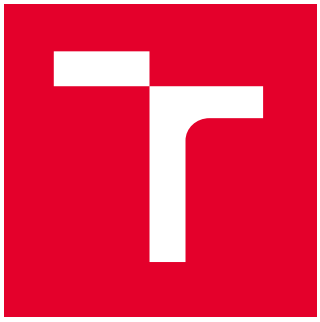


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

VZTAH EMOCÍ A INTONAČNÍCH KŘIVEK

THE RELATION OF EMOTIONS AND INTONATION CURVES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Radka Gavlasová

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. Jana Tučková, CSc.

BRNO 2022

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Bakalářská práce

bakalářský studijní program **Audio inženýrství**
specializace Zvuková produkce a nahrávání
Ústav telekomunikací

Studentka: Radka Gavlasová ***ID:*** 221466
Ročník: 3 ***Akademický rok:*** 2021/22

NÁZEV TÉMATU:

Vztah emocí a intonačních křivek

POKYNY PRO VYPRACOVÁNÍ:

Definujte emoce resp. emoční postoje. Na jejich základě popište intonační křivky vět realizovaných zvolenými vybranými emocemi (emočními postoji). Na základě těchto intonačních křivek určete pasívní a aktivní emoce. Z nahrávek čtyř promluv obsahujících zvolené emoce (emoční postoje) tyto emoce klasifikujte dvěma mluvčími (bude se jednat o hrané, nikoliv spontánní emoce). Práci řešte v prostředí MATLAB. Ke klasifikaci použijte umělé neuronové sítě. Výsledky porovnejte poslechovými testy a analytickým vyhodnocením získaných výsledků.

DOPORUČENÁ LITERATURA:

[1] Vlčková, J. Prozodie, cesta i mříž porozumění. Vyd. 1. Praha: Karolinum, 2006. ISBN 80-246-1266-6.
[2] Crystal, D.: A Little Book of Language. Yale University Press, 2010. ISBN 9780300155334 (ISBN10: 0300155336)

Termín zadání: 7.2.2022 ***Termín odevzdání:*** 31.5.2022

Vedoucí práce: prof. Ing. Jana Tučková, CSc.

doc. Ing. Jiří Schimmel, Ph.D.
předseda rady studijního programu

ABSTRAKT

Tato práce se zabývá intonačními křivkami s jejími vztahy pro různé emoce. Kromě teoretického základu, který pojednává o tvorbě řeči, zpracování signálů a psychologického nastínění rozdělení emocí, obsahuje také tvorbu vlastní emotivní databáze realizované s profesionálními herci. Cílem této závěrečné práce je klasifikace signálu na základě emoce, kterou nahrávka má představovat. Těmito emocemi jsou hněv, radost, nuda a smutek. Klasifikace probíhala pomocí umělých neuronových sítí, konkrétně v aplikaci Classification Learner, kterou poskytuje programovací prostředí Matlab. Použité příznaky pro tuto metodu byly variace fundamentální frekvence a MFCC. Výsledky byly následně porovnány a zanalyzovány poslechovým testem. Tento test pomohl určit, zda jsou výsledky relevantní pro tuto problematiku. Maximální úspěšnost trénování sítě dosáhla přibližně 82 %, testování pak 75 %. Poslechové testy potvrdily, že výsledky odpovídají předpokládanému lidskému vnímání. Pro podrobnější a lepší vyhodnocení, by bylo zapotřebí větší a kvalitnější databáze.

KLÍČOVÁ SLOVA

Intonační křivky, emoce, umělé neuronové sítě, Matlab, Classification Learner, fundamentální frekvence, MFCC

ABSTRACT

This thesis deals with intonation curves and their relation to human emotions. Besides the theoretical part where you can learn about speech production, signal processing and psychological distribution of emotions, there is also a unique database recorded with the help of two professional actors. The main goal of this thesis is to classify created data using artificial neural networks into four classes. Those classes are anger, joy, boredom and sadness. The practical part was implemented in a programming platform called Matlab using Classification Learner app. Features used for this method were variations of fundamental frequency and MFCC. The results were compared with a listening survey so that it could be determined whether the results provided by neural network are relevant to some kind of a human factor. Success rate of the trained models reached 82 %, new data testing reached 75 %. Listening survey confirmed that the results correspond to the assumption of human perception. Better succes rate would be accomplished by using a bigger set of higher quality data.

KEYWORDS

Intonation curves, emotions, artificial neural network, Matlab, Classificiation Learner, fundamental frequency, MFCC

GAVLASOVÁ, Radka. *Vztah emocí a intonačních křivek*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2022, 68 s. Bakalářská práce. Vedoucí práce: prof. Ing. Jana Tučková, CSc.

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Radka Gavlasová
VUT ID autora:	221466
Typ práce:	Bakalářská práce
Akademický rok:	2021/22
Téma závěrečné práce:	Vztah emocí a intonačních křivek

Prohlašuji, že svou závěrečnou práci jsem vypracovala samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autorky*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucí své bakalářské práce paní prof. Ing. Janě Tučkové, CSc. za odborné vedení, trpělivost a podnětné návrhy k práci. Dále děkuji panu Ing. Danielu Kováčovi za odborné konzultace a vstřícnost při řešení problematiky praktické části. Děkuji Vendule Příhodové a Janu Sklenářovi za jejich ochotu a čas, který mi věnovali během nahrávání databáze. V neposlední řadě patří mé díky rodině a přátelům, kteří mě vždy bezmezně podporovali jak ve studiu, tak na cestě za mými sny.

Obsah

Seznam obrázků

Úvod

Emoce ovlivňují náš život denně. Je to jedna z vlastností, která dělá člověka člověkem. V dnešní době je běžné, že se lidská řeč modeluje uměle. Na základě signálové analýzy jsme schopni vytvořit umělou řeč, která se bude podobat té reálné. Samozřejmě tato problematika má několik aspektů, na které je třeba brát při rekonstruování lidského hlasu ohled. Jedním z těchto aspektů jsou nejen lidské emoce, ale právě i intonační křivky, na které může být nahlíženo z různých pohledů. Nejen z hlediska samotného zpracování signálů, ale také například různá intonace způsobená jazykem, ve kterém probíhá komunikace. Cílem této bakalářské práce je vytvořit databázi několika vzorků řečového signálu v různých emocích a tyto nahrávky zanalyzovat a klasifikovat v prostředí Matlab

V teoretické části bude pojednáno o řeči z hlediska její tvorby, ale i o funkci samotných hlasivek. Lidský hlas je nezbytná součást verbální lidské komunikace, ale to platí i pro intonaci, která hraje velice důležitou roli v emoční klasifikaci. Stručně bude zmíněna i různorodost intonace v závislosti na určitém jazyku. V každém jazyce se s intonací totiž pracuje jinak a považuji za důležité a zajímavé to zmínit. Tato práce se zabývá pouze intonačními křivkami, které se uplatňují v českém jazyce. V dalších kapitolách budou také rozebrány jednotlivé konkrétní emoce, které budou předmětem klasifikace. Záměrně jsou voleny pouze dvě dvojice emocí, z nichž obě z dvojice jsou *opačné* povahy.

V praktické části se zaměříme na proces realizace databáze a klasifikaci nahrávek. Pro potřeby této práce byla vytvořena vlastní databáze pomocí dvou herců. Stále se jedná o emoce předstírané, nikoli spontánní, ale z tohoto důvodu byli záměrně voleni lidé s hereckou profesí. Nahrávat skutečné emoce by bylo nereálné, ne-li až nelegální a nebylo by takovým způsobem možné vytvořit plnohodnotnou databázi s dostatečným počtem vzorků. Testovací část databáze bude následně předmětem testu poslechového, který určí, jak moc odpovídá použitá metoda lidskému vnímání emocí.

1 Lidský hlas a řeč

1.1 Tvorba řeči

Řeč je u člověka velmi důležitým dorozumívacím prostředkem. Patří mezi verbální druh komunikace, kterým dokážeme vyjádřit co chceme, jak se cítíme, apod. Řečí, konkrétně jazyky, se zabývá věda s názvem lingvistika, o které bude blíže pojednáno v kapitole 1.2. Avšak v této kapitole se zaměříme na řeč z hlediska její tvorby. Přiblížíme si jednotlivé aspekty a modely znázorňující proces, který při tvorbě probíhá.

1.1.1 Proces tvorby a hlasivky

Hlasivky jsou párový orgán nacházející se na vnitřní straně hrtanu. Součástí struktury hlasivky je hlasivkový sval a hlasový vaz po obou stranách přichycený k chrupavce. Princip je podobný jako u tvorby zvuku dechových nástrojů, přestože byl původně mylně připodobňován principu rozechvívání struny. Později byl přirovnán ke kmitání vzduchového sloupce v píšťale s dvěma protiráznými jazýčky. [1, str. 47]

Proces začíná v plicích, které dodají potřebný vzduch. Ten poté prochází přes hlasivkovou štěrbinu, která je obklopena samotnými hlasivkami. [2, str. 6] Hlasivky se vlivem proudění vzduchu rozkmitají a začnou produkovat pravidelný signál podobající se pilovitému průběhu. Jedinečnost každého hlasu zajišťuje základní frekvence, na které hlasivky kmitají (více v kapitole 1.3).

Tyto signály jsou následně zpracovávány dalšími částmi našeho těla, zejména ústní dutinou. Vyzařování řečového signálu ale neprobíhá pouze z dutiny ústní, ale také nosní. Poměr tohoto vyzařování se dá korigovat na základě různých technik, a to zejména u zpěvu.

1.1.2 Barva hlasu

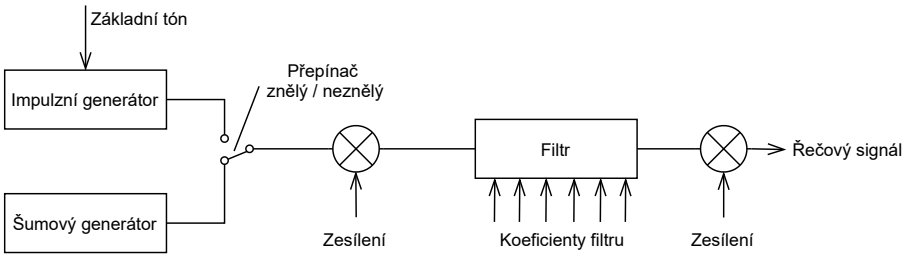
Spektrum signálu, který produkujeme našimi hlasivkami nám pomáhá rozlišovat různé hlasy stejně jako je naše ucho schopno rozlišit jednotlivé hudební nástroje (tedy jejich barvu). Barva našeho hlasu je ovlivněna způsobem kmitání hlasivek, stavbou hrtanu i celého těla, dále pak i dutiny, ve kterých se hlas dál mění. Ani s věkem nezůstává tato vlastnost beze změny, během života dochází ke změnám barvy, která je výraznější u chlapců procházejících pubertou.

1.1.3 Modely

Pro lepší pochopení funkčnosti a možnosti zkoumání bylo vytvořeno několik modelů na tvorbu řeči. Mezi ty nejdůležitější patří model elektronický a model akustický. [2, str. 8] V následujících dvou podkapitolách si tyto jednotlivé modely stručně přiblížíme.

Elektronický model

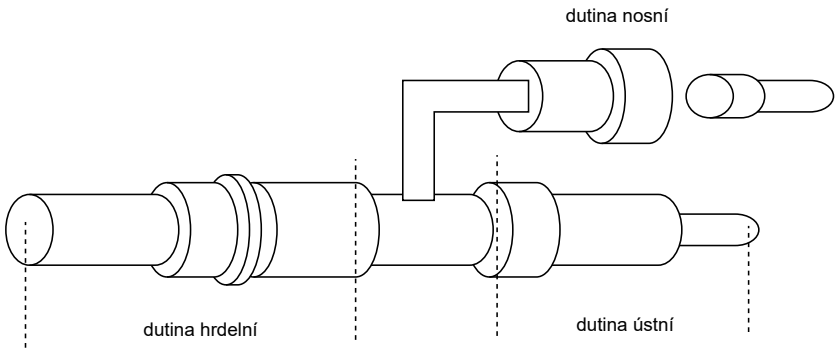
Z hlediska analogie lze hlasový trakt považovat za lineární přenosový systém. Na obr. 1.1.3 je tento systém zobrazen. Základem obvodu je filtr a zdroj buzení. Jsou zde dva typy buzení – jeden pro znělé hlásky, druhý pro neznělé. Znělé hlásky (např. samohlásky) totiž produkují periodický signál podobný pilovitému průběhu, naopak neznělé produkují pouze šum. [2, str. 8]



Obr. 1.1: Elektronický model tvorby řeči (podle [2])

Akustický model

Jedná se o válcový model s několika dutinami (hrdelní, nosní a ústní), který *„vychází z představy dutinového rezonátoru“*. Lze použít ale také zjednodušený model, kde se dutina nosní neuvažuje. [3, str. 12]



Obr. 1.2: Válcový akustický model tvorby řeči (podle [2])

1.2 Lingvistika

Lingvistika, česky *jazykověda*, je obor zabývající se studiem jazyka. Dělí se na spoustu dalších podoborů v závislosti na tom, kterým jazykem či kterou částí se konkrétně zabývá. Jedna část lingvistiky se například věnuje studiu jazyků, které jsou podobné nebo patří do stejné skupiny jazyků příbuzných (např. germanistika). Jiná zkoumá jazyk z hlediska fonologie a fonetiky (a dalších oborů) – tímto odvětvím lingvistiky se budeme zabývat v následujících podkapitolách. Jedná se o zvukovou stránku jazyka. [4, str. 16]

Jazyk má velice mnoho funkcí a hraje významnou roli jak v kultuře jednotlivých zemí, tak i v historii a samotném vývoji lidstva. Velmi zajímavé je také pozorovat, jak se dorozumívají různí živočichové. Existují studie na komunikaci ptáků, šimpanzů nebo jelenů v období říje. Přestože se zvířecí řeč zásadně liší od té lidské, běžně se osvědčuje její pozorování za účelem analýzy a hledání spojitostí.

