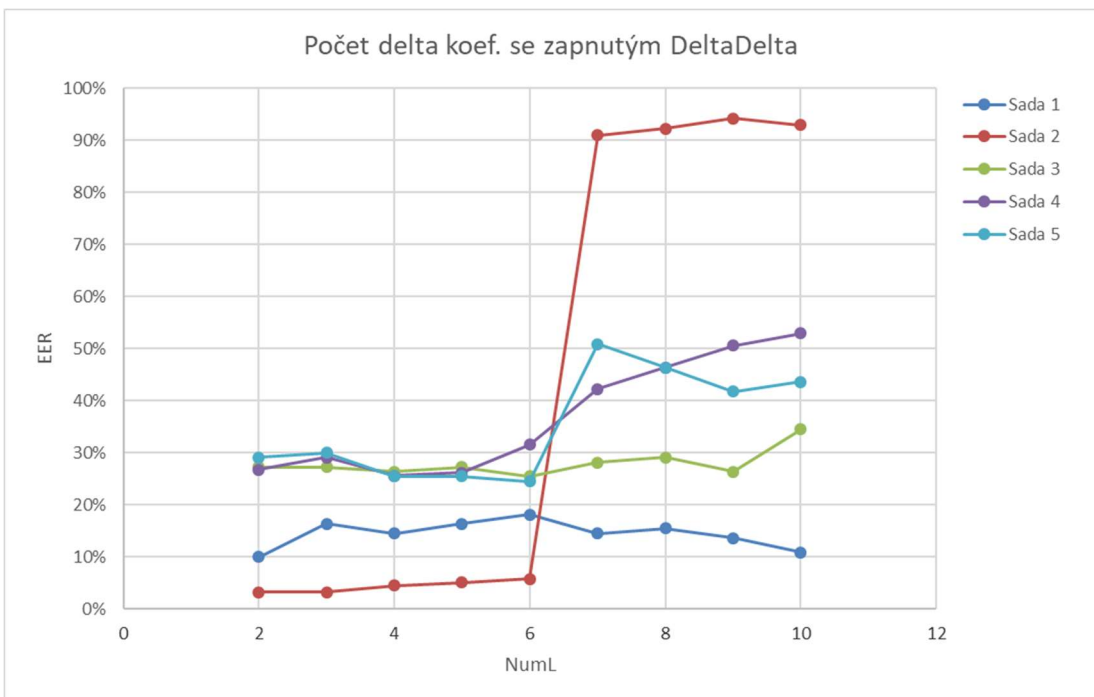
Tabulka 20: Počet Delta koeficientů s vypnutým DeltaDelta EER

NumL	Sada 1	Sada 2	Sada 3	Sada 4	Sada 5	Ø EER
2	10%	3%	27%	27%	29%	19%
3	16%	3%	27%	29%	30%	21%
4	15%	4%	26%	26%	25%	19%
5	16%	5%	27%	26%	25%	20%
6	18%	6%	25%	32%	25%	21%
7	15%	91%	28%	42%	51%	45%
8	15%	92%	29%	46%	46%	46%
9	14%	94%	26%	51%	42%	45%
10	11%	93%	35%	53%	44%	47%
MIN	10%	3%	25%	26%	25%	19%
MAX	18%	94%	35%	53%	51%	47%

Tabulka 21: Počet Delta koeficientů se vypnutým DeltaDelta EER



Graf 7: Počet Delta koeficientů s vypnutým DeltaDelta EER



Graf 8: Počet Delta koeficientů s vypnutým DeltaDelta EER

Dle dostupných dat můžeme počet Delta koeficientů s vypnutým DeltaDelta doporučit v celém rozsahu. Průměrné EER se jen nepatrně zvyšuje a to v rozmezí 16% - 22% viz Tabulka 22.

parametr	doporučené nastavení	dosažená minimální chyba EER
NumL	2	16%
NumL	3, 4	18%
NumL	5, 6	19%
NumL	7	20%
NumL	8, 9	21%
NumL	10	22%

