

**Fakulta elektrotechniky a informatiky
Univerzita Pardubice**

**Doktorský studijní program: P2612 Elektrotechnika a informatika
obor Informační, komunikační a řídicí technologie**

Název tématu disertační práce:
**Analýza biomedicínských signálů pro diagnostiku
poruch hlasu**

Autor: Ing. Milan Jičínský

Školitel: doc. Ing. Jan Mareš, Ph.D.

Školitel specialista: Ing. Libor Kupka, Ph.D., Mgr. Jaroslav Marek, Ph.D.

Pardubice, 20.07.2022

Prohlášení autora

Prohlašuji, že jsem tuto práci vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích 20. 7. 2022

Ing. Milan Jičínský

Poděkování

V první řadě bych rád poděkoval vedoucímu práce doc. Ing. Janu Marešovi, Ph.D. nejen za vstřícný přístup, užitečné rady a trpělivost. Za možnost zpracovávat právě toto jedinečné téma vděčím též svému vedoucímu, který mi poskytl potřebné kontakty na kolegy z FNKV a pomáhal organizovat naše vzájemná setkání. Za FNKV mé díky patří v první řadě doc. MUDr. Martinu Chovancovi, Ph.D., jenž zastřešoval veškerou aktivitu na straně nemocnice. Rovněž si vážím spolupráce s týmem odborníků z FNKV, kteří poskytli nezbytnou zpětnou vazbu při vývoji diagnostického SW a byli ochotni nejen sbírat nahrávky pacientů, ale rovněž přispěli i svými nahrávkami, abychom měli k dispozici i záznamy zdravých osob. Děkuji též vedoucímu katedry Ing. Danielu Honcovi, Ph.D. za pořízení nákladného HW a přátelský a vstřícný přístup během celého studia. Nemohu zapomenout ani na Ing. Libora Kupku, Ph.D., který mi umožnil ve spolupráci s TUL studovat základy zpracování řeči, bez kterých by se tato práce neobešla. Poděkování patří samozřejmě i mé rodině. Všichni mne po celou dobu studia podporovali a stáli při mně. Tuto práci bych chtěl věnovat svojí nejmilejší babičce, kterou mám nade všechno rád. Vždy ve mě věřila a podporovala mě. A za to jí patří velké díky.

Během dlouhých 9 let života v Pardubicích jsem poznal množství úžasných lidí a vytvořila se přátelství na celý život. Místem, které všechny spojovalo se stal pokoj A019, kterému nikdo neřekl jinak než Strojovna Výtahu. O legendární akce nebylo nouze, a tak se zvěsti o Strojovně a jejích členech (já, Čerňas, Raďoš a později Satanda a Suso) rychle šířili a naše parta se stále rozrůstala. Strojovna výtahu se tak navždy stala pojmem v historii Pardubických kolejí. Poslední akce s příhodným názvem The last chlast, tak v roce 2019 ukončila jednu éru života. Ale i když je studentský čas už nenávratně pryč, stále se setkáváme na oslavách, koncertech, festivalech, pijatykách a dalších akcích. Není to totiž o místě, ale o lidech. Letos dokonce poprvé vyrážíme po celé republice navštěvovat naše přátele v rámci Strojovna on tour. A to vše díky vám přátelé. Bez všech těch báječných lidí, které jsem během studia poznal, by to nešlo. A za to vám z celého srdce děkuji.

Na závěr bych chtěl poděkovat Márdimu a Vypsané fiXe, jejichž hudbu jsem si v San Piegu zamiloval, provázela mě celým studiem a po zbytek života mi bude připomínat ty skvělé roky, které jsem v Pardubicích strávil. Je to hudba, která lidi spojuje a vybízí k dalším a dalším setkáním s mými přáteli.

Obsah

Seznam zkratk	8
Seznam obrázků	13
Seznam tabulek	16
Úvod	17
1 Hlas a jeho tvorba	18
1.1 Tvorba lidského hlasu	19
2 Poruchy hlasu	22
2.1 Organické poruchy hlasu	23
2.1.1 Zánětlivá onemocnění	24
2.1.2 Vývojové vady hrtanu	24
2.1.3 Traumatická poškození hrtanu	24
2.1.4 Benigní hrtanové léze	26
2.1.5 Maligní hrtanové léze	28
2.2 Funkční poruchy hlasu	28
2.2.1 Poruchy z přemáhání hlasu	28
2.2.2 Psychogenní poruchy	29
2.2.3 Hlasové neurózy	30
2.3 Metody pro diagnostikování poruch hlasu	30
2.3.1 Hlasové pole	30
2.3.2 Spektrální analýza	32
2.3.3 Multidimenzionální analýza	32
2.3.4 Percepční metody	33
2.3.5 Voice Handicap Index	34
2.3.6 Laryngoskopie a laryngostroboskopie	34
2.3.7 Videokymografie a videolaryngoskopie	35
2.3.8 Ostatní metody	36

2.4	Léčba poruch hlasu.....	37
2.4.1	Hlasová hygiena.....	38
2.4.2	Hlasová reedukace	38
3	Nemoci štítné žlázy a jejich vliv na tvorbu hlasu	39
3.1	Anatomie a funkce štítné žlázy	39
3.2	Nemoci štítné žlázy (tyreopatie)	42
3.2.1	Hypotyreóza.....	44
3.2.2	Hypertyreóza (tyreotoxikóza).....	44
3.2.3	Struma, uzlíky a nádory štítné žlázy	45
3.3	Léčba onemocnění štítné žlázy	46
3.3.1	Operace štítné žlázy	46
3.3.2	Aplikace radioaktivního jódu.....	48
4	Cíle práce	48
5	Zpracování řeči	49
5.1	Obecná metodika zpracování řečového signálu.....	50
5.1.1	Předzpracování.....	50
5.1.2	Segmentace	51
5.1.1	Parametrizace.....	53
5.1.2	Klasifikace	54
5.2	Metody a postupy uplatňované při zpracování řeči a extrakci příznaků.....	55
5.2.1	Frekvenční spektrum	55
5.2.2	Kepstrální analýza.....	60
5.2.3	Aplikace okna	62
5.3	Extrakce příznaků z řeči	64
5.3.1	Energie hlasu a šumu pozadí	65
5.3.2	Střední krátkodobá intenzita	66
5.3.3	Počet průchodů nulou.....	67

5.3.4	Spektrální energie.....	68
5.3.5	Základní tón (frekvence)	69
5.3.6	Základní akustické veličiny	73
5.3.7	Mel-frekvenčních keprální koeficienty.....	76
5.3.8	Rychlost řeči	78
5.4	Měření hladiny akustického tlaku (SPL).....	79
5.4.1	Časové vážení a průměrování	80
5.4.2	Filtrace – váhové frekvenční filtry (A, B, C, D a Z).....	81
5.5	Určení podobnosti nahrávek	83
5.5.1	Princip borcení časové osy	83
5.5.2	Lineární časová transformace	84
5.5.3	Nelineární časová transformace – dynamické borcení časové osy.....	84
6	Zařízení pro záznam hlasu	86
6.1	Základní charakteristiky mikrofonu.....	87
6.1.1	Dělení mikrofonů	87
6.1.2	Frekvenční rozsah a frekvenční odezva	89
6.1.3	Směrová charakteristika	90
6.1.4	Ostatní parametry	92
6.2	Doporučení pro volbu mikrofonu	93
6.3	Kalibrace mikrofonu	95
6.3.1	Dvoustupňová (nepřímá) kalibrace	95
7	Metodika použitá pro sběr diagnostických dat	97
7.1	Měřicí řetězec.....	97
7.1.1	Měřicí mikrofon.....	97
7.1.2	Zvukoměr.....	99
7.1.3	Kalibrátor.....	99
7.1.4	Externí zvuková karta.....	100

7.1.5 Zapojení a nastavení přístrojů.....	101
7.2 Sběr dat.....	102
7.2.1 Rozbor metodiky.....	103
8 Aplikace Voice disorder diagnostician.....	104
8.1 Režimy pro sběr a analýzu dat.....	106
8.1.1 Zadávání údajů o novém pacientovi.....	107
8.1.2 Import dat z databáze pacientů	110
8.1.3 Databáze pacientů	113
8.1.4 Režim nahrávání	114
8.1.5 Kalibrace.....	117
8.1.6 Analýza zaznamenaných dat.....	118
8.1.7 Náповěda	120
8.2 Diagnostika	120
8.2.1 Metodika pro objektivní posouzení změn hlasu	120
8.2.2 Určení reprezentativních příznaků na základě srovnání DTW skóre.....	121
8.2.3 Diagnostický režim aplikace VDD	131
8.3 Možné směry dalšího rozvoje tématu.....	134
Závěr.....	135
Literatura.....	137
Seznam publikací autora.....	147

Použité symboly

$A(i, j)$	– akumulovaná vzdálenost v bodě $[i, j]$,
α	– koeficient filtru
B_k	– spektrální energie k -tého pásma
$c(n)$	– reálné kepstrum řečového segmentu
$c_g(n)$	– kepstrum artikulačního ústrojí (filtru)
c_L	– centrální klipování
c_m	– kalibrační konstantu mikrofonu
$c_{mf}(n)$	– n -tý Mel-frekvenční kepstrální koeficient
$c_x(n)$	– kepstrum buzení (zdroje)
c_z	– kalibrační konstanta zvukoměru
$d(x_i, r_{w(i)})$	– lokální vzdálenost mezi příznakovými vektory x a r v bodě $[i, w(i)]$
$D(\mathbf{X}, \mathbf{R})$	– nejmenší globální vzdálenost mezi \mathbf{X} a \mathbf{R}
Δp_i	– příznakový vektor i -tého dynamického (delta) příznaku
$\Delta \Delta p_i$	– příznakový vektor i -tého akceleračního (delta-delta) příznaku
E	– krátkodobá energie
E_S	– střední krátkodobá energie
$F(i\omega)$	– Fourierův obraz
$F(\omega)$	– Fourierovo (komplexní) spektrum
$ F(\omega) $	– amplitudové spektrum
$\arg F(\omega)$	– fázové spektrum
$F(k)$	– Fourierův obraz při použití N -bodové DFT
$f(n)$	– původní diskrétní signál před transformací (diskrétní originál)
$f(t)$	– spojitý signál před transformací (originál)
f_{Hz}	– frekvence udávaná v Hz

f_{Mel}	– frekvence udávaná v Mel
f_{vz}	– vzorkovací frekvence původního signálu
$g(n)$	– impulsní charakteristiku filtru
$H(z)$	– přenosová funkce filtru
i_k	– index první hodnoty příslušící ke k-tému frekvenčnímu pásmu
I	– akustická intenzita
I_0	– akustická intenzita odpovídající prahu slyšitelnosti
I_k	– index poslední hodnoty příslušící ke k-tému frekvenčnímu pásmu
L	– dimenze příznakového vektoru
L	– hladina intenzity
L_0	– lag
L_p	– hladina akustického tlaku (SPL)
M	– střední krátkodobá intenzita
f_0	– základní frekvence kmitání hlasivek
f_t	– délka rámce
f_s	– vzorkovací frekvence
L_{vz}	– délka rámce (počet vzorků)
K	– počet rámců signálu
p, p_A	– akustický tlak
p_0	– referenční akustický tlak, který odpovídá prahu slyšitelnosti
\mathbf{p}	– příznakový vektor
\mathbf{p}_i	– příznakový vektor i -tého statického příznaku
P	– akustický výkon
q	– Itakurův koeficient
$r(m)$	– normovaný autokorelační koeficient při posunu signálu o m vzorků
\mathbf{R}	– matice příznakových vektorů \mathbf{r} (reference)
$R(m)$	– autokorelační koeficient při posunu signálu o m vzorků

$R_{CCF}(m)$	– cross-korelační koeficient při posunu signálu o m vzorků
R_{\max}	– maximální hodnota ACF
RMS	– časově průměrovaná hodnota signálu (tlaku)
$s(n)$	– výsledný řečový signál
S	– velikost plochy, kterou vlnění prochází
$S(f)$	– spektrum původního signálu (v délce okna) před aplikací okna
t	– čas
t_w	– akumulovaný čas od počátku nahrávky do konce w -tého slova
T_0	– perioda základního tónu
τ	– exponenciální časová konstanta
v	– akustická rychlost
v_a^w	– akumulovaná rychlost řeči určená pro w -té slovo
\bar{v}_s	– průměrná rychlost řeči
w	– akumulovaný počet slov
$w(i)$	– transformační funkce (cesta)
$w(n)$	– okénková funkce
W	– celkový počet slov vyřčených v rámci promluvy
$W(f)$	– spektrum použité okénkové funkce
$\hat{x}(i)$	– vzorky po aplikaci Hammingova okna a FFT
$x(n)$	– původní signál
$x'(n)$	– ustředěný signál po odebrání stejnosměrné složky
X	– matice příznakových vektorů x (reprezentace)
$X(f)$	– výsledné spektrum signálu po aplikaci okénkové funkce
ξ	– pomocná proměnná časové integrace
Z	– vlnový odpor prostředí
ZCR	– počet průchodů nulou

Seznam zkratek

- ACF – Autokorelační funkce (Autocorrelation function)
- ANSI – Americký normalizační institut (American National Standards Institute)
- AOP – Horní hranice pro záznam signálu mikrofonom (Acoustic Overload Point)
- ASR – Automatické rozpoznávání plynulé řeči (Automatic Speech Recognition)
- CCF – Křížová korelační funkce (Cross-correlation function)
- CNS – Centrální nervová soustava
- DCT – Diskrétní cosinová transformace
- DFT – Diskrétní Fourierova transformace
- DIF – Decimace ve frekvenci (Decimation in frequency)
- DIT – Decimace v čase (Decimation in time)
- DTW – Dynamické borcení času (Dynamic Time Warping)
- EKG – Elektrokardiografie
- EMG – Elektromyografie
- FT – Fourierova transformace
- FT3 – Volný trijodtyronin
- FT4 – Volný tyroxin
- FFT – Rychlá Fourierova transformace (Fast Fourier Transform)
- FNKV – Fakultní nemocnice Královské Vinohrady
- GB – Graves-Basedowova choroba (Graves-Basedow disease).
- HMM – Skryté Markovovy modely (Hidden Markov Model)
- HW – Hardware
- HPV – Lidský papilomavirus (Human papilloma virus)
- HSV – Vysokorychlostní videolaryngoskopie (High-speed videolaryngoscopy)
- HTE – Hemithyreoidiektomie
- HTK – Hidden Markov Model ToolKit
- IEC – International Electrotechnical Commission
- IFT – Inverzní (zpětná) Fourierova transformace
- IDFT – Inverzní (zpětná) Diskrétní Fourierova transformace
- LPC – Lineárně prediktivní koeficienty (Linear predictive coefficients)
- LTW – Lineární borcení času (Linear Time Warping)
- MDVP – Multidimenzionální analýza
- MFCC – Mel-frekvenční keprální koeficienty (Mel-Frequency Cepstral Coefficients)

