VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

DETEKCIA STRESU V REČI

STRESS DETECTION IN SPEECH

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE — **SAMUEL ŠOLTÉS**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. FRANTIŠEK GRÉZL, Ph.D.**
SUPERVISOR

BRNO 2021

Vysoké učení technické v Brně
Fakulta informačních technologií

Ústav počítačové grafiky a multimédií (UPGM) Akademický rok 2020/2021

Zadání bakalářské práce

Student: **Šoltés Samuel**
Program: Informační technologie
Název: **Detekce stresu v řeči**
Detecting Stress in Speech
Kategorie: Zpracování řeči a přirozeného jazyka

Zadání:

1. Seznamte se s problematikou detekce stresu (psychického stavu) v řeči
2. Seznamte se s parametry relevantními pro detekci stresu
3. Implementujte výpočet těchto parametrů
4. Proveďte porovnání úspěšnosti detekce stresu z individuálních parametrů a jejich kombinace na nahrávkách řeči

Literatura:

- skripta ZRE
- Young, Steve & Evermann, Gunnar & Gales, M.J.F. & Hain, Thomas & Kershaw, Dan & Liu, Xunying & Moore, Gareth & Odell, James & Ollason, Dave & Povey, Daniel & Ragni, Anton & Valtchev, Valtcho & Woodland, Philip & Zhang, Chao. (2015). The HTK Book (version 3.5a).
- Hansen J.H.L., Patil S. (2007) Speech Under Stress: Analysis, Modeling and Recognition. In: Müller C. (eds) Speaker Classification I. Lecture Notes in Computer Science, vol 4343. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74200-5_6

Pro udělení zápočtu za první semestr je požadováno:

- 10 stran technické zprávy
- implementovaný výpočet 4 stresových parametrů

Podrobné závazné pokyny pro vypracování práce viz https://www.fit.vut.cz/study/theses/

Vedoucí práce: **Grézl František, Ing., Ph.D.**
Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.
Datum zadání: 1. listopadu 2020
Datum odevzdání: 30. července 2021
Datum schválení: 30. října 2020

Abstrakt

Stres vplýva na človeka viacerými spôsobmi a môže viesť k poklesnutiu kvality výkonu či kritickým chybám. Detekcia stresu v reči sleduje ako sa prejavujú vplyvy stresu na reči. Cieľom tejto bakalárskej práce je priblížiť vplyvy stresu, zvoliť vhodné parametre rečového signálu, na ktorých by sa vplyvy prejavili, implementovať výpočet týchto parametrov a porovnať ich úspešnosť detekcie stresu. V práci je opísaný stres a vplyv stresorov na človeka; glotálny pulz, spektrum, základná frekvencia a formanty ako parametre rečového signálu vhodné na analýzu a detekciu stresu; návrh a implementácia výpočtu týchto parametrov a dosiahnuté výsledky výpočtov parametrov na dvoch rôznych databázach.

Abstract

Stress influences people in several ways and can lead to decrease in performance and / or critical mistakes. Stress detection in speech measures the influence of stress in speech. The goal of this thesis is to offer a closer look at the impacts of stress, choose adequate parameters of speech which would manifest these impacts, implement their estimation and compare their results. The thesis contains description of stress and its effects on humans; glottal pulse, spectrum, fundamental frequency and formants as the parameters chosen for stress estimation; design and implementation of parameter value estimation from speech signal and obtained values of given parameters on two different databases.

Kľúčové slová

Stres v reči, analýza rečového signálu, spektrum, glotálny pulz, formanty, základná frekvencia

Keywords

Stress in speech, speech analysis, spectrum, glottal pulse, formants, fundamental frequency

Citácia

ŠOLTÉS, Samuel. *Detekcia stresu v reči.* Brno, 2021. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. František Grézl, Ph.D.

Detekcia stresu v reči

Prehlásenie

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne pod vedením Ing. Františka Grézla Ph.D. Uviedol som všetky literárne pramene, publikácie a dalšie zdroje, z ktorých som čerpal.

.....
Samuel Šoltés
29. júla 2021

Poďakovanie

Chcel by som sa poďakovať doktorovi Grézlovi za zdieľané rady a skúsenosti.

Obsah

Kapitola 1

Úvod

V dnešnej uponáhľanej dobe sa ľudia častokrát strmhlav ženú za splnením svojich povinností. Máme k dispozícii mnoho technológii, ktoré sa neustále vyvíjajú a zjednodušujú nám každodenné činnosti. No aj napriek tomu sa stále ocitáme pod rôznymi tlakmi. Nápor okolia alebo tenzia na seba samého vyúsťuje v stres, ktorý môže pôsobiť motivačne, no častejšie môže negatívne ovplyvniť výkonnosť jednotlivca a/alebo jeho rozhodnutia. Stres je príčinou nepozornosti, či nesústredenosti, ktoré na nás dokážu zapôsobiť pri triviálnych činnostiach a privodiť kritické následky. Chodec by mal pri prechádzaní cez cestu skontrolovať premávku a pozrieť sa na obdive strany, keď je však v strese, môže na tento veľmi jednoduchý úkon zabudnúť (a úplne ho vynechať), a to môže viesť k zrážke s vozidlom.

Ešte katastrofálnejšie následky môže stres zapríčiniť v povolaniach, ktoré kvôli svojej náročnosti vyžadujú vyššiu mentálnu aktivitu, ako napríklad vodiči hromadnej dopravy, dispečing, piloti, atď. Zlyhania v dôsledku stresu v môžu mať v takomto prípade fatálne následky, či už finančného alebo aj spoločenského charakteru. Mohlo by sa predchádzať incidentom pomocou monitoringu týchto osôb? Prejavy stresu na ľudskom organizme sú pomerne preskúmanou oblasťou, a jeho detekcia je možná senzormi telesných funkcií. Analýza zvukovej stopy je výpočtovo nenáročná, neinvazívna voči rečníkovi a je možné ju vykonávať v reálnom čase, stres by teda malo byť možné odhaliť týmto spôsobom.

Kapitola 2

Teoretická část

2.1 Stres

Stres je v slovníku cudzích slov definovaný ako funkčný stav živého organizmu, kedy je tento organizmus vystavený mimoriadným podmienkam [9].

O stres sa ľudia zaujímajú už dlhšiu dobu, boli vypracované rôzne štúdie a analýzy, a jednou z nich je aj [4], z ktorej je ďalej čerpané. Na tému stres je v dnešnej dobe vedených mnoho diskusií s istou príchuťou trpkosti, čo môže byť zapríčinené jej zložitosťou a nejednoznačnosťou. Dalo by sa povedať, že stres je psychologický koncept a ako taký nie je úplne konkrétny - nemôže byť rozoberaný priamo. Mnoho ľudí sa preto radšej tejto téme v každodennom živote vyhýba. Na stretnutí o strese spojenom s výkonmi v armáde padli rozličné názory od dôstojníkov, ako napríklad „V mojej jednotke stres nie je – nedovolím to!“, alebo „Chcem, aby moji ľudia boli v strese, drží ich to v strehu.“ a podobné, iné. Už len na základe týchto dvoch tvrdení je možné usúdiť, že stres môže mať pre rôznych ľudí rôzne významy. Sú ľudia (v tomto prípade s armádnym pozadím), ktorí tvrdia, že im stres pomohol pri vylodení na Iwo Jima, či ľudia, čo tvrdia že stres im skoro pokazil pristávací manéver pri pilotovaní lietadla. Diskusie na túto tému majú pozitívnu, ale aj negatívnu konotáciu, pričom niektorí ľudia sa takýmto diskusiám úplne vyhýbajú.

Rozhodnutie nezúčasťňovať sa konverzácie o strese a jeho vplyvoch na základe nejednoznačnosti tohto konceptu je pochopiteľné, je však diskutabilné, či je toto rozhodnutie správne. Porozumenie stresu a jeho vplyvu na človeka by malo byť dôležité nie len pre ľudí z armády, ale aj pre vedcov, či širokú verejnosť.

Jedným z dôvodov prečo sa o to snažiť by mal byť fakt, že žijeme v dobe, kedy je kladený dôraz na podávanie výkonov. Postupne sa zvyšuje zložitosť aspektov života, moderné technológie hýbu svetom a potenciál nechcených kritických chýb rastie. Výletné, či nákladné lode sa často v minulosti potýkali s incidentmi v podobe zrážky či narazenia na plytčinu, dnes to však sú 300 metrov dlhé supertankery prevážajúce milióny litrov ropy pozdĺž pobreží. Tieto lode sú vybavené najnovšími elektronickými či navigačnými „zázrakmi“, to však, zdá sa, im umožňuje len rýchlejšie vybavenie v prístave a plavbu v hazardnejších podmienkach.

Najhoršia ropná škvrna v histórii Spojených štátov amerických, nehoda Exxon Valdez, zanechala v aljašskom zálive Princa Williama 11 miliónov galónov ropy, no napriek tomu sa neradí ani medzi 25 najhorších havárií ropných tankerov vo svete. Ak vezmeme do úvahy momentálny stav námornej prepravy (kapitáni lodí sú pod enormným produkčným tlakom, posádky nie sú dostatočne vytrénované a obmedzenia na medzinárodnej úrovni sú inkonzistentné), je pravdepodobné, že tieto nehody budú pribúdať. V kontraste k lodnej doprave

však komunita komerčného letectva drží skutočne závideniahodné štatistiky. Stáva sa však čoraz populárnejším spôsobom cestovania, piloti sú tlačení k dodržiavaniu harmonogramov a možné nehody by si vyžiadali veľké množstvo obetí vo vzduchu či na zemi. 19. júla 1989 letu 232 aerolinky United Airlines zlyhal motor a v kabíne stratili hydraulický tlak, čo vyústilo k strate ovládacích prvkov. Náraz lietadla počas improvizovaného pristátia v Sioux City v Iowe zapríčinil smrteľné zranenia 111 z celkových 296 ľudí na palube. Fakt, že letová posádka bola schopná vykonať pristávací manéver s lietadlom s minimálnymi ovládacími prvkami bol považovaný za čiastočný zázrak.

V armáde sa nachádza mnoho príkladov zložitých a náročných pracovných prostredí. Systémy protivzdušnej obrany na palubách námorných lodí umožňujú vojakom detekovať lietadlá na veľké vzdialenosti. Keď je však cieľ identifikovaný, množstvo a komplexnosť informácií, ktoré musia byť spracované za krátku dobu je enormné. Kombinácia týchto faktorov vytvára potenciál pre chyby, ako napríklad zostrelenie iránskeho komerčného letu americkou loďou USS Vincennes v roku 1988. Armáda, petrochemický priemysel, ťažba, námornítcvo, doprava, či nukleárny priemysel sú príklady odborov, v ktorých súčasťou náplne práce zamestnancov je pracovať so zložitými nástrojmi a robiť kritické rozhodnutia v minimálnom čase pod obrovským tlakom. Chyby týchto zamestnancov môžu mať katastrofické následky či už na ľudských životoch, na životnom prostredí, alebo na ekonomike. Pri vyšetrovaní týchto chýb je často ako príčina uvedená „chyba operátora“.

Podať efektívny výkon pod vplyvom stresu bolo dôležité už v časoch dávno minulých, avšak je pravdepodobné, že moderná doba a jej technológie zvýšili stres, pod ktorým musíme byť schopní pracovať, a taktiež následky v prípade, že je táto práca neefektívna. To by nás však nemalo motivovať k zanechaniu vyspelej technológie a návratu k dobám jednoduchším, no skôr k vylepšeniu nášho prístupu a zabezpečenie náročných procesov. Začalo sa teda hovoriť o strese ako o negatívnom vplyve v priemysle, ozbrojených silách, letectve, a iných odvetviach so zvýšenou náročnosťou pracovnej náplne.

Stres však nie je prítomný len v už vyššie spomínaných zamestnaniach, ale dalo by sa povedať, že všade okolo nás. Členovia záchranných zložiek sa nachádzajú vo vypätých situáciách v teréne každý deň. Čo však ľudia v kanceláriách či na cestách, za volantom automobilov? Aj v týchto prostrediach sme vystavení stresorom v podobe hluku, výkonnostného tlaku, predvídania hrozieb, časového tlaku, náročnosťou vykonávaných úloh, nátlaku skupiny a iných. Všetky tieto stresory môžu mať za následok trvalé škodlivé účinky na zdravie človeka. Zrýchlené bitie srdca, namáhavé dýchanie či chvenie; emočné reakcie ako strach, úzkosť či frustrácia, strata motivácie; otupenie poznávacích schopností v podobe zúženej pozornosti, zníženia schopnosti hľadania, dlhší reakčný čas na periférne podnety a znížená ostražitosť, zhoršená schopnosť riešit problémy, strnulosť výkonu, zmeny v sociálnom správaní (strata tímovej perspektívy) alebo zníženie prejavov prosociálneho správania (pomáhanie); a v neposlednom rade znížená imunita voči chorobám – toto všetko môžu byť následky vplyvu stresorov na ľudskú bytosť. Výskum ukázal, že stres z výkonu môže zvýšit chybovosť pri procedúrach až trojnásobne. Naviac bolo dokázané, že čas potrebný na vykonanie manuálnych úloh sa pri stresových podmienkach zdvojnásobil [4].

