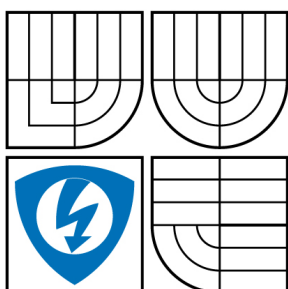




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

DATABÁZE EMOČNÍ ŘEČI

DATABASE OF VOCAL SAMPLES OF HUMAN EMOTIONS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

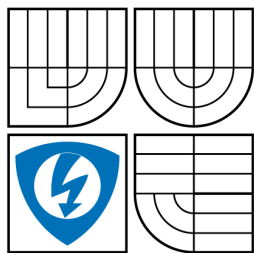
AUTOR PRÁCE
AUTHOR

MICHAL HLAVICA

VEDOUcí PRÁCE
SUPERVISOR

Ing. HICHAM ATASSI

BRNO 2009



**VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ**

**Fakulta elektrotechniky
a komunikačních technologií**

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Michal Hlavica

ID: 98054

Ročník: 3

Akademický rok: 2008/2009

NÁZEV TÉMATU:

Databáze emoční řeči

POKYNY PRO VYPRACOVÁNÍ:

Vytvořte databázi emoční řeči obsahující řečové nahrávky od mluvčích různého věku a pohlaví pro různé emoční stavy. Prodiskutujte různé možnosti a metody jak technické, tak i praktické, které mohou vést ke zefektivnění práce. Dále vytvořte softwarový nástroj s uživatelským rozhraním pro subjektivní hodnocení vytvořené databáze. Součástí bakalářské práce by měly být i výsledky analýzy některých prosodických rysů promluv vytvořené databáze.

DOPORUČENÁ LITERATURA:

- [1] Psutka J., Muller L., Matoušek J., Radová V., Mluvíme s počítačem česky. Academia Praha 2006.
- [2] Vlčková-Mejvaldová J., Prozodie, cesta i mříž porozumění. Karolinum, Praha 2006.

Termín zadání: 9.2.2009

Termín odevzdání: 2.6.2009

Vedoucí práce: Ing. Hicham Atassi

prof. Ing. Kamil Vrba, CSc.

Předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Anotace

V této bakalářské práci je rozebrána teorie emocí, to jak emoce vznikají a jak jsou fyziologicky vyjádřeny lidským tělem. Jak se tyto fyziologické projevy a emoce vůbec promítají do lidské řeči. Dále je popsán proces tvorby řeči a základní prozodické a akustické parametry hlasového projevu relevantní pro výzkum. Je zde také popsána teorie tvorby databází zvukových nahrávek, což je kvalitní podklad pro databázi samotnou. Ta je totiž také součástí této práce a jedná se o nahrávky stříhané z televizních pořadů a seriálů.

Dalším velice důležitým bodem je popsání softwarového nástroje pro subjektivní hodnocení databází, který byl vytvořen jako součást této práce. Byl vytvořen v jazyce C++ za pomoci kompilátoru Builder C++.

Také je zde provedena krátká analýza ukázkových nahrávek pro každou emoci. Tato analýza se zabývá základní frekvencí, intenzitou a prvními třemi formanty.

Klíčová slova: Emoce, fyziologické projevy, řeč, prozódie, databáze, nahrávky, nástroj, subjektivní, programování.

Abstract

In this bachelor work is analyzed theory of emotions, how emotions arise and how they are physiologically expressed by human body. How these physiological expressions and emotions reflect into the human speech. Then is described process of creating of speech and basic prosodic and acoustic parameters relevant for research. Theory of creating of databases is described here as well, which is quality ground for database itself. The database is also part of this thesis and they are records cut from television programmes and serials.

The next very important issue is description of software tool for subjective evaluating of databases, which was created as a part of this thesis. It was created in C++ language with help by compiler Builder C++ .

Also a short analysis of exemplary records for every emotion is done here. This analysis deals with basic frequency, intensity and first three formants.

Key words: Emotion, physiological expressions, speech, prosody, database, records, tool, subjective, programming.

HLAVICA, M. *Databáze emoční řeči*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. 52 s. Vedoucí bakalářské práce Ing. Hicham Atassi.

Prohlášení

Prohlašuji, že svou bakalářskou práci na téma „Databáze emoční řeči“ jsem vypracoval samostatně pod vedením vedoucího bakalářské 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 autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

podpis autora

Poděkování

Mé poděkování bych chtěl vyjádřit hlavně svému vedoucímu bakalářské práce Ing. Hichamu Atassimu. Chtěl bych mu tímto poděkovat za věcné a trpělivé rady při vypracovávání bakalářské práce. Dále bych chtěl poděkovat všem, kteří mě podporovali jak při celém studiu, tak při psaní bakalářské práce.

Obsah

1.	Úvod:	11
2.	Emoce:	12
2.1.	Teorie emocí:	12
2.1.1.	Psychoevoluční teorie Roberta Plutchika	12
2.1.2.	James – Langeova teorie	14
2.1.3.	Cannon – Bardova teorie	14
2.2.	Význam emocí	14
2.3.	Způsoby vyjádření emocí	15
2.3.1.	Zvukové vyjádření emocí	15
2.3.2.	Obličejové vyjádření emocí	16
2.3.3.	Vyjádření emocí pomocí gest	16
2.4.	Fyziologické projevy jednotlivých emocí	16
2.4.1.	Radost	16
2.4.2.	Smutek	17
2.4.3.	Vzteky	17
2.4.4.	Strach	17
3.	Tvorba řeči	18
3.1.	Proces tvorby řeči	18
3.2.	Věkové rozdělení	19
4.	Prozodie a emoce	20
4.1.	Základní prozodické parametry	20
4.1.1.	Intonace	20
4.1.2.	Intenzita	21
4.1.3.	Trvání	21
4.2.	Další zvukové vlastnosti řeči	21
4.2.1.	Artikulace	21
4.2.2.	Barva hlasu	21
4.2.3.	Dýchání	22
4.2.4.	Pauzy	22
4.3.	Akustické parametry relevantní pro výzkum	22
4.4.	Prozodické rysy konkrétních emocí	22
5.	Databáze nahrávek	24
5.1.	Tvoření databáze	24
5.1.1.	Databáze podle typu nahrávání	24
5.1.2.	Důležité faktory při nahrávání	26
5.2.	Dostupné databáze	27
5.3.	Vlastní databáze	28
5.4.	Program Elan	29
6.	Analýza	30
6.1.	Základní frekvence	30
6.2.	Intenzita	32
6.3.	Formanty	33
7.	Aplikace pro subjektivní hodnocení nahrávek	34
7.1.	Požadavky a řešení programu	34

7.1.1.	Vstupní formulář	34
7.1.2.	Admin Menu.....	36
7.1.3.	Příprava testu před jeho spuštěním.....	38
7.1.4.	Hlavní formulář – testování.....	40
7.1.5.	Vyhodnocování testu	43
7.1.6.	Výstupní soubor	44
7.2.	Shrnutí.....	48
8.	Závěr:.....	50

Seznam obrázků

Obr. 1: Plutchikovo kolo.....	13
Obr. 2: Emoční těleso podle Plutchika.	13
Obr. 3: Sekvence událostí vzniku emoce podle James - Langeovy teorie.....	14
Obr. 4: Model hlasového ústrojí člověka.	18
Obrázek 5: Graf věrohodnosti nahrávek.....	25
Obrázek 6: Ukázka z programu Elan.....	30
Obrázek 7: Ukázkové průběhy základní frekvence pro všechny emoce: a) ironie, b) překvapení, c) radost, d) vztek, e) smutek, f) strach.....	31
Obrázek 8: Ukázkové průběhy intenzity pro všechny emoce: a) ironie, b) překvapení, c) radost, d) vztek, e) smutek, f) strach.....	32
Obrázek 9: Ukázkové průběhy prvních třech formantů pro všechny emoce: a) ironie, b) překvapení, c) radost, d) vztek, e) smutek, f) strach.....	33
Obr. 10: Vstupní formulář – CZ.	35
Obr. 11: Vstupní formulář – EN.....	36
Obr. 12: Formulář pro zadání hesla pro Admin Menu.	37
Obr. 13: Formulář Admin Menu.....	37
Obr. 14: Výstupní soubor po prvním zápisu.....	38
Obr. 15: Ukázka struktury nedokončeného testu.	39
Obr. 16: Informační okno před spuštěním testu.....	40
Obr. 17: Hlavní formulář – CZ.....	42
Obr. 18: Hlavní formulář – EN.	42
Obr. 19: Příklad „Confusion matrixu“.	44
Obr. 20: Příklad struktury informací jedné nahrávky ve výstupním souboru.....	45
Obr. 21: Ukázka výstupního souboru.....	46
Obr. 22: Ukázka výstupního souboru testu rozděleného na 3 části.	48

Seznam tabulek

Tabulka 1: Prozodické rysy různých expresivních realizací 1 věty podle Léona.	23
Tabulka 2: Seznam významných emočních databází.	27
Tabulka 3: Přehled zkratk jednotlivých emocí použitých v databázi.	29

1. Úvod:

Cílem této práce je vytvořit rozsáhlou databázi emoční řeči v českém jazyce, která bude sloužit pro rozpoznávání emočního stavu mluvčího z jeho řeči, popřípadě jiným vědeckým účelům přidruženým k této problematice, jako je například syntéza řeči nebo studium prosodických charakteristik řeči. To znamená, že v prvním kroku se bude jednat o shromáždění co největšího počtu nahrávek z televizních pořadů a seriálů, ve kterých je čitelná některá ze základních lidských emocí. Tento typ nahrávek se blíží více ke skupině přirozených nahrávek, protože aktéři, kteří tento materiál vytváří, sice hrají svou roli, ale nesoustředí se na emoce obsažené v jejich řeči. Této problematice se dále věnuje celá kapitola. Ze základních emocí, které jsou obecně známy, bylo vybráno následujících šest: Radost, strach, smutek, vztek, překvapení a ironie. Samozřejmě i rozdělení nahrávek do jednotlivých skupin podle emočního zabarvení si tato práce klade za cíl. Dále se při zařazování nahrávek bere ohled na věk účastníků, respektive věkovou skupinu, do které je zařazen. Tohle však bude rozebráno podrobněji v další části.

Neméně důležitým cílem je nahlédnout do teorie emocí a tvorby databází, aby vznikla kvalitní práce, která se o tyto informace opírá. Zvláště teorie tvorby databází je zde velice důležitá, aby vznikl kvalitní nástroj, což je bod, kam celá tato práce směřuje.

Tím se dostáváme k dalšímu cíli projektu, a tím je vytvořit softwarový nástroj s uživatelským rozhraním, který bude sloužit k subjektivnímu hodnocení vytvořené databáze. Tento nástroj měl být navržen tak, aby mohl pracovat i s jinými databázemi než pouze s databází přidruženou k této práci, což znamená dynamické načítání zvukových nahrávek. Dále aby bylo možno upravovat charakter a některé vlastnosti testu pouze pomocí administrátorského menu a nemuselo být zasahováno do samotného kódu programu. Samozřejmě nástroj musí být co nejvíce uživatelsky přívětivý a intuitivní, protože subjektivní testování budou provádět nejrůznější lidé, nehledě na jejich znalost problematiky nebo počítačovou gramotnost. Výstupem nástroje musí být textový soubor, který obsahuje jak některé osobní údaje uživatele, který test prováděl, tak hlavně data relevantní pro posouzení správnosti zařazení nahrávek. Nástroj pak bude sloužit k subjektivnímu zhodnocení, zda nahrávky, které byly zařazeny do skupin podle emocí v nich obsažených, a to ať ručně nebo pomocí automatického algoritmu, byly tímto zařazeny správně a obsahují opravdu emoce korespondující se skupinami, ve kterých se nachází.

Součástí práce by měla být i analýza některých nahrávek vytvořené databáze, která se zabývá hlavně těmito prosodickými rysy promluvy: Základní frekvence, intenzita a první tři formanty.

2. Emoce:

Co je to vlastně emoce? Odpověď na tuto otázku není vůbec jednoduchá a už vůbec ne určitá. První problém vyvstává hned při zamyšlení se nad tím, co za emoci můžeme ještě považovat a co už ne. Pro celkové pochopení je možná lepší se pozastavit nad tím, k čemu nám emoce slouží, na co je potřebujeme a jak vznikají. V následující kapitole se pokusím tuto velmi komplexní problematiku rozebrat pomocí teorií a názorů známých myslitelů, vědců a filozofů z této oblasti.

Čistě z fyziologického hlediska můžeme říct, že se jedná o mentální stavy, které se projevují reakcí na nějaký podnět a to jak fyziologicky (změna rychlosti tepu, dechu, vyšší aktivita potních žláz, červenání se,...), tak chováním (smích, pláč, atak,...). Tahle definice je sice nevyvratitelná, ale z daleka ne úplná, protože nám neříká nic o vzniku emocí, proč se nám v určité situaci například zvýší tep, apod. Avšak je v ní popsán základní princip emoce, a to že emoce je reakcí na aktuální situaci.

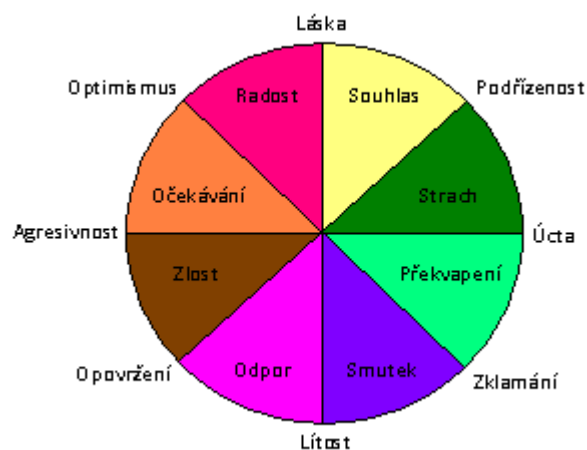
2.1. *Teorie emocí*

2.1.1. Psychoevoluční teorie Roberta Plutchika

Podle psychoevoluční teorie Roberta Plutchika[1] jsou emoce vrozené funkce, které jsou důležité pro přežití. Vyvolávají je podněty, které jsou pro nás významné. Podle jeho teorie existuje u každého člověka základní sada emocí, které člověk pravidelně zažívá. Tyto emoce jsou vrozené a jsou v přímém spojení s chováním pro naše pokračování v životě, ve smyslu, že například útěk nebo boj je přímo spojená reakce, která nám pomáhá přežít.