Tabulka 22: Doporučený počet Delta koeficientů s vypnutým DeltaDelta

Se zapnutým DeltaDelta se průměrné EER pohybuje na rozsahu 2-6 v rozmezí 19% - 21% viz *Tabulka 23*.

parametr	doporučené nastavení	dosažená minimální chyba EER
NumL	2, 4	19%
NumL	5	20%
NumL	3, 6	21%

Tabulka 23: Doporučený počet Delta koeficientů se zapnutým DeltaDelta

3.7 Experiment II.: Verifikace jedinců s optimálními parametry

Cílem Experimentu II bylo ověření, jak automatický systém dokáže rozpoznat jedince rypoše na základě vokalizace. Jde o stěžejní experimenty, neboť ověření úspěšnosti automatického rozpoznávání je základním bodem mé bakalářské práce. Pro experimenty je využito nastavení parametrů, které byly nalezeny v Experimentu I.

Experiment II byl nakonec prováděn ve dvou fázích. Pro jednotlivé fáze používám označení Experiment II. fáze 1 a fáze 2. Původně bylo zamýšleno provést pouze jeden Experiment II. Ale jak uvidíme později, dosažené výsledky měly pro některé jedince chybu EER až kolem 50%, což bylo samozřejmě naprosto nevyhovující; takový výsledek se rovná čisté náhodě. Bylo proto rozhodnuto provést více fází Experimentu II., ve kterých byly některé parametry pozměněny.

V Experimentech II. fáze 1 i fáze 2 byly pro každého jedince použity 2 testovací sady. Tyto sady měly podobný počet nahrávek, ale jinou kombinaci jedinců použitých pro UBM a

TST (TRIAL). U obou sad byly použity nahrávky se všemi třemi mírami stresu. Sadu 1 a Sadu 2 si představíme na jedinci A (*Tabulka 24, Tabulka 25*). U ostatních jedinců B až T jsou použity obdobné sady.

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	2	0	22	21
B	66	50	2	0	0	41
C	67	45	2	39	0	0
D	68	49	2	38	0	0
E	69	48	2	0	0	0
F	70	47	2	41	0	0
G	71	49	2	40	0	0
H	72	47	2	0	0	39
I	73	48	2	0	0	0
J	74	47	2	0	0	39
K	75	48	2	0	0	41
L	76	47	2	0	0	42
M	77	47	2	42	0	0
N	78	49	2	0	0	0
O	79	48	2	38	0	0
P	80	46	2	0	0	41
Q	81	48	2	42	0	0
R	82	50	2	0	0	38
S	83	45	2	42	0	0
T	84	47	2	0	0	42

Tabulka 24: Vzorová Sada 1 jedince A

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	2	0	22	21
B	66	50	2	40	0	0
C	67	45	2	0	0	39
D	68	49	2	40	0	0
E	69	48	2	0	0	0
F	70	47	2	0	0	0
G	71	49	2	0	0	0
H	72	47	2	44	0	0
I	73	48	2	0	0	42
J	74	47	2	0	0	0
K	75	48	2	44	0	0
L	76	47	2	0	0	39
M	77	47	2	0	0	42
N	78	49	2	0	0	0
O	79	48	2	0	0	42
P	80	46	2	40	0	0
Q	81	48	2	43	0	0
R	82	50	2	0	0	39
S	83	45	2	0	0	40
T	84	47	2	38	0	0