Součástí lingvistiky je i samotná intonace, které je vzhledem k jejímu obsáhlému pojetí věnovaná samostatná kapitola 1.3.

1.2.1 Prozodie

Prozodie je oblast lingvistiky, která popisuje vlastnosti zvukové stavby jazyka. Někdy se jako synonymum používá pojem tzv. suprasegmentálních rysů. *„Celek suprasegmentálních neboli prozodických vlastností promluvy je definován jako souhrn intonace, přízvuku, rytmu, sylabizace a distribuce pauz a může být rozdělen do dvou velkých skupin: rytmus a intonace“* [5, str. 19]

Všechny tyto jevy jsou dány změnou základní frekvence. Např. zvýšení základního tónu bude naznačovat důležitost určité části věty. Naopak pokles hlasu či promluva nízkým hlasem dává najevo hlasité vyjádření myšlenek mluvčího nebo zkrátka sděluje méně podstatné informace. Tyto změny výšky hlasu se nazývají intonace a bude podrobněji probrána v podkapitole 1.3. Otázkou stále zůstává, zda je prozodie univerzální pro všechny jazyky. Některé studie zastávají názor, že ano, jiné naopak, že se to jazyk od jazyka liší. J. Vlčková ve své knize *Prozodie, cesta i mříž porozumění* píše, že spousta studií došla k závěru, že některé prozodické rysy se v různých jazycích opakují, tím pádem jsou jistým způsobem *univerzální*. U většiny světových jazyků zpravidla platí princip melodie klesavé (věta oznamovací) a stoupavé (věta tázací). Toto se nám zdá být přirozené už od dětství a jednou z příčin je svalové napětí hlasivek. *„Na konci deklarativní věty pozornost ochabuje a svalové napětí klesá; to s sebou přináší pokles aktivity hlasivek a snížení výšky základního tónu hlasu. Naopak napětí provázející zájem manifestuje zejména v koncové části interogativní věty a finální melodie stoupá.“* [5, str. 29 – 30].

1.2.2 Tempo

Tempo mluveného slova, neboli časový průběh řeči, se liší v závislosti na konkrétním mluvčím a podílí se na *vnímání rytmu promluvy.* Individuální tempo se neliší jen u jednotlivých mluvčích, ale také v rámci jazyka a je velmi nestabilní a proměnlivé. *„Změny tempa uvnitř promluvy, jeho nepravidelnosti a prodlužování jednotlivých segmentů, jsou výrazným rysem expresivity promluvy.“* Výzkumy dnešní češtiny potvrzují, že toto prodlužování se stalo přirozeným projevem mluvy.[5, str. 37]

1.2.3 Intenzita

Hlasitost řečového signálu je vnímána jako intenzita. Velice záleží na citovém stavu mluvčího, ale i dalších fyziologických příznacích, které tento stav mohou ovlivňovat. Ve spoustě studií se ani neuvažuje, protože je považována za zcela náhodnou nebo je její hodnota ovlivněna podmínkami z nahrávání. Z tohoto důvodu se případně pracuje pouze s *„relativními hodnotami v rámci krátkých úseků“*. [5, str. 22]

1.3 Intonace

Intonace je velice důležitou součástí promluvy. Pomáhá nám rozlišit otázku od oznamovací věty nebo rozkazu. Krom toho hraje klíčovou roli v klasifikaci emocí, protože pomocí intonace dokážeme rozpoznat citové zabarvení dané věty. *„Průběh základního tónu se v promluvě manifestuje jako melodie řeči. Změny výšky tónů jsou při vnímání řečové prozodie považovány za percepčně nejvýznamnější.“* [5, str. 21]

Nejprve je nutné si definovat fyzikální veličinu frekvence, která se značí písmenem f a má jednotku Hz. Vyjadřuje počet kmitů za jednu sekundu a platí pro ni vztah

$$f = 1/T\,, \tag{1.1}$$

kde f je frekvence a T je perioda (doba jednoho kmitu). Výšku hlasu definuje základní frekvence, na které kmitají naše hlasivky. U lidského hlasu se tato frekvence pohybuje přibližně v rozmezí 80 až 450 Hz. Mužský hlas se pohybuje na nižších frekvencích něž ženský, nejvyšší frekvence jsou přiřazovány dětskému hlasu. Tyto hodnoty samozřejmě závisí na věku, pohlaví, emočním rozpoložení dané osoby apod. [5, str. 21] Tuto základní frekvenci nazýváme základním tónem řeči a je jedním z hlavních parametrů, který nám pomáhá rozpoznat daného mluvčího podle sluchu. Není však samozřejmě jediným parametrem a její určení může být někdy problematické. [3, str. 56 – 57] Nejpružnější hlasivky má člověk v mladém věku, ale také třeba po ránu, kdy jsou ještě čerstvé, a ne tolik unavené jako třeba večer po celém dni

mluvení. Zejména zmíněná únava může stát za špatnou interpretací daných emocí a nepřesné intonace, která může vést ke špatné klasifikaci.

Autesserre a Di Cristo klasifikovali tři typy změn prozodických rysů v řeči [5, str. 22]:

1. akustické změny lidským uchem nezachytitelné
2. změny slyšitelné, ale nezpůsobující žádný posun ve vnímání významu promluvy
3. změny, které způsobují posun v interpretaci výpovědi.

1.3.1 Melodémy a jejich využití

Základní typ melodického průběhu promluvy se nazývá *melodém*. V českém jazyce máme tři základní typy: *klesavý ukončující* (typickým znakem je pokles intonace, který značí konec věty a mluvčí hlasem klesne až na spodní hranici svého rozsahu), *stoupavý ukončující* (představuje základní formu zjišťovacích otázek, vcelku stabilní) a *neukončující* (má více variant a objevuje se na konci souvětí).
[5, str. 35 – 37]

1.4 Fonetika

Disciplína, která je na pomezí lingvistiky, přesto do ní zcela nezapadá. Zkoumá *„zvukové projevy, zejména zvukovou stránku lidské řeči, fyziologický způsob artikulace (tvorby) těchto zvuků, jejich akustickou stránku a jejich vnímání“*.

Původně, také nazývána *fyziologie mluvy*, byla vnímána jako věda popisující hlásky a jejich tvoření. Od 20. století zkoumá celou oblast řečového signálu včetně jeho prvků, které tvoří plnohodnotnou komunikaci. Pro funkci jednotlivých hlásek, jejich rozdíly a vztahy mezi nimi, se později oddělila samostatná vědní disciplína zvaná *fonologie*. [6, online]

1.4.1 Akustická fonetika

Akustická fonetika je založena na analýze samotného signálu z hlediska frekvence, přenosu zvuku a jednotlivých aspektů řečového signálu. Vzhledem k tomu, že struktura lidské řeči je složitá, tak se při popisu signál dělí na *jednotlivé řečové zvuky ze souvislého proudu*, které jsou návazné a vzájemně se ovlivňují. [7, online]

1.4.2 Auditivní fonetika

Toto odvětví fonetiky se zabývá analýzou sluchem a je nepřímo spojené i s fonetikou artikulační. Nemůže být zcela objektivní, protože je tu faktor lidského sluchu

– tím pádem není tak přesná jako fonetika akustická. Ovšem hraje významnou roli, protože umožňuje vyzdvihnout prvky lidské komunikace, které jsou důležité právě pro dorozumívání.

Pro interpretaci a vnímání řečového signálu je důležité i optické vnímání, jak popisuje M. Krčmová:

„Akustický signál řeči je vnímán posluchačem také opticky. Okrajově si těchto složek všímá artikulační fonetika, detailněji se stává optický signál artikulace důležitý pro neslyšící, ale také např. při dabingu.“ [7, online]

2 Lidské emoce

Lidské emoce patří mezi velice komplexní pojem. Ovlivňují naše psychické i fyzické fungování a nedají se jednoznačně definovat. Zahrnují prožívání vnějších podnětů, ať už pozitivně nebo negativně. Jako jedinec jsme každý jedinečný v tom, co jak cítíme a prožíváme. Také záleží v jaké životní fázi se zrovna nacházíme. Dá se říct, že emoce je mentální stav, který doprovází určitá forma cítění vztažená k nějaké situaci nebo objektu. Nicméně v odborné literatuře se můžeme seznámit s různými významy a vysvětleními, co tento konkrétní pojem vyjadřuje – ať už z hlediska psychologického nebo fyziologického.

„*Výrazným znakem je diferencovanost emocionálních reakcí: novorozenec prožívá jen libost a nelibost, (...) nemluvně již vykazuje určité vrozené druhy emocí a dalším vývojem se emocionální život člověka dále diferencuje až do bohaté škály emocionálních a zejména citových reakcí (...).*“[8, str. 19]

Dalším znakem je například *polarita emocí*. Jedná v podstatě o to, že každá emoce, resp. cit má k sobě i opačnou polaritu – něco jako protiklad, (např. radost a smutek).

2.1 Emoce a cit

Tato krátká podkapitola se zabývá rozdílem mezi city a emocemi. Milan Nakonečný ve své knize *Lidské emoce* [8] popisuje velice zajímavé postřehy různých psychologů, kteří se vyslovili k této problematice. Někteří zastávají názor, že cit a emoce nejsou synonyma, protože pojem *emoce* může být použit ve dvou odlišných kontextech. Jedním z nich je vztažen k vnitřnímu prožitku, ten druhý zahrnuje i tělesný stav a tzv. výraz. Jiní rozlišují emoce a tzv. *smyslové dojmy organického původu*, mezi které patří např. pocit hladu.

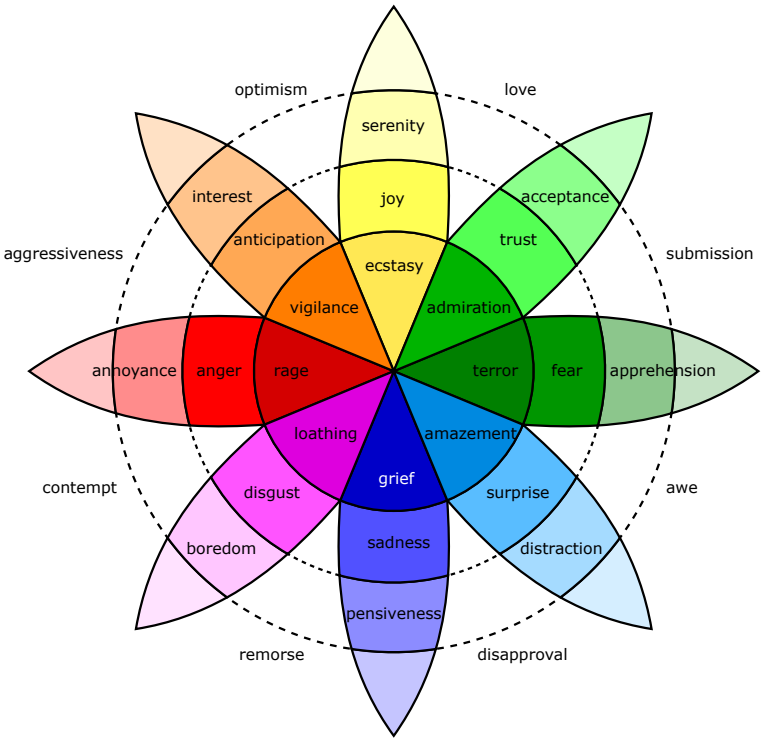
2.2 Druhy emocí

Psycholog Paul Ekman definoval šest základních emocí podle výrazu v obličeji:

- smutek
- štěstí
- strach
- hněv
- překvapení
- znechucení

„Podobně psycholog Robert Plutchik v 80. letech 20. století identifikoval osm základních emocí, které rozdělil do dvojic protikladů, mezi něž patří radost a smutek, hněv a strach, důvěra a znechucení a překvapení a očekávání. Tato klasifikace je známá jako kolo emocí a lze ji přirovnat ke kolu barev v tom smyslu, že určité emoce smíchané dohromady mohou vytvořit nové komplexní emoce.“ [9]

Poté Robert Plutchik definoval tzv. *kolo emocí*, které je zobrazeno na obr. 2.1.



Obr. 2.1: Kolo emocí (převzato z [10])

2.2.1 Aktivní a pasivní emoce

Lidé se často domnívají, že emoce jsou pouze jakési projevy, které jsou neovlivnitelné a zkrátka se nám dějí. Některé procesy nebo vztahy mezi lidským jednáním a emocemi však ukazují, že tomu vždy tak není. A to vede k rozlišení aktivních a pasivních emocí.

„Když jsme aktivní, naše emoce jsou určeny naší přirozeností, kdežto když jsme pasivní, tak jsou do určité míry ovlivněny něčím nebo někým jiným.“ [11]

2.3 Vybrané konkrétní emoce ke klasifikaci

V této kapitole se zaměříme na konkrétní emoce, které byly vybrané ke klasifikaci a tvorbě databáze. Byly vybrány celkem čtyři, z čehož vždy jedna dvojice vykazuje

opačné znaky či význam (*radost - smutek, vztek - nuda*). Je to z toho důvodu, že emoce, které jsou projevem podobnější, by se hůře klasifikovaly a mohly by způsobit nejasnosti i v samotném průzkumu subjektivního hodnocení. Umělá neuronová síť by musela být velice dobře natrénovaná, aby takové emoce dokázala rozlišit. To vzhledem k velikosti naší databáze ani není možné.

2.3.1 Radost

Radost je pozitivní a příjemná emoce, často doprovázená smíchem. Může to být reakce na náš úspěch či zisk, je však obtížné mluvit o konkrétních příčinách. Kromě smíchu radost často vyjadřujeme i svými pohyby, ovšem každý jedinec se bude projevovat jinak. Existují různé formy tohoto prožitku, např. slabší formou radosti může být pouhá *spokojenost* nebo *pocit štěstí*. Za vrcholnou formu je považována *extáze*, což je stav mysli. Jde z větší části o vliv omamných látek nebo o duchovní rovinu spojenou s náboženstvím.