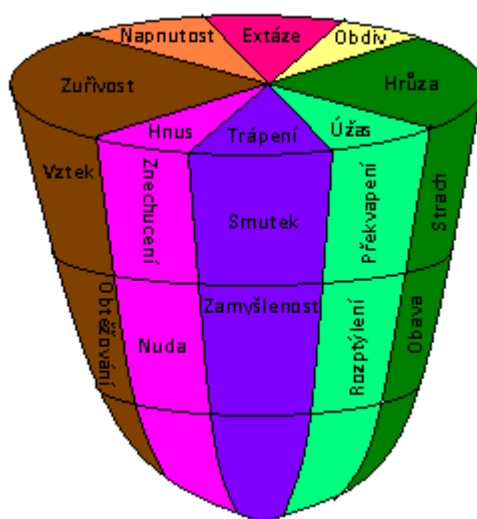
Plutchikův model je založen na "emocionálním kole" (Obr.1). To ukazuje 8 základních emocí, které jsou tvořeny 4 páry opačných emocí: Radost a smutek; souhlas a odpor; strach a zlost; překvapení a očekávání; Podle Plutchikovy teorie nemůžeme zažívat opačné emoce současně. Emoce se dělí na pozitivní (radost, souhlas, překvapení a očekávání) - ty mají blahodárny vliv na náš zdravotní stav a negativní (strach, smutek, odpor a zlost) - ty mají zdravý poškozující vliv.

Jeho dalším předpokladem je, že můžeme zažívat kombinaci emocí. Výsledné kombinace můžeme vidět také na "emocionálním kole". Například radost a souhlas ve výsledku tvoří komplexní emoci lásky.



Obr. 1: Plutchikovo kolo.

Putchik rovněž předpokládal, že každá emoce může být vyjádřena s různou intenzitou. To je vyjádřeno na emočním tělese (Obr.2). Ve vrchní části tělesa můžeme vidět nejintenzivnější verzi emoce, ve spodní naopak verzi emoce nejméně intenzivní. Čím je emoce slabší, tím je obtížnější rozpoznat o jakou emoci se jedná.

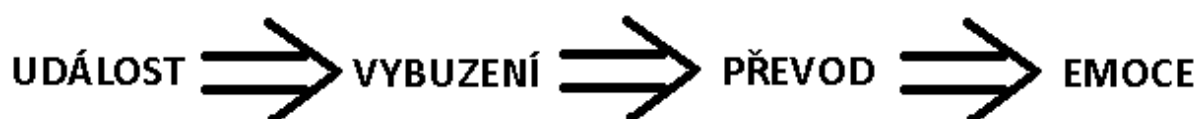


Obr. 2: Emoční těleso podle Plutchika.

2.1.2. James – Langeova teorie

Jedná se o jednu z nejrannějších teorií o funkci a vzniku emocí. Tato teorie [1] je založena na tom, že když jsme vystaveni nějakému hrozícímu podnětu, bezděčný autonomní systém vytváří určité fyziologické události, které mají za následek pocítění příslušné emoce. Například když potkáme v lese medvěda, zvýší se nám krevní tlak, tep a rychlost dechu a to vyvolá emoci strachu a my se dáme na útěk.

Sekvence událostí je tedy znázorněna na Obr.3



Obr. 3: Sekvence událostí vzniku emoce podle James - Langeovy teorie.

Laicky řečeno jde o to, že emoce je důsledkem fyziologických změn vyvolaných danou událostí a ne naopak jak je tomu u většiny ostatních teorií a jak by člověk intuitivně očekával.

2.1.3. Cannon – Bardova teorie

Cannon – Bardova teorie[1] vychází z James – Langeovy teorie, avšak cesta počátečního stimulu je jiná. Tato teorie říká, že když člověk čelí události, která v důsledku ovlivní jeho emoční stav, tak zpráva zaznamenaná smyslovým ústrojím putuje do thalamu v mozku, kde se dělí na dvě části. Jedna část putuje do kůry mozkové, která je spojena s emocemi (strach, radost, zlost, překvapení, smutek,...), zatímco část druhá jde do hypothalamu, který spustí fyziologické změny, jako jsou změna tepu, stahy svalů apod. To znamená, že emoce a fyziologické změny jsou spjaté, ale jedná se o paralelně probíhající procesy.

Tato myšlenka je pořád uznávána a má své příznivce dodnes, i když se i zde vyskytují určité nepřesnosti a problémy.

2.2. Význam emocí

V předchozím textu bylo shrnuto několik nejznámějších teorií o vzniku emocí. Nyní se pozastavím u otázky, k čemu emoce vlastně máme.

V souladu s mnohými autory[2],[3] emoce mají několik funkčních rolí a mezi nejvýznamnější z nich patří tyto:

- Zaručují příslušnou reakci na podněty a situace, které se staly nečekaně a vyžadují okamžitou reakci;
- Je to určitá známka toho, co se v příštím okamžiku chystáme udělat. Z tohoto hlediska se dají emoce považovat za určitou ochranu proti napadení, ničení, apod;
- Mají také komunikativní funkci. Například když s někým mluvíme, tak to v jaké je náladě (emočním stavu) nepoznáme ani tak z toho co říká, ale spíš z toho jak to říká;
- Emoce slouží k tomu, abychom se dokázali psychicky vyrovnat s nastalou situací;

Avšak s přibývajícím věkem člověk dokáže své emoce více ovládat vlivem zkušeností. Tím jak člověk jednotlivé situace v životě zažívá a poznává své reakce a emoce, dokáže postupně ovládat své chování. Proto se emoce objevují mnohem více u dětí, které se ocitají častěji v neočekávaných a neznámých situacích, tam, kde není zaběhnutý vzorec chování.

Emoce jsou spojeny také s kulturním prostředím a přijatelností určitého chování, což má také vliv na komplexnost celé problematiky.

2.3. Způsoby vyjádření emocí

Termín vyjádření emoce není úplně správný a přesný, protože jak je zmíněno výše, vyjádření emoce nemůže být odděleno od emoce jako samostatný proces. Dále jsou z psychologického hlediska v termínu vyjádření emocí zahrnuty také pocity a psychické změny.

Avšak v této podkapitole se pokusím popsat to, jak můžeme “vyjádření” emocí vidět z druhé strany pomocí našich smyslových orgánů. Konkrétně půjde o zvukové a obličejové vyjádření emocí a vyjádření emocí pomocí gest[3].

2.3.1. Zvukové vyjádření emocí

Jedná se o způsob vyjádření emocí, který se dá asi nejméně naučit a ovládat na rozdíl od ostatních dvou způsobů zmíněných v předchozí podkapitole.

Emoce působí změny v dýchání, intonaci a artikulaci, což ovlivňuje mnohé parametry řeči, z čehož budeme vycházet při analýze a následném rozpoznávání emocí. Mezi tyto parametry patří hlasitost, výška, časování (délka mezer mezi jednotlivými slovy, délka slabik) a rychlost řeči samotné. Nyní se tím ale nebudeme zabírat příliš do hloubky, protože této problematice je v další části práce věnována celá kapitola.

Určení emocí z řeči je velice rozsáhlé téma a jeho jednotlivé rysy jsou závislé na mnoha faktorech odrážejících se v mluveném projevu a stejně tak jako psychologie člověka, s tímto tématem spojená, se jedná o velice hlubokou a dosud dokonale neprobádanou oblast.

2.3.2. Obličejové vyjádření emocí

Obličejové vyjádření emocí je nejvíce univerzální vyjádření emocí vůbec. Lidé z různých kultur mohou lehce rozpoznat šťastný nebo smutný obličej vlivem vrozeného emočního programu.

Většina prací zaměřených na toto téma pracuje se statickým vyjádřením obličejových emocí, protože tak se dá zachytit emoce ve své nejvyšší intenzitě. Avšak v poslední době se začal považovat za důležitý celkový průběh obličejového vyjádření. Ten mnohem více reprezentuje reálnou situaci, kdy se obličejové vyjádření mění z neutrálního stavu až po vyjádření s nejsilnější intenzitou. To však není předmětem této práce, a tak nebudu zacházet do podrobností.

2.3.3. Vyjádření emocí pomocí gest

Vyjádřením emocí pomocí gest se myslí jakýkoliv pohyb těla nebo jeho části, který nějakým způsobem vyjadřuje myšlenku nebo pocit. V podstatě se jedná o jakoukoliv grimasu obličeje, pohyb končetin, držení těla, oční kontakt, ale také vzdálenost s jakou stojíme od druhého, dotyky, apod.

Názory na to, na kolik mohou gesta podat věrohodnou informaci o emočním stavu mluvčího se různí a zdaleka na ně v dnešní době neexistuje určitá odpověď. Avšak pro naši práci, která se tímto typem vyjádření emocí nezabývá, bude stačit fakt, že existuje spojení mezi emocemi a gesty, které člověk běžně používá.

2.4. Fyziologické projevy jednotlivých emocí

Nyní bude rozebráno, jaký vliv mají jednotlivé základní emoce, které byly zvoleny pro tuto práci, na fyziologické projevy těla[3].

2.4.1. Radost

Radost se řadí do skupiny pozitivních emocí a je to reakce na úspěch nebo zisk. Podle podnětů, kterými je vyvolána dosahuje různých stupňů intenzity od spokojenosti až po extázi. Fyziologicky se radost projevuje zrychlováním srdeční činnosti a dechu, zvyšováním krevního tlaku a teploty těla, někdy se objevuje i červenání. Mimický výraz je komunikativní.

2.4.2. Smutek

Smutek spadá do skupiny negativních emocí a naopak od radosti je reakcí na neúspěch nebo ztrátu. Jedná se o utlumující cit, kdy dotyčný ztrácí zájem o cokoli a obecně dochází ke snížení aktivity. Fyziologicky se smutek projevuje zpomalením srdeční činnosti a ke změně v rychlosti dechu. Projevy smutku jsou vyjádřením potřeby soustrasti a v tomto smyslu se jedná o komunikativní emoci.

2.4.3. Vztek

Obecně se dá říci, že se jedná o emocionální reakci na nějakou překážku. Vztek se řadí mezi negativní emoce. Pohlížíme na něj jako na cit aktivační, zvyšuje se fyzická síla subjektu a organismus je uveden do pohotovosti. Vztek je spojen s agresí, obranou reakcí a sociálním postojem nepřátelství. Fyziologicky je tato emoce doprovázena zvýšeným svalovým napětím a zlepšenou motorikou. Dále se zvyšuje frekvence tepu a dechu, zvyšuje se krevní tlak, snižuje se pohyb střev a sekrece žaludečních šťáv. Může být doprovázen i mimovolným zaujetím bojového postoje.

2.4.4. Strach

Strach je emocionální reakcí na hrozbu. Stejně jako vztek a smutek spadá do kategorie negativních emocí. Můžeme jej dělit podle typu aktivace na aktivující a deaktivující. Aktivující strach vyvolává akci, což může být například pokus o útěk. Deaktivující se projevuje ztuhnutím organismu, inhibicí tělesných pochodů (dělání se neviditelným, stavění se mrtvým). Fyziologicky je strach provázen vzestupem frekvence nervových impulzů a dalšími změnami, v závislosti na typu aktivace - zvýšením frekvence tepu i dechové činnosti, změna svalového napětí.

V další kapitole se ještě vrátíme k akustickým parametrům jednotlivých emocí, které jsou velmi úzce spjaté s fyziologickými projevy emoční řeči. Ale k tomu je nejprve potřeba znát alespoň zhruba nárys procesu tvorby řeči a také si osvětlit některé pojmy z této oblasti.

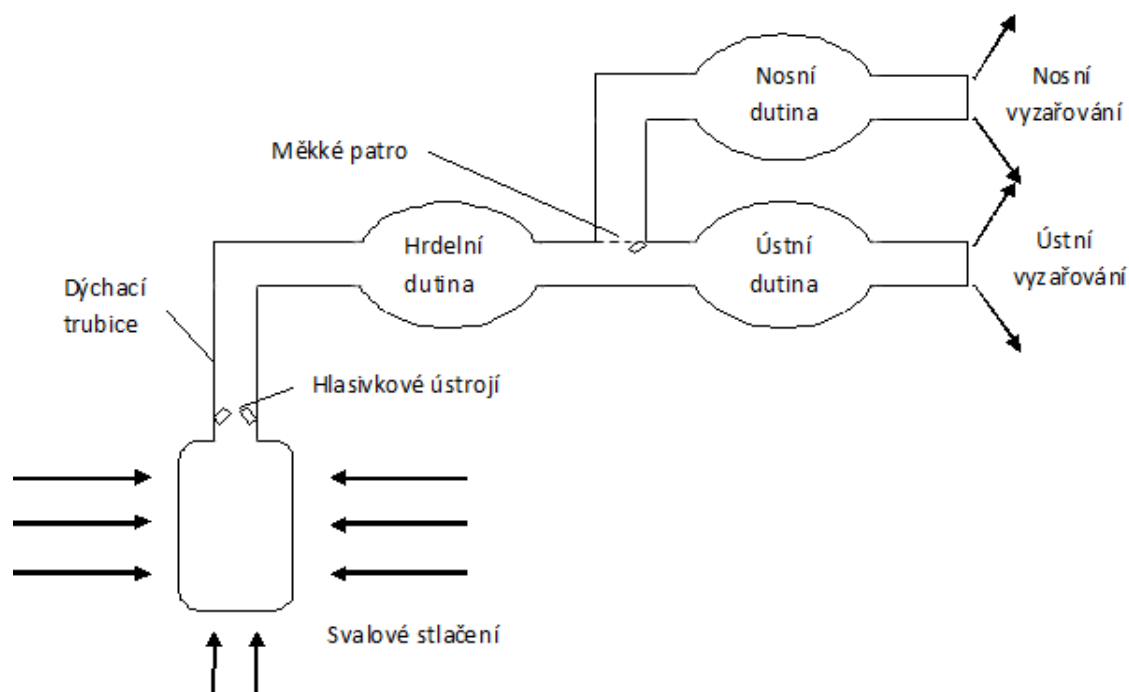
3. Tvorba řeči

Řeč je jedním z nejsložitějších procesů probíhajících v lidském těle. Řeč, mluvení i chápání jsou koordinovány mozkem. V mozkové kůře jsou oblasti zvané řečová centra, kde jsou slova rozluštěna a signály a instrukce jsou odeslány do mnoha svalů v plicích, hrdle a ústech, které se podílejí na tvorbě řeči. Celý dýchací systém a všechny svalové struktury od břicha až po nos mají nějakou roli v tvorbě řeči, ale nejdůležitějšími jsou hrtan, jazyk, rty a měkké patro.

Řeč je tvořena samohláskami a souhláskami. Samohlásky jsou plné zvuky vytvořené fonací. Odlišná rezonance prostoru v ústech a dýchacích cestách je odpovědná za individuální odchylky hlasu. Lebka také rezonuje - když mluvíme, zachycujeme svoji vlastní řeč zčásti přenosem zvuku rezonujícími lebečními kostmi a ušima[4],[6].