NCCF – Normalizovaná cross-korelační funkce (Normalized cross-correlation function)
NN – Neuronová síť (Neural Network)
ORL – Otorinolaryngologie
PA – Pacient
PCA – Analýza hlavních komponent (Principal component analysis)
PDF – Přenosný formát dokumentů (Portable Document Format)
PKN – Počet komplexních násobení
PKS – Počet komplexních sčítání
PSD – Spektrální hustota výkonu (Power Spectral Density)
RE – Reinkeho edém
RMS – Časově průměrovaná hodnota (Root Mean Square)
SNR – Signal to noise ratio
SPL – Hladina akustického tlaku (Sound pressure level)
RP – Hlasové pole (Speech range profile)
SW – Software
T3 – Trijodtyronin
T4 – Tyroxin
THD – Celkové zkreslení (Total Harmonic Distortion)
TPOAb – Tyreoidální peroxidáze
TRH – Tyreoliberin
TSH – Tyreoidální stimulační hormon
TTE – Totální thyreoidektomie
TTS – Z textu na řeč (Text to speech)
UEP – Unie evropských foniatrů
USB – Univerzální sériová sběrnice (Universal Serial Bus)
USG – Ultrasonografie
VDD – Voice Disorder Diagnostician
VHI – Voice Handicap Index
VKG – Videokymografie
VRP – Hlasové pole (Voice Range Profile)
VUT – Vysoké učení technické v Brně
XLR – Audio konektor (External Line Return)
ZCR – Počet průchodů nulou (Zero Crossing Rate)
ZD – Zdravý subjekt

Seznam obrázků

Obrázek 1.1 – Histologické vrstvy hlasivek	19
Obrázek 1.2 – Uložení hlasivek v hrtanu	20
Obrázek 1.3 – Ukázka umístění formantů při fonaci samohlásek „a“ a „i“	21
Obrázek 2.1 – Patologicko-anatomické změny hlasivek u organických poruch hlasu	23
Obrázek 2.2 – Místa s vysokým rizikem poranění průdušnice při intubaci	25
Obrázek 2.3 – Vyšetření hlasového pole (fonetogram)	31
Obrázek 2.4 – Ukázka výstupu multidimenzionální analýzy hlasu	33
Obrázek 2.5 – Laryngoskop s vyměnitelnou lžící	35
Obrázek 2.6 – Speciální kamera pro HSV a zařízení pro její ovládání	35
Obrázek 2.7 – Vyšetření hlasivek pomocí HSV a VKG	36
Obrázek 2.8 – Schéma faktorů ovlivňujících kvalitu hlasu	37
Obrázek 2.9 – Vizualizace hlasového cvičení pro měkké hlasové začátky	39
Obrázek 3.1 – Uložení štítné žlázy v přední části průdušnice	40
Obrázek 3.2 – Řízení činnosti štítné žlázy dle principu negativní zpětné vazby	41
Obrázek 3.3 – Diagnostický algoritmus vyšetřování funkčních poruch štítné žlázy	43
Obrázek 3.4 – Viditelně zvětšená štítná žláza (struma)	45
Obrázek 3.5 – Hlavní rizika při operaci štítné žlázy	47
Obrázek 5.1 – Dělení počítačového zpracování řeči.....	50
Obrázek 5.2 – Promluva „Vážení spoluobčané“ z projevu V. Havla v roce 1991	50
Obrázek 5.3 – Dělení signálu na rámce délky 20 ms s překryvem 10 ms.....	53
Obrázek 5.4 – Základní dělení strojového učení.....	54
Obrázek 5.5 – Postup výpočtu FFT s využitím algoritmu DIT pro decimování v čase	57
Obrázek 5.6 – Spektrogram promluvy: „Přeji vám všem klidný a šťastný Nový rok“.....	58
Obrázek 5.7 – Původní signál, širokopásmový a úzkopásmový spektrogram	59
Obrázek 5.8 – Teorie zdroje a filtru: model používaný u syntetizátorů řeči	60
Obrázek 5.9 – Schéma výpočtu reálného kepstra segmentu řečového signálu	62

Obrázek 5.10 – Ukázka aplikace pravoúhlého okna (dole) na původní signál (nahore).....	62
Obrázek 5.11 – Porovnání frekvenčního spektra obdélníkového a Hammingova okna	63
Obrázek 5.12 – Ukázka ZCR pro rámeček délky 20 ms	67
Obrázek 5.13 – Anatomie lidského ucha	73
Obrázek 5.14 – Frekvenční charakteristiky váhových kmitočtových filtrů	83
Obrázek 5.15 – Globální omezení při použití metody DTW	85
Obrázek 6.1 – Příklad frekvenční odezvy mikrofonu s ledvinovou směrovostí	89
Obrázek 6.2 – Přehled směrových charakteristik mikrofonů	91
Obrázek 6.3 – Dynamický rozsah mikrofonu.....	92
Obrázek 6.4 – Převod hlasu do digitalizované podoby včetně ovlivňujících faktorů	93
Obrázek 6.5 – Kalibrace v řetězci úlohy zpracování řeči.....	95
Obrázek 7.1 – Hlavový mikrofon DPA d:fine použitý pro záznam hlasu pacientů	97
Obrázek 7.2 – Frekvenční charakteristika (odezva) mikrofonu DPA d:fine	98
Obrázek 7.3 – Směrová charakteristika zvoleného mikrofonu DPA d:fine	98
Obrázek 7.4 – Zvukoměr EXTECH 407780A	99
Obrázek 7.5 – Kalibrátor EXTECH 407744	100
Obrázek 7.6 – Externí zvuková karta Focusrite Scarlett 2i2 2nd Generation	100
Obrázek 7.7 – Schéma zapojení měřicího řetězce včetně použitých redukcí.....	102
Obrázek 8.1 – Adresářová struktura aplikace zobrazena přes příkazovou řádku	105
Obrázek 8.2 – Nastavení práv správce pro VDD je vyžadováno pro správnou funkčnost	105
Obrázek 8.3 – Hlavní nabídka aplikace Voice Disorder Diagnostician.....	106
Obrázek 8.4 – Hlavní nabídka aplikace Voice Disorder Diagnostician.....	107
Obrázek 8.5 – Ukázka vygenerovaného identifikátoru na základě údajů o pacientovi	108
Obrázek 8.6 – Dialogové okno zobrazované při shodě identifikátoru.....	109
Obrázek 8.7 – Upozornění na chybějící informace při vyplňování údajů.....	109
Obrázek 8.8 – Upozornění, že údaje o pacientovi nebyly uloženy do databáze.....	110
Obrázek 8.9 – Režim načítání údajů o pacientech v aplikaci VDD.....	111

Obrázek 8.10 – Dialogové okno zobrazované při shodě identifikátoru	112
Obrázek 8.11 – Obrazovka upravování záznamů pacienta hned po stisku tlačítka upravit ...	112
Obrázek 8.12 – Potvrzení zobrazené při automatickém návratu do hlavního menu.....	113
Obrázek 8.13 – Ukázka obrazovky databáze pacientů s řazením příjmení podle abecedy	113
Obrázek 8.14 – Export databáze do souboru a podporované typy souborů	114
Obrázek 8.15 – Hláška upozorňující na zvolení nevhodného zvukového zařízení	115
Obrázek 8.16 – Režim nahrávání po volbě nahrávacího zařízení (před kalibrací)	115
Obrázek 8.17 – Režim nahrávání po úspěšné kalibraci.....	116
Obrázek 8.18 – Režim kalibrace umožňuje v reálním čase náhled signálů z obou kanálů....	117
Obrázek 8.19 – Režim analýzy naměřených dat a výsledky extrakčního algoritmu	119
Obrázek 8.20 – Změny logaritmické energie u zdravých a nemocných	122
Obrázek 8.21 – Změny spektrální energie v pásmech 0-1 kHz u zdravých a nemocných.....	123
Obrázek 8.22 – Změny spektrální energie v pásmech 1-3,5 kHz u zdravých a nemocných..	124
Obrázek 8.23 – Změny spektrální energie v pásmu 3,5-4 kHz u zdravých a nemocných	125
Obrázek 8.24 – Změny ZCR u zdravých a nemocných.....	126
Obrázek 8.25 – Změny základního tónu u zdravých a nemocných	126
Obrázek 8.26 – Změny hladiny akustického tlaku u zdravých a nemocných	127
Obrázek 8.27 – Změny koeficientů keprální analýzy (C1-C3) u zdravých a nemocných...	128
Obrázek 8.28 – Změny koeficientů keprální analýzy (C4-C13) u zdravých a nemocných.	129
Obrázek 8.29 – Změny rychlosti řeči u zdravých a nemocných.....	130
Obrázek 8.30 – Nejvýznamnější změny DTW skóre vztažené k částem nahrávek.....	131
Obrázek 8.31 – Diagnostický režim.....	132
Obrázek 8.32 – Nejvýznamnější změny DTW skóre vztažené k částem nahrávek.....	133
Obrázek 8.33 – Oblasti s vyšším rizikem negativních změn hlasu.....	134

Seznam tabulek

Tabulka 1.1 – Subsystemy tvorby hlasu.....	19
Tabulka 2.1 - Hodnocení hlasu metodou GRBAS.....	33
Tabulka 2.2 – Zásady hlasové hygieny	38
Tabulka 5.1 – Hodnoty použité při segmentaci v disertační práci.....	52
Tabulka 5.2 – Porovnání výpočetní náročnosti (počet operací) DFT a FFT	58
Tabulka 5.3 – Pojmy užívané v keprální analýze.....	60
Tabulka 5.4 - Příklady hodnot základních akustických veličin	76
Tabulka 5.5 – Hodnoty váhových funkcí A, C a Z dle IEC	82
Tabulka 6.1 – Doporučení pro volbu mikrofonu určeného pro měření hlasu a řeči	94
Tabulka 7.1 – Seznam komponent měřicího řetězce.....	97
Tabulka 7.2 – Klíčové parametry mikrofonu DPA d: fine	98
Tabulka 7.3 – Parametry zvukoměru EXTECH 407780A	99
Tabulka 7.4 – Parametry zvukové karty Focusrite Scarlett 2i2 2nd Generation	101
Tabulka 8.1 – Prahové hodnoty příznaků 4. části nahrávek	132

Klíčová slova

Poruchy hlasu, diagnostika, biomedicínské signály, příznakový vektor, extrakce příznaků řeči, kalibrace mikrofonu, keprum, MFCC, základní frekvence, Voice disorder diagnostician.

Keywords

Voice disorders, diagnostics, biomedicine signals, feature vector, feature extraction, microphone calibration, cepstrum, MFCC, pitch detection, Voice disorder diagnostician.

Úvod

Zpracování hlasu se v současnosti stalo velmi populárním tématem s poměrně různorodou paletou možných aplikací. Lidstvu přináší celou řadu výhod a přispívá tak i ke komfortu současných i budoucích generací. V poslední době se již objevují chytré domácnosti ovládané hlasem, je možné komunikovat s mobilním telefonem pomocí slovních příkazů a vytvářet tak záznamy v kalendáři apod. Na mnoha internetových serverech zprostředkujících přehrávání videa jsou k dispozici automaticky vytvořené titulky, v některých státech je na zvukový záznam z místa činu pohlíženo jako na plnohodnotný důkaz a vyšetřovatelé tak mohou podle nahrávek hlasu identifikovat pachatele, počítače se učí mluvit za pomoci syntézy hlasu, objevují se zprávy o překladu jazyků v reálném čase a mnoho dalších neuvěřitelných aplikací. Jedním z mnoha odvětví, kde se může zpracování hlasu uplatnit, je také medicína. Projevy mnohých nemocí lze identifikovat na základě hlasu pacientů. Přitom se vůbec nemusí jednat o nemoci spojené s tvorbou hlasu.

Tato práce je však zaměřena na využití technik pro zpracování řeči v oblasti biomedicíny a diagnostiky poruch hlasu. Konkrétně se zabývá převážně onemocněními štítné žlázy, jenž vyžadují chirurgický zákrok. V těchto případech pak může dojít k rozvoji patologických změn hlasu. Účelem je najít charakteristické rysy, prvky či příznaky, které významně souvisí s výskytem poruch hlasu a na jejichž základě může být takováto porucha diagnostikována. Vzhledem k tomu, že tyto metody v oblasti medicíny prozatím nemají přílišné zastoupení, nabízí se možnost vytvořit aplikaci, která bude umožňovat evidenci pacientů, záznam jejich hlasu včetně okamžité analýzy a diagnostikovat poruchu hlasu. Při tom je nutné dodržet stanovené standardy a používat výhradně kalibrovaná zařízení k dosažení požadovaných výsledků. Pokud by taková aplikace existovala, významně by přispěla v oblasti diagnostiky poruch hlasu, neboť v současnosti se pro tyto účely využívá pouze několik metod, které jsou v podstatě subjektivně vyhodnoceny lékařem, a tudíž nelze hovořit o automatické diagnostice.