Na druhej strane, stres v oblasti rozpoznávania reči by sme mohli vnímať ako výkyvy reči zapríčinené emočným stavom, únavou, vysokou pracovnou záťažou, hlučným prostredím alebo nedostatkom spánku. Na základe týchto príčin sa vyskytujú štyri typy prejavov stresu ako záťaže v reči:

1. neutrálny
2. nahnevaný

3. hlučný

4. Lombardov efekt

Tieto prejavy odzrkadľujú emócie rečníka a asi každý človek už počul alebo rozprával každým z týchto prejavov. Neutrálna reč nepreukazuje žiadne príznaky stresu. Je to teda rečový prejav za obvyklých, pokojných podmienok. V roku 1911 otorinolaryngológ[1] Étienne Lombard vydal článok, v ktorom popisoval jeho pozorovania pri práci v nemocnici. Všimol si, že ak bol pacient pohrúžený do konverzácie vystavený intenzívnym zvukom, zdvihol hlasitosť jeho reči. Naviac, pacient si toho nebol vedomý, z čoho Lombard vydedukoval, že ide o mimovoľný reflex. Lombardov efekt ale však okrem zvýšenia hlasitosti rečníka popisuje aj iné zmeny v jeho prejave, ako napríklad zvýšenie základnej frekvencie, či predĺženie trvania signálu. S vyššou pravdepodobnosťou budú pri vyslovovaní predĺžené samohlásky a Lombardov efekt sa bude prejavovať intenzívnejšie u mužov, než u žien. Narozdiel od vedomého, úmyselného hlučného prejavu sa však parametre reči pri Lombardovom efekte budú líšit [21].

2.2 Analýza rečového signálu

Veľká časť fyziky sa zaoberá vibráciami a vlnením rôznych druhov. Či už je to akustika, mechanika tekutín, optika, elektromagnetika či kvantová mechanika, všetky tieto disciplíny sa zaoberajú signálom a jeho spektrom. Dobrým príkladom by mohol byť mikrofón zaznamenávajúci muzikanta hrajúceho tón na husliach alebo akomkoľvek inom nástroji. Tento mikrofón vyprodukuje napätie proporčné k okamžitému tlaku vzduchu. Osciloskop nám v ideálnom prípade zobrazí periodickú funkciu, $F(t)$. Frekvencia tohto tónu bude, povedzme 440 Herzov – tón ladenia pre orchester.

Zaznamenaný signál však nebude číra sínusoida, bude obsahovať takzvané harmónie alebo alikvotné tóny – násobky základnej frekvencie s rôznymi amplitúdami v rôznych fázach[2], závislými na zafarbení tónu, hudobnom nástroji a hrajúcom muzikantovi. Táto krivka môže byť analyzovaná pre nájdenie amplitúd alikvótnych tónov, či nájdenie amplitúd a fáz sínusoíd, z ktorých pozostáva [8].

Rečový signál môže byť taktiež analyzovaný na viacero účelov. Syntéza reči, ktorej snahou je prirodzene znejúca umelo generovaná reč. Detekcia hlasovej aktivity, kedy je cieľom vyhľadať časti signálu, v ktorých sa vyskytuje reč. Ďalej je možné zvýšiť kvalitu signálu filtrovaním a odstránením šumu, či analyzovať obsah signálu – získať textový prepis hovorenej reči, či detekovať kľúčové slová. S tým je spojené aj rozpoznávanie a identifikácia rečníka, či informácie o emočnom stave rečníka.

Z rečového signálu je možné získať mnoho parametrov. Nie všetky sú však vhodné na analýzu prítomnosti emócií a je preto potrebné vybrať tie, na ktorých by sa efektívne odrazili fyziologické vlastnosti rečníka pod stresom.

Prvým krokom analýzy signálu býva proces emfázy. Ide o zlepšenie prenosových parametrov, a väčšinou sa vykonáva filtrovaním. Filtrovanie je proces redukovania šumu spôsobeného rozruchmi v prostredí počas nahrávania rečového signálu. Účelom preemfázového

[1] Otorinolaryngológia sa zaoberá diagnostikou a liečbou ochorení ucha, nosa, nosovej dutiny a prínosových dutín, hltana a hrtana.

[2] Fáza popisuje uhol spomalenia jednej vlny či vibrácie s ohľadom na inú. Spomalenie o jednu vlnovú dĺžku je ekvivalentné fázovému posunu 2π a každá harmónia má svoju vlastnú fázu ϕ_m popisujúcu jej pozíciu v vrámci periódy.

filtra je zosilniť energiu signálu vo vyšších frekvenciách, ktoré sú oslabené vo vokálnom trakte pri tvorbe reči.

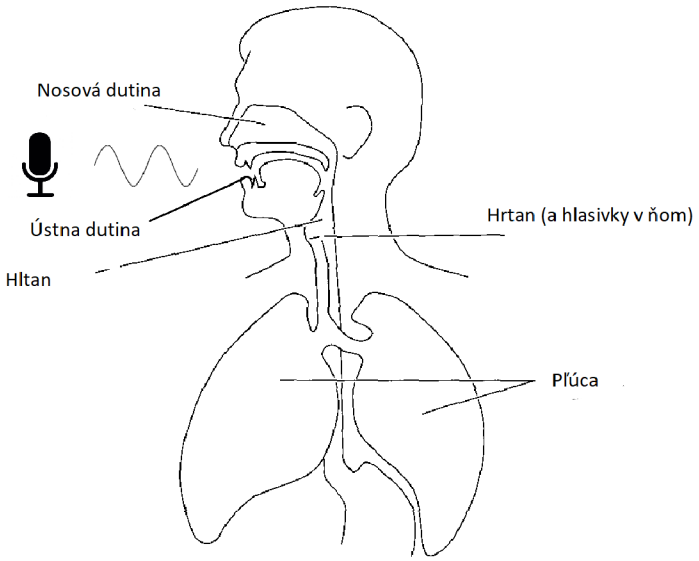
Rečový signál je nestacionárny, čo znamená, že sa niektoré jeho parametre môžu rapídne meniť v čase, a preto je potrebné ho analyzovať po častiach – tieto časti sa nazývajú rámce, a zvyčajne sú veľké 10 až 25 milisekúnd. Použitie rámca eliminuje prudké zmeny a zlepšuje výpočet hodnôt parametrov. Čiastočné prekrytie jednotlivých rámcov znižuje rozdiely medzi týmito rámcami [2].

Sú však aj parametre, ktorých hodnoty sa intenzívne neodrážajú na rámcoch, ale je potrebné ich analyzovať na väčších častiach – napríklad skracovanie slov by bolo ťažko viditeľné na 25 milisekundách signálu.

2.3 Glotálny pulz

2.3.1 Tvorba reči

Drugman [5] opisuje systém produkcie reči, v ktorom značnú rolu hrá glotálny pulz. Tento systém umožňuje rečníkovi vydávať široký rozsah zvukov. Pozostáva z mnohých orgánov podieľajúcich sa na procese fonácie, ktoré môžu byť rozdelené do troch skupín: pľúca, hrtan a hlasový trakt. Z fyziologického pohľadu, prúd vzduchu z pľúc je tlačený hrtanom, kde je modulovaný vibrovaním hlasiviek. Toto vibrovanie prevádza prúd vzduchu na akustické pulzy a poskytuje excitačný signál hlasovému traktu. Hlasový trakt pozostáva z orálnej, nosovej, a hltanovej rezonančnej komory, ktoré tento modulovaný signál ešte viac tvarujú alebo filtrujú. Výsledný prúd je šírený perami. Manipulovaním vibrácií hlasiviek, či konfiguráciou hlasového traktu nad hrtanom artikuláciou jazyka, sánky, mäkkého podnebia či pier je možné vyprodukovať rôzne druhy zvukov.



Obr. 2.1: Preložený obrázok tvorby reči prebraný z webovej stránky[3].

Zdroj hlasu či glotálny zdroj slúži ako excitácia reči tvorenej hlasivkami. Počas znelých zvukov, oscilácia hlasiviek pravidelne prerušuje prúd vzduchu z pľúc a vytvára zmeny

[3] https://www.azlifa.com/phonetics-phonology-lecture-2-notes/

v tlaku vzduchu. Hlasivky sú zvyčajne otvorené, ked nie je fonovaný žiaden zvuk. Pri produkcii neznelých zvukov, napríklad „s“ alebo „ch“, sú hlasivky taktiež otvorené, umožňujúc prúdu vzduchu prechod, no v istom bode hlasového traktu je vytvorené zúženie, ktoré produkuje turbulenciu. Pri produkcii znelých zvukov, napríklad „a“ alebo „e“, svaly kontrolujúce hlasivky (adduktorové) spoja hlasivky spolu, aby vytvorili odpor tlaku vzduchu prichádzajúcemu z pľúc. Tento tlak pod hlasivkami (subglotálný tlak) núti hlasivky k otvoreniu, čo umožňuje vzduchu prejsť cez glotis. K spätnému uzatvoreniu glotisu prispievajú dva faktory. Jedným je elasticita tkaniva, ktorá núti hlasivky k opätovnému nadobudnutiu ich pôvodnej konfigurácie blízko stredu (zatvorenej pozície). Druhým faktorom sú aerodynamické sily. Jednou z nich je Bernoulliho zákon zachovania energie, ktorý zapríčiňuje pokles tlaku medzi hlasivkami keď sa rýchlosť vzduchu zvýši. Iná aerodynamická sila sa prejavuje, keď sa vo vzduchu pri opúšťaní glotisu vytvoria víry, čo má za následok viac negatíveho tlaku medzi hlasivkami. Po zatvorení hlasiviek sa pod nimi opäť vytvorí tlak a sú znovu roztlačené, reštartujúc celý proces. Tento cyklus sa opakuje mnohokrát za sekundu, a trvanie tohto cyklu sa označuje „základná perióda“. Frekvencia tohto cyklu sa označuje ako „základná frekvencia“, alebo výška hlasu, a viac o nej v časti 2.4.

2.3.2 Spracovanie glotálneho pulzu

Modálny hlas predstavuje zväčša neutrálny typ fonácie s malými odchýlkami od periódy k perióde v po sebe idúcich glotálnych cykloch. Taktiež predpokladá, že okolo okamžitého uzavretia glotisu sa nachádza významná excitácia. Nemodálne typy fonácie zahŕňajú značnú rozdielnosť charakteristiky glotálneho zdroja. Vysokorýchlostné fotografovanie pomáha pri porozmení rozdielom medzi vibrovaním hlasiviek medzi rôznymi fonačnými typmi. Vizuálne zobrazenie hlasiviek však nie je možné počas prirodzenej produkcie plynulej reči, keďže aj hlasový trakt je časovo premenlivý. Naviac, vizuálne viditeľné vlastnosti glotálnej vibrácie nemusia byť až také podstatné pri pohľade na produkciu rečovej akustiky alebo vnímania reči. Elektroglotografia (EGG) sa využíva na skoro neinvazívne meranie časovo premenlivej impedancie medzi elektródami umiestnenými na opačných stranách krku počas produkcie plynulej reči. Síce nemôže analyzovať prúd vzduchu v glotise, no môže byť použitá ako referencia.

Metódy používané k odvodeniu fyzických či analytických modelov glotálneho zdroja pomáhajú porozumieť niektorým vlastnostiam glotálneho vibrovania, no nepomáhajú už pri samotnom získaní samotných hodnôt týchto vlastností z rečového signálu, a teda nepomáhajú s identifikáciou vlastností týkajúcich sa rozpoznaia emócií či špecifických zvukov. Podľa modelu lineárnej produkcie reči z roku 1970 sú rečové signály generované filtrovaním hlasového zdroja funkciou hlasového traktu. Počas histórie boli testované modely rôznej náročnosti odvodené od fyziologických meraní. Jeden z týchto modelov analyzoval periódu glotálneho zdroja. V tomto modeli sa pojem „otvorená fáza“ používa na popis časového úseku, počas ktorého sa hlasivky otvárajú, pričom „spätná fáza“ pomenúva čas návratu hlasiviek do pôvodného stavu. Otvorená fáza je ďalej rozdelená na otváraciu a zatváraciu fázu, ktoré sú definované priechodom glotálneho pulzu maximom. Po spätnej fáze hlasivky ostávajú zavrené počas takzvanej „zavrenej fázy“ [5].

2.3.3 Analýza glotálneho pulzu

Matematicky by znelá samohláska mohla byť modelovaná ako výstup lineárneho a časovo nemenného procesu filtrovania. Kvázi periodický glotálny signál v tvorbe reči na mieste hlasivkovej štrbiny, $g(n)$, slúži ako akustický zdroj a vzbudzuje filter reprezentujúci vokálny

trakt, $h(n)$. Výsledná tlaková vlna reprezentujúca reč by pred perami mohla byť vyjadrená ako:

$$s(n) = g(n) * h(n), \tag{2.1}$$

kde $*$ predstavuje konvolúciu [15].