3.1. Proces tvorby řeči

Prostor pohrudnice a ústní dutiny spojuje pohyblivá dýchací trubice, složená z několika chrupavkovitých částí, jíž říkáme hrtan. V místě, kde je hrtan nejužší, je uloženo hlasivkové ústrojí, které se skládá z hlasivkových svalů (ty jsou umístěny mezi chrupavkou štítnou a hlasivkovou) a hlasivek. Hlasivky jsou tvořeny dvojicí pružných výchlipek uvnitř hlasivkové chrupavky, viz obr.4.



Obr. 4: Model hlasového ústrojí člověka.

Při aktivaci hlasivkových svalů se hlasivková chrupavka zúží, hlasivky se přiblíží těsněji k sobě a tvoří tak překážku vzduchovému proudu, který vzniká činností plic. Pružné plátky hlasivek se v pravidelných časových intervalech zároveň ohýbají ve směru proudícího vzduchu a zároveň se vrací zpět do klidové pozice, proud vzduchu je tímto způsobem rozkmitáván v pravidelných intervalech a tím vzniká tón. Ten pak postupuje do hrdelní, nosní a ústní dutiny, kde se pohybem řečových orgánů zpracuje na hlásky. Při deaktivaci hlasivek je umožněn volný průchod vzduchu hrtanem, zvuk se nevytváří[4].

Při průchodu zvuku nadhrtanovým prostorem se zužuje prostor, kterým může proudit vzduch, nesoucí s sebou zvukové vlny. V takovém případě se původně neutrální charakter zvuku mění podle tvaru prostoru, skrz nějž proudí, nebo ve kterém je umístěn a vzniká nápadný dynamický přírůstek v některé oblasti frekvenčního spektra zdroje zvuku. Tomuto jevu říkáme souhrnně rezonance a místo, kde je harmonická složka oproti ostatním částem spektra nápadně zesílena, nazýváme formant.

3.2. Věkové rozdělení

Parametry promluvy jsou do jisté míry ovlivňovány i věkem mluvčího, a tak bylo po zvážení psychologického a fyziologického hlediska utvořeno těchto 8 věkových skupin, které budou použity pro rozdělení mluvčích naší databáze.

- | | |
|-----------------------|-------------|
| 1. Předškolní věk: | 0 - 7 let |
| 2. Školní věk: | 7 - 12 let |
| 3. Pubescence: | 12 - 14 let |
| 4. Adolescence: | 15 - 20 let |
| 5. Mladá dospělost: | 20 - 30 let |
| 6. Střední dospělost: | 30 - 50 let |
| 7. Pozdní dospělost: | 50 - 60 let |
| 8. Stáří: | nad 60 let |

4. Prozodie a emoce

Prozodie popisuje veškeré zvukové vlastnosti jazyka (změny základního tónu, intenzity, trvání,...), které se uplatňují na vyšší úrovni, než je rovina segmentální (hláska, segment)[5].

Emoce jsou v hlase vyjádřeny určitými prozodickými vzorci za přítomnosti paralingvistických informací. Hlasový projev emocí je velice úzce spjat s konkrétními fyziologickými změnami (např. u vyjádření smutku je fyziologickou změnou celková deaktivace, tzn. i ochabnutí svalstva v oblasti hrtanu a to se odráží na prozodických manifestacích této emoce), které jsou u člověka a zvířat podobné[3].

4.1. Základní prozodické parametry

Mezi významné parametry pro popis prozodických změn v řeči patří základní frekvence, intenzita a trvání.

Jednotlivé akustické složky mohou být rozděleny do tří kategorií. V první kategorii se nachází složky, které jsou kontrolované mluvčím, například napětí a subglotální tlak, který ovlivňuje intenzitu hlasu. Druhá kategorie obsahuje složky, které naopak mluvčím kontrolované nejsou, nebo jen okrajově. Do třetí kategorie spadají parametry, které jsou dány věkem, pohlavím a takovými vlastnostmi mluvčího, které jsou z fyziologického hlediska po určitou životní etapu konstantní a tudíž mluvčím neovlivnitelné

Nyní se zaměříme na jednotlivé parametry prozodie[2].

4.1.1. Intonace

V případě znělých segmentů produkuje fonační ústrojí tzv. laryngální neboli hlasivkový tón, někdy nazývaný také základní tón. Ten rezonuje v nadhrtanových dutinách. Průběh základního tónu se v řeči projevuje jako melodie řeči. Změny výšky tónu jsou z hlediska vnímání řečové prozodie považovány za percepčně nejvýznamnější. Mimo jiné rozdíl mezi větou deklarativní a interogativní je dán právě změnami výšky tónů. Průběh a změny frekvence základního tónu umožňuje zjistit pitch detektor, který je součástí programu Praat.

Základní frekvence je sluchově vnímána jako výška hlasu. Jak bylo naznačeno, jedná se o počet hlasivkových kmitů za sekundu. Příčinou zvyšování F_0 je zvyšování napětí hlasivek, což je důsledek celkového napětí mluvčího.

Frekvence základního tónu se u člověka pohybuje v rozmezí od 80 do 450 Hz. Tyto hodnoty se liší podle věku, pohlaví a dalších fyziologických a citových stavech mluvčího. Průměrná hodnota u mužů se pohybuje okolo 100 až 150 Hz/sek, u žen se tato hodnota pohybuje okolo 200 až 300 Hz/sek[6].

4.1.2. Intenzita

Intenzita řečového signálu je vnímána jako síla hlasu - obecně hlasitost. Její úroveň závisí hlavně na velikosti subglotálního tlaku a je spojena s funkcí dýchacího a fonačního systému. Jak už intuitivně vyplývá, se zvyšujícím se subglotálním tlakem, se zvyšuje i intenzita řeči a naopak. Subglotální tlak je přímo spojen s emočním stavem mluvčího a stejně jako změny v dýchání, srdečním rytmu, krevním tlaku a jiné, jsou jedním z fyziologických projevů emocí.

4.1.3. Trvání

V pracích, které se vyloženě nezabývají vlastnostmi hlásek, se měření trvání temporálních změn vztahuje k vyšším celkům, jako je slabika nebo přízvukový takt a stejně tak tomu je i u naší práce.

To je však hluboce svázáno s konkrétním jazykem, avšak finální prodlužování, pozorované u více jazyků bez ohledu na pozici jejich pravidelného přízvuku, je vysvětlováno fyziologickou potřebou různě dlouhé doby potřebné k vytvoření finální melodické kontury.

Celek časových složek, trvání slabik a pauz, je vnímán jako tempo řeči.

Tím byly shrnuty 3 nejvýznamnější prozodické parametry, avšak pro popsání emocionálně příznakové promluvy je nutné uvážit i jiné než pouze prozodické složky promluvy. Proto budou nyní rozebrány ještě další zvukové vlastnosti řeči.

4.2. *Další zvukové vlastnosti řeči*

4.2.1. Artikulace

Stavba artikulačního ústrojí umožňuje měnit velikost a tvar některých akustických dutin. Umožňují jednak rozlišit jednotlivé hlásky, ale hlavně jemné rozdíly v jejich realizaci, které odpovídají buď fyzickému a psychickému stavu mluvčího nebo jeho komunikačnímu záměru.

Vlivem emočních prožitků může dojít ke zvýšení nebo snížení napětí i v oblasti artikulačního svalstva, což má za následek změnu tentokrát nikoliv prozodické, nýbrž segmentální stránky promluvy. Tyto modifikace způsobují akustické změny ve spektru.

4.2.2. Barva hlasu

Stejně jako se emoční stav mluvčího promítá do artikulace vlivem změny napětí v oblasti artikulačního svalstva, promítá se i do barvy hlasu změnou napětí svalů v oblasti hlasového

ústrojí. Hlas je potom vnímán jako přiškrcený, skřípavý, nebo může obsahovat větší podíl šumové složky.

Akustická analýza těchto spektrálních vlastností je však nepoměrně komplikovanější.

4.2.3. Dýchání

Dýchání je ovlivněno činností srdce i celkovým svalovým napětím. Pokud dojde vlivem emočního stavu mluvčího ke změně jednoho nebo obou zmíněných parametrů, je dopad na dechovou strukturu nepopiratelný.

Při souvislé promluvě by měly nádechové pauzy oddělovat logické celky. Změní-li se rozsah dechových celků, pak je délka logických celků nevyrovnaná. Působením některých emocí, se stává dýchání mělčím. Absence dechové podpory je patrná i na kvalitě hlasu. Jeho základní frekvence kolísá, hlas je vnímán jako roztřesený.

4.2.4. Pauzy

Pauzy představují nedílnou součást časové struktury řeči a jsou zde uvedeny pro úplnost celku, avšak kvůli tematickému zaměření zde nebudou podrobněji rozebrány.

Tím byly shrnuty veškeré vlastnosti promluvy jako takové a nyní se podíváme na konkrétní parametry vhodné pro akustický rozbor.

4.3. *Akustické parametry relevantní pro výzkum*

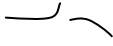
Pro identifikaci emočního stavu mluvčího v hlasovém projevu se ukázaly jako nejvhodnější tyto parametry[2]:

- průměrné F_0 , průměrné F_0 v rámci určitého úseku nahrávky, rozdíl mezi maximálním a minimálním F_0 , tvar a proměnnost intonační křivky;
- co se intenzity týče – její variabilita a proměnnost;
- v oblasti vlastností spektra hodnoty prvních dvou formantů, šířka formantového pásma, šířka spektra, energie v horní části spektra a šum ve spektru;
- tempo promluvy a časové vlastnosti promluvy;

4.4. *Prozodické rysy konkrétních emocí*

Nyní si ukážeme, jak se prozodické rysy (pro nás důležité akustické parametry) projevují u konkrétních emocí[2]. Tyto údaje se však studie od studie mění, ale i tak nám hodnoty můžou sloužit jako orientační a hodně napovědět. Spíše je nutné brát v úvahu vzájemný vztah neutrální promluvy a promluvy vyjadřující různé emoce než absolutní hodnoty parametrů.

Tabulka 1: Prozodické rysy různých expresivních realizací 1 věty podle Léona.

	neutralita	vzteky	smutek	radost	strach	překvapení	obdiv	ironie
průměr F0	120Hz	180 Hz	110 Hz	200 Hz	120 Hz	120 Hz	150 Hz	190 Hz
min. a max. hodnoty F0 a jejich rozdíl	100 ; 170 E = 70 Hz	100 ; 130 E = 30 Hz	100 ; 130 E = 30 Hz	100 ; 250 E = 150 Hz	100 ; 130 E = 30 Hz	100 ; 340 E = 240 Hz	120 ; 200 E = 80 Hz	100 ; 400 E = 300 Hz
tvar intonační křivky								
průměrný rozdíl intenzity od neutr promluvy	0 dB	11 dB	0 dB	16 dB	0 dB	-1 dB	3 dB	3 dB
min., max. intenzita a rozdíl	+4 ; -2 E = 6 dB	+16 ; +3 E = 13 dB	+2 ; -3 E = 5 dB	+18 ; +8 E = 10 dB	+6 ; -3 E = 9 dB	0 ; -3 E = 3 dB	+10 ; +2 E = 8 dB	+5 ; -3 E = 8 dB
pauza	0 ms	0 ms	0 ms	750 ms	0 ms	250 ms	0 ms	0 ms
trvání věty	1700 ms	1650 ms	1950 ms	2700 ms	1850 ms	2300 ms	1920 ms	1800 ms

5. Databáze nahrávek

5.1. Tvoření databáze

5.1.1. Databáze podle typu nahrávání

Základní souboj při tvoření zvukové databáze tkví v duelu kvalita nahrávek vs. autenticita nahrávek. Dále je potřeba samozřejmě zvážit i další faktory jako jsou dostupné prostředky a možnosti nahrávání – zda máme k dispozici nahrávací studio, profesionální herce,.... Nyní se však zaměřím na prvně zmíněnou záležitost[2],[7].

Nahrávací studio

Jedna nebo více vět jsou čteny s různými emocemi. Nahrávky jsou velice kvalitní, rozsah a složení účastníků široké, avšak věrohodnost emocí není ideální ani v případě, že se jedná o profesionální herce (nejedná se o reálné emoce a i herec může být pod tlakem mnohem více než při hraní ve filmu když ví, že je kladen důraz na emoce a tím ztrácí nahrávka na autenticitě). Je to nejvíce užívaná metoda. Speciální odnož této metody - jedná se o navození vyšší autenticity poněkud drastickou formou - např. podáním psychoaktivních látek účastníkům, zadáním náročných požadavků nebo arogantním a hrubým chováním k účastníkům - avšak zde není jistota, že tohle chování odpovídá chování v reálných situacích.

Mezi tyto databáze patří například databáze „TALKAPILLAR“.

Televizní nebo rádiové záznamy

Jedná se o vystřížení částí pořadů nebo filmů, kde je čitelná některá z námi zvolených emocí. Autenticita se zde zvětšuje, ale na úkor kvality nahrávek. Dále je obtížnější nahrávky získat, protože se musíme vyvarovat nahrávek, kde je hluk, hudba v pozadí nebo řeč ostatních účastníků. Metoda také zabere hodně času z důvodu nutnosti shlédnutí velkého množství materiálu.

Příkladem této databáze je „Belfast Naturalistic Database“.

Reálné prostředí

Tato metoda se realizuje např. připevněním mikrofону na určitý subjekt, který pak s nahrávacím zařízením normálně žije a tak je nahráván materiál pro tento typ databáze. Pak samozřejmě musí dojít k “přebrání” materiálu (přeposlouchání desítek hodin materiálu, vybrání emocí, sestřihání,...). Zde je autenticita největší, ale vyvstávají mnohé problémy: nekvalita nahrávky, obtížnost zachycení emoce, která trvá jen pár sekund a nahrávací zařízení

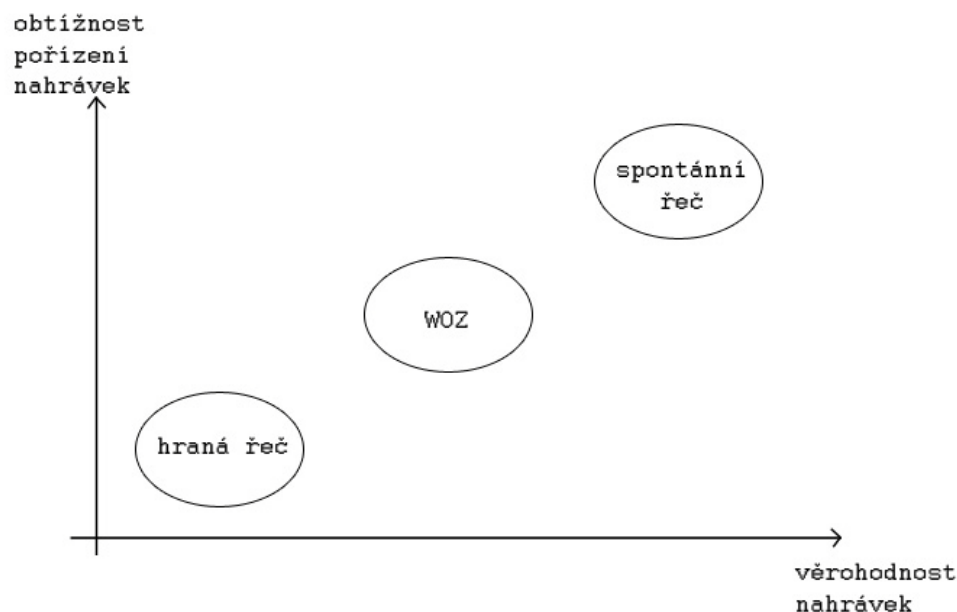
musí být neustále v dostatečné blízkosti od mluvčího a mnohé další. Nahrávací zařízení musí být schované, aby nebyla porušena autenticita nahrávky, kvůli které byla metoda vybrána,... Jedná se o nejobtížnější metodu a téměř není užívána, ale po překonání překážek se dle mého názoru jedná o databázi nejkvalitněji provedenou.