Tabulka 25: Vzorová Sada 2 jedince A

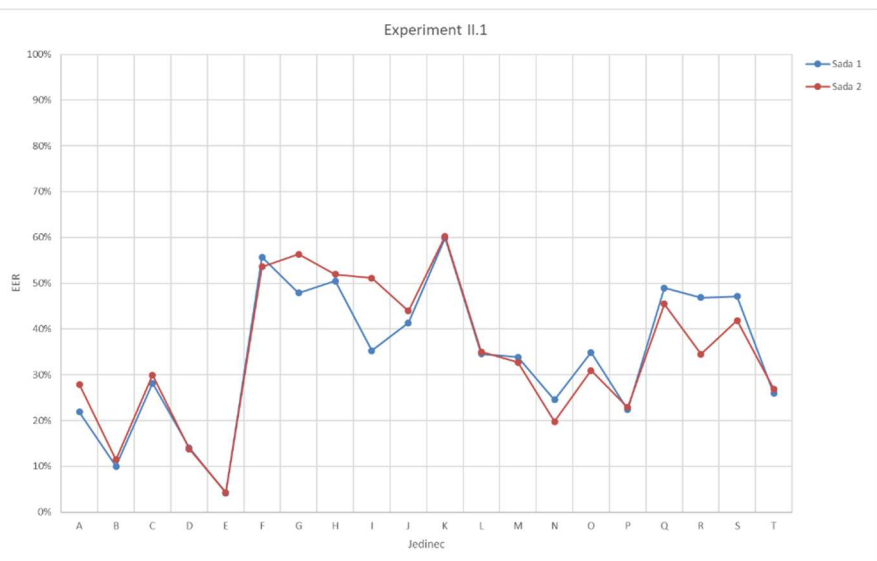
3.7.1 Experiment II. fáze 1

Parametry Experimentu II. fáze 1 byly nastaveny podle výstupu z předchozího Experimentu I:

- MinFreq: od 1500Hz
- MaxFreq: omezeno o 10000Hz
- NumFilter: 20
- TypFiltru: 0 – lineární
- Delta: 1 – zapnuto
- DeltaDelta: 0 – vypnuto
- NumCepst: 20
- NumL: 2

Výsledky tohoto experimentu jsou bohužel nevyhovující pro 9 rypošů (F, G, H, I, J, K, Q, R, Š). Chyba EER se ve 14x případech pohybuje v oblasti 40%-60%, což znamená, že u nich SW nedokázal uspokojivě rozeznat identitu jedince (*Tabulka 24, Graf 9*).

Jedinec	Sada 1	Sada 2
A	22%	28%
B	10%	11%
C	28%	30%
D	14%	14%
E	4%	4%
F	56%	54%
G	48%	56%
H	50%	52%
I	35%	51%
J	41%	44%
K	60%	60%
L	35%	35%
M	34%	33%
N	25%	20%
O	35%	31%
P	22%	23%
Q	49%	45%
R	47%	35%
S	47%	42%
T	26%	27%



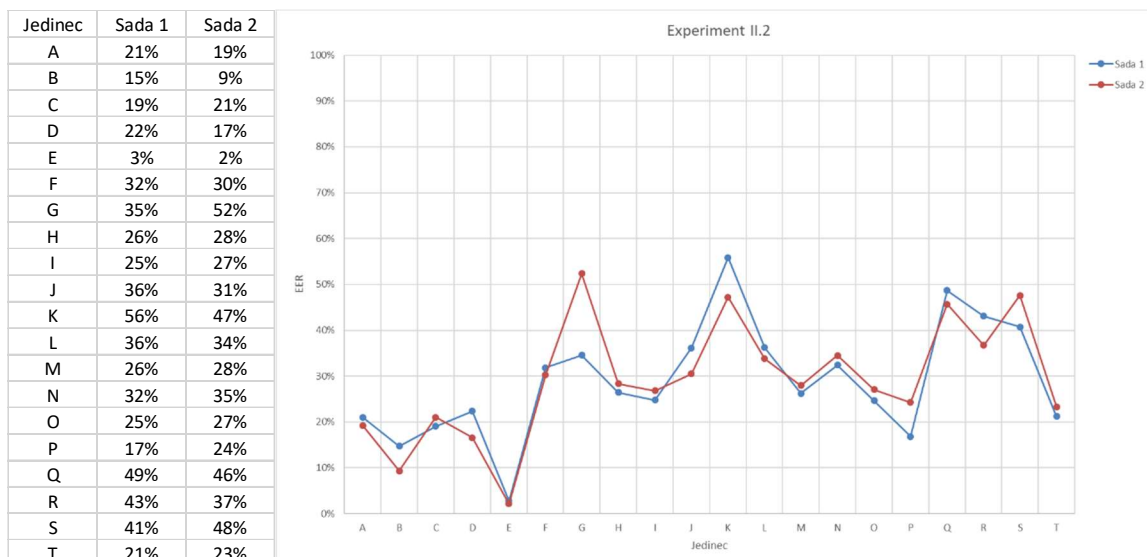
Tabulka 24: EER 1 Graf 9: EER Experimentu II.1