Radost má několik důležitých funkcí, a to jak biologickou, tak tu sociální. V knize [8] se dočteme, že biologická funkce se projevuje hlavně tím, jak se člověk chová k okolnímu světu. Radostný člověk nemá problém být více extrovertní, kdežto smutný člověk se před světem uzavírá. Vývoj tohoto citu a schopnost ho vyvolat je klíčová už od raného věku dítěte. Dává dobrý základ pro schopnost člověka lépe komunikovat a vytvářet vztahy v širším okruhu lidí.

Dále by měl být také zmíněn fakt, že odborné výzkumy potvrzují propojení štěstí a fyzického zdraví. Robert Holden vytvořil dotazník, pomocí něhož zjistil, že 65 ze 100 lidí by si vybralo štěstí než zdraví, přestože obě možnosti mají velkou hodnotu, avšak jdou vždy ruku v ruce. V této souvislosti bychom mohli zmínit i optimismus, který, i když se nejedná o emoci, může zajistit delší a spokojenější život jedince, protože dokáže koukat na svět odlišným pohledem - a tím pádem lépe pracovat se svým emočním stavem. [12, online]

2.3.2 Smutek

Emoce vyvolávající silný nářek doprovázený pláčem, po kterém následuje tiché uzavření před světem, který v tu chvíli působí temně a beznadějně. Člověk touží po návratu k tomu, co bylo ztraceno a chce to zpět - především při ztrátě milované osoby. Smutek má mnoho forem a často záleží na tom, jak silnou a hlubokou ztrátu prožíváme.

„Je řazen k tzv. primární emoci, tj. má u člověka vrozený základ a můžeme ho pozorovat i u zvířat.“ Objevuje se zde i jistá podobnost s fyzickou bolestí, protože smutek často provází i fyziologické změny jako je třeba srdeční arytmie. Mezi ty

další projevy kromě těch fyzických patří i poruchy spánku, nechutenství, apod.[8, str. 252 – 253]

V souvislosti s událostí, která nás tak hluboce zasáhne, může smutek velice často vést až k depresi nebo s ní být zaměňován. Deprese patří mezi psychické poruchy, kdy už nejde o pouhý emoční pocit, ale stav, který má vliv na fungování v běžném životě i na tělesné zdraví. V tomto momentu se z člověka vytrácí jakákoli chuť žít a je potřeba vyhledat odbornou pomoc. V dnešní době je téma a obecně pojem *deprese* méně tabu a terapie může být dobrým způsobem, jak tyto problémy včas podchytit.

2.3.3 Hněv

Hněv je emoce vyvolaná neúspěchem nebo překážkou, která se nám staví do cesty při snaze dosáhnout požadovaného cíle. Stejně jako jiné emoce, i tato má svoje formy - slabší (*pocit rozzlobení*) a silnější (*vztek*). Vztek jako takový je ale *„pokládán za afekt“*. [8, str. 259] Někdy může být spojen až s agresivitou způsobenou touhou po útoku za účelem odstranění dané překážky.

Fyziolog W. B. Cannon uvádí následující fyziologické změny provázející prožívání hněvu [8, str. 260]:

- snížení nebo zastavení procesů v digestivním traktu
- posun krve z abdominálních orgánů ke kosternímu svalstvu
- zesílení kontrakcí srdce
- hlubší respirace
- dilatace bronchoidů
- mobilizace cukru v krevním oběhu

Nebylo by však správné hněv vykreslovat jako pouhou agresi, která má jen stinné stránky. Právě naopak hněv nám pomáhá stanovovat hranice a je to zcela přirozená emoce, stejně jako každá jiná. Problém nastává v momentě, kdy člověk není schopný s tímto stavem správně nakládat. Přílišné cholerické výbuchy ani neustálé potlačování vzteku není řešením a mohou vést k závažným psychosomatickým potížím. [13, online]

2.3.4 Nuda

S tímto pojmem se setkáváme velice často. V extrémních případech by se dalo říct, že může vést i k závažným problémům jako je např. syndrom vyhoření. V psychologickém slovníku se dočteme tuto definici: *„Duševní stav podmíněný monotónností podnětů, projevující se omrzelostí, pocity zbytečnosti, nezajímavosti, nespoko-*

jenosti, vynucenou pasivitou, ztraceným časem, oslabením pozornosti, pocity únavy, depresívními náladami“ [14, str. 124]

Slovní název pro tento stav se poprvé objevil ve slovníku Francouzské akademie v roce 1762. Avšak v naší moderní době tato emoce je trochu jiná, než jak ji zažívali naši předkové. Vnímáme ji jako nepříjemný pocit, kdy člověk není schopný se donutit dělat, co by chtěl. Koncentrace na nějakou činnost se v takovém stavu zdá být naprosto zcestná a nemožná a snaží se najít způsob, jak se tohoto pocitu zbavit. Avšak toto řešení nemívá příliš dlouhodobý účinek. [15, online]

Sociologové uvádějí tři momenty, kdy se k nám nuda dostavuje:
1) Opakující se průběh života nás vyhodí z rytmu a my se najednou nemáme „čeho chytit“.
2) Neschopnost člověka vytvořit něco sám, protože je příliš vtažen do světa plného pohodlí, který je posedlý vizuálním vnímáním.
3) Úbytek pocitů a zdrojů nejistot v moderním světě, který se tak stává příliš stejným a nezajímavým.

2.4 Interpretace emocí

2.4.1 Emoční fráze

V každém jazyce jsou zažité fráze, které se obvykle opakují a jsou emočně zabarvené. Některé byly tak nadužívané, že se z nich stalo *klišé*. I jednoslovné věty dokážou mít náladu určité emoce, proto byly věty pro herce vymýšleny tak, aby tuto zabarvenost pokud možno neměly. Některé jsou inspirovány informacemi z průvodního dokumentu k tzv. *Berlínské databázi* [16], která byla vytvořena profesionálně v bezodrazové komoře.

2.4.2 Úspěšnost poslechových testů

Součástí výzkumu emocí jsou poslechové testy, které slouží k zjištění kvality laboratorně pořízených vzorků. Setkat se s materiálem, který byl pořízen spontánně je spíše výjimečné a jen velmi málo studií pracuje s autentickou emoční podobou.

Procento úspěšnosti těchto testů vždy závisí na velikosti výběru, který je posluchači poskytnut. Průměrně se volí mezi 4 a 10 emocemi. Avšak výsledky nikdy nejsou zcela úspěšné. Pokud bychom se ocitli v reálné situaci, emoce vyhodnotíme mnohem lépe a zdá se nám to být přirozené.

Vlčková [5] uvádí, že na prvních příčkách žebříčku úspěšnosti se pohybují emoce negativní – např. vztek, smutek, protože jsou mnohem snáze interpretované a tedy i klasifikovatelné. Mezi další emoce, které si drží vysokou úspěšnost, patří lhostejnost a radost.

Je nutno zmínit, že výsledky takových testů mohou být ovlivněny konkrétním sdělením, které mělo danou emoci vyjadřovat. Z tohoto důvodu se využívá *maskování* nebo *filtrace*, kdy dojde k cílenému odstranění vyšších frekvencí ze spektra – tedy signál se stává nesrozumitelným, ale zůstane daná intonace i další prvky důležité pro určení emoce.

MOS testy

K určení kvality nahrávek v databázi se používají tzv. MOS testy (mean opinion score). Toto skóre se obvykle vyjadřuje číslem v 1 až 5, kde 1 je nejhorší a 5 nejlepší. Pro emotivní databázi to znamená posuzování jednotlivých nahrávek, jak moc odpovídají emoci, kterou mají reprezentovat. Samozřejmě se jedná o subjektivní posuzování, respondenti se většinou skládají z posluchačů s různou úrovní znalostí a jazykovou informovaností. [17, online]

3 Zpracovávání řečového signálu

Číslicové zpracování signálů má několik vlastních disciplín podle toho, za jakým účelem s řečovým signálem pracujeme. Tyto disciplíny mají velkou skupinu podkategorií s širokou škálou využití v praxi. Patří mezi ně kromě klasifikace emocí například také rozpoznávání mluvčích, syntéza řeči nebo zvýraznění řeči (odstranění šumu, restaurace starých a poškozených nahrávek), apod.

Prvním krokem zpracování signálu je tzv. *pre-processing*, během kterého je signál převeden ze signálového prostoru do prostoru příznakového. Příznakem (angl. *feature*) rozumíme *„vlastnosti obrazce vyjádřené kvantitativně.“* Tedy důležité informace a aspekty řeči (pro každý druh zpracování jsou tyto informace různé, záleží na účelu) nám zůstanou, ale vyhneme se komplikacím jako je příliš velký počet vzorků časového průběhu signálu.

3.1 Preemfáze

Preemfází se nazývá úprava kmitočtové oblasti signálu, konkrétně vyšších frekvenčních pásem. Vztah pro tento druh filtrace je uveden níže.

$$s''(n) = s'(n) - \lambda s'(n-1) \tag{3.1}$$

Nežádoucí efekty způsobené vlastnostmi spektra řečového signálu lze také částečně potlačit kmitočtovým filtrem typu horní propust, který charakterizuje přenosová funkce

$$H(z) = 1 - \lambda z^{-1}. \tag{3.2}$$

3.2 Segmentace

Řečový signál se kvůli svým vlastnostem zpracovává metodami krátkodobé analýzy. Velice důležitou složku předzpracování signálu je jeho rozdělení na jednotlivé úseky. Toho se může docílit více způsoby. Jedním z nich je segmentace pomocí okna, která rozdělí signál na úseky o stejné délce v závislosti na volbě typu okna.

Dále poté např. fonémová segmentace, kde se hledají hranice mezi jednotlivými *řečovými jednotkami* (může se jednat o fonémy, slabiky nebo celá slova).

3.2.1 Segmentace pomocí okna

Časová segmentace se provádí pomocí oken. Ve zpracování řečového signálu se nejčastěji setkáme s Hammingovým oknem a pravoúhlým oknem. Signál rozdělíme

na N úseků o délce 10 – 30 ms. Jednotlivé segmenty se mohou překrývat, ale není to podmínkou. Běžně se volí překryv 50 %, což při vysoké vzorkovací frekvenci vede k náročnému výpočtu. Definice obou oken je uvedena níže (převzato z [18]:

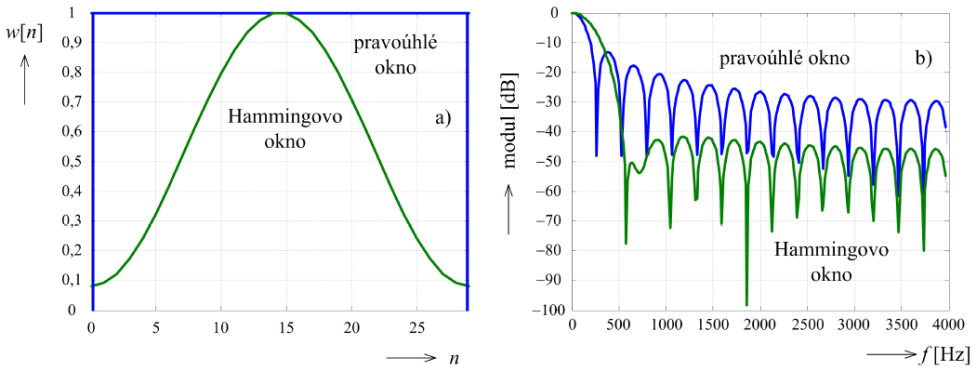
Rovnice pravoúhlého okna:

$$\begin{aligned} w[n] &= 1, \quad \text{pro } n = 0, 1, \ldots, N-1, \\ w[n] &= 0, \quad \text{pro ostatní } n. \end{aligned} \tag{3.3}$$

Pro Hammingovo okno platí:

$$\begin{aligned} w[n] &= 0,54 - 0,46 \cos\left[n\tfrac{2\pi}{N}\right], \quad \text{pro } n = 0, 1, \ldots, N-1, \\ w[n] &= 0, \qquad\qquad\qquad\qquad\quad \text{pro ostatní } n. \end{aligned} \tag{3.4}$$

Při použití pravoúhlého časového okna může dojít k tzv. *spectral leakage*. Jde o prosakování spektrálních složek z vedlejších laloků okna. Na obr. 3.1 je zobrazen časový průběh obou oken spolu s modulovými kmitočtovými charakteristikami. Z těchto průběhů je patrné, že postranní laloky pravoúhlého okna mají mnohem menší útlum oproti Hammingovu oknu. Proto je Hammingovo okno v aplikacích více vhodné a méně náchylné na tuto chybu. [18, str. 51 - 52]



Obr. 3.1: Časový průbeh (a) pravoúhlého okna a Hammingova okna a jejich modul spektra (b) (převzato z [18])

3.3 Fourierova transformace

Výpočet, který zajišťuje převod signálu z časové oblasti do kmitočtové se nazývá Fourierova transformace a je definována integrálním vztahem (podle [19])

$$S(\omega) = \int_{-\infty}^{\infty} s(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t, \tag{3.5}$$

kde $S(\omega)$ je obraz transformace a $s(t)$ její předmět. Stejně tak je možné z kmitočtové oblasti přejít do časové pomocí inverzní Fourierovy transformace.

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}\omega \tag{3.6}$$

Oba vztahy uvedené výše platí pro FT signálů se spojitým časem. Pro diskrétní čas jsou vztahy popsány níže. Tento typ transformace si získal název *Discrete-time Fourier transformation*, tedy DTFT, pro kterou platí

$$S(\Omega) = \sum_{\mathrm{k}=-\infty}^{\infty} s(k) \mathrm{e}^{-\mathrm{i}\Omega k}, \tag{3.7}$$

inverzní DTFT spočítáme jako

$$s(k) = \frac{1}{2\pi} \int_{0}^{2\pi} S(\Omega) \mathrm{e}^{\mathrm{i}\Omega k} \mathrm{d}\Omega. \tag{3.8}$$

3.4 Základní frekvence

Základní kmitočet se v odborné literatuře označuje jako F0. Pro výpočet základního tónu je nezbytná segmentace (popsáno v 3.2).

Mezi základní metody určení parametru základního tónu řeči patří [18, str. 67]:

- detekce základního tónu v časové oblasti,
- detekce základního tónu v kmitočtové oblasti,
- detekce základního tónu v reálném kepstru.