Na zistenie odhadu glotálneho signálu je vhodné použiť metódu IAIF – teda iteratívne adaptívne inverzné filtrovanie [15], [16]. Táto metóda odstraňuje efekty vokálneho traktu efektívnym spôsobom a umožňuje získať presný odhad glotálneho pulzu.

Úspešnosť až 88% pri detekovaní reči pod stresom sa podarilo pri analýze glotálneho pulzu dosiahnuť akademikom z FEKT VUT a FEL CTU [15]. Ich cieľom bolo nájsť podobné rečové charakteristiky reči pod stresom na základe distribučných matíc glotálnych pulzov. Pre účelné porovnávanie matíc našli vhodné vlastnosti, ako napríklad priamé rezy urobené na referenčnej pozícii. Matice obsahujúce glotálny pulz reči pod stresom sa javili byť čiernejšie než matice obsahujúce neutrálnu reč. Upozorňujú však, že spoľahlivosť analýzy glotálneho pulzu je veľmi závislá na analyzovaných dátach – vyskytujúce sa stresory, podmienky pri získavaní dát, či vek rečníkov.

2.4 Základná frekvencia

2.4.1 Význam

Reč sa skladá z náhodných fluktuácií vzduchu vydychovaných z pľúc, ktorý je časovo a frekvenčne modulovaný a časovo-spektrálne tvarovaný po smere z pľúc až k perám. Pri glotise, alebo hlasivkovej štrbine, rýchlosť otvárania a zatvárania hlasiviek určuje základnú frekvenciu reči a časová variácia základnej frekvencie primárne určuje intonáciu reči. Základná frekvencia je ľudským uchom vnímaná ako výška hlasu, nízka hodnota základnej frekvencie je teda hlboký hlas, a vysoká základná frekvencia je vysoký hlas, pričom intonácie sú trajektórie zmien základnej frekvencie.

Rečový signál je viacvrstvový, teda obsahuje rôzne druhy informácií. Segmentové, sprostredkované hlavne hovorenou formou postupnosti slov a nadsegmentové, sprostredkované a úmyselnou intonáciou [10].

Podľa Nishia [11] úroveň artikulácie slov taktiž závisí na osobnom štýle rozprávania, rýchlosti reči, prízvuku či emočnom stave rečníka.

Funkcia výšky hlasu a intonácie sa líši s hovoreným jazykom. V tonálnych jazykoch, ako napríklad čínština alebo niektoré africké jazyky a taktiež do istého rozsahu japončina, zmena tónu či výšky hlasu môže úplne zmeniť jeho význam, teda zdanlivo rovnako znejúce slová (obzvlášť pre zahraničných či nerodilých hovoriacich) môžu mať absolútne rozdielne významy [18].

V netonálnych jazykoch, ako napríklad angličtine, výška hlasu slúži na sprostredkovanie doplňujúcich informácií k obsahu. Môže slúžiť napríklad na:

- rozlíšenie medzi otázkou a oznamovacou vetou,
- naznačenie úmyslu rečníka,
- zdôraznenie určitej časti preslovu,
- naznačenie reakcií, ako súhlasu, prekvapenia, či nudy,
- naznačenie emócií, ako hnevu, radosti, spokojnosti či ľahostajnosti,
- naznačenie hranice viet

2.4.2 Charakteristika základnej frekvencie

Rozmedzie hodnôt základnej frekvencie, výšky hlasu, sa líší prakticky od nízkych hodnôt okolo 20 Hertzov, až po vysoké hodnoty hlasov detí, a to okolo 3000 Hertzov. Rozdiely môžu byť zapríčinené:

- fyziologickými vplyvmi veku, pohlavia a charakteristikou rečníka,
- vedomou intonáciou,
- emóciami, štýlom, prízvukom či spevom.

Základná frekvencia sa zdá byť taktiež najlepším parametrickým indikátorom pohlavia. Typický dospelý muž má priemernú hodnotu výšky hlasu okolo 120 Hertzov (v rozmedzí od 85 do 180 Hz) a typická dospelá žena má priemernú hodnotu výšky hlasu približne 210 Hertzov (s možnými hodnotami od 165 do 255 Hz) [17].

2.4.3 Problémy pri odhade základnej frekvencie

Avšak aj pri extrakcii základnej frekvencie sa môžu vyskytnúť isté komplikácie. Časovozávislá povaha výšky hlasu – naznačuje to, že perióda, alebo jej opak, základná frekvencia, určená z časového rámca je prinajlepšom priemerná hodnota periódy, základnej frekvencie, v rámci. Skutočná perióda sa môže podstatne líšiť medzi rámcami alebo môže na základe emočného stavu hlasu oscilovať v rámci.

Nejednoznačná povaha kvázi-periodickej reči – pre prechodné úseky reči, konkrétne pre nástupnú časť a koniec znelého segmentu, môže byť aj pre experta problém vizuálne určiť správnu hodnotu výšky hlasu, keďže sa perióda mení nepravidelne či drasticky. Naviac sa niekedy môže stať, že v znelom rámci sa amplitúda signálu a / alebo pomer harmónie k šumu výrazne znížia.

Chýbajúca základná frekvencia – sporadicky sa môže základná frekvencia zhodovať s korytom, alebo anti-rezonanciou spektrálnej obálky tak, že prvá pozorovateľná harmónia je v skutočnosti až druhá alebo tretia harmónia.

Polovičný a dvojnásobný odhad výšky hlasu – periodický signál s periódou T ukazuje korelácie vrcholov na kladných celočíselných násobkoch T. To môže viesť k „polovičnej výške hlasu“, chybnému odhadu výšky hlasu, ktorý je o oktávu nižšie než skutočná hodnota výšky hlasu, v prípadoch kde odhad na hodnote $2T$ je vyšší než na hodnote T. Pre niektoré segmenty sa môže periodicita ukazovať na polovici periódy, čo zapríčiňuje odhad „dvojnásobnej výšky hlasu“, a teda hodnotu o oktávu vyššie než skutočná výška, kde hodnota na $T/2$ je vyššia než na T.

Chyba znelosti – pri odhade základnej frekvencie je reč rozdelená do dvoch stavov, znelého a neznelého, pričom znelý stav obsahuje harmonickú štruktúru, a neznelý štruktúru podobnú šumu. Chyba pri detekcii znelých a neznelých stavov ovplyvňuje presnosť odhadu výšky hlasu.

Pomer harmónie k šumu – všeobecne sú znelé signály zložené zo zmesi harmónií a šumov. Odhad výšky hlasu sa zlepšuje so zvyšujúcim pomerom harmónie k šumu, a naopak zhoršuje keď tento pomer klesá, ako napríklad pri zadychčanom, trasúcom či chrapľavom hlase.

Šum – pri odhade výšky hlasu, obzvlášť v prostredií mobilnej komunikácie, tak aj ako pri iných metódach venujúcich sa analýze signálov, presnosť odhadu je ovplyvnená šumom či hlukom v pozadí, a znižuje sa pri zvýšení týchto vplyvov.

Rečové vady a prekážky – koktanie, apraxia[4], dyzartria[5], hlasové poruchy, nemota, či iné môžu skomplikovať odhad výšky hlasu [13].

2.4.4 Výška hlasu a stres

Základná frekvencia a jej chovanie pri strese boli skúmané aj vedcami z univerzity v Texase [3]. Podarilo sa im zaznamenať nárast strednej hodnoty základnej frekvencie pri vedení vozidla a vykonávaní sekundárnych úkonov, rozhovorom s automatickým dialógovým systémom a rozhovore so spolujazdcom. Keďže skúmali aj formanty, ich experiment je spomenutý aj v časti 2.6.

Stredná hodnota a smerodajná odchýlka základnej frekvencie boli skúmané aj medzinárodným tímom z Francúzska a Kréty [16].

2.5 Spektrálna analýza

2.5.1 Diskrétna Fourierova transformácia

Rozklad signálu na kombináciu iných signálov je využívaný vo viacerých oblastiach. Veľkosť digitálnych súborov môže byť zmenšená odstránením niektorých nedôležitých signálov kombinácie – rovnako uľahčenie porovnávania zvukových signálov, či filtrovanie rádiových vĺn na vyčistenie od šumu.

Rečový signál je vzorkovaný, čo znamená že sú jeho hodnoty v diskrétnom čase. Preto sa na reálne signály používa diskrétna Fourierová transformácia, alebo:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{k}{N} n}, \tag{2.2}$$

kde $k = 0 \ldots N-1$, a počíta Fourierovú transformáciu pre diskrétne vzorky. Algoritmicky je implementovaná pomocou FFT, alebo rýchlej Fourierovej transformácie, ktorá je výpočetne násobne rýchlejšia.

Distribúcia spektrálnej energie vyslovenej reči závisí na emočnom obsahu. Bolo pozorované, že vysoko vzrušujúce pocity ako šťastie či hnev zapríčiňujú vyššie hodnoty energie na vyšších frekvenciách, zatiaľ čo napríklad smútok má v týchto vyšších frekvenciách hodnoty energie nižšie [2].

Typicky sa počíta spektrum signálu po rámcoch, z dôvodu nestacionarity signálu.

Fourierovu transformáciu, respektíve jej výsledok – spektrum a jeho vlastnosti, bolo analyzované tímom zaoberajúcim sa detekciou kognitívnej záťaže a frustrácie u vodičov [3]. Hlavnými vlastnosťami spektra, ktoré pozorovali, boli spektrálne ťažisko a spektrálny rozptyl. Očakávali, že oba parametre budú preukazovať zmeny po zvýšení kognitívnej záťaže.

2.6 Formanty

Slovo formant môže v akustike nadobúdať viacero významov, v roku 1994 však Acoustical Society of America definovala formant ako rozsah frekvencií komplexného zvuku, na ktorých sa nachádza absolútne alebo relatívne maximum v zvukovom spektre [19]. Formanty sa líšia od rečníka k rečníkovi, vzhľadom na rozdielne vlastnosti hlasiviek. Pri analýze rečového

[4] potiaž s radením zvukov v slabikách a slovách
[5] paralýza rečového svalstva

signálu sa využívajú na viacero účelov, z ktorých jedným je aj detekcia stresu. Informácie obsiahnuté v najbližších odstavcoch sú čerpané z knihy *Mluvíme s počítačem česky* [14].

2.6.1 Problémy pri identifikácii formantov

Je známe, že prvé tri formantové frekvencie nesú dôležité informácie o charaktere samohlások a znelých spoluhlások a podľa priebehu formantových frekvencií v čase je možné určiť aj miesto artikulácie susedných hlások. Informácia o formantoch je najlepšie obsiahnutá v spektrálnej obálke analyzovaného úseku reči. Väčšina postupov identifikácie frekvencií formantov preto buď implicitne, alebo explicitne využíva práve spektrálnu obálku. Pri identifikácii však môžeme naraziť aj na problémy, z ktorých dva hlavné predstavujú:
◇ Výskyt nepravých vrcholov v spektrálnej obálke. Maximá v spektrálnej obálke sú normálne spôsobené iba formantmi. Môžu sa však objaviť aj ďalšie, nepravé vrcholy, ktoré sú zapríčinené rôznymi poruchami. V prípade, že je spektrálna obálka určovaná metódou lineárne prediktívneho kódovania (LPC)[6], môžu byť tieto poruchy zapríčinené tým, že rád prediktoru je predimenzovaný a prediktor môže nadbytočné póly využiť v určitých prípadoch k vytvoreniu nepravých vrcholov.
◇ Splývanie formantov. Jeden z najobtiažnejších prípadov pri identifikácii formantov nastáva, keď sú dve formantové frekvencie tak tesne blízko seba, že individuálne špičky v spektrálnej obálke splývajú a nie je možné ich od seba jednoducho odlíšiť.