3.7.2 Experiment II. fáze 2

U Experimentu II. fáze 2 byly použity stejné sady jako u fáze 1. Při snaze o optimalizaci a zlepšení výsledného EER byly změněny některé parametry. Tyto změny nebyly voleny náhodně, ale až po provedení několika dílčích experimentů, kdy jsme se zaměřili na ty sady, jenž dávaly nejhorší výsledky. Následně jsme dospěli ke změnám těchto tří parametrů:

- MinFreq: od 500
- MaxFreq: 0 – bez omezení
- Delta: 0 – vypnuto

Snížil se počet experimentů s chybou EER pohybující se v oblasti 40%-60% a to z původních 14x na 8x. Nevyhovující výsledky se nyní týkají místo osmi jen pěti rypošů (G, K, Q, R, S), viz. Tabulka 25, Graf 10. Došlo tedy ke zlepšení, nicméně s výsledky stále nejsme spokojeni.



Tabulka 25: EER 2 Graf 10: EER Experimentu II.2

3.7.3 Experiment II. fáze 3

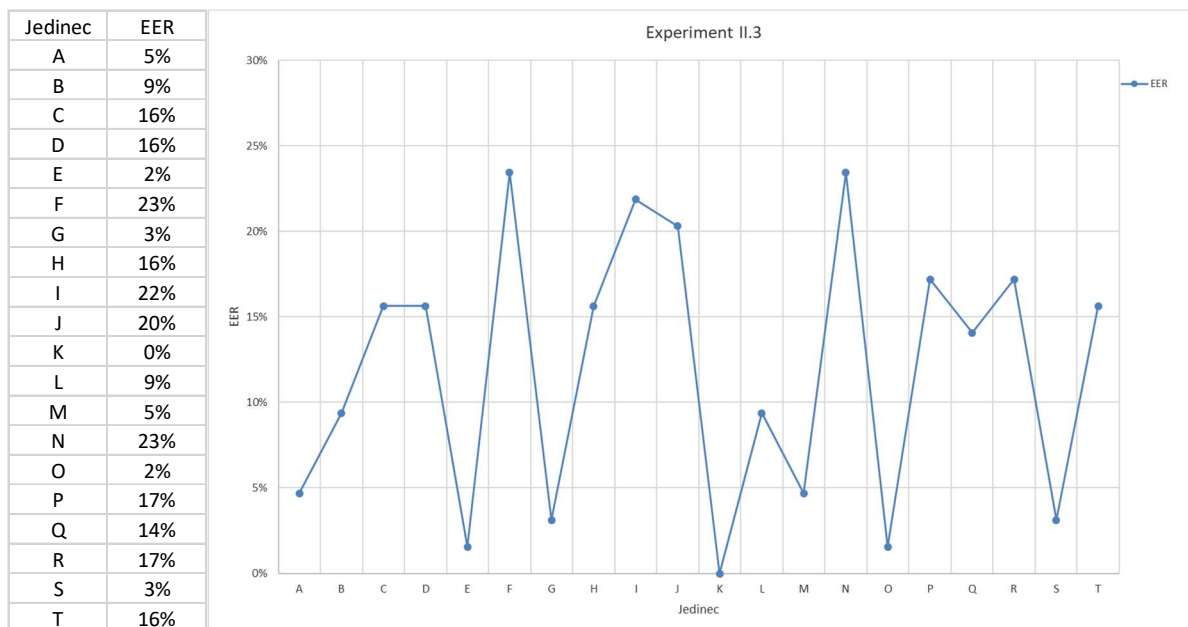
Pro Experiment II. fáze 3 bylo rozhodnuto provést rozpoznávání se stejně koncipovanými (namixovanými) sadami GMM, UBM a TST (TRIAL), avšak pouze s nahrávkami se stejnou nebo blízkou mírou stresu (1 a 2). To je hlavní rozdíl oproti fázi 1 a fázi 2, kde jsme kombinovali nahrávky se všemi mírami stresu dohromady. Jako příklad uvádíme sadu pro jedince A (Tabulka 26). Od hlavního jedince bylo pro GMM a TST použito 8 nahrávek pouze s nízkou úrovní stresu. Pro UBM a TST bylo od ostatních jedinců použito 15 nahrávek s nízkou a 5 se střední úrovní stresu. Pro jedince B až T je princip namíchání sad obdobný. Podrobnější popis sad pro Experimenty II jsou v příloze této práce.