WOZ – Wizard of Oz

V uměle vyvolané řeči je používán systém Wizard of Oz. Znamená to využití programu nebo osoby, která vzájemně komunikuje s mluvčím a snaží se ho uvést do určitého emočního stavu, který je pak zaznamenán. Pokud interakci vytváří stroj, pak tato metoda vyžaduje kvalitní program, který dokáže účastníka přimět říct něco v námi očekávaném emočním stavu. Zkonstruování takového programu ovšem není vůbec jednoduché a ani levné. Avšak po překročení těchto problémů jsou produkovány velice kvalitní, co se akustické stránky týče, a i poměrně autentické nahrávky.

Zde je příkladem databáze „SMARTKOM“.

Věrohodnost jednotlivých metod pořizování nahrávek a obtížnost jejich pořízení je přehledně zobrazena na obrázku 5.



Obrázek 5: Graf věrohodnosti nahrávek.

5.1.2. Důležité faktory při nahrávání

Výzkum v oblasti řeči a emocí se posunuje z etapy, kterou bychom mohli nazvat “výzkumnou” do etapy, kde se objevují robustní aplikace, zejména v komunikaci člověk-počítač. A pokrok v této nastupující etapě je do značné míry dán i vývojem příslušných databází. Proto se nyní zaměřím na 4 hlavní body, které je potřeba zvážit při vytváření jakékoliv databáze. Těmi jsou oblast, přirozenost, kontext a popis[10].

Oblast

Termín “oblast” je použit pro pokrytí více parametrů, které je nutné v databázi zvážit. V první řadě se jedná o počet různých mluvčích, jazyk nahrávek, ale i typ dialektu, pohlaví mluvčích, typy emočních stavů, počet příspěvků daného emočního stavu a funkční nebo společenská funkce. Tyto parametry jsou velice důležité, protože mnohé vlastnosti nahrávek se hodně liší třeba jen změnou jednoho z těchto parametrů.

Přirozenost

Tohle již bylo dostatečně probráno v předchozí kapitole, tudíž zde pouze zmíním, že je potřeba při analýze nahrávek brát ohled na fakt, že není jasné, nakolik důvěryhodně zrcadlí připravená řeč spontánní vyjádření emocí. Připravená řeč je obvykle čtená a jedná se o monolog na rozdíl od reálné situace, což může mít také vliv na akustické parametry promluvy.

Kontext

Existuje přímá evidence, že posluchač používá kontext k rozhodnutí se, o jakou emoci se jedná. Proto je důležité, aby databáze obsahovala záznam o typu kontextu. Rozlišujeme tyto typy kontextů:

- sémantický kontext – řeč může obsahovat slova, která značí, o jakou emoci se jedná (např. láska, prohra, smrt);
- strukturální kontext – emoce způsobují i variace ve stylu promluvy, které se odrážejí v délce frází, přerušení mezi věty,...
- způsobový kontext – za normálních okolností vidíme i slyšíme mluvčího a samozřejmě i vizuální stránka nám podává určité informace, které vypovídají o emočním stavu mluvčího, tudíž je nutné brát v ohled i tento kontext;
- dočasný kontext – u přirozené promluvy se emoce s časem mění a kolísají, a tak je třeba sledovat, zda nedochází při poklesu emoce k vyhodnocení nahrávky jako emocionálně neurčitě;

Popis

Konstruování databáze vyžaduje techniku pro popis jak lingvistického a emocionálního obsahu, tak řečového. Jsou zde zahrnuty položky jako kategorie emoce (radost, smutek,...), kvalita nahrávky, pokud se jedná o multimodální nahrávku, pak i to je nutností uvést.

5.2. Dostupné databáze

Jako poslední část této kapitoly přikládám kolekci emočních databází. Jedná se o databáze jak zvukové, tak vizuální. Není to samozřejmě seznam všech existujících databází, pouze těch významnějších nebo něčím zajímavých. Dále slouží následující tabulka jako obecný přehled toho, co už bylo vytvořeno[5],[7],[8].

Tabulka 2: Seznam významných emočních databází.

název	typ	metoda nahrávky	obsah	obsazení	jazyk
HUMAINE database	AV	přirozená (některé klipy z Belfast databáze, reality show)	50 nahrávek (1 audiovisual)	široký rozsah	EN,FR,HE
Belfast Naturalistic Database	AV	přirozená (diskuzní pořady, rozhovory)	239 TV nahrávek, 209 rozhovoru	31M, 94Ž	EN,FR,HE
Chung	AV	přirozená (televizní rozhovory)	široký rozsah	61korea,6USA	KO, EN
SMARTKOM	AV	WOZ metoda*	448 nahrávek	224 mluvčích	GE
TALKAPILLAR	A	hraná (neutrální věty čtené s různými emocemi)	26 nahrávek ke každé emoci (ve třech úrovních aktivace)	1 herec	FR
Reading-Leeds database	A	přirozená (nepřipravené rozhovory-znovuprožití emocionálních událostí)	široký rozsah	široký rozsah	EN
France et al.	A	přirozená (terapie a telefonní hovory)	široký rozsah	67M, 48Ž	EN
Campbell CREST database	A	přirozená (nahrávky dobrovolníků ohledně sociálních problémů)	1000 hodin nahrávek	široký rozsah	EN,JP,CH
SYMPAFly	A	lidský automat rezervování letů	110 rozhovorů	široký rozsah	GE
DARPA Communicator corpus	A	lidský automat "call centra"	13187 vyjádření, z toho 1750 emocionálních	široký rozsah	EN
AIBO(Erlangen database)	A	lidský automat - interakce s robotem	9,2hod německých a 1,5 hod anglických dětí	51GE, 30EN (děti)	GE, EN
Fernandez et al.	A	nepřipravené matematické otázky	široký rozsah	4 mluvci	EN
Tolkmitt and Scherer	A	reakce na připravené "slejd" obsahující stresující materiály	3 odezvy na každý "slejd" následovány krátkým	33M, 27Ž	GE
Iriondo et al.	A	hraná (čtení různých textů s příslušným emocionálním zabarvením)	nepřipravené 20-40 sec dlouhé nahrávky	8 mluvčích	SP
Mozziconacci	A	hraná (neutrální věty, čteny s různými emocemi)	každý mluvčí 8 vět, každá 3krát opakována	3 mluvčí	NE
McGilloway	A	hraná (různé pasáže, čteny s různými emocemi)	každý mluvčí 5 pasáží	40 mluvčích	EN
Belfast structured Database An extension of McGilloway Datab.	A	10 pasáží ve stylu McGilloway, 10 připravených ve stylu Belfast ND	každý mluvčí čte 20 pasáží	50 mluvčích	EN

Danish Emotional Speech Database	A	hraná (věty, pasáže a slova čteny s různými emocemi)	každý mluvčí čte 2 slova, 9 vět a 2 pasáže s různými emocemi	4 mluvci	DA
Groningen ELRA corpus number S0020	A	hraná	každý mluvčí čte 2 krátké připravené texty	238 mluvčích	NE
Berlin database	A	hraná (neutrální věty čtené s různými emocemi)	každý mluvčí čte 10 připravených vět	5M, 5Ž	GE
van Bezooijen	A	hraná (neutrální věty čtené s	každý mluvčí čte 4 fráze	8 mluvčích	NE
Abelin	A	hraná (neutrální věty čtené s různými emocemi)	1 neutrální fráze čtena s různými emocemi	1 mluvčí	SW
Yacoub et al	A	hraná (prohlášení čteny s různými emocemi)	2433 prohlášení	8 mluvčích	EN
Cantonese Speecon database	A	hraná i spontánní (kancelář, veřejné místa,.../čtení řeči)	5nahrávacích míst pro spontánní, široký rozsah pro čtení	550dosp., 50dětí	Kanton-ština
Thai Speecon database	A	hraná i spontánní (kancelář, veřejné místa,.../čtení řeči)	5nahrávacích míst pro spontánní, široký rozsah pro čtení	552dosp., 50dětí	TH
MediaTeam Speech Corpus	A	hraná (čtení neutrálních vět s různými emocemi)	každý mluvčí čte věty se sedmi různými emocemi	8M, 6Ž	FI
Mandarin Emotional Speech Database	A	hraná (20 smysluplných vět pro různé emoce)	839 nahrávek	18M, 16Ž	Mandarinština

Význam zkratk uvedených v tabulce 2:

A	databáze obsahující pouze zvukovou stopu (audio)
AV	databáze obsahující jak zvukovou, tak i obrazovou stopu (audiovisual)
EN	angličtina
FR	francouzština
GE	němčina
SW	švédština
DA	dánština
FI	finština
TH	thajština
JP	japonština
SP	španělština
CH	čínština
HE	hebrejština

5.3. Vlastní databáze

Vlastní databáze českých nahrávek bude po zvážení problematiky tvoření databází, která je mimo jiné popsána v této práci, obsahovat nahrávky z televizních pořadů a seriálů. Rozhodli jsme se tak na základě vyšší autentičnosti nahrávek a většího rozsahu jak účastníků, tak celkového materiálu, ze kterého lze čerpat. Jedná se o vystřížené zvukové úseky z jednotlivých dílů seriálu (1). Bylo rozhodnuto, že se bude pracovat pouze se stopou zvukovou, a to také odpovídá charakteru softwarového nástroje, který byl zvolen pro vyznačení jednotlivých úseků pořadů, které byly označeny jako emočně zabarvené. Jedná se o volně šiřitelný program Elan, ten bude podrobněji popsán v následující podkapitole.

Databáze obsahuje k tomuto datu přes 200 nahrávek s popisem mluvčího (jméno, pohlaví, věk) přiřazené příslušné emoci. Jelikož se jedná o více seriálů a pořadů formát i kvalita nahrávek se liší podle zdroje nahrávání. Některé byly nahrávány z internetem streamovaných seriálů, další pak byly nahrány přes televizní kartu. Aktéry nahrávek jsou jak muži, tak ženy a jedná se ve velké většině o věkovou populaci středního a vyššího věku. Několik aktérů je však také mladšího věku, a tak databáze obsahuje nahrávky aktérů celé věkové populace. Průměrná délka nahrávek se pak pohybuje okolo 3 sekund a nepřesáhne hodnotu 5 sekund.

Jak už bylo předesláno v úvodu, budeme se soustředit hlavně na těchto 6 základních emocí: Radost, strach, smutek, vztek, překvapení a ironie - popřípadě doplněné o neutrální promluvu. Pro označení emocí jsem z důvodu přehlednosti zvolil zkratky písmenné (viz Tab.3). Číselné zkratky jsou použity pro označení věkové skupiny a ID mluvčího.

Tabulka 3: Přehled zkratk jednotlivých emocí použitých v databázi.

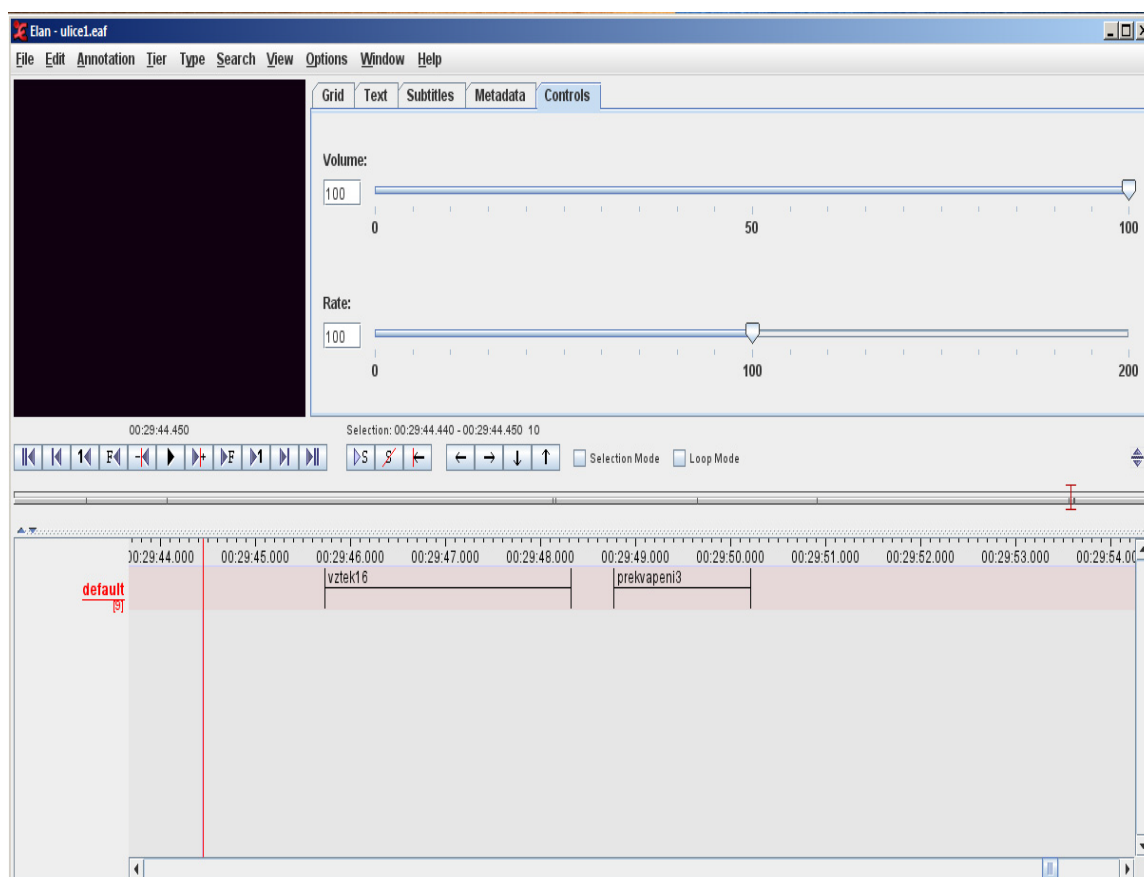
emoce	zkratka
radost	RA
strach	ST
smutek	SM
vztek	VZ
překvapení	PR
ironie	IR
neutralita	NE

Pozn. 1: Jedná se o seriály „Velmi křehké vztahy“ a „Ulice“, dále pak byl vybrán film „Vratné láhve“.