2.6.2 Spôsoby identifikácie formantových frekvencií

Väčšina postupov identifikácie formantových kmitočtov pracuje vo frekvenčnej oblasti a vychádza z analýzy spektrálnej obálky stanovenej metódou LPC. V podstate existujú dva postupy ako určiť zo spektra LPC hodnoty formantových frekvencií. Jeden zisťuje korene polynómu A(z) – teda póly prenosovej funkcie H(z), a druhý hľadá na spektrálnej obálke odvodenej z lineárneho prediktoru lokálne maximá. Spektrum (zelenou) a prenosová funkcia (modrou) sú viditeľné na obrázku 2.2.
• Výpočet pólov prenosovej funkcie. Korene polynómu A(z) zistíme riešením rovnice

$$z^Q + a_1 z^{Q-1} + \ldots + a_{Q-1} z + a_Q = 0. \tag{2.3}$$

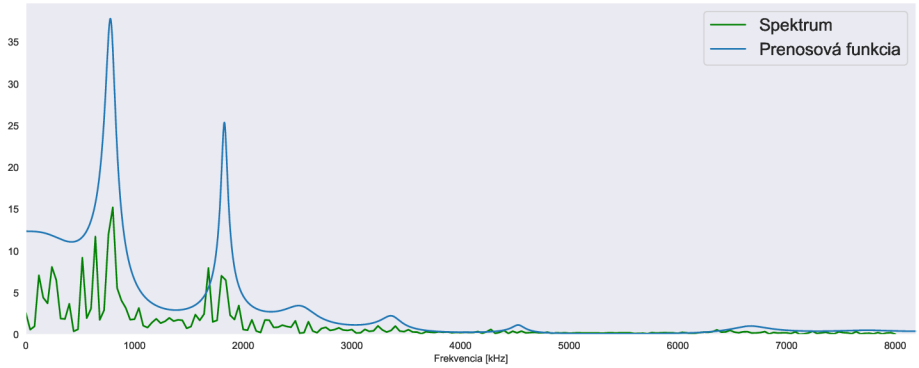
Ide o rovnicu Q-tého rádu s reálnymi koeficientmi, v jej riešení prevažujú páry komplexne združených koreňov. Na riešenie rovnice sa využívajú najčastejšie štandardné počítačové programy na hľadanie koreňov polynómu, ktoré sú obvykle založené na aplikácii Newton-Raphsonovho či Bairstowovho algoritmu. Ak uvážime jednu dvojicu komplexne združených koreňov $z_i = |z_i|e^{j\varphi_i}$ a $\overline{z_i} = |z_i|e^{-j\varphi_i}$ rozloženú v z-rovine, odpovedajúcu formantovú frekvenciu F_i a šírku pásma formantu B_i pre pokles charakteristky o 3 dB je potom možné vyjadriť vzťahmi

$$F_i = \frac{\omega_i}{2\pi} = \frac{arg(z_i)}{2\pi T} \quad [\text{Hz}], \tag{2.4}$$

$$B_i = -\frac{ln|z_i|}{\pi T} \quad [\text{Hz}], \tag{2.5}$$

kde T je perióda vzorkovania pôvodného akustického signálu.

[6] Metóda LPC sa snaží na krátkodobom základe odhadnúť parametre modelu vytvárania reči



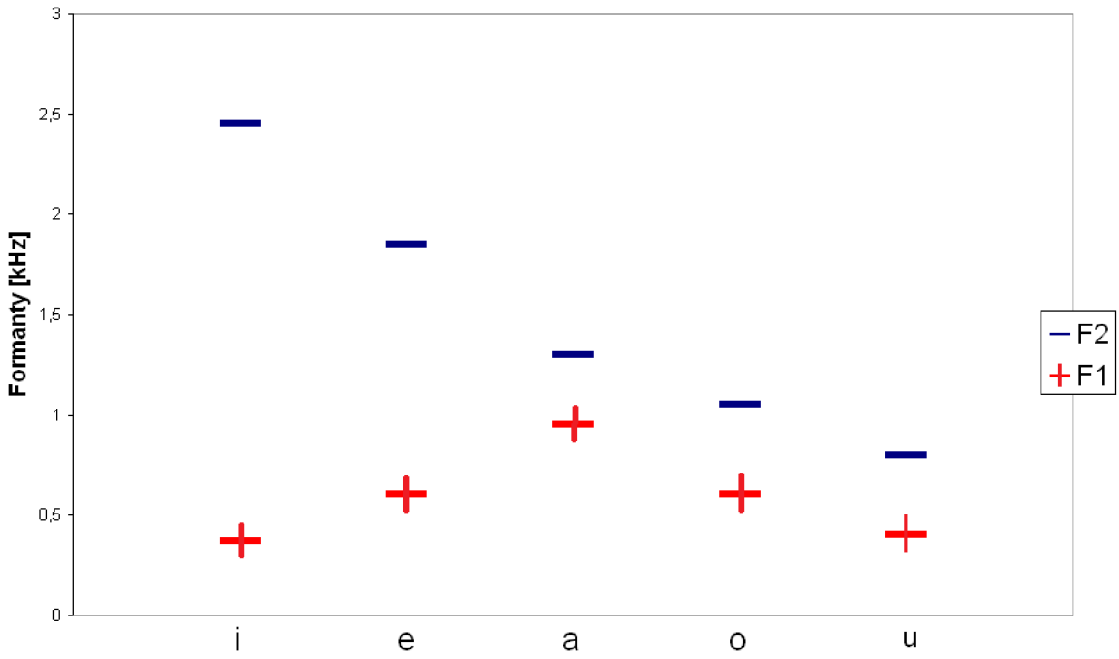
Obr. 2.2: Pozície formantov zakreslené na krivke prenosovej funkcie.

• Vyhľadávanie vrcholov spektrálnej obálky. Ďalším spôsobom zisťovania formantov je výpočet spektrálnej obálky. Formanty môžu byť následne nájdené prehľadávaním obálky a zistením hodnoty frekvencií pre lokálne maximá. K nájdeným formantom je vhodné dolpniť ešte šírku pásma vrcholov, a zamietnuť kandidátov na formanty tie vrcholy, ktorých šírka pásma je väčšia než 500 Hertzov (šírka pásma reálnych formantov nedosahuje túto hodnotu). I keď tento jednoduchý spôsob umožňuje potlačiť veľké množstvo prípadných chybných odhadov, je možné, že keď budú frekvencie susedných formantov veľmi blízko a vytvoria na spektrálnej obálke iba jediný vrchol, je veľmi pravdepodobné, že z dôvodu prekročenia šírky pásma bude tento vrchol ignorovaný. Jednou z možností, ako redukovať mieru chýb v odhadoch je požiadavok istej kontinuity v identifikovaných hodnotách formantových frekvencií susedných rámcov. Nasleduje jednoduchý algoritmus pre odhad hodnôt troch formantov v pásme do 3 kilohertzov. Táto metóda predpokladá, že v rámci $k-1$ boli stanovené hodnoty troch formantových frekvencií $F_j(k-1)(j=1,2,3)$. V k-átom rámci teda môžu nastať tieto prípady:
— Ak sú v spektrálnej obálke nájdené presne 3 vrcholy (85 až 90% prípadov), potom je možné previesť priamy prevod základných odhadov $\hat{F}_i(k)$ na formantové frekvencie $F_i(k) = \hat{F}_i(k), i = 1,2,3$.
— Ak je v spektrálnej obálke nájdený len jeden vrchol na frekvencii $\hat{F}(k)$(asi 1% prípadov), je i-té poradie odhadnutého formantu určené pravidlom najbližšieho suseda

$$i = \underset{j}{\operatorname{argmin}} \, |\hat{F}(k) - F_j(k-1)|, \tag{2.6}$$

teda $F_i(k) = \hat{F}(k)$, a zvyšné dve formantové frekvencie sú doplnené odpovedajúcimi hodnotami predošlého segmentu.
— Ak v spektrálnej obálke analyzovaného rámca reči nájde algoritmus dva, poprípade štyri vrcholy (obvykle sa jedná okolo 10 až 15% prípadov), budú poradia odpovedajúcich formantov určené analogicky podľa pravidla najbližšieho suseda a chýbajúci, respektíve prebývajúci formant bude doplnený hodnotou z rámca $k-1$, respektíve bude zanedbaný.
— Prípady s viac než štyroma vrcholmi v spektrálnej obálke nastávajú výnimočne a je možné ich ošetriť rovnako aplikáciou pravidla najbližšieho suseda a zanedbaním. Ak sú v spektrálnej obálke počiatočného rámca nájdené presne tri formanty, potom je možné počiatočné hodnoty formantov $F_i(0)$ v estimačnom algoritme jednoducho položiť $F_i(0) = \hat{F}_i(0), i = 1,2,3$. V opačnom prípade, teda že v spektrálnej obálke počiatočného rámca

nie je možné nájsť tri vrcholy, je vhodné naštartovať prehľadávanie s a priori nastavenými hodnotami formantových frekvencií $F_1(0) = 0,5$ kHz, $F_2(0) = 1,5$ kHz a $F_3(0) = 2,5$ kHz. Na obrázku 2.3 môžeme vidieť priemerné hodnoty prvého a druhého formantu českých samohlások i, e, a, o, u.



Obr. 2.3: Obrázok prebraný z Wikipedie[8]označujúci priemerné hodnoty formantov českých samohlások.

Existujú aj ďalšie efektívne algoritmy detekcie formantových frekvencií, ktoré sú založené napríklad na aplikácii skrytých Markovových modelov pri modelovaní hlások alebo na sínusovom rozklade rečových kmitov a tak ďalej [14].

2.6.3 Príklady použitia spojené so stresom

Výskumníci z Univerzity New South Wales v Sydney sa rozhodli využiť vlastnosti formantov na meranie úrovne kognitívnej záťaže[9]. Túto metódu analýzy rečového signálu zvolili na základe jej neinvazívnosti, jednoduchom vypočítaní a možnom vykonaní v reálnom čase. Z vlastností formantov vybrali frekvenciu, šírku pásma a prvý rád lineárnej regresie koeficientov frekvenčných trajektórií. Kognitívnu záťaž v ich výskume získavali z databázy nahrávok ľudí podrobujúcich sa Stroop testu. Test je založený na texte obsahujúcom pomenovanie farby a rozdielnej farbe tohto textu. Testovaný je vyzvaný buď k prečítaniu textu alebo farby textu. Ako príklad by teda mohol byť uvedený scenár, kde by bol napísaný text "červená"modrou farbou textu. Správna odpoveď na vyzvanie k prečítaniu textu by bola "červená", no k farbe textu "modrá". Príklad Stroopovho testu je na obrázku 2.4. Výskum-

[8] https://cs.wikipedia.org/wiki/%C4%8Cesk%C3%A9_samohl%C3%A1sky#/media/Soubor:Formant_values_of_Czech_vowels.png

[9] Pojem kognitívna záťaž popisuje množstvo mentálnej snahy vyžadovanej na splnene úlohy.

níci zaviedli pri testovaní umelý časový limit na splnenie. Z výsledkov zistili, že formanty obsahujú užitočné informácie o kognitívnej záťaži [20].

Obr. 2.4: Príklad Stroopovho testu.

Formanty na detekciu stresu sa rozhodli použiť aj výskumníci z Texaskej univerzity v Dallase. Analyzovali rečové signály vytvorené vodičmi osobných vozidiel pri vykonávaní dvoch rôznych sekundárnych kognitívnych činností: interakciu so spolujazdcom a hovormi s automatizovaným dialógovým systémom (neskoršia považovaná za kognitívne náročnejšiu). Pozorovaním zmien centrálnych frekvencií prvých 4 formantov zaznamenali zvýšenie pri komunikácii s dialógovým systémom v porovnaní s neutrálnou rečou [3]. Formanty obsahujú údaje o kognitívnej záťaži a aj napriek rozdielom medzi rečníkmi by mali byť pri referencii k neutrálnej reči vhodné na analýzu pre detekciu stresu.

2.7 Energetické parametre

2.7.1 Energia

Energia je súčet druhých mocnín hodnôt signálu na určitom intervale, a v tejto bakalárskej práci je počítaná jej priemerná hodnota pre daný interval vzorcom:

$$\overline{E} = \frac{1}{n_2 - n_1 + 1} \sum_{n=n1}^{n2} |x[n]|^2 \tag{2.7}$$

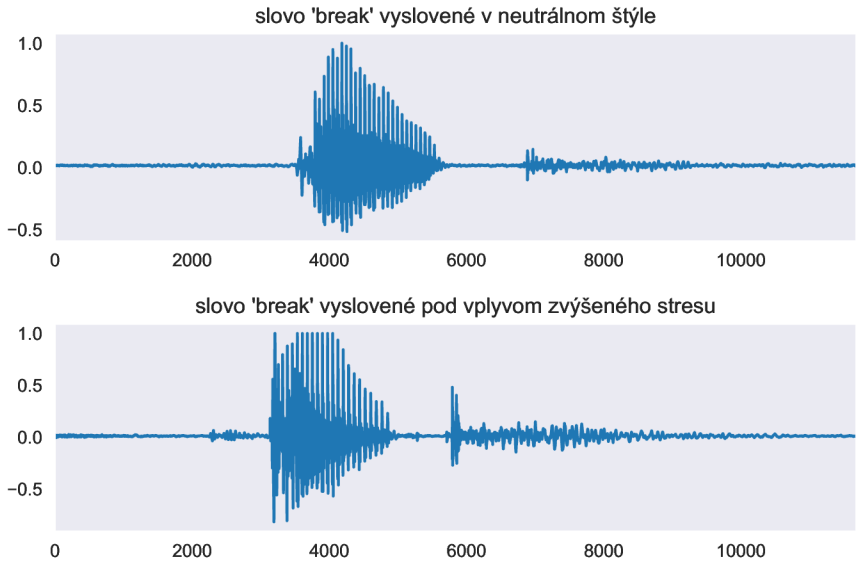
Energia signálu by mala byť pri reči pod stresom vyššia, než pri reči neutrálnej [12]. Na obrázku 2.5 dosahuje viditeľne vyššie hodnoty signál slova vysloveného pod zvýšeným stresom, teda by mala byť vyššia aj jeho energia.