Výsledky Experimentu II. fáze 3 jsou uváděny Tabulka 27 a Graf 11. Je patrné, že EER u žádného jedince nepřekročilo 23%. Došlo tedy k výraznému zlepšení. Software rozpoznal individualitu jedince s úspěšností mezi 77% (jedinci F, N) až 100% pro jedince K.

Je zřejmé, že míra stresu významně ovlivnila úspěšnost rozpoznání; software na její změnu citlivě reagoval. Lze proto usuzovat, že by měla být automaticky rozpoznatelná. Tento teoretický předpoklad potvrdí nebo vyvrátí až budoucí experimenty, které proběhnou mimo rámec této práce.

ID	ASCII	RecsCelk	Typ	UBM	GMM	TST
A	65	46	1	0	8	8
B	66	50	1	0	0	20
C	67	45	1	20	0	0
D	68	49	1	20	0	0
E	69	48	1	0	0	0
F	70	47	1	20	0	0
G	71	49	1	20	0	0
H	72	47	1	0	0	20
I	73	48	1	0	0	0
J	74	47	1	0	0	20
K	75	48	1	0	0	20
L	76	47	1	0	0	20
M	77	47	1	20	0	0
N	78	49	1	0	0	0
O	79	48	1	20	0	0
P	80	46	1	0	0	20
Q	81	48	1	20	0	0
R	82	50	1	0	0	20
S	83	45	1	20	0	0
T	84	47	1	0	0	20

Tabulka 26: Vzorová sada jedince A



Tabulka 27: EER3 Graf 11: EER Experimentu II.3

4 Závěr

Téma mé práce mě zaujalo z hlediska práce se zvukem. Již před lety jsem se na výstavě ForArch setkal s ukázkou domácnosti ovládané pouze hlasem. Tato technologie mě nadchla a začal jsem se o tuto problematiku zajímat blíže.

V rámci mé bakalářské práce bylo prvním krokem seznámení se zvířetem, které jsem do té doby vůbec neznal. Protože nejsem zaměřen na biologii, byla pro mě práce se zvířaty nová zkušenost. Rypoš je zvíře výjimečné svou dlouhověkostí, způsobem života a odolností proti tzv. civilizačním chorobám. Při práci s nimi jsem musel překonávat různá úskalí. Například někteří jedinci nechtěli reagovat na mé podněty, nebo byli agresivní bez příčiny, či jsme je s dr. Hrouzkovou nemohli ani odchytit. Nahrávání se ukázalo jako časově náročné, pro každého jedince to znamenalo opakovat jeho odchyt, přípravu prostředí a vlastní záznam. Často docházelo k prodlevám, kdy zvíře nechtělo vokalizovat. Nahrávání nakonec vyžadovalo několik týdnů, kdy jsem získával potřebný počet nahrávek.