Výpočet v časové oblasti se provádí např. pomocí autokorelační posloupnosti, což je nejjednodušší a také nejméně přesná metoda. Ve spektrální oblasti je možné se setkat s FFT, která je použita k výpočtu spektrogramu signálu. Dochází zde k několika zejména filtračním krokům, které vedou k získání F0. Výpočetní náročnost je vysoká a přesnost bohužel ne příliš dobrá. Při kepstrálních metodách se signál segmentuje a cílem je získat reálné kepstrum každého segmentu, ze kterého jsou odstraněny hodnoty kmitočtů, které nepatří mezi hodnoty F0 lidského hlasu. [18, str. 67-72]

4 Vlastní databáze emotivních promluv

Existuje mnoho profesionálních databází, které by se daly pro tuto práci použít. V této bakalářské práci bude zahrnuta vlastní databáze, která byla vytvořena za pomoci dvou profesionálních herců, kteří zahráli předem určené emoce na několika neutrálních větách. Plán realizace byl inspirován diplomovou prací [20, str. 30-32].

4.1 Realizace

Nahrávání bylo realizováno pomocí velkomembránového kondenzátorového mikrofonu značky Behringer B-1, který má vysokou citlivost a umožňuje úpravu kmitočtové charakteristiky na Low cut. Tato funkce je však v tomto případě nežádoucí a použita nebyla. Právě díky své citlivosti a vyrovnané kmitočtové charakteristice je vhodný pro záznam lidského hlasu, jelikož nám poskytne všechny žádoucí složky spektra. Zmíněný elektroakustický měnič má kardioidní směrovou charakteristiku a vyhovující poměr signál/šum neboli *SNR* (dle výrobce 81 dB). Jako zdroj Fantomového napájení + 48 V byla použita USB zvuková karta Behringer UMC204HD.

Ještě před setkáním s herci byly vytvořeny menší plakáty s větami, které poté herci četli. Byly zvoleny co nejvíce neutrální věty - jednoslovné i víceslovné. A tyto věty poté zahrány v různých emocích a zaznamenány do programu Cubase, který slouží jako *Digital Audio Workstation* k záznamu a úpravě zvukového, příp. audiovizuálního obsahu.

Konkrétní věty jsou sepsány níže v tab. 4.1

Tab. 4.1: Věty zahrané v různých emocích

VĚTY	
Jednoslovné	**Víceslovné**
Vážně?	*Na stole je váza s květinou.*
Jé.	*Donese to ve středu.*
Super.	*Co je v té tašce?*
Dobře.	*Uvidíme se příští týden.*
Poslouchám.	*Vrátím se až večer.*

4.1.1 Proces nahrávání

Nahrávání se zúčastnili dva profesionální herci (muž a žena). Pro nahrávání byla plánovaná technika kontaktního snímání, aby bylo možné zanedbat vliv akustiky místnosti. To znamená, že herci byli v blízké vzdálenosti od mikrofonu. Vše probíhalo v provizorních podmínkách v místě bydliště jednoho z herců z důvodu jejich časové vytíženosti. Dozvuk místnosti tedy nešlo úplně zanedbat, ale bylo snahou ho co nejvíce eliminovat dostupnými prostředky.

Za mikrofon byl před herce umístěn seznam vět napsaný na papíře formátu A3 (viz tab. 4.1). Cílem bylo zajistit co nejpřirozenější prostředí pro všechny projevy emocí, které herec chtěl vyjádřit bez přílišné soustředěnosti na samotný text. Vzhledem k samotné akustice místnosti se musely tyto projevy mírně omezit, aby se do nahrávky nedostal nežádoucí dozvuk. Zúčastnění herci se tohoto úkolu i přes komplikace zhostili velice profesionálně a samotný proces záznamu trval asi 45 minut.

Nahrávání probíhalo v jednotlivých cyklech, kdy bylo vždy předneseno všech 10 vět v jedné emoci. Tento cyklus se opakoval pětkrát pro všechny čtyři emoce. Herci se střídali vždy po jedné emoci a dělali si pauzu po jednom až dvou cyklech podle vlastní potřeby. Na jednoho herce tedy připadlo 200 vzorků - v celkovém součtu 400 vzorků. Některé věty byly opakovány víckrát po sobě (v rámci jednoho cyklu) dle cítění samotného herce. Přesný počet nahrávek v databázi je 461.

4.1.2 Postprodukce a střih

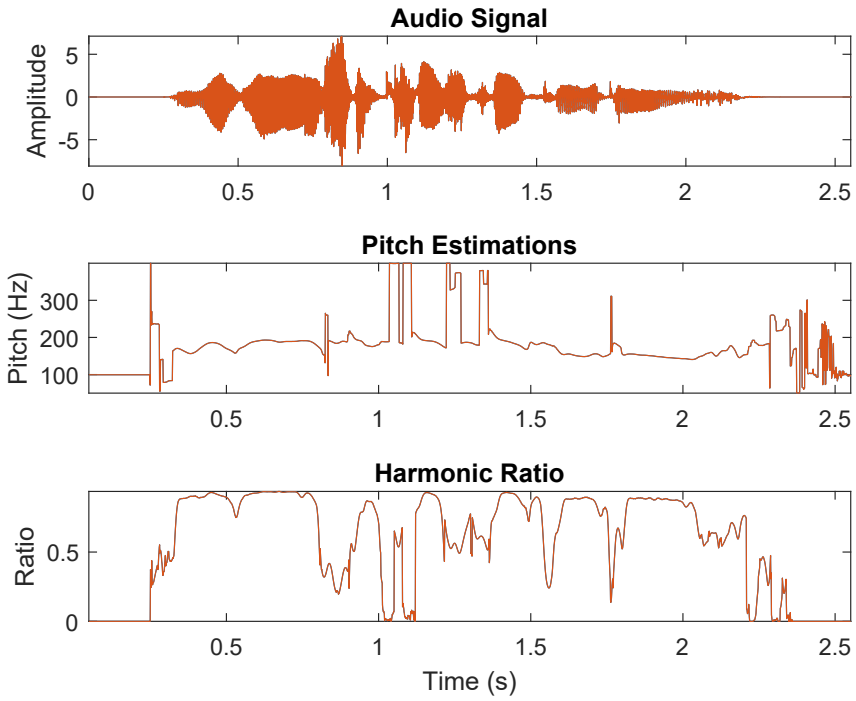
Po získání základního materiálu bylo zapotřebí tyto stopy sestříhat do jednotlivých vět, tedy jedna věta = jeden zvukový soubor z datasetu. Vzhledem k množství zvukových stop, by byl manuální střih velmi časově náročný a zdlouhavý. Z toho důvodu byl v prostředí Matlab vytvořen skript s názvem `strih.m`, který je součástí přílohy této práce. Jde o jednoduchý program, který dokáže najít začátek a konec promluvy pomocí funkce `detectSpeech`. Program tímto způsobem projde, sestříhá všechny soubory ve složce a každou větu uloží do nového souboru s předem definovaným způsobem pojmenování. Avšak po následné kontrole byly u některých vět ustřiženy konce - zejména u věty *Vrátím se až večer*, která končí souhláskou *r*. Proto bylo nutné střih realizovat až o několik desetin sekundy před a po detekovaných hranicích promluvy. Po poslechové kontrole celé složky byly ještě dodatečně odstraněny nepovedené záběry jako např. přeřeknutí.

Uspořádání složky je popsáno v kapitole A o přílohách.

4.2 Intonační křivky

Intonační křivky hrají klíčovou roli nejen v určení nálady mluvčího. Jak již bylo řečeno, liší se v závislosti na konkrétním jazyku – přesto ale snadno najdeme určité podobnosti. V této sekci se nachází rozbor nahrávek lidského hlasu pořízených během nahrávání vzorků do databáze. Vlastní grafy byly vytvořeny v prostředí Matlab. Pro analýzu intonace byla použita stejná věta pro všechny emoce ve znění *„Uvidíme se příští týden“*.

Byl naprogramován vlastní skript s názvem `int.m`, kde byl nejprve načten a zobrazen časový průběh signálu. Následně byl použit příkaz `pitch`, který vrací odhadovanou hodnotu základní frekvence signálu podle uživatelem zvoleného vzorkovacího kmitočtu a `harmonicRatio`, jež udává poměr harmonických složek a šumu v signálu. V místech, kde se HR pohybuje v blízkosti hodnoty 1, lze předpokládat, že v tom místě je znělá hláska. Jinak se hodnoty skokově mění nahoru a dolů - tedy se jedná o místo v časovém průběhu, kde je samotný šum nebo neznělá hláska, (která má charakter šumu, např. *s, š, p,...*). Podle průběhu harmonického zkreslení byla nastavována hodnota `treshold` (neboli tzv. rozhodovací úroveň), aby tyto nepřehledné skokové změny byly vyfiltrovány a v grafu byl ponechán pouze průběh základní frekvence řeči.

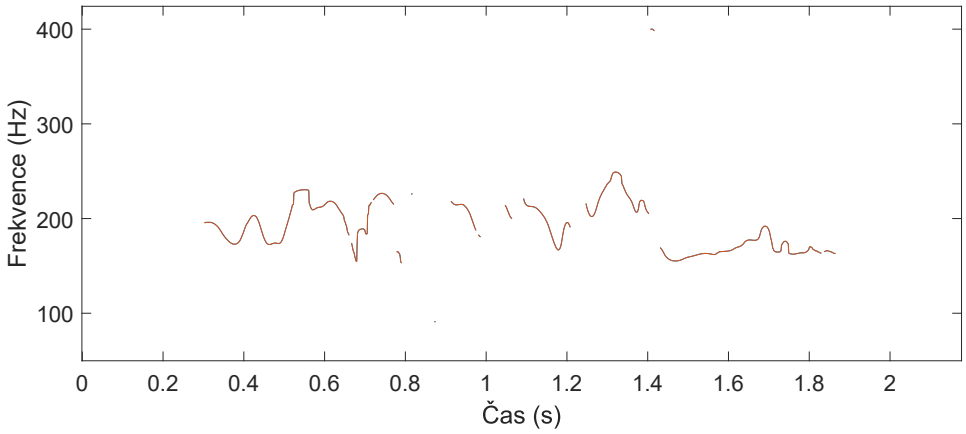


Obr. 4.1: Časový průběh signálu pro nudu (nahoře), odhadovaná hodnota F0 v čase (uprostřed), poměr harmonických složek a šumu v signálu (dole)

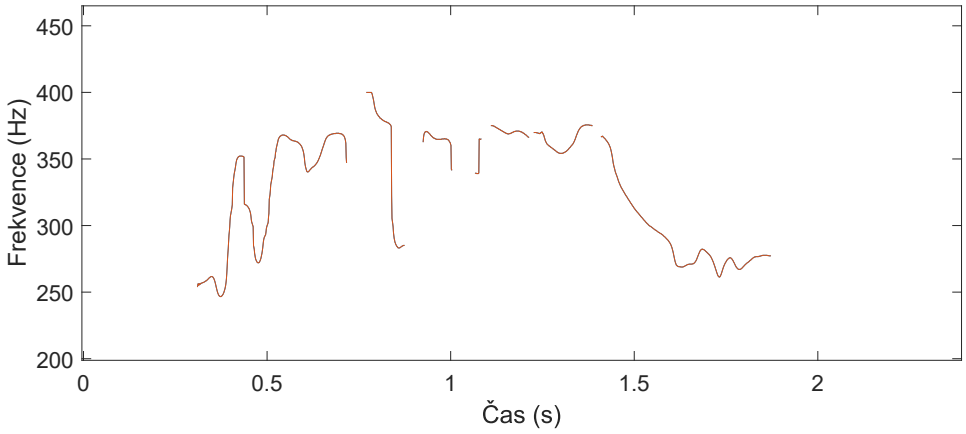
4.2.1 Intonační křivky zvolených emocí

Pro mužský hlas bylo obecně náročnější dané křivky zobrazit. Vina může být ve vlastním skriptu, ale také v interpretaci herce, která bude vždy svým způsobem jedinečná. Přesto byla snaha o to, najít mezi křivkami vztahy, ze kterých můžeme vyvodit závěry o průběhu intonace jednotlivých emocí.

Radost



Obr. 4.2: Intonační křivka pro radost - mužský hlas

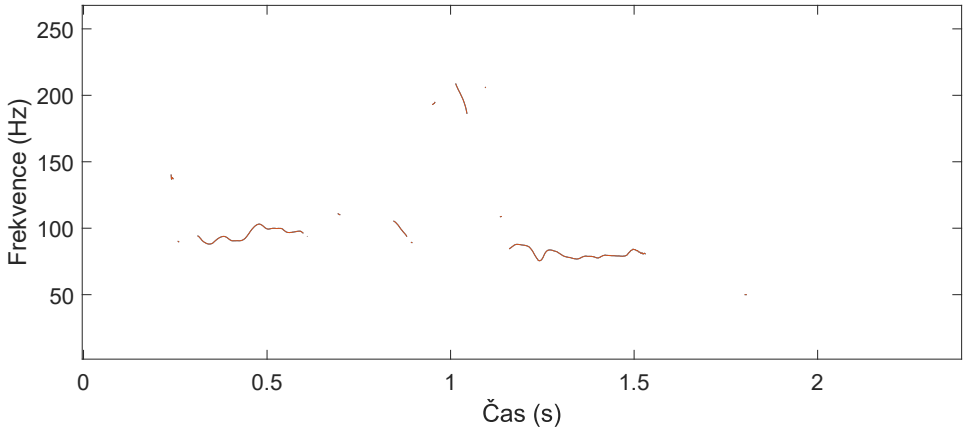


Obr. 4.3: Intonační křivka pro radost - ženský hlas

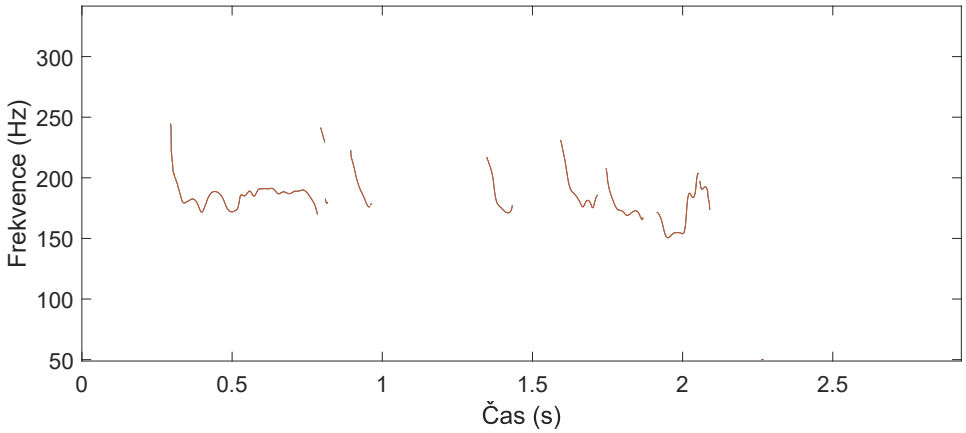
Na prvních dvou obrázcích 4.2 a 4.3 jsou zobrazeny intonační křivky hrané v radosti. Rozdíl mezi nahrávkami je zde patrný, ale obecně lze říct, že se zde objevují intonační výkyvy. Kmitočtový rozsah zde dosahuje až 150 Hz v rámci jedné věty.