1 Hlas a jeho tvorba

Lidská řeč je už od pradávna tím nejpřirozenějším nástrojem pro předávání informací. Tento způsob komunikace umožňuje rychle, přesně a efektivně sdělit vše potřebné. Proto je velmi pohodlné a zároveň přirozené formulovat požadované příkazy právě za pomoci hlasu. Hlas se pro nás stal nepostradatelným nástrojem v běžném i pracovním životě. I proto se každodenně setkáváme s dalšími a dalšími aplikacemi využívajícími zpracování přirozené řeči. Ať už se jedná o tzv. chytré domácnosti, hlasové ovládání aut, překlad jazyků v reálném čase, automatizovanou infolinku či diagnostické aplikace, účel je však stále stejný. Za tím vším snaha zefektivnit, ale především zjednodušit jednotlivé úkoly či zastoupit člověka v běžných a repetitivních záležitostech každodenního života.

Problém nastává, je-li tvorba hlasu narušena ať už zraněním nebo nemocí. Typickými projevy je pak bolest vnímaná pacientem při tvorbě hlasu, omezený rozsah hlasu, chraptění apod. V nejzávažnějších stavech pak dochází k úplné ztrátě hlasu. Naštěstí v současné době již i v těchto velice závažných případech existuje řešení. Lze použít moderní metody syntézy hlasu, které dokáží psaný text převést na fonémy a následně vygenerovat promluvy. Pro převod textu na řeč (syntézu řeči) se často užívá v literatuře zkratka TTS pocházející z angličtiny. Také je možné pořídit tzv. hlasové protézy. Pravdou však stále zůstává, že se dosud nepodařilo vytvořit umělé hlasivky, a to nejen kvůli jejich velmi komplikované struktuře. Hlasivky proto zůstávají nenahraditelnou částí lidského těla a hlasového traktu. Komplexitě tvorby modelu hlasivek se věnuje např. [1].

Hlas a řeč jsou natolik komplexní prostředky pro sdělování informací, že lze nalézt nespočet možností využití a aplikací. Mnoho akademických prací, časopisů, knih a jiných publikací se zabývá hlasem ze všech možných hledisek. Od roku 2002 byl ustanoven 16. duben jako Světový den hlasu. Je to zejména z důvodu, aby si lidé uvědomovali, jak důležitý pro ně hlas je, co všechno jej může negativně ovlivnit a jak tomu předcházet. V nemocnicích po celém světě se mohou zájemci dozvědět potřebné informace a absolvovat kontrolní vyšetření [2]. Řeči a jejímu technickému zpracování je věnována konference celosvětového formátu – Interspeech, která se v roce 2021 konala i v České republice [3].

Před zabýváním se samotnými poruchami hlasu, jejich dopady a možnostmi diagnostiky a léčby, je vhodné nejprve zmapovat funkci ústrojí odpovědných za tvorbu hlasu a znát podstatu vzniku zdravého hlasu člověka.

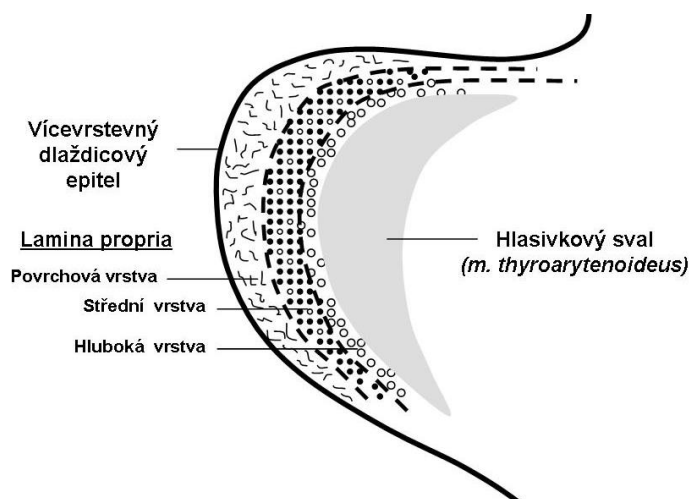
1.1 Tvorba lidského hlasu

Tvorbou hlasu se rozumí celý proces vzniku jednotlivých fonémů tak, jak jsou následně vnímány lidským sluchovým ústrojím. Poměrně nedávno vznikl vědní obor s názvem vokologie [4]. Vokologie se zabývá hlasem, jeho správným využíváním a trénováním. Na tvorbu hlasu lze nahlížet ze dvou směrů. Z pohledu medicíny je klíčová anatomie, která umožňuje popis jednotlivých ústrojí a jejich vlivu na tvorbu hlasu, kdežto z pohledu techniky můžou být jednotlivé komponenty substituovány a celý řetězec s určitou mírou zjednodušení překreslen pomocí funkčních bloků do schématu. Přehled orgánů důležitých pro tvorbu hlasu a jejich příslušnost k jednotlivým ústrojím shrnuje tabulka 1.1.

Tabulka 1.1 – Subsystémy tvorby hlasu [5]

Subsystém	Orgány
Dechové ústrojí	Bránice, hrudní svaly, žebra, břišní svaly, plíce
Hlasové ústrojí	Hrtan, hlasivky a přilehlé chrupavky
Artikulační ústrojí	Hlasový trakt – nadhrtanové dutiny, artikulační orgány

Z hlediska medicíny rozlišujeme tzv. dechové ústrojí, jehož základem jsou plíce a přilehlé svaly. Úkolem dechového ústrojí je zajistit dostatečný proud vzduchu a jeho akustický tlak, který je nezbytný pro další fáze tvorby hlasu. Tvorba zvuku v oblasti hlasového ústrojí se souhrnně označuje jako fonace, proto i tento subsystém bývá často nazýván fonační ústrojí, jehož klíčovým prvkem jsou hlasivky [6].

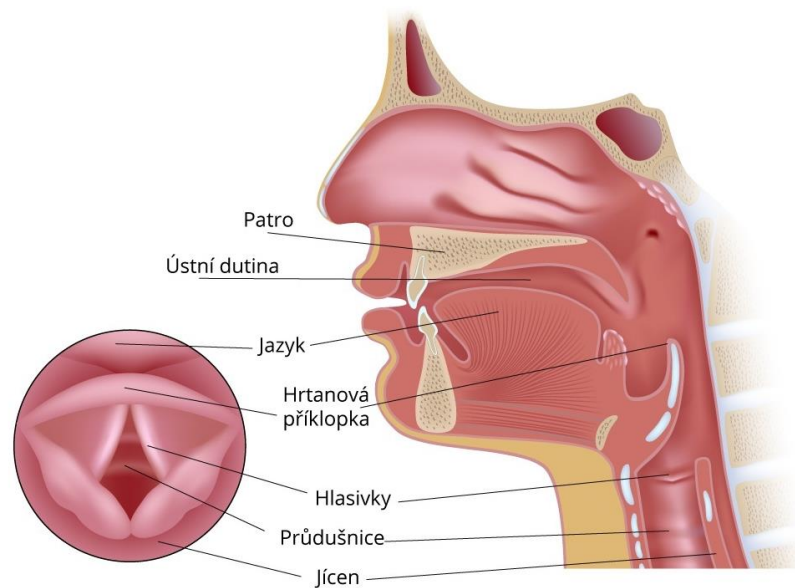


Obrázek 1.1 – Histologické vrstvy hlasivek [7]

Jedná se o párový orgán s třívrstvou strukturou, jenž tvoří velmi komplexní a dosud neúplně popsáný systém, což komplikuje vývoj jejich co možná nejdokonalejší náhrady.

Jak již bylo uvedeno, obecné dělení uvádí 3 základní vrstvy – epitel, podslizniční vrstvu (lamina propia), která je též složena ze 3 částí, a hlasivkový sval. Jednotlivé tkáňové vrstvy hlasivek jsou vyobrazeny na obrázku 1.1. Podrobnosti o stavbě hlasivek a jejich vazbách na jednotlivé svaly a chrupavky uvádí např. [7].

Hlasivky jsou orgánem, který v podstatě omezuje průchod vzduchu hrtanem a tím umožňuje vznik řeči. Hlasivky jsou uloženy v horní části hrtanu v horizontální poloze proti sobě. Jejich umístění je znázorněno na obrázku 1.2. Přiléhají k sobě v oblasti štítné chrupavky a jsou napjaty k hlasivkovým chrupavkám. Mezi hlasivkami vzniká tzv. hlasivková štěrbina (glottis). Právě pomocí hlasivkové štěrbiny může být regulován průchod vzduchu hrtanem. Například při dýchání je hlasivková štěrbina rozšířena a umožňuje tak průchod vzduchu při nádechu a výdechu. Kdyžto při běžné řeči je hlasivková štěrbina zúžena, hlasivky přiléhají k sobě a proudem vzduchu vytlačovaným z plic hlasivky vibrují a vytvářejí tak základní tón daného řečníka.



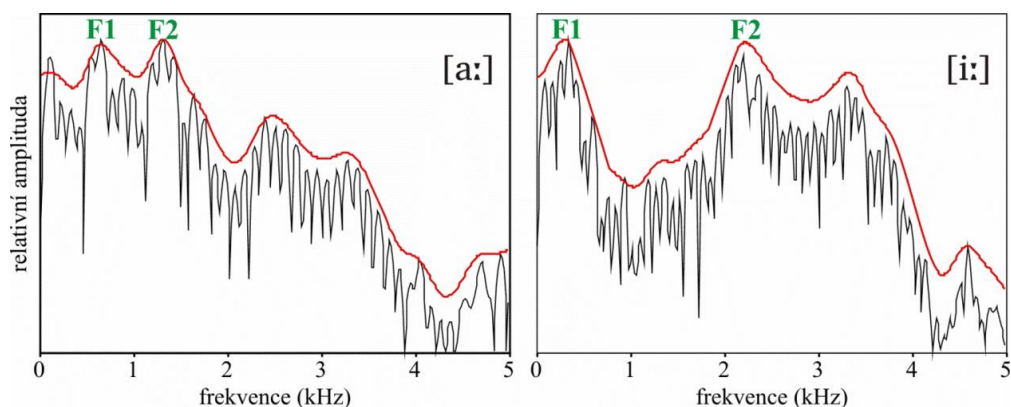
Obrázek 1.2 – Uložení hlasivek v hrtanu [8]

Podrobněji lze celý proces popsat tak, že, vzduch z plic se nahromadí pod semknutými hlasivkami, vytváří tak přetlak, jehož působením se hlasivky oddálí a část vzduchu prochází hlasivkovou štěrbinou, zatímco se hlasivky opět semknou a opětovně na ně působí přetlak sloupce vzduchu a celý proces se opakuje. Rychlost, s jakou probíhá periodické otevírání a uzavírání hlasivkové štěrbiny přímo ovlivňuje výšku hlasu. Nižší rychlost kmitání hlasivek se projevuje u mužů a výsledkem je hluboký hlas, zatímco

rychlejší kmitání hlasivek a vyšší tóny jsou typické pro ženy. Nejvyšších naměřených hodnot při běžné řeči pak dosahují děti.

Tón, který hlasivky generují, je fundamentální částí řeči, která je následně dále modulována do výsledné podoby artikulačním ústrojím. Artikulační ústrojí se skládá z pohyblivých i nepohyblivých částí. Patří sem ústní, nosní a hrdelní dutiny, které se souhrně označují také jako nadhrtanové dutiny. Dále je to skupina tzv. artikulačních orgánů, mezi které se řadí měkké patro, jazyk, zuby a rty. Základní tón vzduchu vydechaného z plic je při průchodu artikulačním ústrojím formován do podoby jednotlivých hlásek, lépe řečeno fonémů. Dochází k rezonanci v jednotlivých dutinách a zbytek dotváří pohyblivé orgány. Například vliv pootevření úst (změna parametrů ústní dutiny) a polohy jazyka vzhledem k zubům je neoddiskutovatelný.

Tyto rezonance v nadhrtanových dutinách mají za následek vznik tzv. formantů, které ovlivňují zabarvení hlasu a umožňují odlišit jednotlivé samohlásky. Při analýze frekvenčního spektra hlasu se formanty objevují jako lokálních maxima. V literatuře se formanty běžně označují písmenem F a číslem formantu. Příkladem může být fonace samohlásek na obrázku 1.3.



Obrázek 1.3 – Ukázka umístění formantů při fonaci samohlásek „a“ a „i“ [8]

Zajímavostí je, že například během šeptání se hlasivky vůbec nezapojují a hlasivková štěrbina zůstává otevřena a neklade tak odpor proudu vydechaného vzduchu stejně jako při dýchání. Výsledné hlásky jsou pak tvořeny pouze zapojením artikulačního ústrojí. I v běžné řeči se vyskytují hlásky, které hlasivky nevyužívají. Označují se jako neznělé. Zatímco samohlásky jsou vždy znělé a vznikají při pravidelném kmitání hlasivek, některé souhlásky jako například š, c či t jsou neznělé. Obecně souhlásky vznikají tak, že proud vzduchu naráží v hlasovém traktu na překážku, kterou tvoří artikulační orgány, a výsledná hláska má pak šumový charakter.

Technický pohled umožňuje substituovat jednotlivé části hlasového traktu a vznikla tak teorie zdroje a filtru [9], jakožto nejpodstatnějších částí. Hlasivky svým periodickým kmitáním generují tón, takže je možné uvažovat na počátku řetězce zdroj periodických kmitů. Pro neznělé hlásky by pak vstupem byl šum. Druhým prvkem je filtr, který upravuje signál generovaný zdrojem a zastupuje tak artikulační ústrojí. Parametry filtru zcela logicky nezůstávají konstantní. Stejně tak jako se mění nastavení artikulačních orgánů při tvorbě zvuků do podoby jednotlivých hlásek, se mění právě parametry filtru. Mnoho výzkumů se zabývalo hledáním vhodného tvaru a parametrů přenosové funkce filtru k modifikaci vygenerovaných kmitů, tak aby výsledky mohly posloužit pro uměle vytvořený hlas – syntézu řeči. Ačkoliv se zdroj a filtr mohou jevit jako oddělené za sebou řazené bloky a v [9] jsou tak i prezentovány, později se ukázalo, že mezi nimi existují nelineární vazby. Obě části se tedy vzájemně ovlivňují a nelze je považovat za nezávislé.