2.7.2 Teagerov energetický operátor

Tím na univerzite v Melbourne spracoval vedomosti o Teagerovom energetickom operátore (TEO) a prakticky ho využil v detekcii stresu [7]. Uvádza, že Teagerov energický operátor pracuje s energiou signálu a označuje sa písmenom Ψ. Prúd vzduchu pri tvorbe reči je vo vokálnom trakte rozdelený do rôznych iných traktov, pričom každý má svoju vlastnú energiu. Vzorec pre TEO je

$$\Psi[x(n)] = x^2(n) - x(n-1)x(n+1) \tag{2.8}$$

A o signále nám môže naznačiť, že ak má signál $x(n)$ iba jednu harmóniu s konštantnou amplitúdou a okamžitou frekvenciou, hodnota TEO $\Psi[x(n)]$ by mala byť rovnakou konštantou pre všetky n – obrázok 2.6. Ak pozostáva $x(n)$ z viacerých harmónií, TEO hodnota sa



Obr. 2.5: Signály slova vyslovené rôznymi štýlmi.

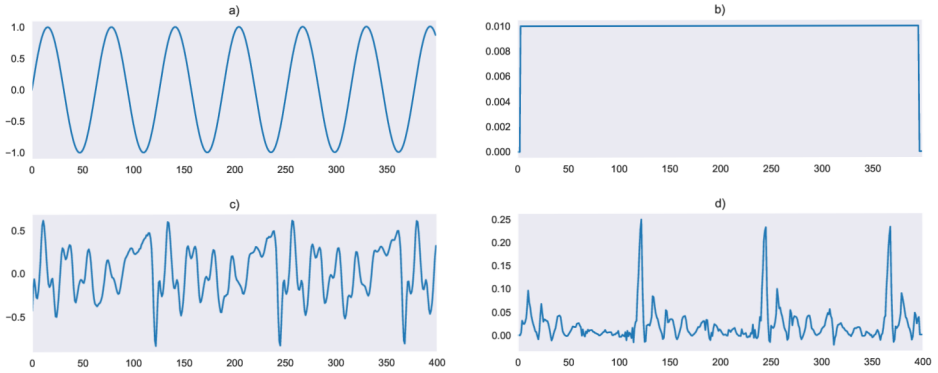
v čase mení a $\Psi[x(n)]$ je funkciou pre n. V skutočnosti však signály vždy budú pozostávať z viacerých harmónií.

TEO sa používa na detekciu emócii a stresu v reči, teda Lombardovho efektu, nahnevanej či hlasnej reči. V práci [2] je TEO získané z rečového signálu ešte pred procesom preemfázy, aby zvýšilo energie v reči, ktoré nie sú dokonale detekované v klasickej MFCC extrakcii. Táto extrakcia je teda vykonaná až nad TEO.

V Melbourne [7] najprv detekovali v rečovej nahrávke znelé a neznelé časti, a následne znelé časti akceptovali k analýze, ktorá začínala filtrovaním pomocou filtra pásmovej priepusti. Potom počítali nad rámcami TEO a normalizovanú obálku TEO autokorelácie a obsah pod touto obálkou. Detekovať stres sa im podarilo s úspešnosťou okolo 90 percent.

2.8 Zhrnutie parametrov

Z rečového signálu sa teda dajú získať mnohé parametre. Všeobecne by sme ich mohli rozdeliť do dvoch skupín, a to krátkodobé a dlhodobé. Jednu skupinu sa teda budeme snažiť získavať z rámcov, a hodnoty parametrov z druhej skupiny si budeme všímať na úsekoch väčších.



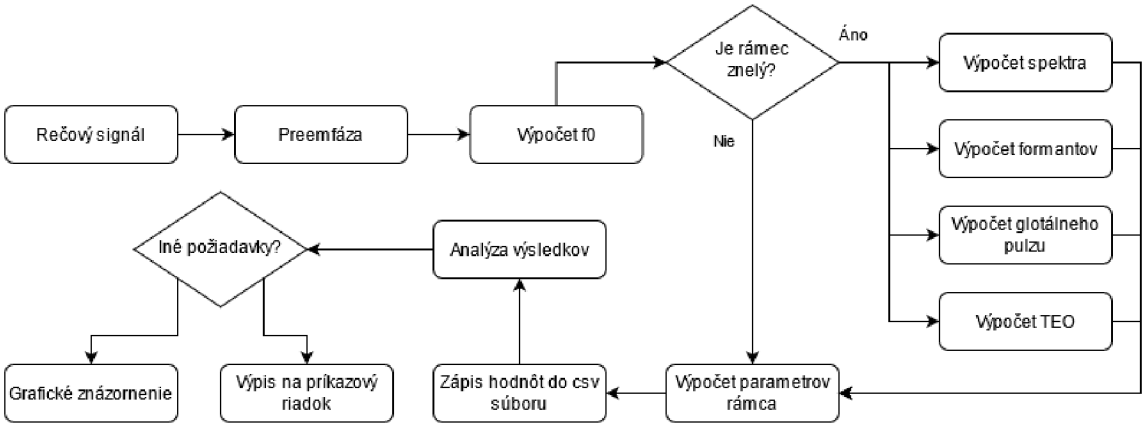
Obr. 2.6: Obrázok ukazuje: a) funkciu sínus, b) TEO tejto funkcie, c) úryvok rečového signálu, d) TEO tohto úryvku

Kapitola 3

Návrh a implementácia

3.1 Návrh

Zistili sme teda, že sú parametre krátkodobé a parametre dlhodobé. Preto je potrebné analyzovať rečový signál systematicky. Ak sa krátkodobé parametre prejavujú len na znelých častiach signálu, je potrebné tieto časti detekovať.



Obr. 3.1: Diagram návrhu analýzy signálu

3.2 Implementačné detaily

3.2.1 Funkcia analyze_src

– hlavná vyhodnocovacia funkcia. Prijíma na vstup argument *source*, a vzorkovaciu frekvenciu daného wav súboru, a vracia vypočítané hodnoty parametrov. Na začiatku je funkciou `rapt` z knižnice `pysptk` vypočítaná základná frekvencia pre celý signál po častiach – rámcoch, avšak daná funkcia nepodporuje prekrývanie rámcov, preto sú na základe veľkosti posunu dopočítané hodnoty aj pre prekrývajúce sa rámce.

Z nenulových hodnôt základnej frekvencie sú vypočítané informácie o nej: priemerná hodnota, maximálna hodnota a smerodajná odchýlka. Následuje výpočet priemernej dĺžky znelého a neznelého úseku. Po získaní týchto hodnôt následuje normalizácia signálu na interval $< -1, 1 >$, filtrovanie hornou priepusťou a použitie Hammingovho okna (emfáza).

Výpočet rýchlosti prechodu nulou pre celý signál bude slúžiť k vyhodnocovaniu znelosti rámca.

Krátkodobé parametre sú získané z rámcov signálu vo *while* cykle, ktorý na začiatku každej iterácie kontroluje, či rámec nepresahuje veľkosť vyhodnocovaného wav súboru. Ak áno, cyklus je ukončený. Inak vypočíta pre rámec hodnotu rýchlosti prechodu nulou, ktorú použije spolu s hodnotou základnej frekvencie k vyhodnocovaniu znelosti rámca. Ak je rámec znelý, je preň vypočítaná hodnota formantových parametrov funkciou `get_formants`, spektrálne parametre funkciou `spektrum`, parametre glotálneho pulzu funkciou `glottal_analysis` a stredná hodnota TEO funkciou `np.mean` z vráteného poľa funkciou `teo`.

Po skončení iterovania sú pre získanie ich priemernych hodnôt parametre vypočítané zo znelých rámcov vydelené ich počtom, a ostatné parametre vydelené celkovým počtom rámcov. Posledným počítaným parametrom je *voiced_percent*, ktorý je stonásobkom podielu počtu znelých rámcov a počtu všetkých rámcov. Funkcia vracia slovník *parameter:hodnota*.

3.2.2 Funkcia spektrum

Cieľom tejto funkcie je získať o signále spektrálne informácie. Kód funkcie bol prebraný z projektu v predmete Signály a systémy z akademického roku 2019/2020, a signál prevádza do frekvenčnej domény pomocou funkcie `rfft` z knižnice `NumPy`. Z hodnôt spektra je vypočítaná stredná hodnota, ktorá slúži na odseknutie malých hodnôt vo vysokých frekvenciách. Z týchto hodnôt sú následne vypočítané parametre spektra: medián, smerodajná odchýlka, minimum a maximum spektra, frekvencia maxima a rozsah frekvencií. Všetky tieto parametre sú následne aj so strednou hodnotou zapísané do slovníka a vrátené ako návratová hodnota funkcie.

3.2.3 Funkcia get_formants

Výpočet formantov a ich šírka je vypočítaná pomocou získania pólov prenosovej funkcie, už spomenutej v časti 2.6. Zdrojový kód bol získaný zo StackOverflow[1], avšak to je len prepis MatLab príkladu z oficiálnej webovej stránky[2] do `Python3`. Dôležité je zmieniť, že pre dosahovanie výsledkov bolo potrebné na určovanie stupňa prediktoru lineárnej prediktívnej analýzy (LPC) použiť vzorec prebraný z webu[3], $order = 2 + (fs/1000)$, kde $order$ je stupeň a fs je vzorkovacia frekvencia signálu. Návratovou hodnotou funkcie je slovník s hodnotami prvých troch formantov a ich šírkami.

3.2.4 Funkcia glottal_analysis

Odhad glotálneho pulzu je získaný podobne ako v článku [15]. Na rečový signál je aplikovaná metóda iteratívneho adaptívneho inverzného filtrovania. Táto metóda využíva lineárne prediktívnu analýzu (LPC) k zisteniu polynómu následne dosadeného do inverzného filtra, funkcie `deconvolve` z knižnice `scipy.signal`. To sa deje opakovane, pričom je stupeň LPC v každom kroku iný – 1, 12, 4, 12 a signál je normalizovaný na interval $< -1, 1 >$. Výstupom tejto metódy je odhad glotálneho pulzu, na ktorom sú následne počítané parametre: minimum, maximum, stredná hodnota, smerodajná odchýlka a rýchlosť prechodu nulou.

[1] https://stackoverflow.com/q/25107806
[2] https://www.mathworks.com/help/signal/ug/formant-estimation-with-lpc-coefficients.html
[3] https://stackoverflow.com/a/27352810

3.2.5 Funkcia teo

Hodnoty Teagerovho energetického operátora sú získané použitím kódu z repozitára[4]. V tomto repozitári je verzia všeobecného energetického operátora, s dosadením hodnôt potrebných pre získanie vzorca 2.8.

3.3 Uživateľské rozhranie a spustenie

Keďže bola programovacia časť bakalárskej práce napísaná v programovacom jazyku `Python 3`, spustenie je jednoduché. Program neobsahuje žiadne grafické rozhranie a je spúšťaný pomocou príkazového riadku. Je potrebné overiť, či sú nainštalované všetky potrebné knižnice pre beh programu, ktoré sa nachádzajú v súbore `requirements.txt` – to je možné zabezpečiť spustením príkazu `pip`[5] v podobe:

```
pip install -r requirements.txt
```

Následne je spustenie možné v príkazovom riadku zavolaním

```
$ python3 stredet.py evaluation_path
```

kde `evaluation_path` značí cestu k súboru / súborom určeným k výpočtu parametrov.

Program podporuje voliteľné argumenty,

- -h, --help: pre nápovedu k použitiu programu
- --ref=ref_path: cesta k súborom slúžiacim ako referencia
- --annot=annot_path: cesta k textovým súborom obsahujúcim výstup diarizácie
- -k, --keepcsv: pre ponechanie .csv súboru obsahujúceho výsledky

Po spustení sú spracované argumenty, cesta / cesty k analyzovaným súborom, a podľa voliteľných argumentov spustený výpočet. Výsledné hodnoty parametrov sú zapísané do .csv súboru.

[4] https://github.com/otoolej/envelope_derivative_operator/blob/1b3395f6f36f32084e65d37a2bba5323b18320c0/energy_operators/general_nleo.py

[5] pip je inštalátor balíkov pre jazyk `Python` a je taktiež potrebné ho mať nainštalovaný

Kapitola 4

Testovanie a výsledky

4.1 Databáza SUSAS

4.1.1 Pôvod

SUSAS, alebo Speech Under Simulated and Actual Stress, pozostáva z piatich častí, zahŕňajúcich širokú škálu stresorov a pocitov. K databáze bola vydaná aj dokumentácia [6], z ktorej je v tejto sekcii čerpané.

Databáza sa delí na tieto časti:

1. štýly rozprávania,
2. úloha s jedným zameraním alebo reč v hlasitom prostredí,
3. úloha s dvomi zameraniami,
4. úlohy vykonávané pod skutočným strachom zapríčineným pohybom,
5. dáta z psychiatrickej analýzy.

Časť databázy obsahujúca reč pod vplyvom simulovaného stresu pozostáva z dát v desiatich stresových štýloch (rečové štýly, úloha s jedným zameraním, Lombardov efekt), zatiaľ čo reč pod skutočným stresom bola získaná od subjektov vykonávajúcich počítačovú úlohu s dvomi zameraniami, alebo podstupujúcich jazdu na horskej dráhe. Neskôr boli k databáze pridané ešte zvukové záznamy štyroch pilotov bojových vrtuľníkov. Ak by sme považovali $s(n)$ za normálny, nezašumený rečový signál, tak signál, ktorý dorazí k mikrofónu bude zahŕňať skreslenie zapríčinené stresom, Lombardovým efektom a pridaným šumom. Iné skreslenie, napríklad zlé spojenie mikrofónu, telefónny kanál, či kódovanie hlasu môže taktiež znehodnotiť rečový signál.