5.4. Program Elan

Jedná se o profesionální nástroj pro vytváření komplexních poznámek v obrazovém a zvukovém záznamu. Do jednoho záznamu může být přidáno neomezené množství nahrávek, a to např. slov, vět, překladů nebo popisů jakýchkoliv poznatků z nahrávky. Tyto poznámky mohou být přiřazené k určitým velice přesným časovým úsekům a toho také bylo využito pro naše účely – označit část nahrávky, kde se nachází některá emoce. Čas počátku poznámky v nahrávce a také čas ukončení poznámky je vyznačen s přesností na tisíciný sekundy. Program pak umožňuje tyto poznámky zapsat do výstupního textového souboru, čehož bylo také využito, a tak bylo získáno časové vyznačení všech zvolených emočních úseků. Tyto údaje lze pak poměrně jednoduše aplikovat např. do programu Matlab, který nahrávky vystřihá.

Ukázka z tohoto programu je zobrazena na obr. 6.



Obrázek 6: Ukázka z programu Elan.

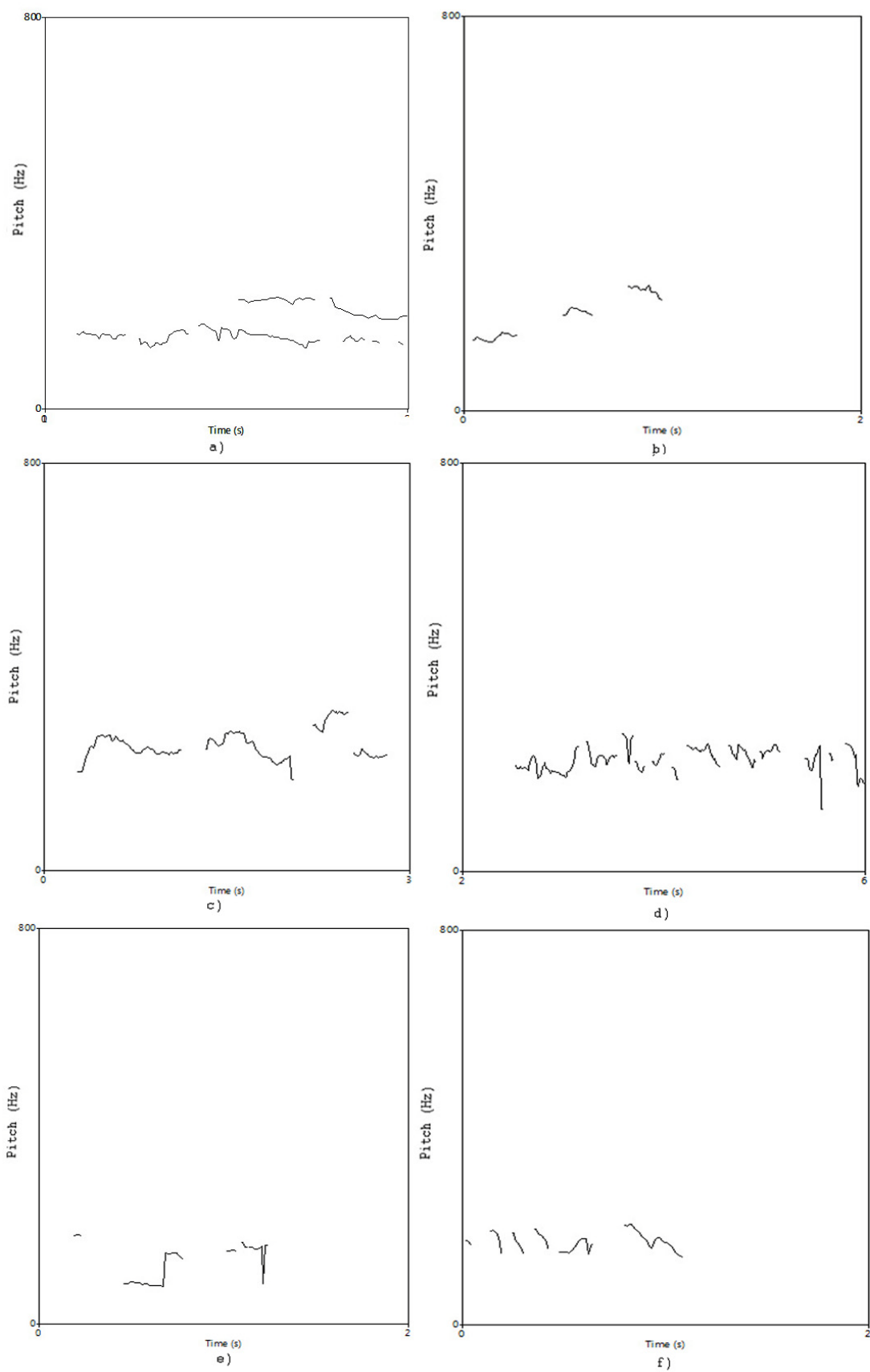
6. Analýza

Součástí této práce je také obecná analýza ukázkových nahrávek pro každou zvolenou emoci. Pro analýzu jsme využili bezplatný program Praat, který je vytvořen přímo pro tyto účely. Prostředí programu je velice intuitivní, a tak zde nebudou jednotlivé kroky analýzy nijak detailněji popisovány. Při analyzování vybraných nahrávek jsme se soustředili na základní frekvenci F_0 , intenzitu I a první tři formanty F_1 , F_2 a F_3 .

6.1. Základní frekvence

Základní frekvence je sluchově vnímána jako výška hlasu. Program Praat ji umožňuje zaznamenat v čase pomocí funkce „Periodicity“.

Na obrázku 7 jsou zobrazeny ukázkové průběhy pro všech 6 zvolených emocí.

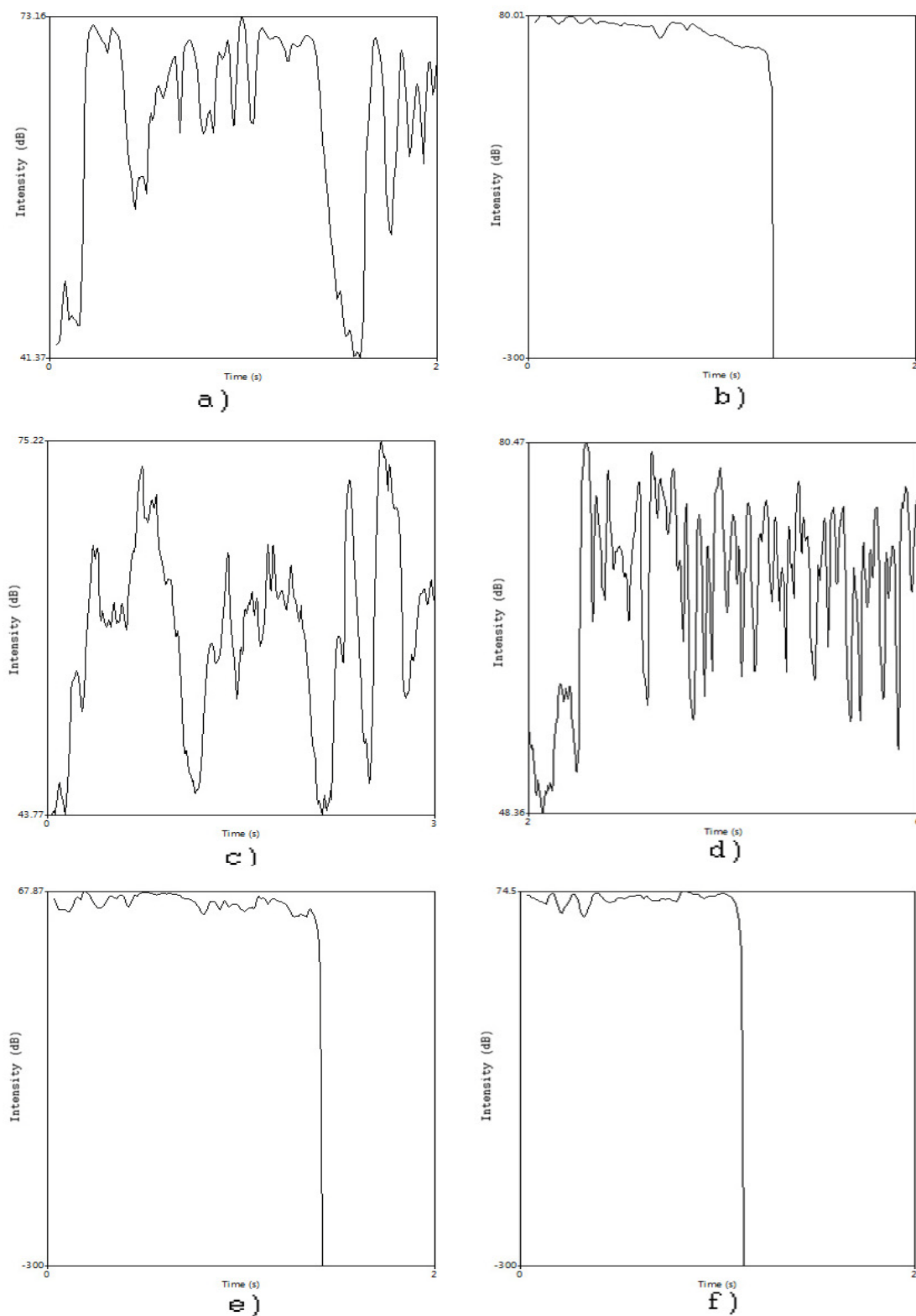


Obrázek 7: Ukázkové průběhy základní frekvence pro všechny emoce: a) ironie, b) překvapení, c) radost, d) vztek, e) smutek, f) strach.

6.2. Intenzita

Jedná se o množství akustické energie, která projde jednotkovou plochou za jednotku času. Tento průběh umožňuje program Praat vykreslit funkcí „To Intensity“.

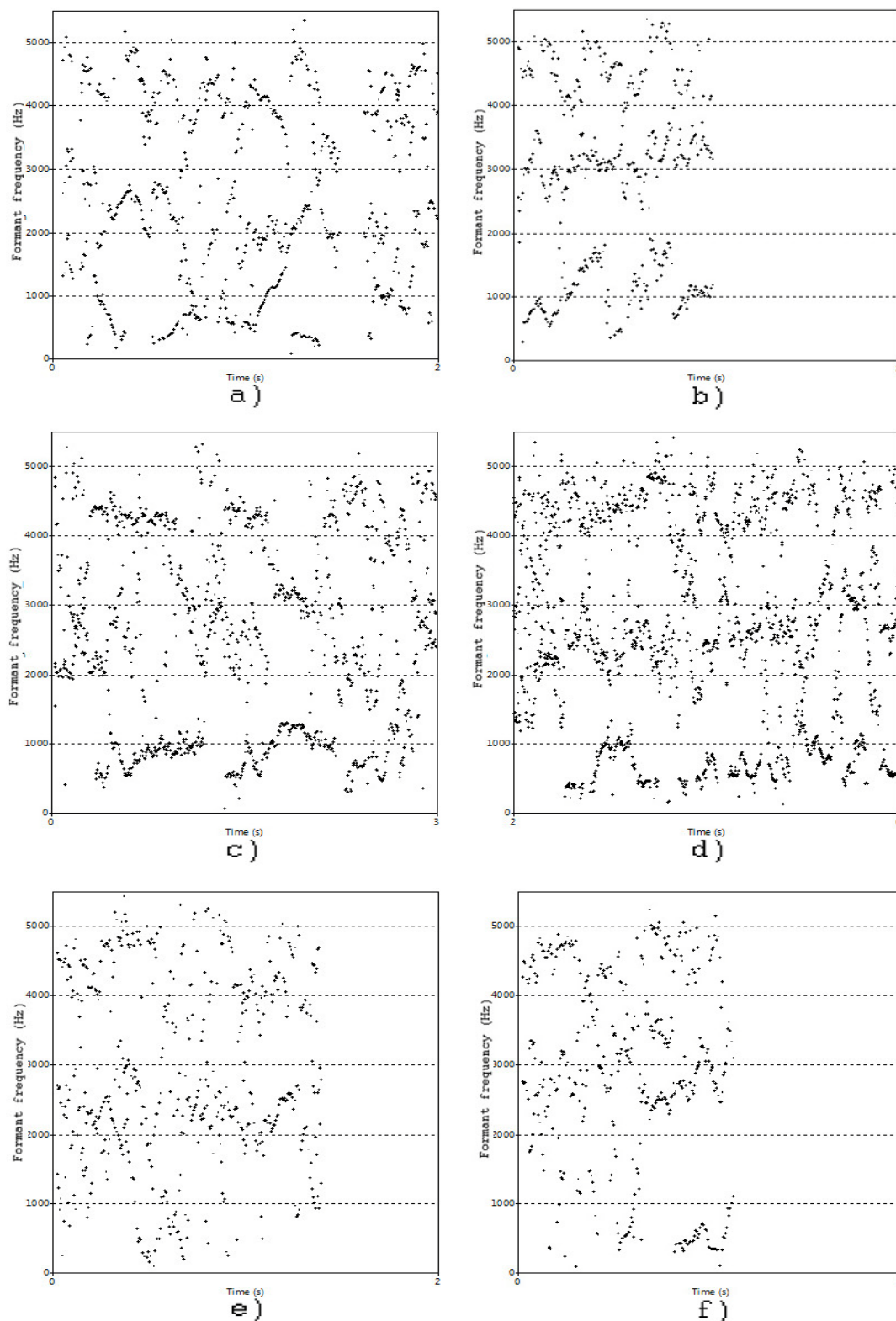
Na následujícím obrázku jsou opět ukázkové průběhy pro všechny zvolené emoce.



Obrázek 8: Ukázkové průběhy intenzity pro všechny emoce: a) ironie, b) překvapení, c) radost, d) vztek, e) smutek, f) strach.

6.3. Formanty

První 3 formanty jsou z hlediska vnímání řeči velice důležité pro určení samohlásek. Program Praat umožňuje vykreslení prvních třech formantů pomocí funkce „Formants&LPC“. Opět následují ukázky pro veškeré zvolené emoce.