2 Poruchy hlasu

Hned na úvod je třeba zmínit, že poruchy hlasu nejsou totéž, co vady řeči (resp. poruchy řeči). Vady řeči se ani mezi poruchy hlasu neřadí. Souvisí totiž se špatnou artikulací. Příkladem může být hypokinetická dysartrie, jejíž příčina je organického původu [10]. V případě alalie se zase jedná o nesprávné vyslovování jedné či více hlásek. Protože řeč se vyvíjí již od dětství, vady řeči vznikají převážně v tomto období a není-li jim věnována dostatečná pozornost, mohou pak pacientům výrazně komplikovat život. Následky pak bývají i psychologického charakteru. Typické pro tyto subjekty je snížené sebevědomí a sebehodnocení, větší uzavřenost a neochota komunikovat. Protože se ale práce nezabývá vadami řeči a logopedickými metodami a doporučeními, nebudou vady řeči v rámci této práce již dále rozebírány.

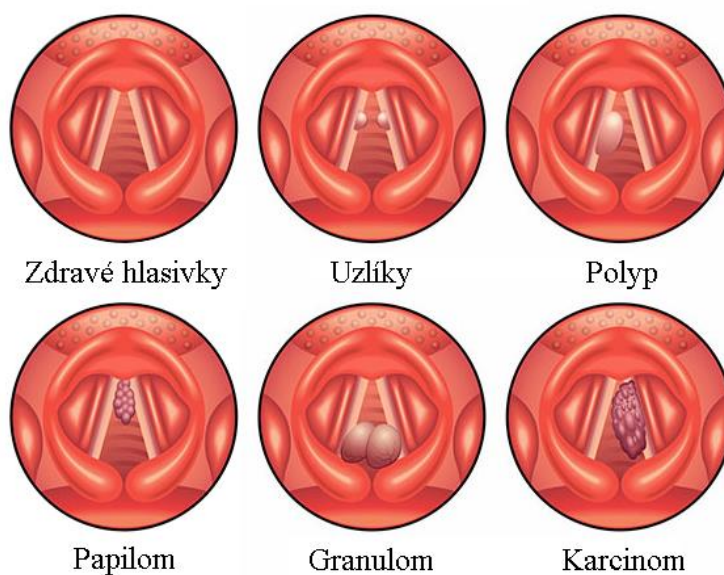
Je-li však tvorba zdravého hlasu jakýmkoliv způsobem narušena (zatímco výslovnost je správná), nazývá se tento stav porucha hlasu. Typické poruchy budou popsány v kapitolách níže, nicméně existují i další v některých případech vzácné poruchy hlasu, jejichž úplný výčet není součástí této práce. Je tu i celá řada dalších onemocnění, které mohou mít vliv na hlasový projev daného subjektu, jako například Parkinsonova nemoc [11] a [12]. Kromě toho existují i další nemoci, které lze ze záznamu hlasu diagnostikovat. Příkladem může být Alzheimerova nemoc či ADHD [13].

Jak již bylo zmíněno dříve, hlasivky jsou klíčovým orgánem pro tvorbu hlasu. Je třeba si uvědomit, že jsou prvním článkem v řetězci, který zpracovává vzduch z plic. Právě hlasové ústrojí postihuje celá řada poruch hlasu od těch krátkodobých, které samy odezní,

až po ty nejzávažnější, které vedou k úplné ztrátě hlasu (afonie). Vznik poruch hlasu může mít různé příčiny. Tvorbě hlasu obecně škodí extrémní podmínky. Typickým příkladem mohou být velké rozdíly teploty. Velmi škodlivé je samozřejmě kouření, či pravidelné požívání alkoholu, který vysouší sliznice podobně jako například nápoje obsahující kofein. Pobyt v nadměrně prašném prostředí také není ku prospěchu. Ale poruchy hlasu mohou vzniknout i poraněním, nadměrným namáháním hlasivek či vrozenými vadami. Na proces tvorby hlasu zkrátka působí celá řada dalších faktorů, mezi nimiž je i psychický stav daného člověka. Některé profese jsou těmito poruchami ohroženy více než ostatní. Jedná se zejména o zpěváky, herce, moderátory, hlasatele či učitele. Právě vzniku poruch hlasu u pedagogů se blíže věnuje [14] či [15]. Problematiku prevence u hlasových profesionálů řeší [16]. Jako obecného průvodce poruchami hlasu a metodami jejich diagnostikování lze doporučit knihu Vokologie [17]. Nejčastěji se v literatuře poruchy hlasu rozdělují do dvou základních skupin – organické a funkční [18].

2.1 Organické poruchy hlasu

Organickými poruchami hlasu se rozumí patologické změny v anatomii zejména v oblasti hrtanu. Jak se takové patologicko-anatomické změny projeví na hlasívkách, je zobrazeno na obrázku 2.1. Tento typ poruch hlasu může být dle [19] způsoben zánětlivými onemocněními hrtanu, vývojovými vadami hrtanu, traumatickým poškozením hrtanu, benigní hrtanovou lézí nebo maligním nádorem.



Obrázek 2.1 – Patologicko-anatomické změny hlasivek u organických poruch hlasu [20]

2.1.1 Zánětlivá onemocnění

Laryngitida je typickým zánětlivým onemocněním hrtanu. Dle doby trvání se rozlišuje akutní či chronický zánět a z pohledu etiologie může jít o neinfekční či infekční variantu [18]. Spíše výjimečně může být bakteriálního původu. Mnohem častěji je však příčinou virová infekce, jak uvádí [21]. Laryngitida obvykle doprovází chřipku, zápal plic či nemoci průdušek. Její samostatný výskyt je méně častý. Jedná se o sezónní onemocnění, které se nejvíce projevuje s příchodem jara a v podzimních měsících. Zánět může vyvolat i nadměrné zatížení hlasivek například při snaze o komunikaci ve velmi hlučném prostředí (hudební kluby apod.), aktivní či pasivní kouření, alergická reakce či dlouhodobý pobyt v prašném prostředí. Typickým projevem je chrapot, zvýšená teplota a obtíže spojené s tvorbou hlasu. Může dojít i ke krátkodobé ztrátě hlasu. Naštěstí v případě akutní laryngitidy potíže většinou během několika dní samy odezní.

Například i běžná rýma působí změny lidského hlasu. V důsledku toho, že jsou sliznice oteklé a vedlejší nosní dutiny, kde by za normálních okolností docházelo k rezonancím, jsou naplněny hlenem, mění se tím i charakteristiky artikulačního ústrojí. Nejvíce je to patrné u tzv. nazálních, jinak řečeno nosových souhlásek. V češtině rozlišujeme celkem tři – m, n a ň.

2.1.2 Vývojové vady hrtanu

Pokud jde o vývojové vady hrtanu, více než na tvorbu hlasu pacientům znepríjemňují život způsobené potíže při polykání a snížení průchodnosti dýchacích cest, zejména u dětí může být situace opravdu kritická. Jedná se však o vzácně se vyskytující onemocnění. V literatuře se uvádí pro tyto nemoci též název vývojové úchylnky hrtanu. Řadí se sem například vrozená laryngomalacie, laryngokéla, diaphragma laryngis, deformity epiglottis či tzv. kočičí křik, nebo poměrně vzácný rozštěp prstěncové chrupavky či atrézie glottis. Detailnější popis některých zmíněných poruch je uveden v [22].

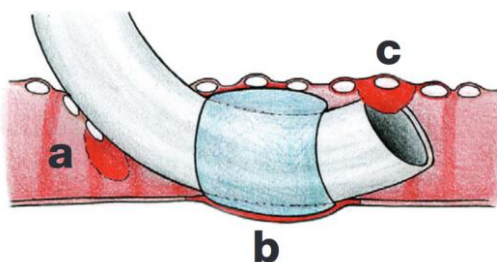
2.1.3 Traumatická poškození hrtanu

K traumatickému poškození hrtanu patří akutní fonotrauma, iatrogenní léze hlasivek a úrazy hrtanu. Akutní fonotrauma vznikne nadměrným zatížením (křik, kašel) fonačního ústrojí, což vede k rozšíření hlasivkových cévek a následnému krvácení do sliznice hlasivek, což je příčinou vzniku hematomu. Hlasivky mohou být náchylnější ke vzniku akutního fonotraumatu například v důsledku již probíhající laryngitidy. Příznakem je chrapot, který může přecházet až v afonii [18].

Iatrogenní léze hlasivek se může objevit, vyskytnou-li se komplikace při operaci hlasivek. Během chirurgického zákroku je poraněn zvrtný nerv, což vede ke zvýšení dušnosti a chrapotu po zákroku.

Působením vnějšího tlaku v oblasti fonačního ústrojí pak může dojít k poranění hrtanu (ať už v důsledku nehody, sebevražedného jednání či při napadení). Poškození hrtanu se rozlišuje na vnější a vnitřní podle příčiny vzniku. Vnější poranění může být buďto tupé (způsobené úderem tupého předmětu na přední stranu krku) nebo ostré, vzniklé následkem řezného, bodného či střelného zranění. Vnitřní poškození hrtanu bývá často způsobeno poleptáním či popálením. Výjimkou není ani vdechnutí cizího předmětu či bodnutí hmyzem.

Speciálním případem vnitřního poranění hrtanu jsou postintubační stenózy. Je to problém zejména dlouho intubovaných pacientů, u kterých zavedená kanyla tlačí na stěnu hrtanu/průdušnice, čímž na tomto místě vzniká zjizvení a stenóza. Granulomy po těchto zákrocích vznikají v zadní části glottis. Příznaky, které může postintubační stenóza vyvolávat, jsou chrapot, který přetrvává i dlouho po intubaci, snížení hlasové výkonnosti a dušnost [23]. Obrázek 2.2 demonstruje místa, na kterých může být průdušnice poraněna při použití tracheostomické kanyly během intubace. Místo „a“ symbolizuje poranění horním okrajem kanyly, zatímco označení „b“ odpovídá poranění, které vznikne tlakem manžety na průdušnici. Poslední místo „c“ může být pohmožděno okrajem tracheostomické kanyly.



Obrázek 2.2 – Místa s vysokým rizikem poranění průdušnice při intubaci [24]

Při chirurgickém zákroku v oblasti hlasivek mohou nastat komplikace, které mohou vyústit v částečnou či úplnou obrnu hlasivek. Jedním z takových zákroků může být operace štítné žlázy [25], kdy je porušen zvrtný nerv (laryngeus recurrens). Webová stránka [26] uvádí, že šance na vznik poškození hlasivek po chirurgickém zákroku je poměrně malá (pouze asi 3 %). Více o tomto zákroku a štítné žláze jako takové bude uvedeno v samostatné kapitole.

2.1.4 Benigní hrtanové léze

Mezi benigní hrtanové léze patří polypy, uzlíky, hlasivkové vředy a granulomy či subepiteliální léze hrtanu [19]. Změny se mohou objevit v kterékoliv části hrtanu, nicméně zpravidla nejčastěji bývají postiženy přímo hlasivky, což se ověřuje opticky. Může být zvolena neinvazivní léčba, kdy je pacientovi pouze doporučeno, aby udržoval hlasový klid nebo případně změnil prostředí, které jeho hlasu škodí. Může být přistoupeno k léčbě kortikoidy, které se v této souvislosti využívají nejčastěji. V případě, že ani medikamentózní léčba nemá požadovaný efekt, může být přistoupeno k léčbě chirurgické [27]. V literatuře existuje několik různých pohledů na to, jak nezhoubné hrtanové nádory dělit. Často se uvádí rozdělení na pravé a nepravé nádory. V rámci této práce bylo zvoleno dělení podle [18] s ohledem na etiologické faktory a stanovení léčby:

- nodulární (uzlíky),
- polypy,
- Reinkeho edém,
- cysty
- granulomy.

Do nodulárních lézí se řadí všechny typy uzlíků, tj. hlasivkový, zpěvácký a kantorský. Z histologického hlediska se nodulární léze dělí na pravé uzlíky, serózní pseudocysty a fusiformní edémy. Příčinou jejich vzniku je nadměrné užívání hlasu či špatná technika užívání hlasu. To zahrnuje mimo jiné i rychlé tempo řeči, vyšší polohu hlasu než obvykle či vyšší intenzitu (hlasitost) promluvy. Typicky tímto působením tedy nejprve vzniká fonotrauma. Uzlíky se mohou vyskytovat i u dětí, kde jsou možnosti léčby většinou omezené [28]. Mladší děti totiž většinou nezvládnou podstoupit hlasovou reedukaci, a navíc jen zřídka dokáží mít pod kontrolou, a hlavně omezit svou hlasovou aktivitu. K chirurgickým zákrokům se přistupuje spíše až v pubertě, kdy jsou tito pacienti schopni dodržet zásady hlasové rehabilitace.

Polypy mohou být v jednoduchosti popsány jako vyklenutí části sliznice v oblasti hlasivky. Objevují se stejně jako další útvary na hlasivkách po nadměrném zatížení hlasového aparátu či nevhodné technice tvorby hlasu. Jejich výskyt samozřejmě narušuje jinak pravidelné kmitání hlasivek a způsobuje chrapot. Po odstranění se polypy analyzují ještě mikroskopicky, aby se ověřilo, zda neobsahují zhoubné buňky. Uvádí se, že riziko výskytu polypu u mužů je dvakrát větší než u žen. Jednou z osobností českého showbyznysu, u které se objevil hlasivkový polyp, je zpěvák Ladislav Kerndl, který

podstoupil chirurgické odstranění polypu a nějakou dobu po operaci udržoval doporučený hlasový klid [29].