U této konkrétní věty je lidsky přirozené klást důraz na slovo *týden*, protože je nositelem důležité informace. U expresivních emocí je tento důraz více znatelný u obou herců - zejména u mužského hlasu na obr. 4.2. U slova *týden*, dosahuje F0 nejvyšší hodnoty na první slabice. Poté nastává prudký spád s koncem (v tomto případě) oznamovací věty.

Smutek



Obr. 4.4: Intonační křivka pro smutek - mužský hlas



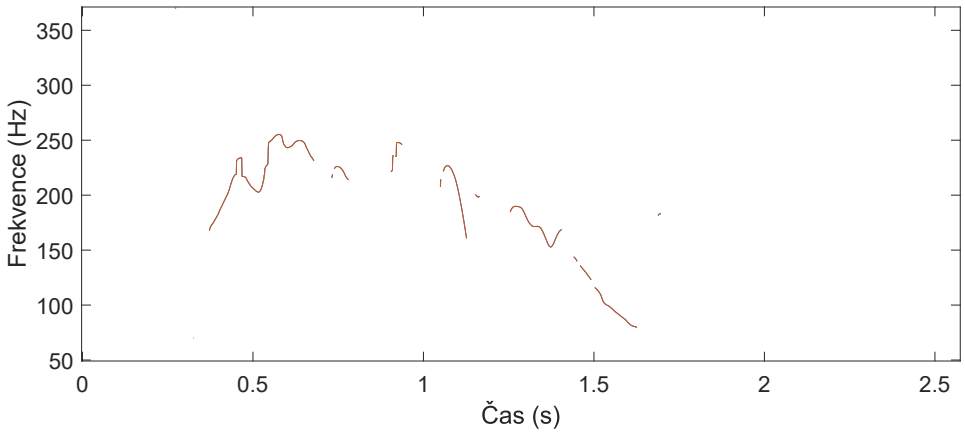
Obr. 4.5: Intonační křivka pro smutek - ženský hlas

V případě smutku je zde jeden výrazný rys. A to jsou pauzy mezi slovy, které mohou dávat najevo, že se dané osobě těžko mluví. Např.:

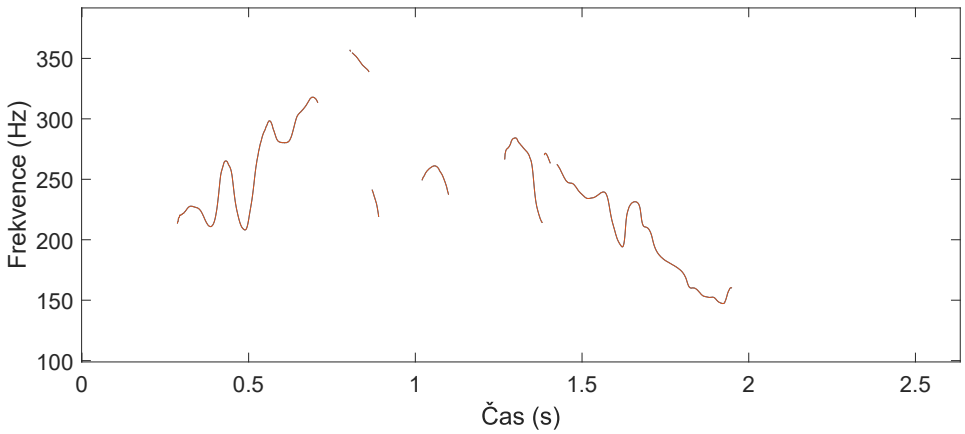
„Uvidíme se - příští týden.“
„Vrátím - se - až večer.“

Promluva obsahuje více dechu a je velice tichá, zpravidla i pomalejší. Kmitočtový rozsah nepřesahuje 50 Hz, charakterem se trochu blíží nudě.

Hněv



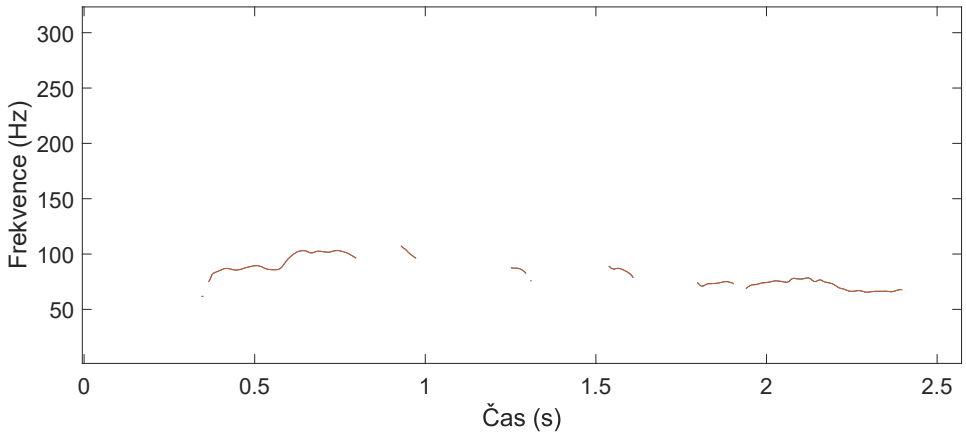
Obr. 4.6: Intonační křivka pro hněv - mužský hlas



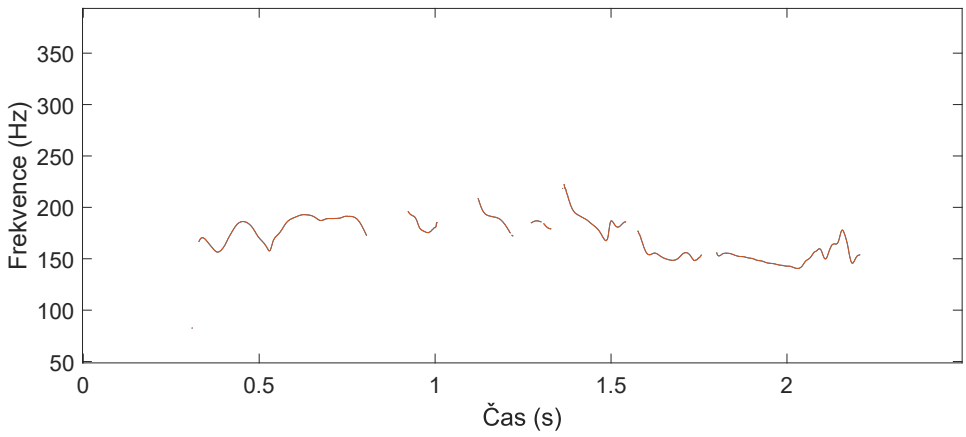
Obr. 4.7: Intonační křivka pro hněv - ženský hlas

Na obrázcích 4.6 a 4.7 vidíme intonační křivky pro hněv. Zde jsou křivky téměř totožné, pomineme-li rozdílné polohy hlasů. Tempo promluvy je velice rychlé. V první polovině je prudký nárůst F0, poté strmý spád.

Nuda



Obr. 4.8: Intonační křivka pro nudu - mužský hlas



Obr. 4.9: Intonační křivka pro nudu - ženský hlas

Co se týče nudy, tak ta je v provedení obou mluvčích velice monotónní. Rozsah intonační křivky nepřesahuje 50 Hz. Rozdíl mezi polohou mužského a ženského hlasu je zde přibližně 100 Hz.

Dalším podstatným znakem této emoce je čas trvání. U jiných emocí délka průběhu zpravidla nepřesahovala 2 s (pro víceslovné věty). Na obr. 4.8 můžeme vidět, že věta trvá téměř 2,5 s, u průběhu na obr. 4.9 jde o 2,25 s.

5 Umělé neuronové sítě

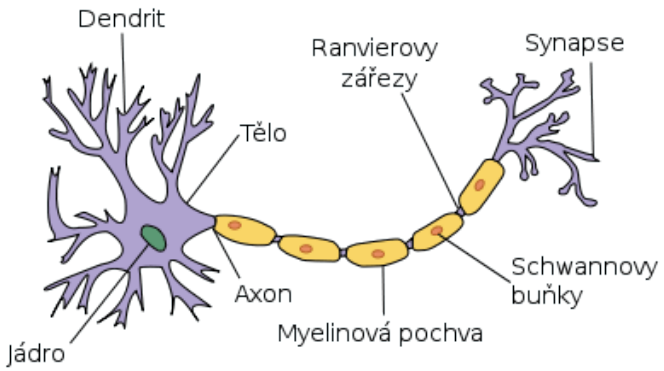
Umělé neuronové sítě jsou v dnešní době nedílnou součástí mnoha operací, které vycházejí z podstaty lidského myšlení a snaží se ho pochopit. Lidský mozek neustále přijímá *vstupní data*, která následně analyzuje. Principem UNS je snaha tento proces napodobit a využít při různých aplikacích. K tomu je ale zapotřebí velké množství dat, na která se UNS musí adaptovat, aby správně vyhodnocovala výsledky. Zpravidla lze říct, že čím větší databáze vstupních informací, tím větší šance na úspěšnost.

Mezi hlavní využití patří predikce a klasifikace problémů, konkrétně rozpoznávání řeči, klasifikace a predikce lidských onemocnění (např. ze snímků orgánů), rozpoznávání obličejů nebo překlady a digitalizace textů.[22, str. 109-110] Jsou to v podstatě matematické modely biologických neuronových sítí. V této kapitole bude krátce zmíněna stavba biologického neuronu, na základě čehož bude možné analogicky navázat na jeho matematický model. Čerpáno bude převážně z publikace [23].

5.1 Biologický neuron

K pochopení principu umělých neuronových sítí je zapotřebí zmínit, jak funguje lidská nervová soustava. Její základní stavební jednotkou je nervová buňka *neuron*, která přenáší vzruchové informace mezi místem vzniku a lidským mozkem. V této kapitole bude stručně zmíněna stavba a funkce biologického neuronu.

Mezi čtyři základní části neuronu patří *soma* (tělo s buněčným jádrem), *dendrity* (krátké výběžky vedoucí vzruchy k buňce), *axon* (výstup neuronu, dlouhý a silnější) a *synapse* (slouží k předávání vzruchů a spojení dvou neuronů). Model biologického neuronu je zobrazen na obr. 5.1 níže. [23, str. 46-47]



Obr. 5.1: Popis částí biologického neuronu (převzato z [21])

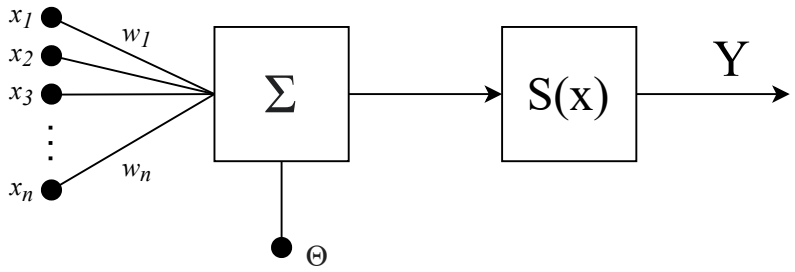
5.2 Matematický model neuronu

Jako je biologický neuron základní stavební jednotkou nervové soustavy, tak je matematický neuron (nazýván též *perceptron*) základem pro umělé neuronové sítě. S prvním matematickým modelem přišli psychiatr Warren McCulloch a matematik Walter Pitts. Každý model neuronu má dvě části:

- obvodová funkce (určuje kombinace vstupních parametrů uvnitř neuronu)
- aktivační funkce (přenosová funkce neuronového modelu)

Volba průběhu aktivační funkce je důležitá pro následné trénování sítě. Nejčastěji se používá např. sigmoida nebo hyperbolická tangenta.

Na obr. 5.2 je zobrazeno jak takový model neuronu vypadá z matematického hlediska. Parametry x_i představují vektor vstupních dat, w_i jsou synaptické váhy a $S(x)$ je výše zmíněná aktivační funkce. Část Y představuje výstup neuronu, Θ je práh. Jak vidíme, zde se z hlediska terminologie obejvuje mnoho pojmů, se kterými je potřeba se v této problematice seznámit. Synaptické váhy určují, jak velkou průchodnost bude mít signál mezi jednotlivými vrstvami. Každý neuron v následující vrstvě je spojen s každým neuronem z vrsty předchozí - každému spoji je přiřazena jedna váha. Během učení UNS se hodnoty těchto vah mění na základě testovací množiny.



Obr. 5.2: Matematický model neuronu (podle [24])

5.3 Algoritmy učení

Abychom mohli pomocí sítě testovat nová data, je nutné nejprve provést trénování. Jedná se o *optimalizační proces*, kdy sledujeme vzájemnou závislost parametrů UNS (také nazývané hyperparametry) a parametrů vstupujících do UNS. Nejčastěji se setkáme s učením

- s učitelem,

- bez učitele.