1 Prvá časť databázy zahŕňa slová vyslovené v rôznych štýloch rozprávania, pričom tieto štýly sú: pomalý, rýchly, jemný, hlasný, nahnevaný, čistý a otázka. Slovník pozostáva z 35 anglických slov používaných v leteckej komunikácii, a obsahuje podmnožiny, ktoré častokrát predstavujú problém s efektivitou pre systémy automatického rozpoznávania reči. Príkladom takejto podmnožiny by mohla byť {go, hello, oh, no}, či {degree, three, thirty, freeze}. Tieto slová boli nahrané deviatimi mužskými rečníkmi hovoriacimi troma populárnymi americkými dialektmi – všeobecným americkým, bostonským a new-yorkským.

2 V druhej časti databázy sa úroveň vyvíjaného stresu na subjektu mierne zdvihla, a to zadaním úlohy s jedným zameraním, alebo reči pod vplyvom Lombardovho efektu.

Rovnakých 9 rečníkov vyslovilo rovnakých 35 slov ako v prvej časti, avšak tentokrát počas vykonávania počítačovej úlohy, ktorá mala simulovať stres. Táto úloha pozostávala z presunu objektu po displeji z počiatočnej do koncovej polohy protichodným pohybom riadiacej páky, pričom boli zaznamenavané dve úrovne stresu – pred a po zvýšení tlaku potrebného na pohyb riadiacou pákou. Subjektívne výsledky, výkonnostné dáta a informácie o zvýšení tepu srdca preukázali, že po zvýšení náročnosti ovládania skutočne vzrástla aj celková náročnosť úlohy. V tejto časti databázy sa nachádzajú aj nahrávky vytvorené simulovaním Lombardovho efektu.

3 Tretia časť pozostáva z nahrávok získaných popri vykonávaní úlohy s dvomi zameraniami. Cieľom tejto úlohy bolo simulovať stres v kokpite. Zamerania tejto úlohy boli simulované ako dva bežné úkony vykonávané pilotmi – pilotovanie, teda kontrolovanie pohybu, a zameriavanie cieľa. Pilotovanie bolo simulované pomocou displeja, na ktorom boli na pravej zobrazené dve rovnobežné sínusovky, ku ktorým bola pričítavaná a odčítavaná konštanta s úmyslom vytvoriť vzhľad kľukatej cesty. Poloha simulovaného kokpitu bola zobrazená ako kruh na spodnej časti obrazovky, ktorý mohol rečník ovládať riadiacou pákou a udržiavať po horizontálnej osi displeja na pomyselnej ceste. Zameriavanie cieľa bolo vytvorené podobne, pričom sa nachádzalo na ľavej strane displeja a miesto sínusoviek musel pilot pohybom druhej riadiacej páky po horizontálnej osi udržiavať trojuholník medzi dvoma vertikálnymi priamkami. Naviac boli k pohybom pripočítavané náhodné čísla, čo malo reprezentovať šum pri ovládaní automatického zameriavacieho systému. Simulácia prebiehala 80 sekúnd, pričom prvých 20 vykonával pilot len pilotovanie, ďalších 40 vykonával činnosti zároveň, a zvyšných 20 len zameriavanie cieľov. Počas simulácie bol pilot niekoľkokrát vyzvaný k prečítaniu rôznych náhodných slov z už spomínaného slovníka zobrazujúcich sa na displeji.

4 V štvrtej časti databázy SUSAS sa nachádzajú nahrávky získané pri dvoch typoch strachu zapríčineného pohybom. V pokuse simulovať rýchle zmeny vo výške a smere bežné pre pilotov v kokpite boli vybrané dva spôsoby. Tieto spôsoby sú horské dráhy – nevyžadujú žiaden tréning, avšak dokážu vytvárať podobný typ stresu. Jazda na prvej horskej dráhe trvá približne 60 sekúnd, pričom pričom okolo 10 sekúnd z toho pozostáva jazda z voľného pádu. Kabína s pasažiermi je vynesená do výšky skoro 40 metrov, kde chvíľu čaká, až následne priamo padne vertikálne okolo 30 metrov. Počas pádu boli rečníci vyzvaní k opakovaniu zvolených slov zo slovníka. Druhá vybraná horská dráha je klasický vlak prechádzajúci po konštrukcii vytvorenej z drevených trámov. Celá jazda pozostáva z dlhých úsekov hore a dolu, s krátkymi úsekmi v rovine. Z dôvodu vyššieho počtu pasažierov sú na nahrávkach počuteľné výkriky.

Neskôr doplnené súčasti obsahujú pilotov bojových vrtulníkov dvoch druhov – v prevádzke, no v pozícii na zemi, a pri pilotovaní. Znovu sa vyskytuje rovnakých 35 slov, od pilota a kopilota, buď zo situácie so zapnutými motormi na pristávacej dráhe (priečinok `medst`), alebo vykonávajúc základné letové manévre (priečinok `hist`).

5 V piatej časti boli nahrávaní pacienti podstupujúci psychiatrickú analýzu. Vyskytujú sa tu preto trochu rozličné emócie, než v predchádzajúcich častiach, a to hlavne depresia, strach a úzkosť. Vzhľadom na to, že sú to však citlivé údaje, nie je táto časť súčaťou databázy.

Databáza SUSAS bola používaná pri viacerých prípadoch analýzy a detekcie stresu v reči, napríklad zmiešaním parametrov TEO (v kapitole 2.7.2) [2], či rôznymi inými TEO vylepšeniami [7].

4.2 Databáza UTDrive

4.2.1 Pôvod

Za databázou UTDrive stoja výskumníci z Univerzity Texas v Dallase. Vydali k nej aj dokumentáciu [1], ktorá popisuje pôvod, vytváranie a výsledný obsah tejto databázy.

Viaceré štúdie dokázali, že vodiči automobilov šoférujú lepšie a bezpečnejšie pri používaní rečovo ovládaných interaktívnych systémov vo vozidle v porovnaní s používaním manuálne ovládaných systémov. Aj napriek lepšej ovládateľnosti a stálej možnosti vývoja kvalitnejšej obsluhy hlasom bude však interakcia vodiča s akýmkoľvek systémom odvádzať jeho pozornosť rôznymi úrovňami od jeho primárnej úlohy – vedenia vozidla. To by sa však ideálne diať nemalo a vodiči by mali venovať stopercentnú pozornosť šoférovaniu. V dnešnej dobe však životný štýl poskytujúci množstvo technologických vymožeností aj na palube motorových vozidiel prakticky znemožňuje venovať sústredenie výhradne šoférovaniu, a pridáva škálu rozptýlení v podobe rôznych asistencií vodičovi či zábavných systémov. Najlepšími príkladmi by mohli byť napríklad telefónne hovory, nastavenie a sledovanie navigácie, vybavovanie emailovej pošty, výber hudby či rôzne nastavenia daného automobilu. Ak vodič na tieto vedľajšie úkony míňa nevyužité kognitívne kapacity, jeho sústredenie na vedenie vozidla nie je nijako ovplyvnené. To by malo byť cieľom návrhov interaktívnych systémov pre vozidlá, ktoré by mali brať do úvahy vodičské schopnosti a potrebnú kognitívnu záťaž a schopnosť vodiča vynaložiť túto záťaž popri šoférovaní. Vediaci tieto faktory by účelným využitím údajov o jazde mohli asistenčné systémy vo vozidle pomôcť vodičovi vyrovnať sa s rôznymi rozptýleniami, ako napríklad obmedziť využívanie aplikácií v hustej premávke.

Vo vozidle nachádzajúcom sa v premávke sa vyskytujú aj zvuky negatívne ovplyvňujúce rečové ovládanie týchto interaktívnych systémov, čo núti vodiča k ich prekonávaniu, a to môže viesť napríklad k Lombardovmu efektu. V takýchto podmienkach teda môžu systémy automatického rozpoznávania reči náražať na problémy s výkonom, a šofér môže pri ovládaní rečou v dôsledku intenzívneho sústredenia na vedenie vozidla vynechať povely, hovoriť gramaticky nesprávne, v prejave používať dlhšie medzery medzi slovami či výplne v podobe „uhm“ alebo iných.

V tejto databáze sa nachádzajú okrem zvukových dát z mikrofónov aj dáta obrazové, zachytávajúce vodiča a pohľad vpred z vozidla, či dáta zaznamenané rôznymi palubnými počítačmi či senzormi o stave a zmenách v automobile a jeho riadení.

4.2.2 Dostupné nahrávky

Zvukové nahrávky boli pri vytváraní tejto databázy zaznamenané pomocou piatich mikrofónov umiestnených v rade, ktorá bola umiestnená na vrchu čelného skla, vedľa slnečných clon. Kvôli rôznym druhom vedľajších zvukov prítomných v kabíne, ako napríklad hluk klimatizácie, tykanie smeroviek, hudba či predbiehajúce vozidlá, bola kvalita rečového signálu z radu mikrofónov vylepšená algoritmami.

Subjekty podieľajúce sa na vytváraní tejto databázy pozostávali zo študentov, zamestnancov a zboru fakulty. Každý zúčastnený šoféroval vozidlo na dvoch rôznych okruhoch, kde prvý prechádzal obytnou oblasťou a druhý priemyselnou oblasťou. Oba okruhy trvajú približne desať až pätnásť minút na prejdenie, počas ktorého sú zúčastnení vodiči vystavení rôznym úlohám s líšiacimi sa úrovňami kognitívnej záťaže po dobu približne jednej hodiny. Najčastejšie vedľajšie úlohy:

1. interakcia s komerčným automatickým rozpoznávačom reči za použitia hands-free. Vodič telefonuje dialógovému systému leteckej spoločnosti za účelom zistenia príletovej

či odletovej brány konkrétneho letu; alebo telefonuje hlasovému portálu za účelom zistenia informácie osobného záujmu – počasie v inom meste...
2. čítanie dopravného značenia, názvov ulíc, štátnych poznávacích značiek
3. ladenie rádia, vkladanie CD do mechaniky, voľba konkrétnej nahrávky z CD
4. všeobecná konverzácia so spolujazdcom
5. informovanie o vykonávaná úkonov spojených so šoférovaním
6. zmena jazdného pruhu

Americká National Highway Traffic Safety Administration označila štyri rozdielne typy rušenia vodiča: vizuálne, zvukové, fyzické a kognitívne. Síce sú posudzované samostatne, navzájom sa však nevylučujú. Dialógový systém na palube a interakcia s ním teda môže predstavovať zdroj všetkých štyroch typov rušenia: zadávanie údajov do telefónu – fyzické, sledovanie telefónu – vizuálne, rozhovor – zvukové, a sústredenie na tému konverzácie – kognitívne.

Databáza UTDrive bola analyzovaná aj pri výskume prejavov kognitívnej záťaže na reči [3]. Základná frekvencia, formanty, či spektrálne vlastnosti napovedali zvýšenými hodnotami pri aktivitách o detekovateľnosti kognitívnej záťaže.

Z tejto databázy sú na internete voľne dostupné nahrávania `dm1007_s1` a `dm1007_s2`, avšak prvé menované je nekompletné, a preto je v tejto bakalárskej práce analyzované nahrávanie druhé. To pozostáva zo štyroch častí – každá označená spôsobom

```
dm1007_s2_pX_2007_03_12_tttttt_AI_1,
```

kde `X` je číslo nahrávky (1-4), a `tttttt` označuje čas zhotovenia nahrávky. Ďalej z dokumentácie dostupnej na webovej stránke[1] projektu vyplýva, že z v názve tejto nahrávky sa nachádzajú nasledujúce informácie: `dm` značí, že rečník je muž, pričom `1007` je jeho identifikácia, a posledný údaj `AI_1`, označuje nahrávacie zariadenie – už spomínaný rad mikrofónov.

V jednotlivých častiach ďalej môžeme vidieť:

- Nahrávka 1: v obytnej oblasti, úlohy B: spontánna jazda, bežné úlohy, informovanie na dialógovom systéme ohľadne letov a čítanie dopravného značenia;
- Nahrávka 2: v obytnej oblasti, voľné jazdenie;
- Nahrávka 3: v priemyselnej oblasti, voľné jazdenie;
- Nahrávka 4: v priemyselnej oblasti, úlohy B: čítanie dopravného značenia, zisťovanie informácií z hlasového portálu, zmena jazdných pruhov a konverzácia so spolujazdcom.

Podľa predpokladu by teda mala byť úroveň stresu vyššia v nahrávkach 1 a 4, keďže mal rečník v týchto častiach vykonávať kognitívne náročné úlohy. V ďalšej časti sa pozrieme, ako sa rečové parametre vyvíjali v priebehu týchto nahrávok.