Reinkeho edém (RE) je často diagnostikovaným benigním nálezem způsobujícím poruchu hlasu. Nese jméno po svém objeviteli. Je to jeden z útvarů, jehož pravděpodobnost výskytu se zvyšuje s pravidelnou konzumací alkoholu nebo kouřením. Na rozdíl od polypů, které postihují spíše muže, RE se mnohem frekventovaněji vyskytuje u žen – nejčastěji ve věku mezi 40 až 70 roky. Objevuje se u žen s hlubokým či hrubým hlasem. Větší riziko výskytu je spojováno i s nadváhou. V literatuře bývá uváděn i pod názvem polypoidní chorditida nebo polypoidní degenerace. Někdy bývá též nazýván „nemoc upovídáných kuřáček“ [18]. Reinkeho prostor se nachází v subepiteliální části hlasivek. Přesněji řečeno odpovídá povrchové vrstvě lamina propria. RE je otok celé membránózní části hlasivky, který může být mírný, ale také velmi výrazný, kdy se objem hlasivek zvětší natolik, že se hlasivky překrývají. Na sliznici se střídavě objevují známky hypertrofie i atrofie epitelu. RE nezasahuje do hlubších vrstev hlasivek. Podrobně se Reinkeho edémem, možnými příčinami jeho vzniku a možnostmi ovlivnění těchto faktorů zabývá [30].

Cysty se na hlasivkách vyskytují v oblasti lamina propria superficialis. Jejich dělení není v literatuře jednoznačně definováno. Podle některých autorů (Bouchayer a Hirano) lze zjednodušeně hlasivkové cysty rozdělit na mukózní retenční cysty a epidermoidní, přičemž druhé zmíněné se dále dělí na zavřené, otevřené a ostatní. Dle těchto autorů lze k otevřeným epidermoidním cystám přiřadit i cystysulcus vocalis či slizniční můstek [18].

Granulom, jak již název napovídá, je granulovitý útvar vyskytující se na chrupavčité části hlasivek. Jedním z významných etiologických faktorů vzniku granulomu je gastrolaryngeální reflux žaludečních šťáv [31]. Před samotnou léčbou granulomu se doporučuje nejprve medikamentózní léčba refluxu žaludečních šťáv. Granulomy vznikají též v souvislosti s traumatickým poraněním hrtanu např. intubací.

Mezi pravé benigní nádory hrtanu se řadí papilom. Jeho výskyt není věkově nijak vymezen a postihuje tedy jak dospělé, tak i děti. Papilomy hrtanu se objevují v souvislosti s virem lidské papilomatózy (HPV). Konkrétně se jedná o typ HPV 6 a HPV 11. Papilomavirus napadá kůži a sliznice. Celkově existuje přes 100 různých druhů, z toho 40 druhů jsou lidské papilomaviry, které se dělí na vysoce rizikové a nízko rizikové podle toho, zda mohou v nakažených buňkách vyvolávat nádorové bujení. Zajímavostí je, že do věku 50 let se virem nakazí 65% populace, přičemž větší část tvoří ženy. Nicméně imunitní systém si s HPV dokáže většinou poradit. HPV se nejčastěji přenáší pohlavním stykem. Léčba se nezaměřuje na HPV, nýbrž na chirurgické odstranění papilomu. HPV není možné

léčit virostatiky tak jako jiná virová onemocnění. Neosvědčila se ani imunoterapie. Jako prevence slouží očkování, které je však dostupné pouze proti některým typům HPV. Více informací o HPV je uvedeno v [32]. Papilom je benigním útvarem v hrtanu, který však časem může degenerovat a asi v 10 % případů přecházet v maligní nádor [22].

2.1.5 Maligní hrtanové léze

Pokud je nález onkologického charakteru, foniatr má za úkol posoudit, o jak závažný nález se jedná. To znamená opticky (za pomoci videostroboskopie) určit, do jaké hloubky karcinom zasáhl tkáň a posoudit, jak je ovlivněna hybnost hlasivek. Mezi nejčastější maligní nálezy patří zejména karcinom hrtanu a hltanu. Důležitá je role foniatra po zákroku, kdy je třeba s pacientem probrat možnosti hlasové rehabilitace [33].

2.2 Funkční poruchy hlasu

Na rozdíl od organických poruch hlasu se u funkčních poruch nevyskytují na hlasivkách žádné patologické změny, které by měly za následek vznik poruchy. U funkčních poruch hlasu je problém ve špatné technice tvoření hlasu a přetěžování hlasotvorného ústrojí. Teprve až následkem těchto nevhodných fonačních návyků vznikají fyziologické změny v oblasti hlasivek jako například již zmíněné uzlíky. Oba druhy poruch hlasu jsou tedy vzájemně propojeny a funkční poruchy hlasu mohou přecházet v organické a naopak. Funkční poruchy hlasu lze rozdělit do třech kategorií:

- poruchy z přemáhání hlasu,
- psychogenní poruchy,
- hlasové neurózy.

2.2.1 Poruchy z přemáhání hlasu

Fakt, že organické a funkční poruchy hlasu se prolínají, potvrzuje i to, že někteří autoři [34] do skupiny poruch vzniklých přemáháním hlasu řadí i uzlíky. Hlavním zástupcem je hyperkinetická dysfonie [35]. Díky laryngoskopii mohou být pozorovány tři fáze tohoto onemocnění. Zatímco v první fázi jsou hlasivky naprosto intaktní, v další fázi dochází k vřetenovitému zhuštění hlasivek. Poslední etapou je pak vznik uzlíků v místě největšího rozkmitu hlasivek (přibližně v jejich polovině). V této části se již pacient dostává do začarovaného kruhu, ze kterého mu může pomoci jedině odborník v oblasti ORL. Čím větší námahu člověk vyvine pro tvorbu hlasu, tím více svůj stav ještě zhoršuje. V prvních dvou etapách postačí dodržení zásad hlasové hygieny, a především omezení

nadměrného užívání hlasu. Vše se po čase vrátí do původního stavu. Kdežto vazivově organizované uzlíky v poslední fázi jsou již ireverzibilní.

Tato porucha hlasu se může objevit u neškolených jedinců, ale také u hlasových profesionálů [36]. Hyperkinetická dysfonie postihuje v dospělosti častěji ženy. Její vznik může podpořit kouření, fonace v hlučném prostředí, chronická onemocnění dolních cest dýchacích, nepříznivé prostředí obsahující toxiny či alergeny. Není výjimkou, že hyperkinetickou dysfonií trpí i děti. Na rozdíl od dospělé populace tato porucha hlasu převládá spíše u chlapců. Na vině jsou nejrůznější hlasové aktivity spojené s účastí v divadelních kroužcích či pěveckých sborech, kdy děti nemají zvládnutou správnou techniku tvorby hlasu. Další častou příčinou je křik. Uvádí se, že hyperkinetickou dysfonií trpí přibližně každé šesté dítě [37].

2.2.2 Psychogenní poruchy

Psychogenní poruchy hlasu zahrnují dysfonii či afonii způsobenou psychologickými příčinami. Také se o těchto poruchách hovoří jako o psychicky podmíněných. To znamená, že hlasivky jsou v tomto případě zcela v pořádku, ale na vině je tentokrát psychika. Typickým jevem je, že při smíchu či kašli je zachován zvučný hlas. Psychogenní afonie může být vyvolána intenzivním emočním zážitkem (účastník tragické nehody, oběti trestných činů). Taková událost v životě těchto pacientů může zanechat šrám právě v podobě ztráty hlasu. Leckdy se však stává, že pacient si není vědom události, která tyto potíže vyvolala. Výrazně častěji jsou psychogenní poruchy hlasu diagnostikovány ženám, a to konkrétně ve věku 20-30 let a ve věku 50-60 let. Tento typ poruch hlasu by měl s pacienty spíše než foniatr řešit psychiatr.

K psychogenním poruchám hlasu patří rovněž tzv. fistulový hlas. U chlapců po ukončení puberty již rozměry hrtanu odpovídají podmínkám na tvorbu normálního mužského hlasu. Nicméně může se stát, že zůstává zachován vyšší (fistulový) hlas. Podobný efekt může mít prodloužená mutace, kdy hlas přeskakuje z vyššího dětského hlasu na hlubší mužský. V obou případech se provádí hlasová cvičení pod dohledem odborníka. V knize [38] je mezi psychogenní poruchy hlasu zařazen ještě ventrikulární hlas, který není tvořen přímo účastí hlasivek, nýbrž za pomoci ventrikulárních řas, které se nachází nad hlasivkami. Tento způsob tvorby hlasu generuje hrubý, hluboký a chraptivý hlas.

2.2.3 Hlasové neurózy

Mezi nejčastější hlasové neurózy patří podle [34] spastická dysfonie, která je také zároveň obtížně léčitelná. Projevem je sevřený, tlačенý až chraptivý hlas se zkrácenou fonační dobou. Řeč připomíná koktání. Projevy nejsou přítomny ve zpěvu či při smíchu. Subjekt vyvíjí nadměrný fonační tlak při bezdůvodném svírání hlasové štěrbinu nebo prostoru nad ní (supraglottický prostor). Dochází k trvalým svalovým kontrakcím jednotlivých svalů a svalových skupin. To po čase vyvolává organické změny na hlasivkách, jejichž okraje zesílí stejně tak jako samotné hlasivky. Kromě hlasivek může postihnout ventrikulární řasy, epiglottis a aryepiglottické řasy [18]. Psychoterapie se v tomto případě téměř nevyužívá a léčba spočívá v hlasových cvičeních, která bohužel většinou nepřináší uspokojivý výsledek. Krátkodobě situaci řeší botulotoxinové injekce, jejichž aplikace se ale musí po třech měsících opakovat. Dlouhodobějším řešením může být chirurgický zákrok, kdy po protěti zvratného nervu (nervus laryngeus recurrens) příznaky ustoupí na několik let [23]. K hlasovým neurózám se řadí například ještě fonastenie či dysphonia hysterica.

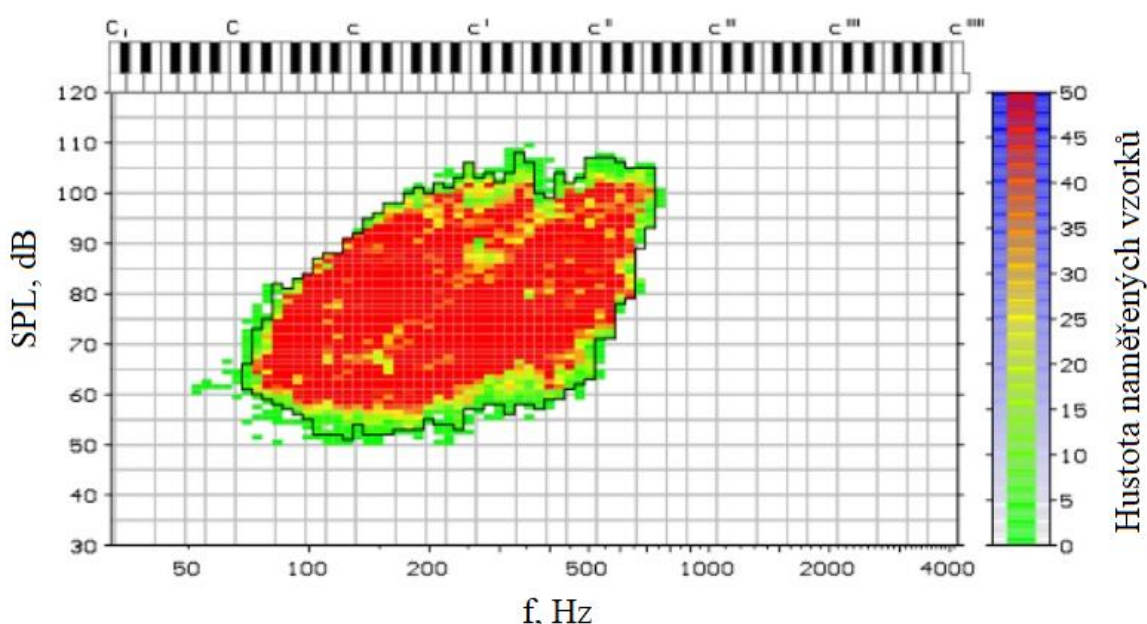
2.3 Metody pro diagnostikování poruch hlasu

Problémem současně používaných metod pro diagnostiku poruch hlasu je převážně orientace na optickou evaluaci, kdy je pacientovi do krku zavedena speciální endoskopická kamera, což může být pro pacienta nepříjemné. Alternativou jsou metody akustické, které se zabývají záznamem a analýzou řeči a charakteristickými jevy, které tvorbu řeči provází. Samostatnou kategorií jsou pak percepční metody. Další nevýhodou je, že se v praxi skutečně využívá pouze malé množství metod, a to i v případě, že pacient podstoupí vyšetření na specializovaném oddělení. Lékař tedy nemá k dispozici mnoho alternativ. Tyto metody v sobě většinou neodrážejí znalosti z oblasti zpracování řeči a často je výsledek subjektivně ovlivněn úsudkem a zkušenostmi lékaře. Srovnání nejpoužívanějších akustických a optických metod je uvedeno v publikaci [39].

2.3.1 Hlasové pole

Jako první byla zařazena metoda vyšetření hlasového pole pacienta (VRP) a to převážně z důvodu, že se jedná o univerzální a velmi efektivní nástroj, který lze považovat za objektivní. Tento postup jako jeden z mála odráží některé prvky z analýzy hlasu, řadí se tedy mezi akustické metody. Nicméně pro diagnostiku jako takovou se v praxi bohužel příliš nepoužívá.

Základem této metody je stanovit, jak již název napovídá, hlasový profil pacienta. Právě k jeho určení jsou využity znalosti zpracování řeči – zejména detekce základního tónu řeči a měření hladiny akustického tlaku (SPL). Podrobnosti o jednotlivých veličinách jsou zmíněny v kapitole extrakce příznaků z řeči. Hlasové pole je interpretováno graficky ve formě tzv. fonetogramu. Tento dvourozměrný graf má na horizontální ose základní tón řeči v jednotkách Hertzů (Hz), kdežto na vertikální osu se vynáší hladina akustického tlaku v decibelech (dB). Zatímco dříve se konstruovaly pouze kontury fonetogramu a hovořilo se o tzv. zpěvním či celkovém hlasovém poli, řečové (mluvní) pole se začalo zaznamenávat až později s rozvojem výpočetní techniky. U zjišťování mluvního rozsahu řeči, někdy též označovaného zkratkou SRP, čte pacient zadaný text, kdežto u zpěvného rozsahu se graf stanovuje z fonace vokálů a, e, i či u.