Během učení s učitelem (*supervised learning*) jsou požadované výsledky předem známy a přivádí se na výstup sítě. Zpravidla počet neuronů ve výstupní vrstvě odpovídá počtu tříd, do kterých klasifikujeme data. Učení bez učitele (*unsupervised learning*) je založené na schopnosti sítě samostatně rozeznat klasifikační třídy na základě určitých podobností mezi vzory.

Při učení je také důležité sledovat tzv. chybovou funkci, která nám značí, kdy je nutné trénovací proces ukončit. Jedná se o závislost chyby klasifikace na počtu iterací. Průběh by neměl být příliš plochý, ani oscilovat. Typicky závislost klesá s počtem provedených iterací. [23, str. 55-57]

Speciální metodou odvozenou zejména pro vícevrstvé sítě je tzv. algoritmus zpětného šíření chyby (*Back Propagation Algorithm*). Jedná se o iterační proces učení s učitelem, kdy je na vstup přiváděna matice vstupních parametrů. Po projití sítí se spočítá chyba vypočtené hodnoty od požadované, tato chyba se *„zpětně přepočítává do předchozích vrstev a synaptické váhy představující paměť, jsou opraveny.“*
[23, str. 58]

5.4 Druhy umělých neuronových sítí

5.4.1 Vícevrstvé neuronové sítě

Vícevrstvé neuronové sítě (někdy také označované MLNN - *Multilayer Neural Network*) jsou tvořeny neurony uspořádanými do vrstev. Tyto vrstvy mohou být různě propojené, proto rozlišujeme sítě s *dopředným šířením informace* a *rekurentní sítě*. U rekurentních sítí jsou navíc neurony propojené i pomocí zpětných vazeb. Úspěšnost trénování takové sítě je závislá na množství vzorků, které máme k dispozici (včetně jejich kvality a kvality samotných příznaků). U dopředných sítí jsou neurony ve vlastní vrstvě nepropojené. Propojení je plně mezi vrstvami. Při učení je třeba znát cílovou hodnotu (neboli třídu) daného vzoru. Jedná se tedy o učení s učitelem. [23, str. 71-75] Nejjednodušší MLNN je již zmíněný perceptron s jednou skrytou vrstvou.

5.4.2 Samoorganizující se neuronové sítě

Samoorganizující se neuronové sítě (SOM) rozpracoval finský vědec Teuvo Kohonen. Vycházejí z poznatku, že člověk během života střádá informace, které mozek zpracovává pouze v oblasti k tomu určené (např. zpracování řeči). Na tomto principu funguje sdružování do shluků s podobnými vlastnostmi. SOM tedy využívá pouze jakési soutěžní strategie bez nutnosti předkládání výstupů (neboli učení bez učitele).

6 Praktická část

Praktická část této práce se zabývá zejména získáním příznaků z nahrané databáze, konkrétním řešením a výsledky jednotlivých metod. Aby byla práce více komplexní, byly výsledky získané pomocí umělé inteligence porovnány s výsledky poslechového testu, který provedlo několik respondentů.

Na úvod je potřeba se krátce seznámit s prostředím, ve kterém byla celá praktická část realizována.

6.1 Programovací prostředí Matlab

Matlab je interaktivní programovací prostředí se sadami různých toolboxů vzhledem k charakteru aplikace. Realizace celé praktické části probíhala pouze ve verzi R2021b. V níže uvedených podkapitolách jsou krátce popsané knihovny funkcí, které byly využity pro účely této bakalářské práce.

6.1.1 Audio Toolbox

Tato knihovna umožňuje analýzu audio signálů včetně různých operací, jako je změna časového měřítka, výpočet akustických veličin, měření impulsní odezvy, apod.

6.1.2 Signal Processing Toolbox

Signal processing toolbox je knihovna funkcí určená k analýze a zpracování signálů. Umožňuje práci s datasety, které je třeba připravit jako vstupní parametry umělé inteligence (nebo i jiných klasifikátorů či aplikací). Samotná knihovna obsahuje několik aplikací, které usnadňují práci s parametrizací, vizualizací a zpracováním signálů v časové i kmitočtové oblasti.

6.1.3 Statistics and Machine Learning Toolbox

Jedná se o knihovnu určenou zejména pro metody strojového učení, kdy je potřeba analyzovat a modelovat data. Opět je zde několik aplikací, které jsou určené ke klasifikaci dat konkrétní metodou.

Classification Learner App

Tato aplikace (dále jen CL) umožňuje klasifikaci dat přímo z pracovního prostředí Matlabu nebo jiného souboru. Obsahuje mnoho metod z oblasti umělé inteligence, které si uživatel sám může zvolit a optimalizovat podle svých potřeb. Tato aplikace

byla vybrána z důvodu jednoduché implementace UNS a srozumitelného grafického prostředí, které umožňuje vykreslovat potřebné grafy, matice záměn, apod. Veškeré dostupné UNS v tomto toolboxu jsou plně propojené a upravovat se dají pouze skryté vrstvy. Počet neuronů ve vstupní vrstvě určuje počet příznaků, které byly získány z dat. Počet neuronů ve výstupní vrstvě je definován počtem tříd, do kolika chceme data klasifikovat. Tento údaj si aplikace sama zjistí z matice příznaků trénovacího datasetu, která vstupuje do aplikace, resp. sítě.

6.2 Parametrizace vstupních dat

Aby bylo možné data klasifikovat, je potřeba ze signálu získat příznaky, které budou vstupovat do systému. Ať už se jedná o umělou neuronovou síť nebo jakékoli jiné metody. Mezi vstupní parametry byly původně vybrány pouze hodnoty F0 v závislosti na čase pro konkrétní emoci. Z vypočtených hodnot bylo potřeba sestavit matici, která bude mít ve sloupcích jednotlivé hodnoty zvolených příznaků a řádky budou reprezentovat jednotlivé nahrávky (resp. věty).

Hlavním problémem tohoto procesu byla odlišná délka každé nahrávky. Aby se mohlo jednat o matici, kde každá nahrávka bude obsahovat stejný počet oken, byly nahrávky uměle prodloužené vektorem nul představujícím ticho. Nejdelší nahrávka tím pádem jako jediná doplnění nulami nepotřebovala. Ovšem databáze obsahuje jak věty delší, tak věty jednoslovné. V takovém případě polovina nahrávek je přepsána na vektory, kde je poměr užitečné informace k výše uvedeným nulám velice nevýhodný. Z tohoto důvodu musela být zejména u intonace zvolena jiná sada intonačních příznaků než pouhý časový průběh.

6.2.1 Výpočet F0

K tomuto účelu byl využit skript `int.m`, který zobrazoval průběhy intonačních křivek. Skript obsahuje klíčovou funkci `pitch`, která zde bude využita znovu. Délka jednoho okna byla nastavena na 30 ms a překrývala se o 15 ms.

Aby mohla být získána kvalitní informace o intonaci v rámci jedné emoce, ale zároveň odpadl problém s doplňováním nulami, byly získány z nahrávek následující příznaky:

1. Průměrná hodnota základní frekvence F0 v rozsahu nahrávky.
2. Maximální hodnota zákl. frekvence F0 v rozsahu nahrávky.
3. Minimální hodnota zákl. frekvence F0 v rozsahu nahrávky.
4. Rozdíl průměrné hodnoty F0 celé databáze a průměrné hodnoty F0 dané nahrávky.

6.2.2 MFCC

K výpočtu mel-frekvenčních kepstrálních koeficientů byla využita funkce `mfcc`, kterou Maltab nabízí. Nejnovější verze 2022a dokonce umožňuje grafickou vizualizaci takto spočítáných hodnot. Přednastavený počet koeficientů. které funkce počítá je 13, což pro tento druh aplikace je vhodná hodnota.

K těmto příznakům bylo přistupováno dvěma způsoby:

1. Průměr k-tého koeficientu pro všechna okna, výsledkem je 13 hodnot pro každou nahrávku.
2. Průměr 13 koeficientů pro jedno okno, výsledkem je 321 hodnot pro jednu nahrávku.

Do finální matice příznaků byl použit způsob číslo 1., který opět umožňuje analyzovat nahrávky navzdory jejich vzájemně různé délce.

6.2.3 Finální matice příznaků

Finální matice má rozměry $m \times n$, kde m je počet nahrávek a n je počet příznaků a zařazení do emoční třídy. Databáze byla procentuálně rozdělena na trénovací a validační set obsahující 90 % nahrávek a testovací set obsahující 10 %. Je běžné, že se dělí data např. v poměru 70:20:10, ale aplikace Classification Learner využívá tzv. křížovou validaci přímo v trénovacím setu. Více v podkapitole 6.3.1.

6.3 Import dat do aplikace CL

Vstupní matice ve formátu .xlsx byla nahrána do aplikace Classification Learner. Jednotlivé soubory jsou součástí přílohy.

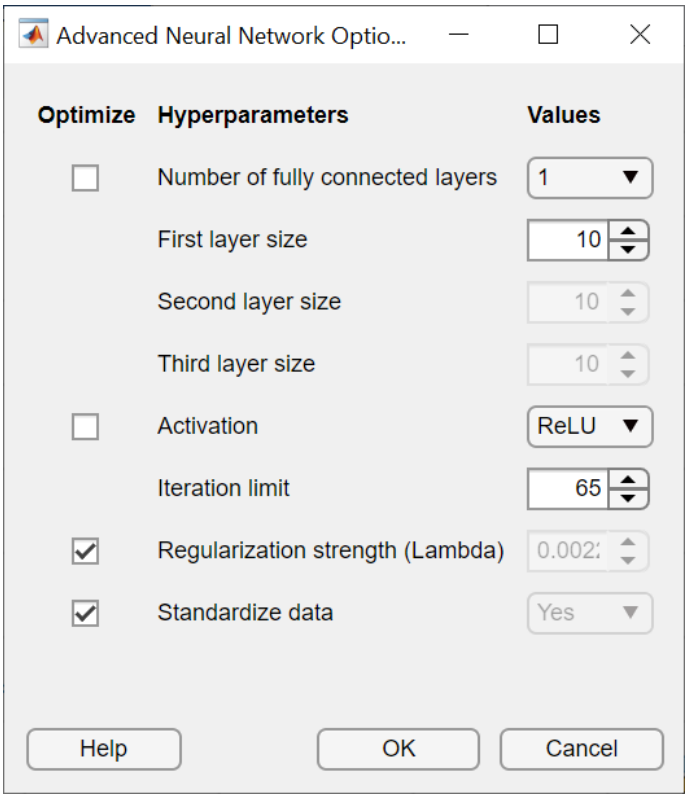
6.3.1 Křížová validace

První parametr, který je uživatel vyzván nastavit, je validace. Aplikace CL umožňuje několik způsobů, včetně pevného procentuálního rozdělení (např. 20 % trénovací množiny). To se během trénování neprojevilo jako vhodná metoda. Nejlepší výsledky poskytla tzv. *cross validation* (v překladu křížová validace). Je to druh validace, která rozdělí trénovací dataset na k podmnožin. Jedna podmnožina je vždy použita na testování, zbytek na trénování. Tento proces se opakuje, dokud algoritmus neprojde všechny podmnožiny. Nejčastěji se volí hodnoty 5 nebo 10. [25, online]

6.4 Výsledky klasifikace

Po úspěšném nahrání všech vstupních dat bylo nutné zvolit umělou neuronovu síť. Aplikace nabízí několik předpřipravených modelů s konkrétním a neměnitelným počtem skrytých vrstev a iterací, které jsou koncipované spíše pro nové uživatele, kteří chtějí vyzkoušet více metod najednou. Pro účely této práce byla zvolena tzv. *Optimizable Neural Network*. Model této sítě byl optimalizován a laděn tak, aby úspěšnost byla co nejvyšší. Parametry, které bylo možné měnit, jsou zobrazeny na obr. 6.1. Po dosažení nejvyšší možné úspěšnosti, byl vždy vyexportován. Následné testování nových dat již probíhalo mimo aplikaci přes funkci `yfit = C.predictFcn(T)`, kde `yfit` je výsledek klasifikace, `C` je název použitého modelu a `T` je matice testovacích dat.

Na následujících obrázcích jsou uvedeny celkem tři modely. Každá UNS má 17 neuronů ve vstupní vrstvě a 4 neurony ve výstupní vrstvě. Byl upravován zejména počet neuronů v jedné skryté vrstvě, protože větší počet skrytých vrstev rapidně snižoval úspěšnost trénování až o 10 % při přidání druhé skryté vrstvy. Dále bylo pracováno s počtem iterací a aktivační funkcí. Lambda zpravidla zůstávala beze změny. Jedná se o hyperparametr, který aplikuje penalizaci za zvyšování hodnot parametrů vstupujících do UNS, a tím tak zabraňuje přetrénování.



Obr. 6.1: Optimalizovatelná UNS v aplikaci CL

U každého modelu se po natrénování vygeneruje matice záměn, která má řádky i sloupce označeny počátečními písmeny jednotlivých emocí - *H*, *N*, *R* a *S*. Na hlavní diagonále jsou modře označená pole, která značí správně klasifikované nahrávky. Všechna pole mimo tuto diagonálu jsou chyby klasifikace. Nejúspěšněji dopadl smutek, těsně za ním hněv. Nejméně úspěšná pak byla radost. Nejčastější chybou byla vzájemná záměna smutku a nudy, protože to jsou emoce intonačně i tempově velice podobné. Ovšem v některých případech docházelo i k záměně např. radosti za smutek, ale nikdy ne naopak.

Následující podkapitoly podrobněji popisují tři nejúspěšnější modely. Jsou očíslované záměrně z důvodu orientace v grafech a maticích, v závorce jsou uvedeny počty neuronů ve vstupní, jedné skryté a výstupní vrstvě.

6.4.1 Model 1 (17-10-4) s křížovou validací 5

Na obr. 6.2 se nachází model UNS, který má jednu skrytou vrstvu obsahující 10 neuronů. Aktivační funkce zde byla zvolena sigmoida a křížová validace byla nastavena na hodnotu 5, počet iterací 70. Úspěšnost při trénování byla 79,8 %. Z matice záměn můžeme vidět, že tento model na testovací množině dat dosáhl pouhých 65 %.