4.2.3 Testovanie a výstup

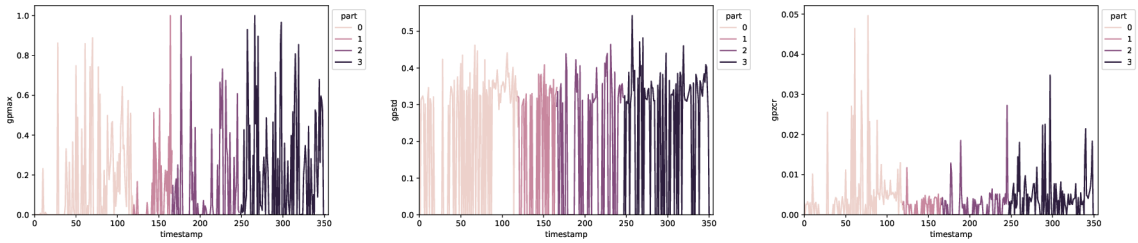
Zložka `dm1007_s2` teda obsahuje štyri nahrávky s dĺžkou približne 15 minút. Keďže je tento signál pridlhý, a súčasťou nahrávania boli aj konverzácie so spolujazdcom, výskumná skupina Speech@FIT aplikovala na tento signál diarizáciu[2], ktorej výstupom boli textové súbory obsahujúce časti nahrávky identifikujúce rečníkov. Avšak identifikácia nebola úplne jednoznačná, a na nahrávke boli rozpoznaní viac než dvaja rečníci, z výstup diarizácie sme teda skôr využili delenie na časti – podľa dokumentácie k databáze sme vedeli, že bol snímajúci mikrofón pripevnený blížšie k vodičovi než spolujazdcovi, a tak by mali mať

[1] Dokumentácia k projektu UTDrive
[2] Proces delenia signálu na časti podľa identity rečníka

časti vyprodukované vodičom vyššiu energiu než časti vyprodukované spolujazdcom. Preto sme pre každú časť označenú diarizáciou vypočítali priemernú energiu, z energií vypočítali strednú hodnotu, a následne vyberali k analýze parametrov len tie časti, ktorých priemerná energia bola vyššia než stredná hodnota energií.

Pre tieto časti sme teda získali parametre z kapitoly 2 a pokúsili sa vizuálne analyzovať, či hodnoty parametrov naznačujú splnený predpoklad zvýšeného stresu na daných úsekoch nahrávok.

Získali sme teda parametre viac než 350 častí nahrávania `dm1007_s2`, ktoré sa nachádzajú na horizontálnej osi nasledujúcich grafov. Na vertikálnej osi sú vykreslené skúmané parametre.



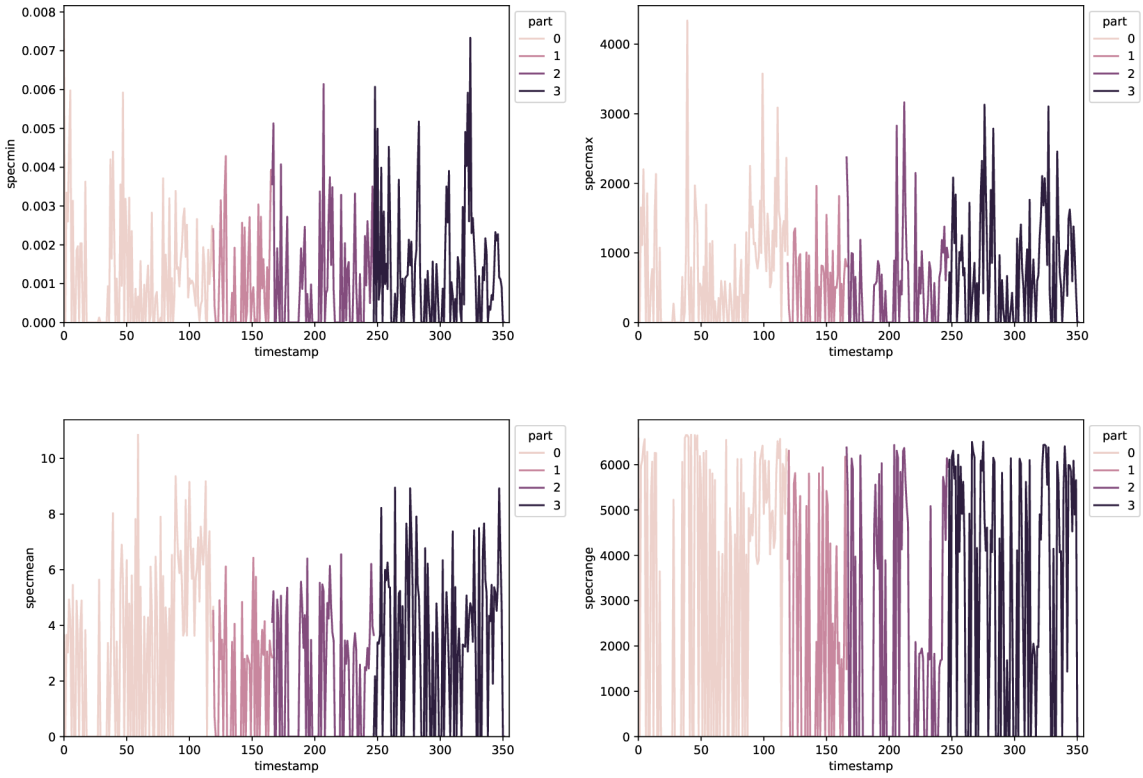
Obr. 4.1: Obrázky vývoja priemerných hôdnot parametrov glotálneho pulzu: maximálnej hodnoty (MAX), smerodajnej odchýlky (STD), a rýchlosti priechodu nulou (ZCR).

Ak sa pozrieme na priemerné hodnoty parametrov z obrázkov 4.1, zobrazené v tabuľke 4.1, zistíme, že hodnoty skutočne sú vyššie v prvej a poslednej časti.

Časť	**Hodnoty**		
	MAX	**STD**	**ZCR**
0	0.14714343844849137	0.20925176600516962	0.0047586640397415565
1	0.10724344260276113	0.18267896950245865	0.002028663843974446
2	0.12500097013284514	0.17545937044235224	0.0024920141496433657
3	0.21571480699864568	0.24291592562708209	0.0045082855820249195

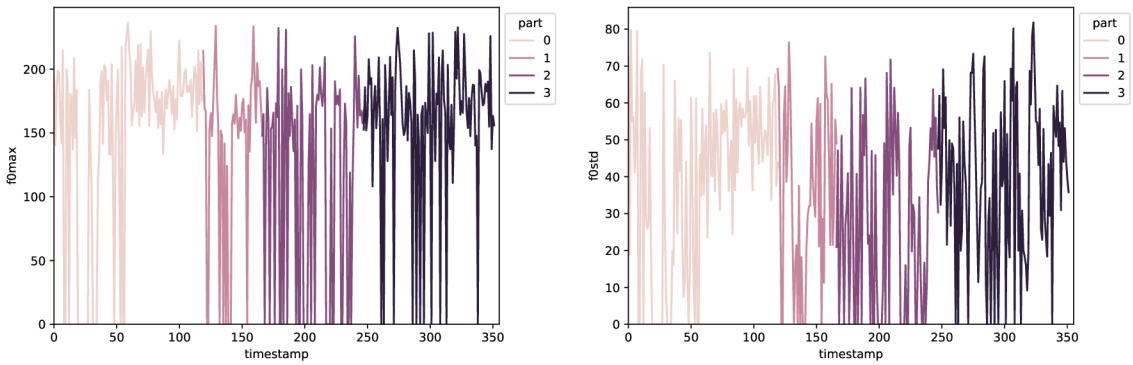
Tabuľka 4.1: Tabuľka priemerných hodnôt vybraných parametrov glotálneho pulzu.

Zo spektrálnych parametrov vizuálne naznačovali zvýšenú úroveň stresu štyri konkrétne – zobrazené na obrázku 4.2. Po spriemerovaní častí hodnoty potvrdili predpoklad a pre prvú a štvrtú časť boli vyššie.

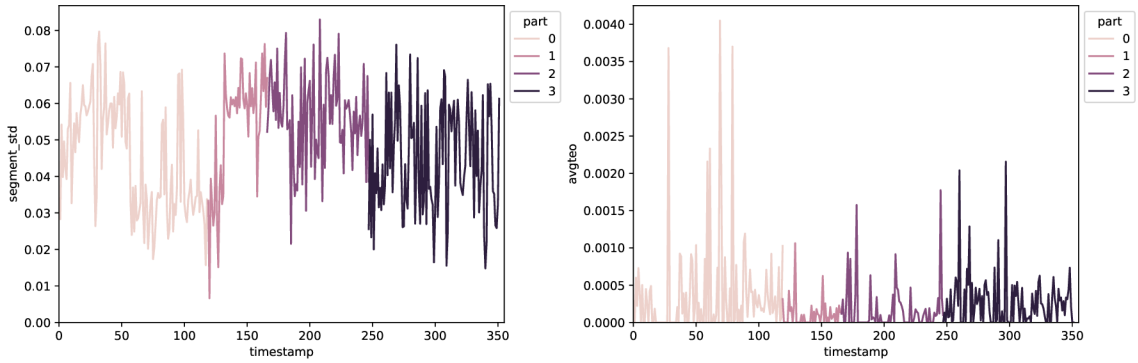


Obr. 4.2: Priemerné hodnoty spektrálnych parametrov: minimum, maximum, stredná hodnota, rozsah

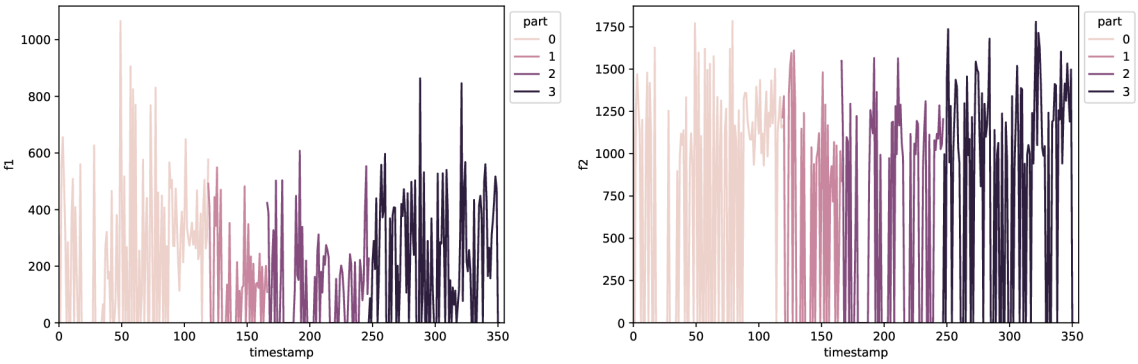
V obrázkoch 4.3 až 4.5 všetky parametre dosahujú vyššie hodnoty v predpokladaných úsekoch nahrávania.



Obr. 4.3: Vývoj hodnôt maxima a smerodajnej odchýlky základnej frekvencie.

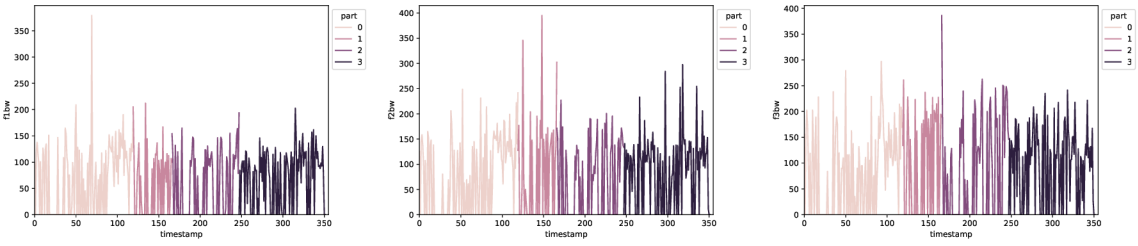


Obr. 4.4: Priemerná hodnota Teagerovho energetického operátora a smerodajnej odchýlky signálu.



Obr. 4.5: Prvé dve priemerné formantové frekvencie v nahrávkach.

Šírky druhého a tretieho formantu sa postupne v častiach taktiež menili, avšak na základe ich priemerných hôdnot z tabuľky 4.2 môžeme povedať, že predpoklad nespĺňajú – aj keď by sa to po pohľade na obrázok 4.6 mohlo zdať.



Obr. 4.6: Priemerné hodnoty šírok prvých troch formantov.

Časť	Hodnoty	
	F2BW	**F3BW**
0	77.6315031921221	84.11095795338811
1	86.00099317257592	98.36683746545253
2	68.72632996602061	86.55557191452233
3	86.50994481029193	93.78509281356486

Tabuľka 4.2: Tabuľka priemerných hodnôt šírok druhého a tretieho formantu.

4.3 Vyhodnotenie

Celkové vyhodnotenie a porovnanie parametrov rečového signálu bolo vykonané na databáze SUSAS. Neutrálna reč, teda reč z prvej kategórie databázy bola použitá ako referenčná hodnota (v nasledujúcich tabuľkách 4.3 až 4.8 sa jej hodnoty nachádzajú v stĺpci *ref*. Nasledujúce stĺpce *medst*, *hist* a *scream* obsahujú hodnoty kategórií SUSAS – teda *medst* označuje priemerné hodnoty parametrov nahrávok vytvorených pri vykonávaní jednej úlohy, *hist* pri vykonávaní dvoch úloh a *scream* na horských dráhach.