Obrázek 2.3 – Vyšetření hlasového pole (fonetogram) [40]

Zatímco pacient čte daný text, jsou po celou dobu zaznamenávány oba parametry a určují tak body, které se následně promítnou ve výsledném fonetogramu. Jak již bylo zmíněno dříve, v běžné řeči dochází k neustálým změnám a přestavování artikulačního ústrojí – jinak řečeno, mění se tedy základní tón. Pokud je text dostatečně dlouhý a rozmanitý, aby dokázal obsáhnout většinu fonémů a artikulační jevů pro daný jazyk, bude i výsledný fonetogram přesnější. Nyní již pro jednotlivé jazyky existují tzv. standardní texty používané právě pro tyto účely. V České republice se, jak uvádí [41], používá text

„Podzim na starém bělidle“. V publikaci [42] se zase uvádí text „Jak v Kocourkově chytali zajíce“, případně se lze setkat i s dalšími alternativami.

Obecně lze říci, že směrodatný je tvar zaznamenané oblasti a její rozměry. Typickým projevem poruch hlasu může být jak snížení intenzity hlasu, tak omezení rozsahu frekvencí, což se na fonetogramu promítne právě na velikosti zvýrazněné oblasti. Právě velikost plochy v sobě ukrývá údaje o tom, jak pacient s hlasem dokáže pracovat. Obecně u hlasatelů, moderátorů či zpěváků má zpravidla hlasové pole větší plochu, než je tomu u netrénovaných řečníků. Podrobný rozbor všech klíčových prvků a parametrů fonetogramu uvádí [41].

Příkladem aplikace vyvinuté pro vyšetření hlasového pole je Fonetogram 4.5 LITE. Soubor kvantitativních parametrů extrahovaných z hlasu a detailní popis této aplikace vyvinuté v Plzni je k dispozici v publikaci [42]. Jako další příklad aplikace vyvinuté v České republice, a to konkrétně ve Výzkumném centru hudební akustiky Hudební fakulty Akademie múzických v Praze, je RealVoiceLab [43].

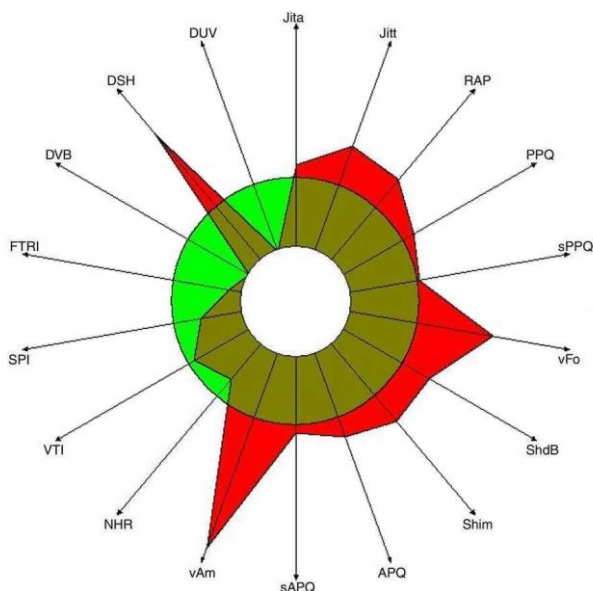
2.3.2 Spektrální analýza

Další z akustických metod je spektrální analýza zvukového signálu. Jedná se o analýzu rozložení energie hlasu ve frekvenčním spektru. Spektrální analýza využívá diskrétní Fourierovu transformaci (DFT) pro určení spektra. Pokud analýza probíhá v pásmech nad 50 Hz (typicky 300 Hz), hovoříme o širokopásmovém zobrazení, které se využívá zejména pro zaznamenání polohy (určení frekvence) a energie formantů. Je-li šířka pásma volena zhruba do 20 Hz, jedná se o úzkopásmové zobrazení, které je vhodnější pro zobrazení harmonických složek a meziharmonických pásem. Spektrální analýzu lze tedy zjednodušeně řečeno považovat za nástroj pro určení polohy formantů a rozložení akustické energie ve zvolených pásmech [43].

2.3.3 Multidimenzionální analýza

Za zmínku stojí také metoda MDVP (Multi-Dimensional Voice Program). Multidimenzionální analýza hlasu je funkční akustická vyšetřovací metoda, která na rozdíl od vyšetření hlasového pole pracuje s kvalitativními parametry a jejich uspořádáním. Výstupem této analýzy je paprskový graf – ukázka na obrázku 2.4. Ten lze dle [44] rozdělit na čtyři kvadranty, které tvoří logické celky. V prvním kvadrantu jsou parametry vypovídající o frekvenčním kolísání. Do druhého se vynáší stupeň neznělých úseků, stupeň subharmonických a tremor. Ve třetím kvadrantu jsou pak zaneseny šumové parametry a

poslední kvadrant zobrazuje amplitudové kolísání. Výpis všech zpracovávaných parametrů je rovněž uveden v [44]. MDVP může být s výhodou použita pro sledování průběžného stavu pacienta jeho uzdravování po zákroku. Hlavní nevýhodou je, že pro vyhodnocení a stanovení výsledné diagnózy je nicméně zapotřebí mít nejen s MDVP velké zkušenosti. Přiřazení diagnózy ke grafu nemusí být vždy jednoznačné.



Obrázek 2.4 – Ukázka výstupu multidimenzionální analýzy hlasu [45]

2.3.4 Percepční metody

Jak již název napovídá, po akustických metodách jsou zařazeny metody subjektivní založené na percepci čili poslechu. Foniatr ohodnotí hlas pacienta dle škály dané Unií evropských foniatrů (UEP) v rozmezí 0–6, kde 0 – normální, 1 – zastřený, 2 – lehká dysfonie, 3 - střední dysfonie, 4 – těžká dysfonie, 5 – afonie a 6 – ztráta hlasu po laryngektomii nebo traumatu. Jedná se tedy o kvalitativní posouzení hlasu bez určení konkrétní diagnózy a toto číslo pouze určuje, do jaké míry porucha hlasu pacienta ovlivňuje [33].

Tabulka 2.1 - Hodnocení hlasu metodou GRBAS

Význam	Interval pro hodnocení	Popis hodnocení
G – celkový stupeň poruchy hlasu	0-4	norma až těžká dysfonie
R – drsnost, hrubost hlasu	0-3	norma až těžká
B – dyšnost	0-3	norma až těžká
A – hlasová slabost	0-3	norma až těžká
S – hlasové napětí	0-3	norma až těžké

Dalším přístupem je metoda, která nese označení GRBAS. Je to opět subjektivní způsob, přičemž tentokrát se lékař nezaměřuje na kvalitu hlasu jako takového, ale hodnotí

chrapot pacienta. Jednotlivá písmena názvu metody (v angličtině) odpovídají kritériím (vlastnostem) na základě, kterých má být chrapot hodnocen. Rozmezí hodnot není pro jednotlivé vlastnosti stejné. Metodu GRBAS přehledně popisuje tabulka 2.1.

2.3.5 Voice Handicap Index

Jedná se o celosvětově rozšířenou metodu subjektivního hodnocení hlasu. Na rozdíl od předešlých percepčních metod však hodnocení neuvádí lékař nýbrž pacient. Jedná se o formu dotazníku, který pacient vyplní a poskytne tak lékaři první informace o projevech poruchy hlasu, kterou trpí, a to včetně subjektivního hodnocení, které svému zdravotnímu stavu přisuzuje. Protože v České republice nebyla přesně stanovena a unifikována podoba tohoto dotazníku, objevovali se různé verze. Nicméně v roce 2009 publikoval Jan G. Švec a kolektiv verzi, která je odborně přeložena z anglického jazyka. Klinikám ORL se doporučuje používání právě této verze [46]. Tato verze VHI dotazníku je zároveň v příloze A disertační práce.

2.3.6 Laryngoskopie a laryngostroboskopie

Základní optickou metodou pro vyšetření hrtanu a průdušnice je laryngoskopie. Tato metoda se soustřeďuje na hledání anatomických změn právě v těchto oblastech. Vyšetření probíhá za fonace hlásek *e* nebo *i*. Při tvorbě těchto hlásek je svaly hrtanu nadzvednuta hrtanová příklopka, díky čemuž je pozorování hlasivkové štěrbiny přehlednější. Laryngoskopie se rozděluje na přímou (direktní) a nepřímou. Nepřímá se provádí za pomoci laryngoskopického zrcátka či optickým zvětšovacím laryngoskopem. Přímá varianta se provádí pomocí flexibilního laryngoskopu nebo rigidního tubusu a mikroskopu.

Jak o laryngostroboskopii uvádí [23], jedná se o v praxi nejvyužívanější metodu pro vyšetření pacientů s poruchami hlasu. Jak již bylo zmíněno v kapitole věnované fonačnímu ústrojí, hlasivky při tvorbě řeči u zdravého člověka periodicky kmitají, přičemž rychlost kmitání odpovídá základní frekvenci hlasu. Protože je však frekvence kmitání hlasivek zhruba 50 Hz (uvažován nejhlubší hlas) a více není možné tento děj pozorovat pouhým okem, které zvládá rozlišit zhruba 20 snímků za vteřinu (20 Hz). U rychlejších dějů již oko není schopno rozlišit hranice jednotlivých snímků, které splývají a rozmazávají se. Tohoto efektu mimo jiné využívá například televize.

Druhá nedokonalost našeho zraku spočívá v určitém zpoždění, kdy na sítnici zůstává poslední spatřený obraz. Je-li periodický děj osvětlován záblesky světla v intervalech shodných s periodou opakování tohoto děje, pak se pozorovaný děj jeví jako statický

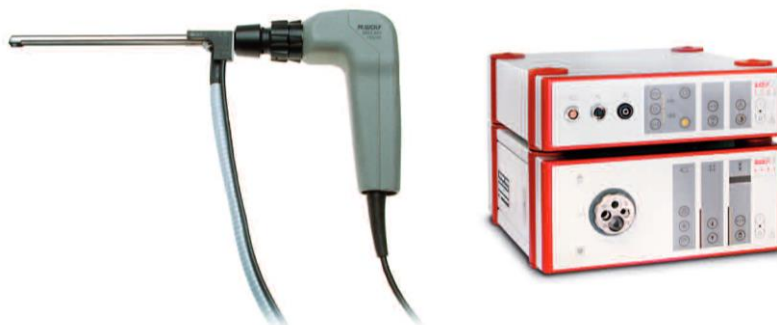
(neměnný). Jinak řečeno, pozorovatel jej vidí stále ve stejné fázi. V případě laryngostroboskopie, kdy lékař pozoruje hlasivky pacienta by to znamenalo, že lze sledovat například pouze fonační fázi, kdy jsou hlasivky semknuty. Pokud se frekvence osvětlování zvýší, děj se bude jevit jako zpomalený, přičemž míra zpomalení závisí na rozdílu mezi frekvencí pozorovaného děje a frekvencí osvětlování. V současnosti se již laryngostroboskopy synchronizují automaticky s využitím mikrofону či elektroglottografie zmíněné níže. Tato metoda selhává u pacientů s výrazným chrapotem, u kterých již kmitání hlasivek není periodické, a laryngostroboskopy tudíž nemohou být úspěšně synchronizovány.



Obrázek 2.5 – Laryngoskop s vyměnitelnou lžící [47]

2.3.7 Videokymografie a videolaryngoskopie

Videolaryngoskopie respektive videolaryngostroboskopie je pouze jiné označení pro laryngostroboskopie jen s tím rozdílem, že se při vyšetření pořizuje video záznam za pomoci endoskopické kamery. Videokymografie (VKG) využívá princip tzv. řádkového snímání, což znamená, že kamera sleduje pouze určitou část (detail) hlasivek. Zároveň je snímací frekvence kamery mnohonásobně vyšší než frekvence kmitání hlasivek. Uvádí se až 8 kHz. To je výhodné pro pozorování velmi rychlých a nepravidelných dějů. Metodě VKG a její aplikaci v klinické praxi se v první části své práce [48] věnuje Jan Švec.



Obrázek 2.6 – Speciální kamera pro HSV a zařízení pro její ovládání [49]

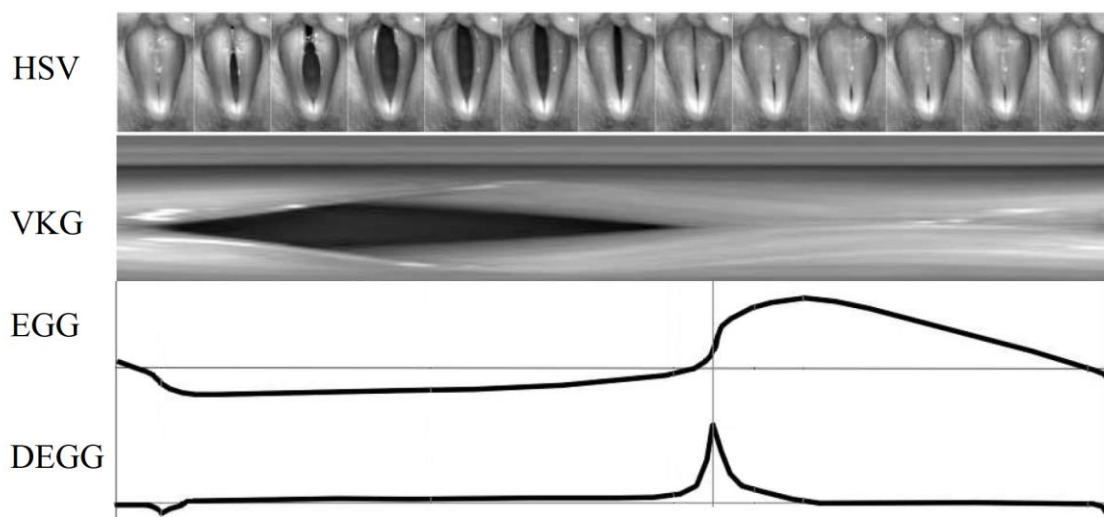
Poslední optická metoda a alternativa k výše zmíněným se objevila s nástupem nové generace kamer. Vysokorychlostní videolaryngoskopie (HSV) umožňuje sledovat celé hlasivky, a přitom zachovat vysokou frekvenci snímání obrazu a to až 20 kHz. Kombinuje tak výhody obou předešlých metod – rychlost záznamu i celoplošné snímání. Ukázka vybavení potřebného pro HSV je na obrázku 2.6.