Správně bylo klasifikováno 80 % smutku a nudy, což jsou emoce intonačně velice podobné, avšak v tomto případě téměř nedocházelo k jejich záměně. Nejhůře pak vyšla radost, která byla chybně přiřazena nejen hněvu, ale také smutku.

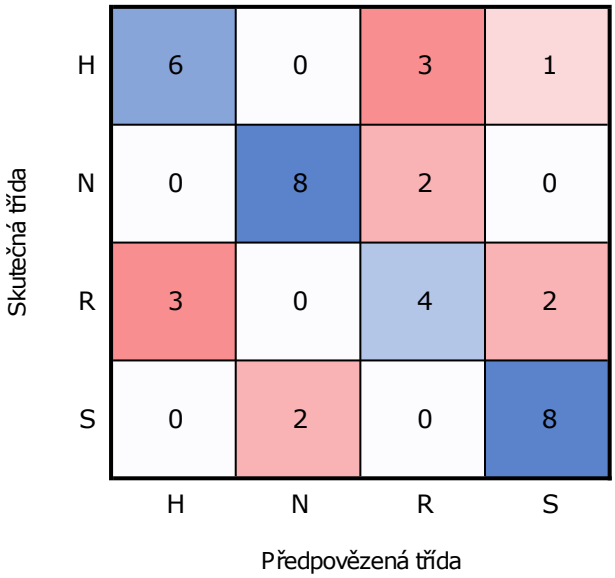
6.4.2 Model 2 (17-10-4) s křížovou validací 10

Dalším modelem je UNS opět s 10 neurony ve skryté vrstvě, ale tentokrát s 10-násobnou křížovou validací. Počet iterací je stejný jako v předchozím případě, včetně aktivační funkce. Úspěšnost modelu vzrostla na 67,5 % a to hlavně díky zmíněné validaci, po které byl výsledek trénování zvýšen na 80 %. Matice záměn je zobrazena na obr. 6.3.

Největší chybu klasifikace zde způsobuje záměna radosti za hněv. Naopak nejúspěšnější byl smutek, který dosáhl 90 %, těsně za tím hněv se 70 %.

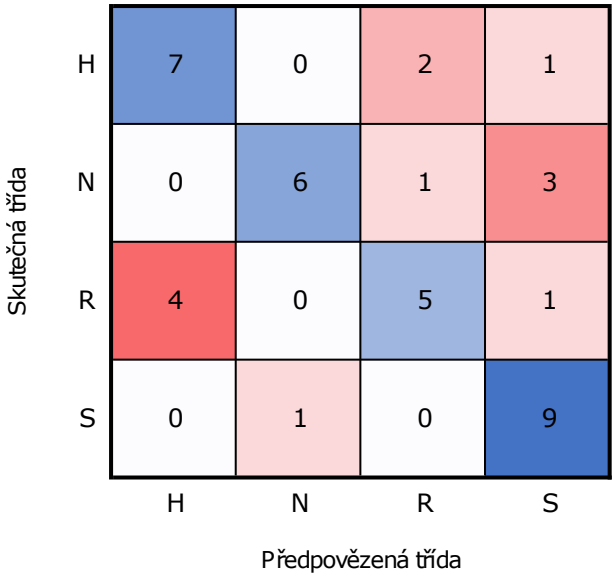
6.4.3 Model 3 (17-12-4) s křížovou validací 10

Nejúspěšnějším modelem se stal model č. 3 (viz obr.6.4),který má ve své skryté vrstvě neuronů 12. Model byl natrénován na 82 %, což je nejvyšší dosažený výsledek v rámci všech vyzkoušených modelů. Jako aktivační funkce zde figuruje ReLU (Rectified Linear Unit). Snížení iterací na 65 zajistilo mírné zvýšení procentuálního výsledku o pár desetin. Testování dosáhlo 75 %.



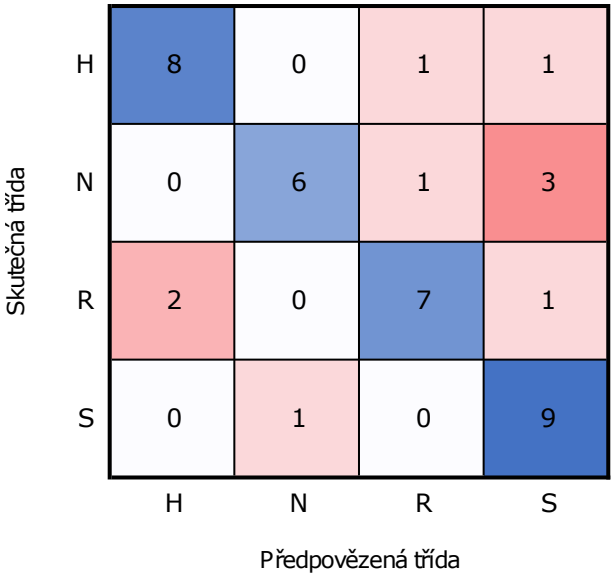
Správně klasifikováno:	26	nahrávek
Chybně klasifikováno:	14	nahrávek
Celková úspešnost:	65	%

Obr. 6.2: Model 1



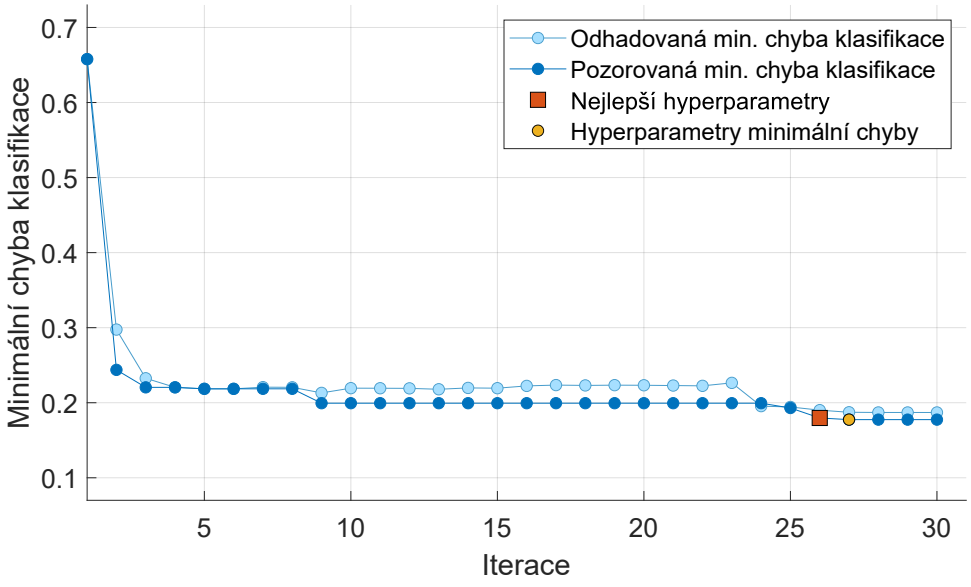
Správně klasifikováno:	27	nahrávek
Chybně klasifikováno:	13	nahrávek
Celková úspešnost:	67,5	%

Obr. 6.3: Model 2



Správně klasifikováno:	30	nahrávek
Chybně klasifikováno:	10	nahrávek
Celková úspešnost:	75	%

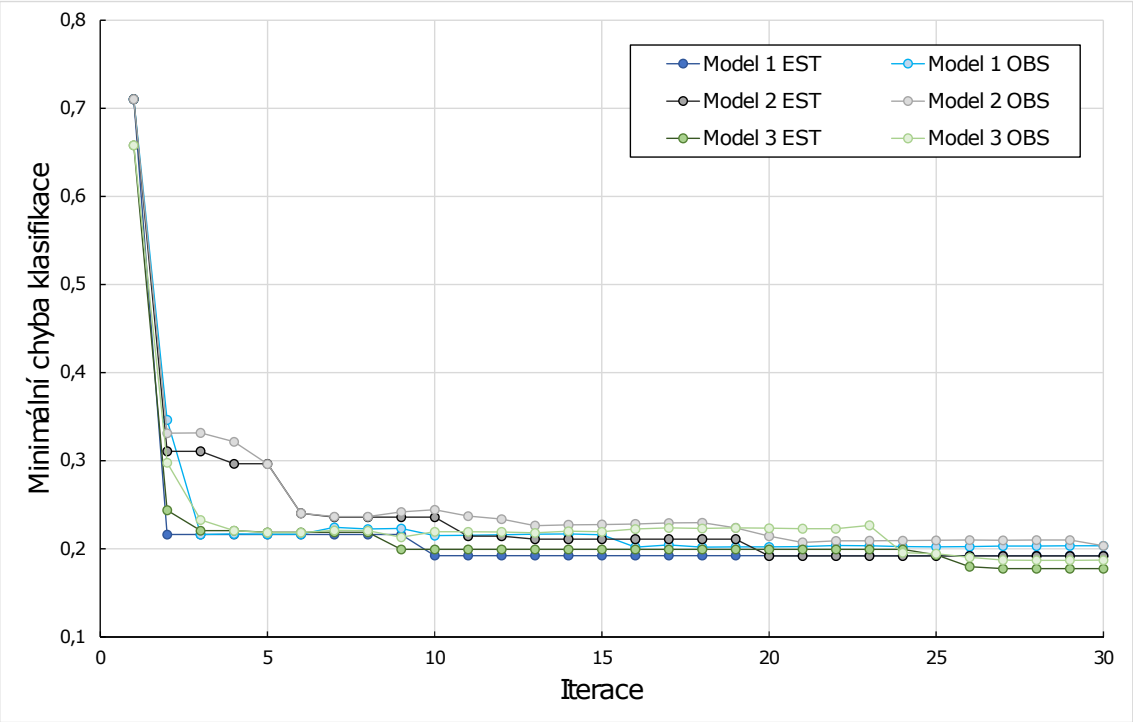
Obr. 6.4: Model 3



Obr. 6.5: Chybová funkce modelu 3

Na obr. 6.5 je uvedena závislost minimální klasifikační chyby na počtu iterací. Tento graf byl vykreslen aplikací CL po dokončení trénovacího procesu. Je závislý na optimalizačním procesu, kdy se projdou různé kombinace hyperparametrů klasifikátoru. Světle modrá křivka reprezentuje odhadovanou minimální chybu klasifikace, která je vypočítaná optimalizačním procesem aplikace. Tmavě modrá křivka značí pozorovanou minimální chybu klasifikace. Červený čtvercový bod značí bod, kde jsou hyperparametry nejvhodnější, žlutý bod označuje iteraci, která odpovídá hyperparametrům udávající pozorovanou minimální chybu klasifikace.

Následující graf vykresluje křivky chybových funkcí pro všechny tři popsané modely. Jedná se celkem o šest průběhů, pro každý model dva. Opět se tu vyskytuje závislost EST (podle anglického výrazu *estimated*) jako odhadovaná minimální chyba klasifikace a OBS (podle *observed*) jako pozorovaná minimální chyba klasifikace. Je zde viditelné, že model 3 dosahoval nejnižších hodnot mezi 25. a 30. iterací.



Obr. 6.6: Chybové funkce pro všechny tři modely

6.4.4 Další modely

Z výsledků je zřejmé, že nejlépe fungovala pyramidovitá neuronová topologie, tzn. počet neuronů ve skryté vrstvě je menší než ve vstupní vrstvě, ale zároveň větší než počet neuronů ve výstupní vrstvě. Pokud byla skrytá vrstva rozšiřována nad

číslo 12, úspěšnost trénování prudce klesala. Stejně tak přidáváním dalších skrytých vrstev. Výsledek se poté pohyboval okolo 60 %, někdy i méně.

Jiné modely byly trénovány také s podobnými parametry, ale byl změněn kmitočtový rozsah funkce `pitch`, která počítá F0. Interní kmitočtový rozsah je nastaven na 50 – 400 Hz, což může zejména u ženského hlasu ovlivňovat výpočet. Rozsah byl tedy změněn na 20 – 600 Hz, aby zde byla dostatečná rezerva pro intonaci v rámci lidského hlasu. Úprava těchto parametrů neměla na výsledek trénování žádný vliv. Úspěšnost byla opět přibližně 80 %.

Volba jiné validace též nepřinesla nic, co by se dalo zařadit mezi hlavní modely. Pevně určený validační dataset (tzv. *Hold-out validation*), je spíš vhodný pro velké databáze.

6.5 Vyhodnocení

Celková maximální úspěšnost testování dosáhla 75 %. Během testování bylo zjištěno, že jednotky % natrénovaného modelu ovlivňovaly úspěšnost testování až o 10 %. Tento fakt je dán i tím, že testovací dataset obsahuje malé množství nahrávek. Nebylo tedy možné aplikovat hluboké neuronové sítě s vysokým počtem skrytých vrstev. Výsledky jsou však vzhledem ke všem výše uvedeným faktům uspokojivé.

Poslední částí této práce je poslechový test, jehož úkolem bylo pomocí několika respondentů zjistit, jak moc relevantní jsou dosažené výsledky.

6.5.1 Poslechový test

O poslechových testech, které se v této oblasti běžně provádějí, bylo pojednáno blíže v kapitole 2.4.2. Z časových důvodů a nedostatku prostředků nebylo možné dělat rozsáhlé testy, které by určily kvalitu databáze. Pro aspoň částečné porovnání zcela uměle vytvořeného systému a lidského faktoru, bylo vybráno několik respondentů, s nimiž byl poslechový test ve *skromnější* formě zrealizován.

Předmětem testu bylo ověřit, zda je člověk schopen rozpoznat emoce z nahrávek pouze z vlastních zkušeností, poznatků, vlastního citu. Test tedy probíhal na identickém datasetu, který byl použit na testování, aby mohly být jednotlivé výsledky porovnány. Respondentům byl poskytnut papír s čísly od 1 do 40, ke každému číslu psali vždy písmenko H, N, R nebo S - podle toho, která emoce podle nich právě zazněla. Výsledky v % jsou uvedeny v tabulce 6.5.1. Nahrávky byly předem uspořádány do náhodného pořadí, aby byl test více autentický a posluchač se musel nad odpovědí zamyslet.