Prvá časť tabuľky obsahuje vypočítané hodnoty, druhá časť obsahuje hodnoty normované referenčnými hodnotami. Na druhej časti je teda možné pozorovať percentuálny nárast alebo pokles hodnôt parametrov. Hodnota v stĺpci *ref* je rovná 1.0, a teda hodnota v zvyšných stĺpcoch, napríklad 1.08 naznačuje 8 percentný nárast v danej kategórii – nárast parametru pri reči na danej úrovni stresu oproti neutrálnej reči.

Parametre glotálneho pulzu:

	Hodnoty parametrov			
	ref	medst	hist	scream
gpmax	0.991017	0.991952	0.991548	0.854312
gpmin	-0.408332	-0.374842	-0.449980	-0.839405
gpmean	0.226218	0.242154	0.204108	0.012009
gpstd	0.363565	0.358861	0.369644	0.398497
gpzcr	0.017314	0.012983	0.016975	0.030826
	ref	medst	hist	scream
gpmax	1.0	1.000943	1.000536	0.862056
gpmin	1.0	0.917984	1.101997	2.055693
gpmean	1.0	1.070442	0.902259	0.053086
gpstd	1.0	0.987064	1.016722	1.096082
gpzcr	1.0	0.749828	0.980370	1.780361

Tabuľka 4.3: Priemerné hodnoty vlasností glotálneho pulzu.

Spektrálne parametre:

	Hodnoty parametrov			
	ref	medst	hist	scream
specmedian	1.569288	1.697781	1.772885	2.067416
specmean	2.825704	3.125420	3.257267	4.068876
specstd	4.051446	4.088014	4.363738	6.024746
specmin	0.016465	0.012247	0.014884	0.028476
specmax	61.200782	66.038693	65.380067	61.322125
specmaxloc	194.181818	137.159091	295.795455	552.318182
specrange	3997.727273	3998.340909	3998.750000	3463.431818
	ref	medst	hist	scream
specmedian	1.0	1.081880	1.129739	1.317423
specmean	1.0	1.106068	1.152727	1.439951
specstd	1.0	1.009026	1.077082	1.487061
specmin	1.0	0.743811	0.903965	1.729477
specmax	1.0	1.079050	1.068288	1.001983
specmaxloc	1.0	0.706344	1.523291	2.844335
specrange	1.0	1.000153	1.000256	0.866350

Tabuľka 4.4: Hodnoty spektrálnych parametrov.

Základná frekvencia:

	Hodnoty parametrov			
	ref	medst	hist	scream
f0avg	113.581219	110.037803	109.634399	132.380005
f0max	162.624582	179.920721	174.593720	180.548171
f0std	51.375860	51.973040	52.846699	60.739756
	ref	medst	hist	scream
f0avg	1.0	0.968803	0.965251	1.165510
f0max	1.0	1.106356	1.073600	1.110215
f0std	1.0	1.011624	1.028629	1.182263

Tabuľka 4.5: Hodnoty základnej frekvencie.

Formantové parametre:

	Hodnoty parametrov			
	ref	medst	hist	scream
f1	335.693307	276.513019	286.439996	330.387940
f1bw	111.540291	110.219175	103.190755	84.959114
f2	1422.727762	1415.213367	1345.944323	1149.147037
f2bw	127.554117	125.765736	130.774502	86.393074
f3	2327.629922	2341.265193	2292.925111	2047.131216
f3bw	129.157118	133.563873	135.870351	91.665500
	ref	medst	hist	scream
f1	1.0	0.823707	0.853279	0.984196
f1bw	1.0	0.988156	0.925143	0.761690
f2	1.0	0.994718	0.946031	0.807707
f2bw	1.0	0.985979	1.025247	0.677305
f3	1.0	1.005858	0.985090	0.879492
f3bw	1.0	1.034119	1.051977	0.709721

Tabuľka 4.6: Priemerné frekvencie a šírky prvých troch formantov.

Dĺžka:

	Hodnoty parametrov			
	ref	medst	hist	scream
avgvoicedlen	12.116941	8.770839	9.191940	10.340567
avgunvoicedlen	11.425864	7.909015	8.324516	12.064286
voicedpercent	24.138259	28.334963	24.791521	14.745385
	ref	medst	hist	scream
avgvoicedlen	1.0	0.723849	0.758602	0.853398
avgunvoicedlen	1.0	0.692203	0.728568	1.055875
voicedpercent	1.0	1.173861	1.027063	0.610872

Tabuľka 4.7: Trvanie znelých a neznelých úsekov.

Energetické parametre:

	Hodnoty parametrov			
	ref	medst	hist	scream
avgteo	0.000723	0.000448	0.000683	0.003431
segment_mean	0.005781	0.006119	0.006639	0.001072
segment_std	0.031123	0.031646	0.035828	0.070010
	ref	medst	hist	scream
avgteo	1.0	0.619324	0.944765	4.747337
segment_mean	1.0	1.058383	1.148424	0.185434
segment_std	1.0	1.016783	1.151151	2.249449

Tabuľka 4.8: TEO, priemer a smerodajná odchýlka rámcov.

Po zhliadnutí výsledných dát je otázne, ako efektívne sú na analýzu reči pod stresom. Hodnotenie parametrov by teda mohlo byť vyjadrené ako podiel:

$$uspesnost = \frac{pocet\_vysledkov\_podla\_ocakavania}{celkovy\_pocet\_vysledkov},$$

pričom by násobením číslom 100 bola dosiahnutá percentuálna úspešnosť. Napríklad pre energiu, kde sme od všetkých parametrov očakávali nárast, by teda úspešnosť bola $\frac{6}{9}$, keďže *avgteo* je väčšie len v stĺpci *scream*, *segment_mean* v stĺpcoch *medst* a *hist*, a *segment_std* vo všetkých troch stĺpcoch.

parameter	úspešnosť
spektrum	0.81
základná frekvencia	0.778
dĺžka	0.333
energia	0.667
glotálny pulz	0.467
formanty	0.389

Tabuľka 4.9: Úspešnosť parametrov v zisťovaní stresu.

Kapitola 5

Záver

V prvej časti tejto práce sme sa pozreli na problematiku stresu. Aké sú na stres všeobecne pohľady, čo stres môže spôsobiť, aký má fyziologický či psychologický vplyv na osobu. Aké sú činnosti a zamestnania, v ktorých sú ľudia pod vplyvom stresu a častokrát musia robiť kritické rozhodnutia. Následne sme uviedli klasifikáciu stresu ako prejavu v reči pri rozpoznávaní.

V druhej časti sme sa pozreli na parametre rečového signálu. Čo je glotálny pulz, aký je jeho význam pri tvorbe reči, ako je spracovaný a následne akými spôsobmi bol a môže byť analyzovaný v súvislosti so stresom. Základná frekvencia, alebo výška hlasu, je taktiež populárnym parametrom pri analyzovaní rečového signálu. Študovali sme jej význam, z nej získateľné údaje, problémy pri jej odhade a práce zaoberajúce sa jej spojením so stresom. Následne sme sa pozreli na diskrétnu Fourierovu transformáciu ako nástroj na spektrálnu analýzu signálov. V ďalšej časti sme sa venovali formantom, spôsobu výpočtu a problémom, na ktoré je pri výpočte možné naraziť. Spomenuli sme aj ich použitie na pozorovanie prejavov stresu. O možnej zvýšenej intenzite signálu pri strese nám z parametrov napovie energia a Teagerov energetický operátor. V tretej časti sme sa venovali návrhu a implementácii výpočtu daných parametrov. Už spomínané parametre sme rozdelili do dvoch skupín: krátkodobé a dlhodobé, čo ovplyvnilo postup ich výpočtu. Implementačné detaily sú popísané pomocou hlavných funkcií programu. Spomenuté je aj uživateľské rozhranie a postup pri spustení programu.

Štvrtá časť sa zaoberá databázami, testovaním a výsledkami. Databázy SUSAS a UTDrive sú približené cez ich pôvod a obsah. Následne je na nahrávaní z databázy UTDrive otestovaný výpočet parametrov, sú zobrazené grafy priebehu a ich popis. V poslednej časti je vyhodnotenie parametrov, ktoré bolo vytvorené použitím databázy SUSAS a rozdielmi medzi neutrálnou rečou a rečou pod rôznymi úrovňami stresu.

Literatúra

[1] Angkititrakul, P., Kwak, D., Choi, S., Kim, J., PhucPhan, A. et al. Getting start with UTDrive: Driver-behavior modeling and assessment of distraction for in-vehicle speech systems. In:. 2007, s. 1334–1337.

[2] Bandela, S. R. a Kumar, T. Stressed speech emotion recognition using feature fusion of teager energy operator and MFCC. In:. Júl 2017, s. 1–5. DOI: 10.1109/ICCCNT.2017.8204149.

[3] Boril, H., Sadjadi, S., Kleinschmidt, T. a Hansen, J. Analysis and detection of cognitive load and frustration in drivers' speech. In:. Január 2010, s. 502–505.

[4] Driskell, J. a Salas, E. *Stress and Human Performance*. Taylor & Francis, 2013. Applied Psychology Series. ISBN 9781134771820. Dostupné z: https://books.google.cz/books?id=I6Va4ay9FSAC.

[5] Drugman, T., Alku, P., Alwan, A. a Yegnanarayana, B. Glottal source processing: From analysis to applications. *Computer Speech & Language*. 2014. DOI: https://doi.org/10.1016/j.csl.2014.03.003. ISSN 0885-2308. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0885230814000229.

[6] Hansen, J. a Bou Ghazale, S. E. Getting started with SUSAS: a speech under simulated and actual stress database. In: *EUROSPEECH*. 1997.

[7] He, L., Lech, M., Memon, S. a Allen, N. Recognition of stress in speech using wavelet analysis and teager energy operator. In:. ISCA, 2008, s. 4.

[8] James, J. *A Student's Guide to Fourier Transforms: With Applications in Physics and Engineering*. Cambridge University Press, 2011. Student's Guides. ISBN 9781139493949. Dostupné z: https://books.google.cz/books?id=_T99VW0ARfkC.

[9] Linhart, J. *Slovník cizích slov pro nové století: základní měnové jednotky, abecední seznam chemických prvků, jazykovědné pojmy, 30,000 hesel*. Litvínov: Dialog, 2002. ISBN 80-85843-61-7.

[10] Miller, J. L. Interactions in processing segmental and suprasegmental features of speech. *Perception & Psychophysics*. Springer Science and Business Media LLC. 1978, zv. 24, č. 2, s. 175–180. DOI: 10.3758/bf03199546. Dostupné z: https://doi.org/10.3758/bf03199546.

[11] Nishio, M. a Niimi, S. Changes in Speaking Fundamental Frequency Characteristics with Aging. *Folia Phoniatrica et Logopaedica*. S. Karger AG. 2008, zv. 60, č. 3, s. 120–127. DOI: 10.1159/000118510. Dostupné z: https://doi.org/10.1159/000118510.

[12] Nwe, T., Foo, S. a De Silva, L. Classification of stress in speech using linear and nonlinear features. In:. 2003, sv. 2, s. 9–12. ISSN 15206149.

[13] Pawi, A. Modelling and extraction of fundamental frequency in speech signals. In:. 2014.

[14] Psutka, J., Müller, L., Matoušek, J. a Radová, V. *Mluvíme s počítačem česky*. Vyd. 1. Praha: Academia, 2006. ISBN 80-200-1309-1.

[15] Sigmund, M., Prokes, A. a Brabec, Z. Statistical analysis of glottal pulses in speech under psychological stress. Január 2007.

[16] Simantiraki, O., Giannakakis, G. a Pampouchidou, M. Stress Detection from Speech Using Spectral Slope Measurements. In: *Pervasive Computing Paradigms for Mental Health*. Springer International Publishing, 2018, s. 41–50. ISBN 978-3-319-74935-8.

[17] Traunmüller, H. a Eriksson, A. The frequency range of the voice fundamental in the speech of male and female adults. In:. 1993.

[18] Wang, L., Ambikairajah, E. a Choi, E. H. C. *Automatic Tonal and Non-Tonal Language Classification and Language Identification Using Prosodic Information*.

[19] Wolfe, J. *Formant: what is a formant?* Dostupné z: http://newt.phys.unsw.edu.au/jw/formant.html.

[20] Yap, T. F., Ambikairajah, E., Epps, J. a Choi, E. H. C. Cognitive load classification using formant features. In: *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)*. IEEE, 2010. DOI: 10.1109/isspa.2010.5605535. Dostupné z: https://doi.org/10.1109/isspa.2010.5605535.

[21] Zollinger, S. A. a Brumm, H. The Lombard effect. *Current Biology*. Elsevier BV. 2011, zv. 21, č. 16, s. R614–R615. DOI: 10.1016/j.cub.2011.06.003. Dostupné z: https://doi.org/10.1016/j.cub.2011.06.003.