2.3.8 Ostatní metody

Mezi další metody používané při vyšetření orgánů odpovědných za tvorbu hlasu mohou být podle [23] zařazeny například:

- elektrolottografie (EGG),
- elektromyografie (EMG),
- pneumografie,
- spirometrie.

Elektrolottografie se používá jako podpůrná metoda pro synchronizaci laryngostroboskopu. Princip metody spočívá ve snímání změn elektrického odporu při uzavření a otevírání hlasivkové štěrbiny (glottis). Na krk v oblasti, kde se nachází hlasivky jsou na obou stranách štítné chrupavky umístěny elektrody. Měří se velikost změny vysokofrekvenčního proudu nízké intenzity, která se převádí na již zmíněný elektrický odpor. Pomocí EGG lze takto určovat dobu uzavření a dobu otevření hlasivkové štěrbiny. Tyto hodnoty se pak využívají pro stanovení příslušných koeficientů.



Obrázek 2.7 – Vyšetření hlasivek pomocí HSV a VKG [41]

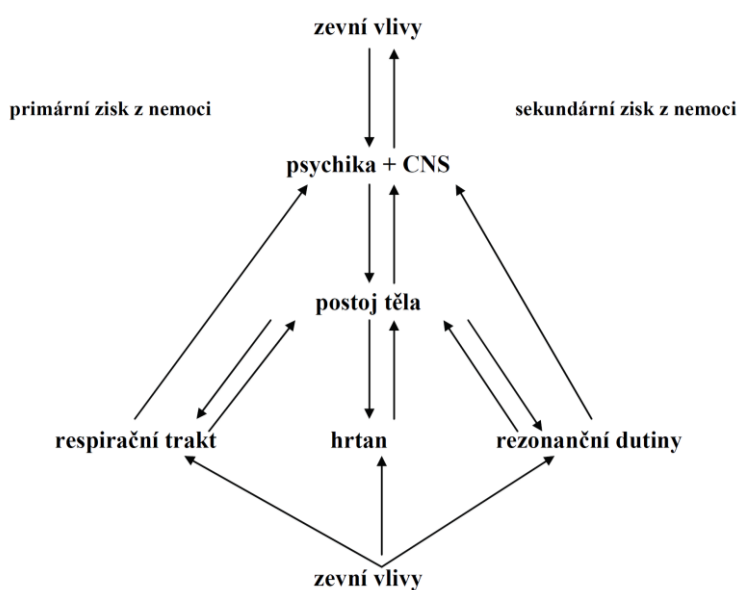
Na obrázku 2.7 je zobrazen jeden cyklus postupného otevírání a uzavírání glottis. V horní části jsou snímky pořízené vysokorychlostní kamerou metodou HSV, kde první

část odpovídá otevírání a od pátého snímku se již glottis uzavírá. Druhý obrázek shora je grafický výstup metody VKG. Níže je signál získaný pomocí metody EGG. Ve chvíli, kdy dojde k semknutí hlasivek, se rychle změní i charakter signálu podle principu popsaného výše. Derivací EGG signálů je obdržen průběh DEEG křivky, která umožňuje jednoznačně identifikovat okamžik, kdy došlo k semknutí hlasivek.

Pokud jde o ostatní přístupy, EMG je neurologická metoda, která spočívá ve sledování elektrické aktivity, která je generována motorickými procesy svalovými vlákny v hrtanu. Pneumografie sleduje pohyby hrudníku a břišní stěny během fonace a spirometrie se zabývá určením kapacity plic, která je klíčovým faktorem ovlivňujícím fonační dobu pacientů.

2.4 Léčba poruch hlasu

Po úspěšné diagnostice poruchy hlasu následuje její léčba. Ta se odvíjí od toho, o jakou poruchu hlasu se jedná. Zcela logicky bude navrhován jiný postup např. u psychogenních poruch než u maligních nádorových onemocněních hrtanu apod. Ne vždy musí být řešením chirurgický zákrok. V některých případech postačí soubor vhodně zvolených cvičení a odstranění škodlivých návyků. Prvním krokem je získání povědomí o tom, že kvalita hlasu je provázána, jak s fyzickým, tak i duševním stavem člověka. Všechny tyto mnohdy obousměrné vazby mají vliv při utváření hlasu a schematicky jsou zobrazeny na obrázku 2.8. Obecně je ale důležitá už samotná prevence před vznikem poruchy hlasu, a právě k tomuto účelu slouží zásady hlasové hygieny.



Obrázek 2.8 – Schéma faktorů ovlivňujících kvalitu hlasu [18]

2.4.1 Hlasová hygiena

Obecně lze říci, že hlasová hygiena představuje soubor pravidel, doporučení a vhodných návyků, jejichž účelem je zvýšit informovanost zainteresovaných osob a minimalizovat tak dopady všemožných negativních vlivů na jejich hlas. Protože osvojení pravidel hlasové hygieny a dodržování vhodných návyků je dobré řešit již u dětí v mateřských školách, některé práce se zabývají i tím, jak hlasovou hygienu přiblížit dětem [50]. Hlasová hygiena tedy představuje jakousi šablonu toho, jak by se měl člověk chovat, aby zachoval svůj zdravý hlas. Nejedná se pouze o doporučení, co by člověk pro kvalitu a zdraví svého hlasu měl dělat. Výchova ke zdravému hlasu má za úkol rovněž informovat o faktorech, které jsou hlasu škodlivé a je lepší se jich vyvarovat. Tyto pravidla hlasové hygieny shrnuje tabulka 2.2.

Tabulka 2.2 – Zásady hlasové hygieny [51]

Hlasu prospívá	Hlasu škodí
kvalitní spánek	prašné, zakouřené a hlučné prostředí
měkké hlasové začátky	alkohol, kouření, návykové látky
hlasový klid (při obtížích)	přetěžování (křik)
správné držení těla a dýchání	nevyléčené infekce dýchacích cest
procvičení hlasivek před projevem	ostrá a dráždivá jídla

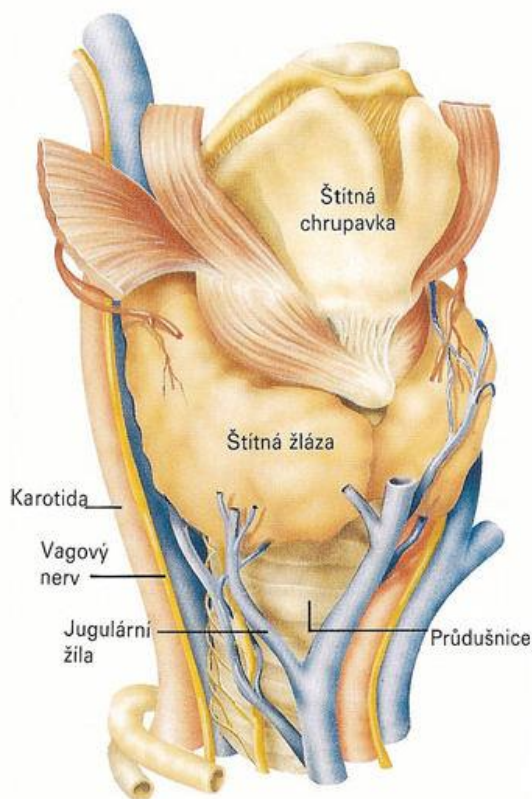
2.4.2 Hlasová reedukace

Zatímco hlasová edukace je určena lidem se zdravým hlasem i fonačním ústrojím, techniky hlasové rehabilitace a reedukace mají sloužit výhradně pacientům s organickým postižením hlasového aparátu, pacientům po chirurgických zákrocích v této oblasti či pacientům, jejichž hlas vykazuje známky patologických jevů. V první řadě je nutno zmínit, že postup není vždy stejný a u každého pacienta se může lišit. Je třeba postupovat individuálně v závislosti na charakteru potíží a míře postižení hlasu. Při sestavování individuálního reedukačního plánu se podle [18] hodnotí:

- chování – temperament a psychický stav pacienta,
- tempo řeči,
- držení těla, napětí krčních svalů a pohyby těla při verbální komunikaci,
- dýchání – způsob nadechování a vydechování, fonační čas,
- hlas – tvrdé hlasové začátky, napětí v hlase, spasticita, stupeň hlasové dyšnosti
- lokální nález – napětí a hybnost hlasivek, organické změny.

Součástí hlasové reedukace tedy nejsou pouze fonační cvičení, ale jedná se o mnohem komplexnější proces. Terapeut by měl tedy pacienta instruovat, jaký je správný

ze tří částí. Největší část tvoří levý a pravý lalok, které jsou spojeny žláznatým můstkem (istmem). Laloky jsou ovoidního tvaru. Pro lepší představu je anatomická struktura ukázána na obrázku 3.1.

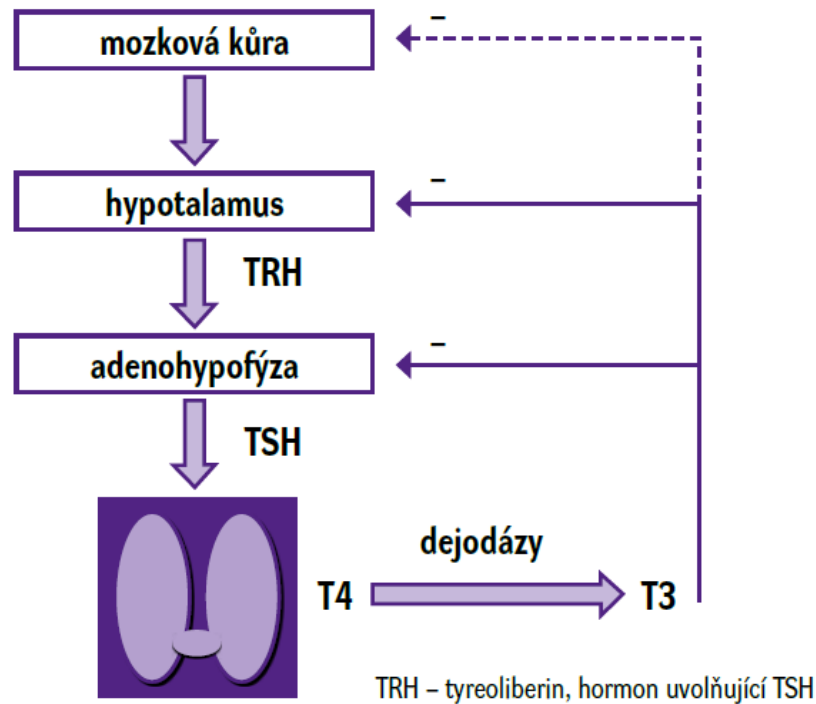


Obrázek 3.1 – Uložení štítné žlázy v přední části průdušnice [55]

Jedná se o žlázu s vnitřní sekrecí (v literatuře též endokrinní), která produkuje dvojici hormonů. První z nich se nazývá tyroxin a v literatuře se též označuje jako T4. Druhým produkovaným hormonem je trijodtyronin, který se zkráceně značí T3. Právě pomocí těchto hormonů dokáže štítná žláza ovlivňovat procesy v našem organismu. Aby však byla zajištěna její správná činnost, je zapotřebí dostatek jódu. V období po druhé světové válce se jód začal celoplošně přidávat do soli, aby populace netrpěla jeho nedostatkem. Jód si pak štítná žláza bere přímo z krve a distribuuje jej dále do hormonů, které vytváří. Značení T4 také znamená, že hormon obsahuje 4 atomy jódu. Působením enzymů zvaných dejodázy na hormon T4 se jeden atom odštěpí a vzniká T3. Pro správnou činnost dejodáz je nutný selen.

Činnost štítné žlázy (tvorba hormonů) je řízena mozkiem, přičemž se zapojuje zejména hypotalamus a přední lalok podvěsku mozkového (adenohypofýza). Hypotalamus vytváří hormon tyreoliberin (TRH) a adenohypofýza pak působí přímo na štítnou žlázu tyreoidálním stimulačním hormonem (TSH), též zvaný tyreotropin. Uplatňuje se zde

princip záporné (negativní) zpětné vazby. Trpí-li organismus nedostatkem T3 a T4, úroveň TSH v krvi roste. Výsledkem pak bude zvýšení produkce těchto hormonů. Podobně to platí i obráceně. Pokud má tělo hormonů nadbytek, hodnota TSH se sníží, což má za následek i snížení produkce hormonů. TSH je proto velmi citlivý a jednoduše měřitelný indikátor stavu štítné žlázy, stanovený na základě odběru krve. Celý proces, jakým mozek řídí činnost štítné žlázy je schematicky znázorněn na obrázku 3.2.



Obrázek 3.2 – Řízení činnosti štítné žlázy dle principu negativní zpětné vazby [53]

Produkované hormony ovlivňují látkovou výměnu. Tím jsou myšleny všechny metabolické děje v organismu. Nadměrná produkce hormonů způsobuje to, že pacienti vnímají zrychlený srdeční rytmus, jsou červení v obličeji, točí se jim hlava a jsou podráždění [56]. V pokročilejších fázích pak trpí průjmem a rapidním snižováním tělesné hmotnosti obvykle o 5-10 kg ale někdy i více. Lidé, kteří mají produkci hormonů štítné žlázy nižší, se potýkají naopak potýkají s nárůstem váhy, který ale není tak markantní. Pohybuje se v hodnotách do 5 kg. Další příznaky, které na sobě mohou tyto lidé pozorovat, jsou např. nechut k jídlu, pocit nedostatku energie a vyčerpanosti, což vede k nadměrnému spánku. Těmto pacientům bývá často zima a obecně lze říct, že jsou velmi pasivní a neprožívají plnohodnotný život.