Snahou bylo získat několik respondentů z různých profesních oblastí, zastoupit obě pohlaví i odlišné věkové kategorie. Toto se povedlo pouze částečně, avšak pro tyto

účely jsou výstupy dostačující. Je zde zastoupena pouze věková kategorie v rozmezí přibližně 20 až 52 let V tabulce jsou vždy vypsány informace o daném člověku, který test podstoupil. Jedná se o pohlaví, věk, zaměstnání, příp. volnočasové aktivity.

Tab. 6.1: Výsledky poslechového testu

Respondent		**Pohlaví**	**Věk**	**Úspěšnost**
A	student, muzikant	muž	21	87,50 %
B	podnikatel	žena	52	65,00 %
C	student informatiky	žena	20	70,00 %
D	programátor, muzikant	muž	24	87,50 %
E	student, muzikant	žena	22	72,50 %
F	student, sportovec	muž	23	87,50 %
G	muzikant	muž	28	82,50 %

Jak je vidět v tabulce, poslechový test byl proveden se třemi respondenty, kteří jsou pro účely zachování anonymity označeni písmeny A až G. Výsledky účastníků B a C se zcela shodují s úspěšností dosažené pomocí UNS. Respondenti A a D dosáhli až 87,5 %. Zde může nastat spekulace, zda je to proto, že se dané osoby aktivně věnují hudbě. Toto tvrzení může částečně vyvracet stejný výsledek respondenta F, resp. E. Test by muselo podstoupit mnoho takových účastníků, aby z toho mohl být určen jasnější závěr. Lze zde pouze uvést, že mužští posluchači byli úspěšnější (nehledě na povoláni nebo zájmy).

Hlavní účel testu byl však splněn. Respondent B nejčastěji zaměňoval nudu za smutek (či naopak), dvakrát také hněv za radost. To samé platí i pro respondenta C. Tyto výstupy zcela potvrzují, že modely UNS z kapitoly 6.4 pracovaly s daty podobně a dělaly naprosto stejné chyby. U nesprávně „klasifikovaných“ nahrávek zpravidla nehrálo roli, zda se jednalo o mužský nebo ženský hlas hrající emoci. Chyby spíš způsobovaly jednoslovné věty jako např. *„Jé.“*, *„Dobře.“* nebo *„Super.“*, které byly zkrátka pro rozeznání emoce příliš nejednoznačné a posluchačům připadaly jako vytržené z kontextu.

Závěr

Cílem této bakalářské práce bylo přiblížit téma lidských emocí z pohledu nejen psychologického ale hlavně z hlediska jejich klasifikace. V první části teoretických kapitol bylo pojednáno o tvorbě řeči samotné, protože poskytuje mnoho důležitých aspektů, do kterých je třeba mít vhled při řešení této problematiky. Dále je tu pak řeč z pohledu jazyka, kterým se zabývá mnoho studií a vědních oborů větvících se podle konkrétního zaměření. Mezi ty patří lingvistika spolu s fonetikou, fonologií a všemi jejich podkapitolami. Intonace a intonační křivky, které jsou v názvu této práce, patří pod již zmíněnou lingvistiku a hrají důležitou roli ve všech světových jazycích. V kapitolách 1.3 a 4.2 byly rozebrány jednotlivé prvky intonace, její funkce v komunikaci a souvislosti nebo naopak rozdíly, se kterými se můžeme setkat v jiných jazycích. Pro názornou ukázku byly vytvořeny grafy intonačních křivek vzorků řečových signálů.

Velkou kapitolu pokryly lidské emoce jako takové. Ty, které byly vybrané pro zpracování této práce, byly podrobně rozebrány v jednotlivých podkapitolách. Nejsou snadno definovatelné, proto se často střetávají různé druhy názorů odborníků, přesto zůstávají fascinující součástí lidského života. Po této sekci se plynule přešlo k jejich interpretaci, problematice poslechových testů a předpokladu úspěšnosti výsledků.

S tím úzce souvisí tvorba databáze, jež byla součástí zadání. Vzhledem k omezeným možnostem a nevyužití laboratorních podmínek při nahrávání z důvodu časového vytížení herců, nebylo vytvořeno tolik vzorků, kolik by bylo vhodné k dosažení uspokojivých výsledků - to však není cílem této práce. Po získání databáze následuje zpracování získaných signálů. Tento proces byl popsán z teoretického hlediska v kapitole 3.

Na tvorbu databáze navazoval rozbor intonačních křivek, které byly podrobně popsány v kapitole 4.2.

Praktickou část této práce pokrývá klasifikace emocí pomocí UNS. Nejprve bylo nutné provést parametrizaci vstupních dat. Zde se vyskytlo několik komplikací, kvůli nimž musela být matice vstupních příznaků přepracována. Tento fakt potvrdil, že je velice důležité pracovat s příznaky pečlivě, protože každá chyba může snižovat kvalitu výsledků. Bylo také zjištěno, že funkce `pitch`, na které stojí celý výpočet intonace a všech jejích příznakových variant, v některých případech nepřinášela zcela vhodné výsledky. Zde ovšem nepomohla ani úprava kmitočtového rozsahu, úspěšnost trénování zůstala stejná. Na vině může být nekvalita nahrávek, které byly bohužel realizovány v nepřívětivých podmínkách.

Během testování různých sad příznaků byl zjištěn významný přínos mel-frekvenčních kepstrálních koeficientů. Vstupní matice tedy obsahovala i prvních 13 těchto koeficientů.

V kapitole 6.4 jsou probrány tři hlavní modely, které přinesly nejvyšší úspěšnost během trénování (i následném testování). Celá databáze byla rozdělena procentuálně v poměru 90:10, kde 90 % tvořil testovací/validační dataset. Všechny 3 modely využívají křížovou validace o hodnotě 5 nebo 10. Změny probíhaly zejména v počtu skrytých neuronů, aktivační funkci a počtu iterací. Nejvyšší dosažená úspěšnost trénování byla 82 %, u testování 75 %. Tyto výsledky jsou uspokojivé vzhledem v povaze databáze, která je pro klasifikaci pomocí UNS poměrně malá.

Praktickou část uzavírá poslechový test, jehož cílem bylo porovnat výsledky uměle vytvořeného systému a přirozeného lidského vnímání emocí. Předpokladem bylo, že úspěšnost bude podobná, včetně chyb klasifikace, které modely UNS provedly. Toto se zcela potvrdilo, ve dvou případech se úspěšnost respondentů shodovala s modely 1 a 2. V jednom případě bylo dosaženo 87,5 %, což je o více než 10% rozdíl nejlepšího modelu 3. Je zde tedy zřejmé, že výsledky jsou individuální a mohou záviset teoreticky i na tom, zda se člověk věnuje hudbě (příp. psychologii) a vnímá intonaci, resp. melodii promluvy čitelněji.

Literatura

[1] HÁLA, Bohuslav, SOVÁK, Miloš. *Hlas. Řeč. Sluch*: 2. vydání Praha: Česká grafická unie a. s. 1947. 297 s.

[2] SIGMUND, Milan. *Analýza řečových signálů*: 1. vydání. Brno: Vysoké učení technické, 2000. ISBN 80-214-1783-8.

[3] SIGMUND, Milan. *Rozpoznávání řečových signálů*: 1. vydání. Brno: Vysoké učení technické, 2007. ISBN 978-80-214-3526-1 .

[4] ČERNÝ, Jiří. *Dějiny lingvistiky*: Votobia, 1996. ISBN 80-85885-96-4.

[5] VLČKOVÁ-MEJVALDOVÁ, J. *Prozodie, cesta i mříž porozumění*: 1. vydání. Praha: Karolinum, 2006. ISBN 80-246-1266-6.

[6] KRČMOVÁ, Marie. *Nový encyklopedický slovník češtiny*: [online] [cit. 2021-12-1] `https://www.czechency.org/slovnik/FONETIKA`

[7] KRČMOVÁ, Marie. *Fonetika a fonologie*: 3. vydání Elportál, Brno : Masarykova univerzita. ISSN 1802-128X. 2009.

[8] NAKONEČNÝ, Milan. *Lidské emoce*: Praha: Academia, 2000. ISBN 80-200-0763-6.

[9] Více autorů. *Věda o emocích: Základy psychologie emocí*: [online] [cit. 2021-11-13]

`https://www.superionherbs.cz/veda-o-emocich-zaklady-psychologie-emoci/`

[10] Autor neznámý. *File:Plutchik-wheel.svg*: [online] [cit. 2021-05-26]. Dostupné z `https://commons.wikimedia.org/wiki/File:Plutchik-wheel.svg#filelinks`

[11] CARLISLE, Claire. *Spinoza, part 6: Understanding the emotions*: [online] [cit. 2021-12-8] `https://www.theguardian.com/commentisfree/belief/2011/mar/14/spinoza-understanding-emotions`

[12] SCOTT, Elizabeth. *The Link Between Happiness and Health*: [online]. [cit. 2021-11-21] `https://www.verywellmind.com/the-link-between-happiness-and-health-3144619`

[13] HOLOUBEK, Jiří. *Hněv je pozitivní emoce, ale neumíme s ním pracovat*: [online]. [cit. 2021-11-21] `https://dvojka.rozhlas.cz/hnev-je-pozitivni-emoce-ale-neumime-s-nim-pracovat-7614961`

[14] HARTL, Pavel, HARTLOVÁ, Helena. *Psychologický slovník*: Praha: Portál, 2000. ISBN 80-7178-303-X. 361 s.

[15] JOSEPHY, Michal. *Nuda je velkým fenoménem moderní doby. Lidé se však nudili odjakživa*: [online]. [cit. 2021-11-21]

https://www.national-geographic.cz/clanky/nuda-je-velkym-fenomenem-moderni-doby-lide-se-vsak-nudili-uz-odjakziva.html

[16] Více autorů. *A Database of German Emotional Speech*: [online]. [cit. 2022-05-24] https://www.researchgate.net/publication/221491017_A_database_of_German_emotional_speech

[17] UNUTH, Nadeem. *Mean Opinion Score (MOS): A Measure of Voice Quality*: [online]. [cit. 2022-05-24] https://www.lifewire.com/measure-voice-quality-3426718

[18] SMÉKAL, Zdeněk. *Zpracování řeči*: Brno: Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací, 2013. ISBN 978-80-214-4896-4.

[19] SMÉKAL, Zdeněk. *Analýza signálů a soustav*: Brno: Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací, 2012. ISBN 978-80-214-4453-9.

[20] ŠRAMKA, Martin. *Klasifikace emocí*: Praha: České vysoké učení technické. Fakulta elektrotechnická. 2010.

[21] MEDALOVÁ, Kristína *Neuron a jeho stavba*: [online] [cit. 2022-02-08]

https://www.mentem.cz/blog/neuron/

[22] PANESAR, A. *Machine Learning and AI for Healthcare: Big Data for Improved Heatlth Outcomes*: 2. vydání. Coventry: Apress, 2021. ISBN 978-1-4842-6536-9.

[23] TUČKOVÁ, Jana. *Vybrané aplikace umělých neuronových sítí při zpracování signálů*: 1. vydání. V Praze: České vysoké učení technické, 2009. 224 s. ISBN 978-80-01-04229-8.

[24] *Umělá neuronová síť*: [online] [cit. 2022-05-23] https://cs.wikipedia.org/wiki/Um%C4%9Bl%C3%A1_neuronov%C3%A1_s%C3%AD%C5%A5

[25] BLAHA, Milan. *K-násobná křížová validace*: [online] [cit. 2022-05-23] https://tinyurl.com/4pr69wb9

Seznam symbolů a zkratek

CL	Classification Learner
F0	Fundamentální frekvence
FT	Fourierova transformace
FFT	Rychlá Fourierova transformace
DFT	Diskrétní Fourierova transformace
MFCC	Melovské frekvenční kepstrální koeficienty
MLNN	Vícevrstvá neuronová síť
MOS	Mean Opinion Square
SOM	Samoorganizující neuronová síť
UNS	Umělá neuronová síť

Seznam příloh

A Obsah elektronické přílohy

A.1 Databáze

V elektronické příloze se nachází složka s názvem *Databáze emotivních promluv*, která obsahuje nahrávky emocí. Složka obsahuje soubory formátu .wav a jsou pojmenovány podle dané emoce. Vzhledem k velikosti souborů byla tato příloha umístěna na Google disk a je dostupná na adrese `https://drive.google.com/drive/folders/18EoimgX26P1stoA-3Ur5hh0NOB7dKRUH?usp=sharing` včetně dalších částí přílohy.

Příklad názvu nahrávky:

- **Emoce_XY.wav-partZ**

  Emoce - název emoce, kterou nahrávka reprezentuje

  X - muž (značeno **H**) nebo žena (značeno **V**)

  Y - číslo sekvence (pouze pro účely nahrávání)

  Z - číslo střižené části (vygenerováno skriptem stříhajícím nahrávky)

  Někdy se za nahrávkou nachází ještě písmeno **A** až **C**, jedná se o ručně střižené vzory z důvodu chyby skriptu.

A.2 Skripty z MATLABu

Další součástí přílohy je řada skriptů, které byly využity pro jednotlivé operace s datasetem.

- `int.m` - tvorba grafů intonačních křivek
- `strih.m` - zautomatizovaný střih nahrávek
- `kepstrum.m` - výpočet MFCC
- `f0.m` - výpočet F0 nahrávek
- `rozsireni_vektoru.m` - doplnění nahrávek nulami
- `int_mn_D.m` - výpočet průměrné hodnoty F0 celé databáze
- `vstup_matrix.m` - finální matice příznaků

A.3 Matice příznaků

Hlavním výstupem výše uvedených skriptů jsou matice příznaků. Jedná se o soubory formátu .xlsx - jeden k trénování a validaci, druhý pro samotné testování.

- VSTUP_17-train70_val20.xlsx
- VSTUP_17-test10.xlsx

A.4 Obrázky

Příloha obsahuje též obrázky výsledků klasifikace umělých neuronových sítí včetně matice záměn jak pro trénování, tak pro testování. Dále chybové funkce všech tří modelů prezentovaných v kapitole 6.4.

Výstupy poslechových testů byly též převedeny na matice záměn, které zobrazují podrobnější výsledky.