Obrázek 9: Ukázkové průběhy prvních třech formantů pro všechny emoce: a) ironie, b) překvapení, c) radost, d) vztek, e) smutek, f) strach.

7. Aplikace pro subjektivní hodnocení nahrávek

Podstatnou součástí této práce bylo naprogramovat softwarový nástroj s uživatelským rozhraním, který bude sloužit k subjektivnímu hodnocení vytvořené databáze.

To znamená, že náhodně vybraní lidé, kteří nemají odborné znalosti z této problematiky, budou poslouchat jednotlivé emočně zabarvené nahrávky a zařazovat je do skupin náležitým emočním stavům podle jejich subjektivního vnímání. Bude se jednat o zpětnou vazbu, kontrolu správnosti jakéhokoli nástroje pro automatické rozřazování emočních nahrávek nebo pouze hodnocení některé konkrétní databáze zvukových nahrávek, protože neexistuje žádné pravidlo, které by nám jednoznačně určilo emoční zabarvení lidské řeči. Jedná se o subjektivní zabarvení informace, a tak i rozpoznání této informace je subjektivní záležitost.

Jelikož budou nástroj používat nejrůznější lidé bez ohledu na zaměření a odbornost, bylo kladeno za cíl, aby byl nástroj co nejvíce uživatelsky přívětivý. Dále bylo na základě požadavků (konkrétně uvedeno dále v této kapitole) rozhodnuto, že se bude jednat o nástroj s grafickým rozhraním. Spolu s požadavkem na programovací jazyk C++ jsem se rozhodl, že použiji překladač Builder C++. Jedná se konkrétně o verzi 6.0 (Build 10.157).

7.1. Požadavky a řešení programu

Nyní budou rozebrány jednotlivé požadavky, které byly kladeny na softwarovou aplikaci, a také jejich konkrétní řešení. Bude se jednat spíše o obecné řešení, konkrétní kód je na příloženém cd. Obecnými požadavky bylo, aby byl nástroj dvojjazyčný (čeština a angličtina), aby byl přehledný a pro uživatele jednoduchý. Dále jsou požadavky členěny do několika podkapitol z hlediska struktury programu.

Těmi jsou Vstupní formulář, kde uživatel zadává vstupní data, Admin Menu, kde může administrátor měnit parametry testu, příprava testu před jeho spuštěním a důležitý kód, který je zde vykonán, dále pak samotný formulář hlavního testu, ve kterém probíhá testování, proces po ukončení testu, kdy jsou data zpracována a jako poslední zápis výsledků do výstupního souboru.

7.1.1. Vstupní formulář

Jedná se o první z několika oken, se kterými se uživatel setká po spuštění programu. V tomto okně uživatel zadává údaje relevantní pro tento test a dále data potřebná pro správný běh programu, viz. obr.5.

Jelikož byl zadán požadavek na dvojjazyčnost, bylo nutné, aby uživatel jako první položku označil jazyk, ve kterém chce s nástrojem komunikovat. To je vyřešeno komponentou typu Radiobutton v nejvrchnější části okna (obr.5), která ihned po přepnutí zajistí změnu veškerých nápisů podle jazyka našeho výběru (obr.6). Dále se však změní také všechny nápisy ve všech dalších oknech, se kterými budeme pracovat a stejně tak text ve výstupním souboru.

Samozřejmě bylo nutné, aby uživatel zadal své osobní údaje a dále také údaje o svém hudebním nadání, což mohou být důležité informace při vyhodnocování výsledků testů. Při postupu v okně od vrchu dolů (obr.5), vidíme v další části okna příslušné položky k těmto požadavkům. Konkrétně jsou to komponenty typu Edit pro vkládání jména, příjmení a věku uživatele, komponenty typu Radiobutton pro vybrání jeho pohlaví a dále komponenty typu CheckBox pro zvolení odpovědi na otázky jestli má uživatel hudební sluch a jestli někdy hrál na

hudební nástroj. Pokud uživatel nezadá své jméno, je automaticky pojmenován jako „anonym“.

Dalším požadavkem pro vstupní formulář bylo zajistit, aby mohl uživatele zadat fyzickou cestu k uložené databázi a také cestu, kam bude po úspěšném dokončení testu uložen výstupní soubor – výsledek testu. Tento požadavek je zprostředkován funkcí SelectDirectory, která je navázána na komponentu typu Button. To můžeme vidět realizováno v další části vstupního formuláře, viz obr.5.

Velmi důležitou částí vstupního formuláře a vůbec celého programu bylo zajistit, aby mohl uživatel test např. v půlce přerušit a poté rozpracovaný test dodělat jindy. Tento požadavek byl zadán z toho důvodu, že některé databáze můžou mít až tisíce nahrávek, a tak jejich celkové přehrávání a zařazení může trvat i několik hodin. Z toho důvodu vyvstala potřeba rozdělit test na několik částí dle potřeby. Proto je v dolní části vstupního formuláře (obr.5.) komponenta typu CheckBox. Pokud je tato zaškrtnuta dojde k pokračování již rozpracovaného testu (podrobněji popsáno dále v této kapitole).

Dále je na vstupním formuláři komponenta MainMenu, na které jsou pouze aktivační prvky Konec, nápověda, pomoc a Admin Menu. Prvními třemi zmíněnými se nebudu nijak podrobněji zabývat, protože jsou obecně známy a pro charakter tohoto nástroje nejsou až tolik podstatné. Prvku Admin Menu byla přiřazena samostatná část později v této kapitole. Poslední nezmíněná komponenta na vstupním formuláři je komponenta typu Button s nápisem „Začít test“, která slouží ke spuštění samotného testu, ale to bude také podrobněji probráno až dále v této kapitole.

Vstupni formular

Soubor Nápověda

Zvolte jazyk
Choose a language

☒ CS
☐ EN

Jméno / Name

Věk / Age

Pohlaví / Sex

☐ M ☐ Z

Hrál/a jste někdy na nějaký hudební nástroj?

☐ ano ☐ ne

Myslíte si, že máte hudební sluch?

☐ ano ☐ ne

Zadejte cestu k nahrávkám:

Procházet

Zadejte cestu kde bude uložen výstupní soubor s výsledky:

Procházet

Pokud již máte rozdělaný test a chcete v něm pokračovat, zaškrtněte prosím toto políčko - ☐

Začít test

Obr. 10: Vstupní formulář – CZ.

Vstupni_formular
File Help

Zvolte jazyk
Choose a language

☐ CS
☒ EN

Name Age Sex

☐ M ☐ Z

Do you think you have musical ear? ☐ yes ☐ no

Have you ever played any musical instrument? ☐ yes ☐ no

Choose a path to your records:

Choose a path where a file with the results of the test will be saved:

If you have worked up a test before and want to continue in it now please check this field - ☐

Obr. 11: Vstupní formulář – EN.

7.1.2. Admin Menu

Formulář Admin Menu slouží pouze pro administrátora, jak už sám název napovídá a běžný uživatel se s ním tak vůbec nedostane do kontaktu. V tomto formuláři administrátor určuje charakter testu, vlastnosti, které budou povoleny a které naopak zakázány.

Jelikož tento formulář má být nepřístupný běžnému uživateli, bylo základním požadavkem toto nějak ošetřit. V tomto případě byla jasná volba, a to chránit Admin Menu heslem. To je realizováno novým modálním formulářem (obr.7), který se objeví po kliknutí na položku Admin Menu v komponentě MainMenu. Uživatel je vyzván k zadání hesla pro vstup do administrátorského menu, což je realizováno komponentou typu Edit, která má nastavenou vlastnost PasswordChar na „*“, aby veškeré znaky, které uživatel, resp. administrátor zadává, byly zobrazeny jako znak „*“. To je zde takto nastaveno z důvodu opatření proti nechtěnému přečtení hesla další osobou. Pokud je heslo zadáno správně, objeví se potvrzovací okno s nápisem „Heslo bylo zadáno správně“, v opačném případě „Heslo nebylo zadáno správně“ a jste opět vyzváni k zadání správného hesla. Toto heslo je pevně zadáno v příkazové části programu a lze ho kdykoliv změnit. Po zadání správného hesla je formulář pro zadání hesla zavřen a je spuštěn další modální formulář, tentokrát již formulář administrátorského menu (obr.8).

Obr. 12: Formulář pro zadání hesla pro Admin Menu.

Po dohodě mělo toto menu prozatím obsahovat pouze možnost zakázat opakované přehrávání nahrávek a časem lze podle potřeby jednoduchou úpravou v kódu přidat další položky odpovídající individuálním potřebám administrátora. Možnost zakázání opakovaného přehrávání nahrávek je realizována komponentami typu CheckBox, kdy po zaškrtnutí možnosti „ne“ a zmáčknutí tlačítka „OK“ je do určité proměnné uložena informace, která je při samotném testu vyhodnocena tak, že lze každou nahrávku přehrát jen jednou a poté je tlačítko pro přehrání nahrávky deaktivováno a tím je zabráněno opakovanému přehrávání nahrávek. Opakované přehrávání nahrávek je implicitně povoleno, a tak pokud je v Admin Menu zaškrtnuta volba „ano“ nedojde k žádné změně parametrů testu, pouze je konkrétní proměnná sloužící pro tuto vlastnost nastavena do implicitního stavu.

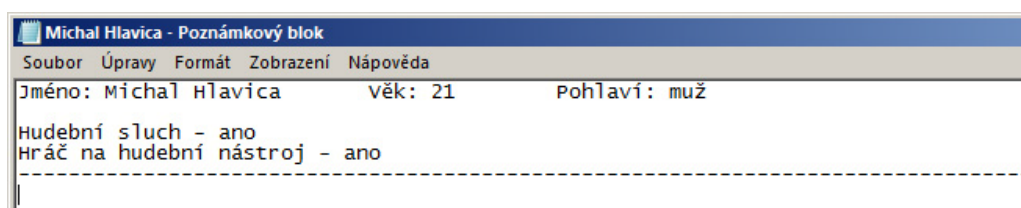
Obr. 13: Formulář Admin Menu.

Oba tyto formuláře, uvedené v této podkapitole, jsou samozřejmě zahrnuty v přepínání českého a anglického jazyka ze vstupního formuláře a jsou tudíž dvojjazyčné, jak již bylo řečeno dříve v této kapitole.

7.1.3. Příprava testu před jeho spuštěním

I když tuto část programu nemůžeme vidět v žádném z formulářů, je to velice podstatná a pravděpodobně nekomplikovanější část celého programu. Jedná se o část kódu, který je vykonán od kliknutí na tlačítko „Začít test“ až po zobrazení a spuštění samotného formuláře testu.

Jelikož v základních požadavcích bylo, aby výstupní soubor obsahoval veškeré osobní údaje vyplněné ve vstupním formuláři, tak první činnost, která se vykoná po kliknutí na tlačítko „Začít test“, je vytvoření textového výstupního souboru, který nese název jména testujícího (obr.9). To je realizováno funkcí fopen s parametrem „w“. I zápis do souboru je v českém nebo anglickém jazyce, podle toho, který byl zvolen ve vstupním formuláři.



Obr. 14: Výstupní soubor po prvním zápisu.

Jak již bylo dříve zmíněno, dalším požadavkem na tuto aplikaci bylo, aby mohl uživatel test kdykoliv přerušit a dokončit jej později. Proto je dalším krokem při spouštění testu zjištění, zda uživatel zaškrtnl komponentu typu CheckBox pro pokračování v dříve rozpracovaném testu. K vysvětlení tohoto bodu je nutno nastínit, jak proces samotného ukládání rozpracovaného testu probíhá.

Uživatel má kdykoliv možnost test ukončit dřív, než vyhodnotil veškeré nahrávky, které databáze obsahuje. Tudíž je po kliknutí na tlačítko vyhodnocení testu testována podmínka, zda byly zařazeny všechny nahrávky databáze. Pokud tomu tak není, do stejného adresáře, který byl nastaven pro uložení výstupního souboru, se kromě samotného výstupního souboru uloží také textový soubor obsahující strukturu samotného testu podle následujících pravidel:

- na prvních x řádků se zapíší fyzické cesty jednotlivých nahrávek v pořadí, v jakém byly náhodně přeházeny pro testování, kde x je celkový počet nahrávek databáze určené k testování;
- na další řádek je zapsána hodnota naposledy zařazené nahrávky před zmáčknutím tlačítka vyhodnotit;
- další řádky obsahují každý po jedné souřadnici confusion matrixu;
- jméno souboru struktury vznikne sloučením jména uživatele a textem „_struktura“;

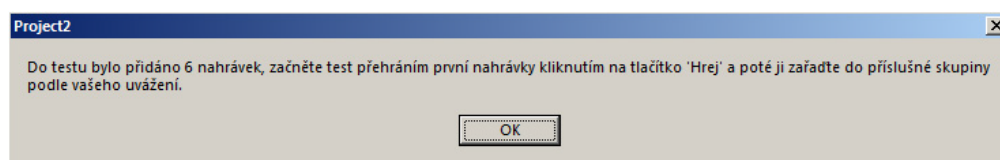
Příklad této struktury můžeme vidět na obrázku 10.

emocí více, jsou nahrány pouze do maximálního počtu 6. Další emoce již nejsou zařazeny k testování.

V dalším bodě tedy dojde, za pomoci stejných funkcí jako při hledání názvů složek, k hledání názvů jednotlivých zvukových nahrávek ve všech nalezených složkách databáze a jejich uložení do alokované struktury. Názvy jsou ukládány ve dvou tvarech a tudíž do 2 struktur. Jednak ve tvaru, ve kterém budou načítány k přehrávání, to znamená i s fyzickou cestou, kde jsou uloženy. Druhým tvarem je pouze samotný název nahrávky, který je používán pro porovnání nahrávek při vyhodnocování testu.

Jelikož je logickou podmínkou, aby byly nahrávky přehrávány při každém spuštění testu v jiném pořadí, je dalším bodem náhodné přeházení všech nahrávek, které byly načteny do struktury pro testování. To je vyřešeno prohozením dvou náhodných položek struktury a tento proces je 1000krát opakován. To je dle mého názoru dostatečná změna pořadí nahrávek pro tento účel.

Poté už je pouze, pro přiblížení se co největší uživatelské přívětivosti, na obrazovku vypsáno, kolik nahrávek bylo zařazeno k testování a jak v samotném testování postupovat (obr.11). I když samotné prostředí pro testování je, myslím si, velice intuitivní.



Obr. 16: Informační okno před spuštěním testu.

Tato podkapitola „Příprava testu před jeho spuštěním“ samozřejmě nepopisuje veškerý kód, který je před spuštěním testu vykonán, jako je všemožné zabezpečení proti pádu programu, skládání fyzických cest z názvů složek a názvů samotných souborů, ukládání různých pomocných proměnných, vypsání českých nebo anglických názvů do oken hlavního testovacího formuláře, atd. Jedná se pouze o důležité body a části programu.