Když jsem postupně získal relevantní data, mohl jsem přistoupit k jejich zpracování, což obnášelo základní vyčištění nahrávek od nežádoucích ruchů a jejich nastříhání. K tomu účelu jsem využil program Audacity. Tento freeware program se mi osvědčil. Je přehledný, uživatelsky přívětivý a pro tuto práci naprosto dostačující. Nahrávky jsem předal svému vedoucímu práce.

Následovalo vytvoření testovacích sad a jejich uložení na definované úložiště. Poté probíhala příprava vlastních experimentů a jejich spouštění. Poslední fází bylo zpracování výstupů, které systém generuje v podobě matic a exportuje do Excelu.

Každá z uvedených fází trvala několik dní. Z tohoto pohledu byla náročná zejména příprava testovacích sad a to s ohledem na nutnost neustálé křížové kontroly. Příprava každé sady zabrala čas mezi pěti až dvaceti minutami plus další čas nutný na ověřování možných duplicit. Bylo využito 105 sad (5x Experimenty I., 100x Experimenty II.), ale celkem jich bylo připraveno více než 150. Vlastní experimenty pak běžely v rozmezí hodin až dnů plus jejich příprava. Experiment I. čítal 560 iterací a trval tři dny. Experiment II. fáze 1 měl 80 iterací a trval den. Experiment II. fáze 2. měl iterací 40 a běžel 12 hodin. A konečně Experiment II. fáze 3. byl nastaven na 20 iterací a jeho trvání bylo dvě hodiny.

Výsledky ukázaly, že automatické rozeznávání jedinců rypoše lysého na základě jejich vokalizace, je při správném zadání parametrů úspěšné. V Experimentu II. fáze 3, který

považuji za finální výstup, se nám podařilo, díky použití omezené míry motivace nahrávek a správně nastavené kombinaci parametrů, rozeznat individualitu jedince s úspěšností mezi 77% až 100%. Pro devět experimentů, což je téměř polovina všech jedinců, bylo dosaženo úspěšnosti rozpoznání $\geq 90\%$. Pro sedmnáct jedinců byla míra rozpoznání $\geq 80\%$.

Z dosažených výsledků zároveň vyplynulo, že míra stresu významně ovlivňuje míru úspěšnosti rozpoznání. Z toho lze usuzovat, že jí bude možné automaticky rozpoznat s využitím vhodně zvolených identifikátorů.

Z rozhodnutí vedoucího práce bylo určeno, že součástí mé bakalářské práce nebudou experimenty rozeznání pohlaví, velikosti a motivace jedince. Pro tyto parametry by bylo nutné porovnávat nejen chyby EER, ale též tzv. matice \mathbb{C} , obsahující parametry nahrávek (energie, cepstrální koeficienty, delta koeficienty, ad.). S těmito maticemi se operuje v Matlabu a slouží pro získání statistických modelů GMM a UBM. Nakonec se totiž ukázalo, že samotné provedení Experimentů I a II bylo tak časově náročné, že by nebylo reálné uведенé cíle v rámci mé práce naplnit. Současně s tím je nutno konstatovat, že teoretická náročnost takových postupů by výrazně přesahovala rámec bakalářské práce.

Do této práce nebyli zahrnuti jedinci, kteří jsou v kolonii rypošů lysých jedinců, kteří se množí. Tedy hlavní samice a samci. Existuje domněnka, že jejich vokalizace se od ostatních ve skupině liší. Toto by mělo být předmětem dalšího zkoumání.

Doufám, že má práce bude moci sloužit jako podklad pro další práce, týkající se rozpoznávání zvířat a tím povede k lepšímu poznání života a způsobu komunikace rypošů, a tím k bližšímu porozumění mezi lidmi a zvířaty.