7.1.4. Hlavní formulář – testování

Jedná se v podstatě o nejdůležitější formulář, o stěžejní formulář celého softwarového nástroje, protože v něm probíhá samotné testování programu. Úkony, které proběhnou při startu samotného okna, byly probrány v předchozí části této kapitoly, a tak se zde bude soustředěno pouze na samotnou funkci a princip testování a požadavky na ně.

Obecným požadavkem bylo, aby byl formulář schopen přehrávat jednotlivé nahrávky databáze a ty pak zařazovat do skupin podle uvážení uživatele. Dále pak bylo samozřejmostí, aby po zařazení jedné nahrávky, byla nabídnuta nahrávka další. Vracení se v seznamu databáze a opětovné zařazování nahrávek bylo v původním návrhu také, ale nakonec z toho bylo upuštěno.

Z toho také vyplývá základní koncept hlavního formuláře. Realizace spočívala v šesti komponentách typu Button - tlačítkách, každé pro jednu emoci (jak už bylo zmíněno dříve, aplikace je v tomto velice pružná, a tak samotné názvy emocí načítá při spuštění testu, tudíž pokud je emocí méně než 6, pak jsou nadbytečná tlačítka zrušena). Při kliknutí na jakékoliv z nich dojde k zařazení aktuální nahrávky do výsledkové struktury. Vícenásobné zařazení jednotlivých nahrávek je ošetřeno deaktivací všech tlačítek emocí po prvním kliku na některou z nich a jejich aktivace je opět provedena až po přehrání nahrávky další.

Dále jsou to tlačítka pro samotné přehrání a zastavení aktuální nahrávky („Hrej“ a „Stůj“). Pokud bylo v Admin Menu před započítáním testu zakázáno opakované přehrávání nahrávek, pak i tlačítko pro přehrání nahrávky je po jednom přehrání deaktivováno až do doby než dojde k načtení další nahrávky. Toto tlačítko („Hrej“) je navázáno na funkci PlaySound, kde jejím parametrem je proměnná, do které se při každém stisku tlačítka nahraje cesta k aktuální nahrávce. Tlačítko „Stop“ je stejně tak vyřešeno funkcí PlaySound, tentokrát ovšem s nulovým parametrem, což není po programátorské stránce ideální řešení, nicméně to na funkčnosti nijak neubírá.

Dalšími tlačítky v základním konceptu jsou tlačítka pro posunutí na předchozí nebo následující nahrávku databáze. Jak už bylo řečeno, z posouvání na předchozí nahrávky seznamu bylo upuštěno, a tak je tlačítko „Předchozí nahrávka“ ponecháno po celou dobu testu deaktivováno.

Posledním nezmíněným tlačítkem je tlačítko pro vyhodnocení testu. To může být zmáčknuto téměř kdykoli v průběhu testu. Téměř kdykoliv znamená, že pokud již uživatel přehrál určitou nahrávku, pak ji musí před vyhodnocením testu zařadit. Test je pak po kliknutí na toto tlačítko vyhodnocen a ukončen. To je však podrobněji popsáno dále v této kapitole, proto to zde nebude více rozebíráno.

Dalším požadavkem bylo, aby uživatel viděl, kolik času již uběhlo od začátku testování, a tak byla do pravého horního rohu formuláře umístěna komponenta, která zobrazuje uplynutý čas od spuštění testu. Toto je realizováno pomocí časovače, který vysílá impuls každou sekundu. Na tento impuls je navázána jednoduchá souslednost několika příkazů přiřazení a výsledné hodnoty jsou zobrazovány na hlavním formuláři ve formě času.

Jak již bylo řečeno dříve, od aplikace se požaduje co největší uživatelská přívětivost a schopnost uživateli co nejvíce usnadnit práci s tímto nástrojem. Tuto snahu zde představují 2 komponenty typu Edit, které zobrazují aktuální počet nahrávek zbývajících v tomto testu a počet nahrávek již zařazených. Další komponenta typu Edit nás po zařazení nahrávky informuje, do jaké skupiny jsme nahrávku zařadili a jak pokračovat dále v testu pro jeho úspěšné dokončení. Posledním bodem tohoto požadavku je aktivování a deaktivování jednotlivých tlačítek podle nastalé situace a tím nasměrování uživatele na správnou cestu. Výsledný návrh vycházející z předešlých požadavků a podmínek můžeme vidět na obrázku 12.

Obr. 17: Hlavní formulář – CZ.

Tento formulář je samozřejmě také možno spustit v anglické verzi (obr.13), a to na základě jazyka zvoleného ve vstupním formuláři. Názvy emocí jsou zde pochopitelně stejné jako ve formuláři českém, protože jsou tyto načítány ze složek databáze, jak již bylo zmíněno dříve.

Obr. 18: Hlavní formulář – EN.

Domnívám se, že tento formulář splňuje jak požadavky technického a funkčního rázu, tak požadavky na uživatelskou přívětivost. Okno je navrženo v jednoduchém stylu a spolu s komponentami typu Edit, které navigují jak pokračovat v testu a dále také spolu s vhodnou aktivací a deaktivací jednotlivých prvků je tento softwarový nástroj určen téměř pro každého, nehledě na jeho počítačovou gramotnost.

7.1.5. Vyhodnocování testu

Nyní přistoupíme k popsání metod, které jsou použity pro vyhodnocování testu, ke způsobu, jakým jsou získány výstupní hodnoty. Nejedná se však pouze o část kódu, která je vykonána po kliknutí na tlačítko „Vyhodnoť“, ale i o vyhodnocování testu, které probíhá již při samotném testování.

Vyhodnocování testu se odvíjí od požadavků, které byly kladeny na charakter výstupního souboru. Tento však bude detailněji popsán až v další kapitole, nyníjší kapitola bude zaměřena pouze na proces vyhodnocování. Hlavním požadavkem, který vychází z principu celého nástroje, bylo zjistit, v jaké emoční skupině se nahrávka původně nacházela (skupina, do které byla nahrávka zařazena osobou nebo nástrojem vyhodnocujícím danou databázi) a do jaké emoční skupiny byla nahrávka zařazena uživatelem, dle jeho subjektivního posouzení. To je vyřešeno následujícím způsobem. Princip vychází z toho, že všechny názvy každé emoční skupiny jsou na začátku nahrány do struktury odpovídající dané emoci. Tyto názvy jsou pak seskupeny do jedné struktury, která je náhodně zamíchána, aby pořadí přehrávání bylo pokaždé jiné. Takže máme po startu testu 6 struktur, kde každá obsahuje názvy nahrávek podle emoční skupiny a dále 1 strukturu, která obsahuje všechny názvy databáze nehledě na jednotlivé emoční skupiny.

V průběhu testu tedy uživatel přehrává nahrávky z jedné struktury společné všem emočním skupinám a při zařazení aktuální nahrávky, tedy po stisku tlačítka náležící zvolené emoci, se vykoná část kódu, který porovnává název aktuální nahrávky se všemi názvy ve všech emočních skupinách, dokud nenajde shodu. Až najde shodu, tak emoční skupina, ve které byla tato shoda nalezena, je původní emoční skupinou a emoční skupina, do které byla nahrávka zařazena, koresponduje s tlačítkem emoce, které bylo právě zmáčknuto.

Nástroj také vyžaduje, aby jeho výstupem byl tzv. „confusion matrix“. Jedná se o matici, tabulku, která posuzuje úspěšnost algoritmu, v našem případě míru shody původního zařazení nahrávek a zařazení výsledného, které odpovídá subjektivnímu posouzení uživatele. Zjednodušeně lze říci, že nám matice záměn, jak se také „confusion matrix“ nazývá, vypovídá o subjektivní úspěšnosti rozpoznání nahrávek, resp. úspěšnosti původního zařazení osobou nebo nástrojem, který databázi vytvářel. Konkrétně můžeme tuto úspěšnost (číselnou hodnotu shodných nahrávek původního a současného zařazení každé emoce) vidět na hlavní diagonále matice. Příklad matice záměn můžeme vidět na obr. 14.

The diagram illustrates a confusion matrix for emotion recognition. It features a 6x6 grid with rows and columns labeled: smutek, radost, vztek, ironie, překvapení, and strach. The main diagonal, representing correct classifications, is highlighted with a line and labeled 'hlavní diagonála - úspěšnost rozpoznání'. The original assignments are circled on the left and labeled 'původní zařazení', while the resulting assignments are circled on the right and labeled 'výsledné zařazení'.

	smutek	radost	vztek	ironie	překvapení	strach
smutek	76	6	2	12	2	1
radost	18	65	21	32	19	9
vztek	15	2	72	12	12	8
ironie	6	3	12	83	11	14
překvapení	5	15	15	2	52	3
strach	13	12	4	2	10	69

Obr. 19: Příklad „Confusion matrixu“.

Jednotlivé koeficienty této matice jsou navyšovány při každém zařazení nahrávky - každém kliknutí na tlačítko určité emoce. Při vyhodnocování jsou tedy koeficienty pouze poskládány do tvaru odpovídajícímu této matici. Dále je spočítána procentuální úspěšnost rozpoznání všech emocí jako podíl součtu všech nahrávek ku součtu nahrávek nacházejících se na hlavní diagonále vynásobený hodnotou 100.

V průběhu vytváření nástroje se ukázalo, že by bylo vhodné, aby u každé nahrávky figuroval i čas potřebný k jejímu zařazení. V praxi to znamená, že při přehrání nahrávky (kliknutí na tlačítko „Hrej“) se uloží hodnota aktuálního času do příslušné proměnné a při zařazení nahrávky (kliknutí na tlačítko zvolené emoce) je znovu uložena hodnota času, tentokrát však do jiné pomocné proměnné. Ihned po zařazení se vykoná kód, který za pomoci několika podmínek odečte hodnoty uložených časů a hodnotu ve formě času je doba potřebná pro zařazení nahrávky.

Stejně tak bylo řešeno i zjištění celkového času, potřebného pro dokončení celého testu. Jedinou změnou zde bylo, že první čas byl zaznamenán při spuštění testu a čas druhý při kliknutí na tlačítko „Vyhodnoť“.

Dalším bodem, který by měl být zmíněn v této kapitole je předčasné vyhodnocení testu, tomu je však věnován dostatek prostoru v předchozí podkapitole, a tak je zde tento bod vynechán. Dále je tomu taky z důvodu, že vyhodnocení rozpracovaného testu probíhá úplně stejně jako vyhodnocení testu až na jeho konci.

7.1.6. Výstupní soubor

Jak už bylo dříve v této kapitole řečeno, vytváření výstupního souboru je úzce spjato s procesem vyhodnocení testu. Výstupní soubor obsahuje veškerá data, která jsou získána při vyhodnocování a mimo to také osobní informace, které uživatel zadal ve vstupním formuláři tohoto nástroje.

Požadavky na výstupní soubor a jejich praktická realizace jsou podrobně popsány v předchozí podkapitole s názvem Vyhodnocování testu. Tato podkapitola se proto bude zabývat spíše vizuální stránkou výstupního souboru a také posloupností dat, kterou tento soubor obsahuje.

Výstupní soubor má nést název uživatele, který daný test vyhodnocuje. To je zabezpečeno jednoduchým předáním obsahu komponenty typu Edit do příslušné proměnné za pomoci funkce strcpy.

Prvním blokem výstupního souboru měl podle požadavků být blok s osobními informacemi, které uživatel poskytne ve vstupním formuláři. Těmi jsou jméno uživatele, jeho věk a pohlaví, dále pak informace o tom, zda uživatel hrál na nějaký hudební nástroj a o tom, jestli má hudební sluch. Tyto informace jsou tedy do souboru zapsány jako první, a to hned při stisknutí tlačítka „Začni test“. Při tomto zápisu je tedy výstupní soubor vytvořen a při jakémkoliv dalším zápisu už je otevřen pouze pro čtení. Pro vytvoření textového souboru a zápis do něj je použita funkce fopen s parametrem „w“. Pro přehlednost jsou jednotlivé tematické bloky výstupního souboru odděleny řádkem znaků „-“.

Hned za blokem s osobními informacemi, bylo požadováno, aby se nacházel blok s informacemi o jednotlivých nahrávkách, které byly zařazeny. Už bylo dříve zmíněno, že tento blok je vyhodnocován v průběhu testu, a to pokaždé při kliknutí na zvolenou emoční skupinu. Tedy po stisku jednoho z těchto náležíčích tlačítek proběhne kód, který zapíše výstupní informace pro danou nahrávku do výstupního souboru. Výstupní soubor není otevírán při každém stisku tlačítka, ale je otevřen pouze jednou při vytvoření formuláře, a to opět pomocí funkce fopen, tentokrát však s parametrem „a+“, což znamená, že soubor je otevřen pro úpravy, na konci souboru. Příklad struktury informací jedné nahrávky i s popisky můžeme vidět na obrázku 15.

Struktura má 5 částí, kde v první části je zapsáno pořadové číslo nahrávky, v části druhé je název nahrávky, přečtený ze složky databáze. Další dvě části, třetí a čtvrtá, slouží k označení skupin původního a současného zařazení nahrávky. Poslední část obsahuje informaci o době potřebné pro zařazení nahrávky od jejího poslechnutí. Struktura je vytvářena formou konzistentní tabulky, takže jednotlivé názvy jsou přesně pod sebou a nedochází k posunutí nahrávek vlivem délky jednotlivých názvů nebo vlivem jiných aspektů. Šířka každého pole je určena podle nejdelšího slova v daném sloupci a ještě je k němu přidána rezerva několika znaků, takže nikdy nemůže dojít k překrývání slov. I počet nahrávek může být až třímístný a struktura stejně nebude porušena. To je vyřešeno posloupností několika příkazů, kdy pořadové číslo nahrávky dělíme několikrát deseti a zjišťujeme, po kolikátém dělení je menší než 1. Podle toho je určen počet mezer nutný k doplnění tak, aby nebyla porušena konzistence struktury tabulky.

pořadí	název	původní	rozpoznáno	čas na určení
1	prekvapeni1.wav	prekvapeni	smutek	0:0:1

Obr. 20: Příklad struktury informací jedné nahrávky ve výstupním souboru.