Seznam literatury

- [1] DON E. WILSON, THOMAS E. LACHER, JR, RUSSELL A. MITTERMEIER. (2016), Handbook of the Mammals of the World . Volume 6. Lynx Edicions, Barcelona. ISBN 978-84-941892-3-4
- [2] NEVO E. (1999), Mosaic Evolution of Subterranean Mammals. Regression, Progression, and Global Convergence. Oxford University Press, Oxford. ISBN 9780198575726
- [3] UHLÍŘ, J., SOVKA, P., POLLÁK, P., HANŽL, V., ČMEJLA, R. (2007): Technologie hlasových komunikací, ČVUT, Praha. ISBN 978-80-01-03940-3
- [4] PSUTKA, J., MÜLLER, L. ET AL. (2006): Mluvíme s počítačem česky. Academia, Praha. ISBN 80-200-1309-1
- [5] HORÁČEK, J., ŠVEC, J. (2008): Modelování lidského hlasu, Vesmír 87, 833, 2008/12
- [6] Handheld Solid State Recorder. Marantz Professional [online]. [cit. 2021-03-27]. Dostupné z: <https://www.marantzpro.com/products/view/pmd661mkii> [7] Sennheiser [online]
- [7] SENNHEISER. *SENNHEISER.cz* [online]. [cit. 2021-03-27]. Dostupné z: <https://www.sennheiser.cz/index.php?id=23&produkt=8440>
- [8] PTÁČEK, Ladislav. *Automatická identifikace a verifikace ptáků*. Plzeň, 2016. Disertační práce. Západočeská univerzita v Plzni Fakulta aplikovaných věd. Vedoucí práce Doc. Ing. Luděk Müller, Ph.D.
- [9] HETH G.; FRANKENBERG E.; NEVO E. 1986: Adaptive optimal sound for vocal communication in tunnels of a subterranean mammal (*Spalax ehrenbergi*). *Experientia* 42: 1287-1289.
- [10] HEFFNER R. S.; HEFFNER E. H. 1993: Degenerate hearing and sound localization in naked mole rats (*Heterocephalus glaber*), with an overview of central auditory structures. *The Journal of Comparative Neurology*. 331(3): 418-433.
- [11] MORTON E. S. 1977: Occurrence and significance of motivation structural rules in some bird and mammals sounds. *American Naturalist* 111(981): 855-869.
- [12] KNOTKOVÁ E.; VEITL S.; ŠUMBERA R., SEDLÁČEK F.; BURDA H. 2009: Vocalization of the silvery mole-rat, comparison of vocal repertoire in subterranean rodents with different social systems. *Bioacoustics*, 18(3): 241-257.

- [13] FREEBERG T. M.; DUNBAR R. I.; ORD T. J. 2012: Social complexity as a proximate and ultimate factor in communicative complexity. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 1785.
- [14] BRANDBURRY, J. W., VEHRENCAMP SANDRA L. (1998). *Principles of Animal Communication*, Second Edition. Sinauer Associates, Inc.; First edition. ISBN 978-0878933761
- [15] HETH G.; TONDRAK J.; BURDA H. 2002: Individual Odor Similarities Within Colonies and Across Species of *Cryptomys* Mole Rats. *Journal of Mammalogy*. 83: 569-575.
- [16] DVOŘÁKOVÁ V. 2013. Individual features in vocalization of the Mashona mole-rat (*Fukomys darlingi*). Mgr. thesis. Department of Zoology, University of South Bohemia, Czech Republic.
- [17] PEPPER J. W.; BRAUDE S. H.; LACEY E. A.; SHERMAN P. W. 1991: Vocalizations of the naked mole-rat. In: *The Biology of the Naked Mole-Rat* (Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D., eds). Princeton Univ. Press, Princeton, NJ.

Přílohy

Součástí této práce je příloha s názvem *Popis sad*, obsahující popis sad nahrávek, které byly použity pro Experimenty II. Příloha je uložena jako samostatný soubor do systému STAG.