Dalším blokem, tedy v pořadí již třetím, který se měl podle návrhu nacházet ve výstupním souboru, je blok dat s „confusion matrix“. Co je „confusion matrix“ nebo také matice záměn, což je český název pro tento výraz, bylo popsáno dříve v této kapitole. Matice záměn je do výstupního souboru zapsána až po stisknutí tlačítka „Vyhodnoť“. Je opět otevřen soubor pro úpravy, což v jazyce C++ znamená funkci `fopen` a parametr „a“. Potom je opět zjištěn nejdelší prvek celé výsledné matice, aby tato mohla být navržena tak, aby nedošlo k překrývání jednotlivých sloupců matice. Potom je dopočítán a doplněn, jednoduchou posloupností, počet mezer k zajištění stejné konzistence v celé tabulce a jednotlivé prvky matice jsou jeden po druhém zapisovány do výstupního souboru. Koeficienty matice jsou načítány z příslušných proměnných, které byly zvyšovány po celou dobu testu, a to při každém kliknutí na tlačítko emoční skupiny (zařazení nahrávky). Matice záměn je dimenzována na 6 skupin, pokud je v testu skupin méně, pak jsou bohužel vypsány i koeficienty pro neexistující skupiny, ty jsou však nulové, a tak to na kvalitě vyhodnocení nic neubírá. Více skupin než 6 být v testu nemůže, jak již bylo zmíněno na začátku této kapitoly.

Samozřejmě součástí bloku „confusion matrix“ musí být i úspěšnost rozpoznání nahrávek (blíže popsáno v podkapitole „Vyhodnocování testu“), která se nachází na hlavní diagonále matice záměn. Tato informace je tedy zapsána do výstupního souboru hned za touto maticí.

Poslední informací, která je zapsána do výstupního souboru, je informace a celkové době potřebné pro celé testování. Ta se tedy nachází úplně na konci výstupního souboru.

Tím byly probírány veškeré bloky a informace, které jsou zapisovány do výstupního souboru. Ukázku, jak by mohl tento výstupní soubor vypadat, můžete vidět na obrázku 16.

Michal Hlavica - Poznámkový blok

Soubor Úpravy Formát Zobrazení Nápověda

Jméno: Michal Hlavica Věk: 21 Pohlaví: muž

Hudební sluch - ano
Hráč na hudební nástroj - ano

pořadí	název	původní	rozpoznáno	čas na určení
1	radost1.wav	radost	smutek	0:0:1
2	ironie1.wav	ironie	prekvapení	0:0:2
3	strach2.wav	strach	vztek	0:0:3
4	ironie2.wav	ironie	smutek	0:0:0
5	strach1.wav	strach	radost	0:0:1
6	prekvapení2.wav	prekvapení	ironie	0:0:13
7	smutek2.wav	smutek	strach	0:0:0
8	vztek1.wav	vztek	radost	0:0:1
9	prekvapení1.wav	prekvapení	vztek	0:0:0
10	radost2.wav	radost	smutek	0:0:1
11	smutek1.wav	smutek	prekvapení	0:0:0
12	vztek2.wav	vztek	prekvapení	0:0:1

	ironie	prekvapení	radost	smutek	strach	vztek
ironie	0	1	0	1	0	0
prekvapení	1	0	0	0	0	1
radost	0	0	0	2	0	0
smutek	0	1	0	0	1	0
strach	0	0	1	0	0	1
vztek	0	1	1	0	0	0

Úspěšnost rozpoznání byla: 0.000%.

celkový čas testu byl: 0:00:37.

Obr. 21: Ukázka výstupního souboru.

Avšak výstupní soubor může docela změnit svoji podobu, pokud byl test jednou nebo vícekrát přerušen a k jeho dokončení došlo až později. Problematika rozdělení testu na více jeho částí byla již probrána dříve v této kapitole, takže tato stránka věci, zde nebude nijak více přiblížena.

Pokud k ukončení testu došlo dříve, než byly zařazeny veškeré nahrávky, které byly vybrány pro toto testování, pak je vytvořen výstupní soubor – struktura testu. Tento soubor nese veškeré informace jak pro úspěšné pokračování v testu, ale také veškeré informace důležité pro finální výstupní soubor. Při předčasném ukončení tedy dojde k vypsání všech bloků výstupního souboru v pořadí, v jakém bylo zmíněno dříve v této podkapitole. Avšak při úspěšném navázání testu, jsou na konec souboru připsány další data, jak bylo v požadavcích na výstupní soubor a rozdělení testu na více částí.

Pokud tedy dojde k požadavku pro pokračování v testu, pak je ihned po stisknutí tlačítka „Začít test“ ve vstupním formuláři zapsáno do výstupního souboru označení, že se jedná o pokračování v testu. K tomu nám opět poslouží funkce `fopen` s parametrem „a“.

Dále probíhá zápis do souboru klasickým způsobem, jak je to popsáno výše. Pořadové číslo nahrávek je samozřejmě navázáno na předchozí test, takže se nepočítají nahrávky opět od čísla 1. Jak vyplývá z textu v podkapitole „Příprava testu před jeho spuštěním“, matice záměn, která je zapsána do výstupního souboru po pokračování (každý další zápis po prvním) obsahuje koeficienty odpovídající všem předchozím průběhům testů, protože při navázání testu jsou tyto koeficienty navýšeny o koeficienty z původního testu.

Tedy výsledný soubor po zařazení všech nahrávek databáze, kdy byl test rozdělen na více částí, obsahuje několik matic záměn, kde každá matice vypovídá o průběhu celého dosud provedeného testu. Takže první matice nese informace o průběhu prvního testování. Matice druhá o průběhu prvního a druhého testování, atd. Matice celková, tedy ta, která nese informace o celém testu, se nachází úplně na konci výstupního souboru.

Také úspěšnost rozpoznání nahrávek se počítá z aktuální matice, takže každá tato informace nacházející se pod maticí záměn, odpovídá této matici.

Jednotlivé časy testů avšak odpovídají časům každého testování zvlášť a nekumulují se, jak je tomu u matic záměn.

Ukázku testu, rozděleného na 3 části můžeme vidět na obrázku 17.

Michal Hlavica - Poznámkový blok						
Soubor Úpravy Formát Zobrazení Nápověda						
Jméno: Michal Hlavica věk: 21 Pohlaví: muž						
Hudební sluch - ano						
Hráč na hudební nástroj - ano						
pořadí	název	původní	rozpoznáno		čas na určení	
1	strach1.wav	strach	smutek		0:0:2	
2	vztek2.wav	vztek	radost		0:0:1	
	ironie	prekvapení	radost	smutek	strach	vztek
ironie	0	0	0	0	0	0
prekvapení	0	0	0	0	0	0
radost	0	0	0	0	0	0
smutek	0	0	0	0	0	0
strach	0	0	0	1	0	0
vztek	0	0	1	0	0	0
Úspěšnost rozpoznání byla: 0.000%.						
Celkový čas testu byl: 0:00:08.						
Pokračování v testu:						
pořadí	název	původní	rozpoznáno		čas na určení	
3	smutek1.wav	smutek	smutek		0:0:1	
4	strach2.wav	strach	prekvapení		0:0:1	
5	ironie2.wav	ironie	radost		0:0:0	
6	vztek1.wav	vztek	vztek		0:0:1	
7	strach1.wav	strach	smutek		0:0:1	
8	radost2.wav	radost	radost		0:0:1	
9	ironie1.wav	ironie	smutek		0:0:0	
	ironie	prekvapení	radost	smutek	strach	vztek
ironie	0	0	1	1	0	0
prekvapení	0	0	0	0	0	0
radost	0	0	1	0	0	0
smutek	0	0	0	1	0	0
strach	0	1	0	2	0	0
vztek	0	0	1	0	0	1
Úspěšnost rozpoznání byla: 25.000%.						
Celkový čas testu byl: 0:00:15.						
Pokračování v testu:						
pořadí	název	původní	rozpoznáno		čas na určení	
10	vztek2.wav	vztek	smutek		0:0:1	
11	vztek1.wav	vztek	radost		0:0:0	
12	ironie1.wav	ironie	prekvapení		0:0:0	
	ironie	prekvapení	radost	smutek	strach	vztek
ironie	0	1	1	1	0	0
prekvapení	0	0	0	0	0	0
radost	0	0	1	0	0	0
smutek	0	0	0	1	0	0
strach	0	1	0	2	0	0
vztek	0	0	2	1	0	1
Úspěšnost rozpoznání byla: 25.000%.						
Celkový čas testu byl: 0:00:07.						

Obr. 22: Ukázka výstupního souboru testu rozděleného na 3 části.

7.2. Shrnutí

Tímto byl rozebrán celý softwarový nástroj postupně po jeho logických celcích. Nástroj byl popsán z funkčního hlediska, z hlediska obecných principů. Konkrétní posloupnost příkazů zde nebyla ukázána, protože z hlediska funkce programu se jedná o nepodstatnou část. Pro

konkrétní řešení, zde rozebraných celků, je celý kód k nahlédnutí na přiloženém DVD, kde lze nalézt také komentáře k většině problémů, kterým bylo čeleno.

Domnívám se, že nástroj splnil veškeré požadavky, které na něj byly kladeny. Dokáže spolehlivě provádět testy a dá se použít i na jinou databázi než je pouze tahle, což je jeden z jeho velkých kladů. Vůbec největší předností tohoto nástroje, je dle mého názoru jeho uživatelská přívětivost. K provedení tohoto testu je za potřebí minimální uživatelské znalosti z oblasti počítačů, z oblasti databází nebo teorie emocí nejsou dokonce znalosti potřeba žádné.

8. Závěr:

Účelem první části této práce bylo nahlédnout do teorie emocí, pochopit jak emoce vznikají a jak působí na lidský organismus. Dále také popsat problematiku tvoření databází zvukových nahrávek, protože i tato je součástí této práce. Pro obecný přehled byla přiložena i kolekce významných databází, které už byly vytvořeny. Dále také bylo započato tvoření databáze zvukových nahrávek. Po zvážení problematiky tvorby databází a prostudování dostupných materiálů již vytvořených databází bylo rozhodnuto, že nahrávky budou získávány z televizních pořadů a seriálů. Důvodem je vyšší autentičnost nahrávek, velký rozsah jednotlivých účastníků a vůbec celého materiálu, který je pro tuto metodu dostupný. Všechny tyto informace sloužili jako kvalitní teoretický podklad pro pokračování v tomto projektu bakalářskou prací.

Dále pak bylo započato vytváření softwarového nástroje s uživatelským rozhraním, který bude sloužit k subjektivnímu hodnocení námi vytvořené databáze. Nástroj lze ale také použít na hodnocení jiných databází, protože je naprogramován tak, aby dokázal při startu testu dynamicky načíst emoční skupiny dané databáze. Charakter testu lze měnit pomocí administrátorského menu, které je také součástí nástroje, a je chráněno heslem, aby k němu běžný uživatel neměl přístup. Jelikož byla jedna ze zásadních otázek, zda nástroj udělat v českém nebo anglickém jazyce, přistoupilo se k realizaci v obou těchto jazycích. Uživatel má tedy možnost zvolit si jazyk podle potřeby. Veškeré další formuláře nacházející se v programu a také výstupní soubor s výsledky testu se pak od toho jazyka odvíjí. I dalšími prvky, které jsou do testu zakomponovány, bylo dosaženo velké uživatelské přívětivosti. Program je velice intuitivní a uživatel je po celou dobu běhu programu naváděn jak postupovat při jednotlivých krocích testu. To je také jedna z největších předností celého nástroje. Uživatel, který bude používat tento nástroj, nepotřebuje žádné odborné znalosti, dokonce znalosti z oblasti počítačů nutné k obsluze tohoto programu jsou na velice nízké úrovni. Nástroj je k uživateli také velice přátelský tím, že je možnost test kdykoliv přerušit a dokončit jej později. To je z důvodu, že některé databáze zvukových nahrávek jsou velice obsáhlé a můžou dosahovat až několika stovek nahrávek. Proto jejich subjektivní testování může zabrat až několik hodin.

Výstupní textový soubor obsahuje osobní informace uživatele, ale hlavně výsledky testování. Druhá zmíněná část – výsledky testování, zahrnují informace o každé zařazené nahrávce (pořadové číslo, původní a zvolené zařazení a čas potřebný k zařazení), dále pak „confusion matrix“ a některé další doplňující informace k testu.

Lze říci, že nástroj byl naprogramován podle všech požadovaných nároků a vhodnými metodami realizuje konkrétní požadavky, které na něj byly kladeny. Nástroj tedy může sloužit k hodnocení jakékoliv emoční databáze zvukových nahrávek.

Jelikož je tato práce součástí většího vědeckého projektu, kde bude také sloučeno více emočních databází, samotný proces hodnocení nahrávek proběhne až po dokončení celého projektu.

Součástí bylo i provedení analýzy ukázkových nahrávek pro každou emoci. Ta byla provedena programem Praat. Konkrétně se analýza zabývá základní frekvencí, intenzitou a prvními třemi formanty nahrávek.

Dále bylo bakalářskou prací pokračováno v rozšiřování databáze zvukových nahrávek. Ve své konečné podobě obsahuje databáze přes 200 nahrávek.

Použitá literatura

- [1] *Laughing Out Loud To Good Health* [online]. [2008] [cit. 2008-11-25]. Dostupný z WWW: <<http://library.thinkquest.org/25500/>>.
- [2] ESPOSITO, Anna. *The Amount of Information on Emotional States Conveyed by the Verbal and Nonverbal Channels: Some Perceptual Data*. [s.l.] : [s.n.], 2007. s. 249-265.
- [3] VLČKOVÁ - MEJVALDOVÁ, Jana. *Prozodie, cesta i mříž porozumění - Experimentální srovnání příznakové prozodie různých jazyků*. [s.l.] : [s.n.], c2006. 203 s.
- [4] *Anamnéza* [online]. [2008] [cit. 2008-12-02]. Dostupný z WWW: <http://www.anamneza.cz/moduly/lidsketelo.php?soubor=kapitola4_8.php>.
- [5] *Wikipedia* [online]. [2008] [cit. 2008-12-04]. Dostupný z WWW: <<http://cs.wikipedia.org>>.
- [6] Zpracování češtiny v soudobé hudbě I. . *A tempo revue* [online]. 2008 [cit. 2008-12-05].
- [7] TSANG-LONG, Pao, et al. *The Construction and Testing of a Mandarin Emotional Speech Database*. [s.l.] : [s.n.], [2008]. 9 s.
- [8] *The HUMAINE Portal* [online]. c2008 [cit. 2008-12-10]. Dostupný z WWW: <<http://emotion-research.net/wiki/Databases>>.
- [9] DOUGLAS-COWIE, Ellen, et al. Emotional speech: Towards a new generation of databases. *Speech Communication*. 2003, no. 40, s. 33-60.

Seznam použitých symbolů a zkratek

F_0	základní frekvence
WOZ	Wizard of Oz – speciální metoda nahrávek
F_{0m}	střední hodnota základní frekvence
I_m	střední hodnota intenzity
F_{1m}	střední frekvence prvního formantu
F_{2m}	střední frekvence druhého formantu
F_{3m}	střední frekvence třetího formantu