Mimo vazby na metabolismus mají produkované hormony prokazatelné oběhové účinky. Štítná žláza totiž primárně ovlivňuje srdce, jehož tep upravuje na základě množství hormonů obsažených v krvi. Působením většího než normálního množství produkovaných

hormonů na srdce, dochází ke zrychlení srdečního rytmu, které je pacienty vnímáno jako nepravidelné srdeční akce či bušení srdce. Pokud je hormonů málo srdeční pulze se zpomaluje.

Neméně důležitou roli hraje štítná žláza v období těhotenství. Přiměřené množství hormonů je naprosto klíčové pro vývoj kostí, svalů, nervů a mozku dítěte. Pro správnou funkci všech zmíněných částí lidského těla, pak štítná žláza přispívá během celého života jedince. Ať už se jedná o nadbytečnou či nedostačující produkci hormonů, je nutná léčba. Obě varianty mají dopad na zdravotní stav pacienta a pokud jsou neléčené, mohou například vyústit až v selhání srdce. Proto jsou-li hodnoty hormonů mimo normální rozsah, jedná se z hlediska medicíny o nemoc štítné žlázy. Odborně se nemoci štítné žlázy označují jako tyreopatie.

3.2 Nemoci štítné žlázy (tyreopatie)

Obecně lze říci, že choroby štítné žlázy jsou poměrně frekventované. Uvádí se, že jimi trpí minimálně 5-10% populace, což rozhodně není zanedbatelné číslo. Zarážející je přitom výskyt tyreopatie ve srovnání žen a mužů. U žen se choroby štítné žlázy vyskytují až osmkrát častěji než u mužů. Odvětví, jenž se specializuje na onemocnění obecně endokrinních žláz, se nazývá endokrinologie. Endokrinolog určí (např. na základě hodnoty TSH) hladinu hormonů v krvi pacienta. Má-li pacient sníženou produkci hormonů, pak trpí hypothyreózou. Naopak je-li funkce štítné žlázy zvýšená, toto onemocnění se nazývá hypertyreóza (někdy též tyreotoxikóza). Nicméně o onemocnění štítné žlázy může být řeč i při její normální funkci (tzv. eutyreóza či eufunkce). Podobně jako na hlasívkách se totiž i na štítné žláze mohou tvořit uzlíky či nádory. Navíc může jít o zvětšení štítné žlázy, jenž se odborně označuje jako struma. Jak uvádí [53], mezi rizikové faktory pro poruchy štítné žlázy se řadí:

- výskyt tyreopatie u pacienta již v minulosti či výskyt v rodině,
- struma či uzlíky dříve diagnostikované u pacienta nebo v rodině,
- cukrovka 1. typu, celiakie a jiné poruchy imunity v rodině či u pacienta,
- pozitivní protilátky proti štítné žláze (autoimunitní onemocnění),
- některé laboratorní nálezy (zvýšený cholesterol apod.),
- příznaky svědčící o možné poruše funkce štítné žlázy,
- ozařování,
- léčba určitými léky (amiodaron, biologická léčba, lithium),
- neplodnost, předčasný porod či prodělaný potrat.

K diagnostickým účelům se převážně využívají hodnoty TSH, ale také volné hormony T3 a T4, které se označují FT3 a FT4. U některých metod je rozhodující množství protilátek proti štítné žláze (TRAK či TPOAb). Onemocněními štítné žlázy se zabývá i celá řada akademických prací. Např. [57] uvádí i doporučení s ohledem na prevenci.

3.2.1 Hypotyreóza

Snížená produkce hormonů (hypotyreóza) je v České republice nejčastějším typem tyreopatie. Odhaduje se, že se vyskytuje až u 5 % populace. Přestože hypotyreóza může být získaná (tzn. v průběhu života), existuje i varianta vrozená, kvůli které se zavedlo plošné vyšetření u novorozenců označované jako screening. Jak již bylo zmíněno, štítná žláza potřebuje pro svoji funkci jód. Vlivem jeho většího nedostatku se u pacienta nemoc může vyvinout. Příčinou vzniku je ale typicky chronický zánět (chronická lymfocytární tyreoiditida). Jedná se o autoimunitní reakci organismu, jejíž příčina dosud nebyla zcela objasněna. Mezi příznaky hypotyreózy patří zimomřivost, únava, ospalost, bolesti svalů a svalová slabost, suchá kůže, padání vlasů či zpomalení srdeční činnosti. U žen způsobuje poruchy menstruačního cyklu a plodnosti. U mužů zapříčiňuje poruchy potence.

3.2.2 Hypertyreóza (tyreotoxikóza)

Hypertyreóza je opakem hypotyreózy a její výskyt se odhaduje zhruba na 0,2-1 % v naší populaci. Pokud jde o mladší generaci, nejběžněji nemoc zapříčiňuje Graves-Basedowova choroba (GB). Protože GB je autoimunitní onemocnění, v organismu se vytvářejí protilátky proti štítné žláze (TRAK), které podněcují štítnou žlázu k nadměrné tvorbě hormonů. Tato zvýšená činnost může být doprovázena zvětšením štítné žlázy (vznik strumy). U starší populace pak hypertyreózu způsobují převážně uzlíky. Při této nemoci pacienti pociťují pocení, bušení srdce, poruchy srdečního rytmu (často fibrilace síní), rychlou srdeční akci, nervozitu, nespavost, úbytek svalové hmoty či únavu. Často pacienti popisují pocity neustálého hladu, což někdy naopak vede ke zvýšení tělesné hmotnosti. Ženám způsobuje poruchy plodnosti a nepravidelnosti v menstruačním cyklu. U mužů se dostávají poruchy potence. Ve velmi vzácných případech může dokonce hypertyreóza vyvolaná GB chorobou způsobovat endokrinní orbitopatii, která působí na oči. Dochází při ní k slzení a pálení očí, vystupování očí z očních bez možnosti dovržení víček, což vede k vysychání rohovky. Někdy se též vlivem snížené hybnosti očí dostavuje efekt dvojitého vidění. V těch nejvážnějších případech vede až k úplné ztrátě zraku zapříčiněné utlačováním očního nervu.

3.2.3 Struma, uzlíky a nádory štítné žlázy

Jedním z patologických jevů, který lze u štítné žlázy pozorovat, je její zvětšení čili struma. Právě kvůli možnosti utlačování dalších struktur krku se často přistupuje k operaci. Nejedná se o funkční poruchu. Struma se může nacházet u pacienta s hypothyreózou, hypertyreózou, ale i eutyreózou (normální funkcí). Vzniky strumy může zapříčinit i mírný nedostatek jódu. Jsou-li na štítné žláze zároveň i uzly, pak je řeč o tzv. retrosternální strumě. Pokud zde žádný uzel není, struma se označuje jako difuzní. Struma je v pozdějších fázích dobře viditelná v oblasti krku. Jako ukázka může posloužit fotografie na obrázku 3.4.



Obrázek 3.4 – Viditelně zvětšená štítná žláza (struma) [58]

Pokud jde o uzlíky, ohroženou skupinou jsou převážně ženy ve věku nad 50 let, u kterých se uvádí výskyt ve 30-40 % případů. Uzly se většinou tvoří v místech, kde je nedostatek jódu. Pozitivem je, že naprostá většina uzlů je benigní. Uvádí se, že nezhoubných je 90-95 %. Uzlíky na štítné žláze pak mohou být příčinou vzniku hypertyreózy.

Poslední skupinou jsou maligní útvary na štítné žláze, které doplňují větší zastoupení benigních uzlů. Vyskytují se tedy zhruba v 5-10 % případů, kdy je takový útvar nalezen. Karcinomy štítné žlázy se dělí na diferencované a nediferencované. Častěji se objevující skupinou jsou diferencované, které jsou v raném stádiu dobře léčitelné. Nejběžnějším typem diferencovaného karcinomu je v rámci České republiky karcinom papilární. Méně časté jsou karcinomy nediferencované, u nichž je pacientova prognóza horší. U včasné nalezených a odstraněných maligních útvarů do velikosti 2 cm, je velká šance, že funkce štítné žlázy nebude vůbec ovlivněna. V lékařské praxi se označují termínem mikrokarcinomy. Mimo to je štítná žláza cílem metastáz karcinomu plic, ledvin a prsu.

3.3 Léčba onemocnění štítné žlázy

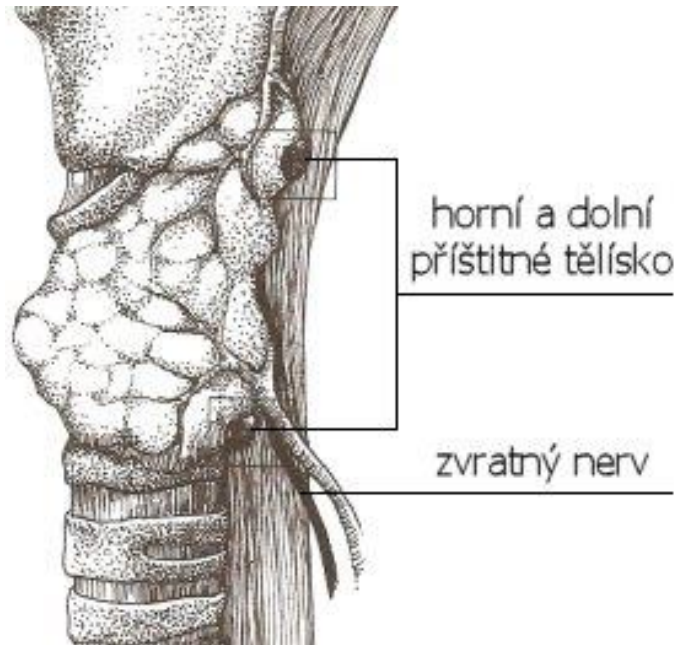
Léčba se samozřejmě odvíjí od typu onemocnění. Obecně lze ale říci, že rozhodující je včasná diagnostika nemoci. Ve většině včasné diagnostikovaných případů je pak léčba efektivní a bez trvalých následků. Speciálním případem je pak léčba v období těhotenství, kterou se ale práce nebude zabývat.

3.3.1 Operace štítné žlázy

Operačním zákrokům při onemocněních štítné žlázy se věnuje [59]. Operace bývá až tou poslední možností, kterou endokrinolog doporučuje, nemá-li jiné východisko. Předchází jí medikamentózní léčba, která upravuje hladinu hormonů v krvi. Přistupuje se k ní v případě strumy, zhoubného nádoru (resp. zvětšujícího se uzlu s podezřením na maligní nález, který nereaguje na léčbu) či hypertyreóze způsobené GB chorobou. Operační zákroky se dělí na dva druhy. Při tzv. hemithyreoidektomie (HTE) je pacientovi odebrán jeden lalok obsahující uzel a istmus (můstek), přičemž se zachová druhý lalok, který je zdravý. Zhruba ve 25 % případů se lalok, jenž zůstal zachován, zvětší a plně nahradí i chybějící část. Tito pacienti po zbytek života nemusí podstupovat žádnou léčbu, pokud se jejich stav nezmění. Ve většině případů je indikována substituční léčba, kdy je pacientovi podáván tyroxin. Druhým typem zákroku je totální thyreoidektomie (TTE), při níž se odstraní celá štítná žláza. Tito pacienti musí stejně jako většina těch, kteří prodělali HTE, po zbytek života podstoupit substituční léčbu. S aplikací pravidelných dávek tyroxinu se začíná hned následující den po zákroku, aby se nezačala rozvíjet pooperační hypotyreóza.

Zárok probíhá v celkové narkóze. V dolní části krku (na přední straně) se provede řez obvykle v rozsahu 3-7 cm. Velikost řezu závisí na rozměrech štítné žlázy, anatomických poměrech (délka a tloušťka krku), ale také na zkušenostech operátora. V některých případech lze k operaci přistoupit i endoskopicky, což zajišťuje pacientovi menší řez a následně i jizvu (zhruba 2,5-3 cm). Dále je nutné podvázat cévy vedoucí ke štítné žláze případně pouze k operovanému laloku. Operovaná část se uvolní tak, aby zůstala fixována již jen k průdušnici. A nyní přichází část, při které může docházet ke komplikacím. Je totiž nutné lokalizovat zvrtný nerv, jenž zajišťuje hybnost hlasivek. Jedním z cílů operace je dbát na to, aby zvrtný nerv nebyl při zákroku porušen. Nerv se vypreparuje podél průdušnice až do místa, kde vstupuje do hrtanu. Druhá riziková část přichází vzápětí. Je nutné identifikovat tzv. příštítná tělíska. Jejich poškození by opět vedlo ke zbytečným komplikacím. Pro představu je umístění zvrtného nervu a příštítných tělísek

pravého laloku zobrazeno na obrázku 3.5. Jakmile jsou všechny rizikové oblasti lokalizovány, může být bezpečně přerušen úpon žlázy k průdušnici. V případě odstranění celé štítné žlázy se tento postup aplikuje na obou stranách (tedy pro oba laloky). Do rány se zavede drenážní hadička k odsávání nahromaděných tekutin, které by zpomalovali proces hojení. Rána se poté zašije. Drenáž bývá 2. den po operaci odstraněna.



Obrázek 3.5 – Hlavní rizika při operaci štítné žlázy [59]

Pokud jde o možné komplikace, které byly již naznačeny výše, jejich výskyt není v současnosti nijak častý. Je to však hlavní důvod obav pacientů, kteří mají operaci podstoupit. Je to dáno zejména historicky, kdy byly operace štítné žlázy často neúspěšné a pacient na následky operace zemřel, respektive čelil pooperačním komplikacím. Historicky první zákrok je datován do roku 952, avšak téměř celé následující tisíciletí byla bilance úspěšnosti zákroků velmi špatná, jak uvádí [60]. Ke zlepšení došlo až v druhé polovině 19. století, kdy vznikly první zásady relativně bezpečné tyreoidektomie.

V současné době se riziko poranění zvratného nervu pohybuje okolo 3 %. Používají se různé nejmodernější metody a technologie ke zviditelnění nervu. Například je možné do hlasivky během operace umístit jehličkovou elektrodu, který bude během operace snímat valovou činnost hlasivky. Dále se pak používá speciální stimulátor, který při stimulaci zvratného nervu rozpohybuje hlasivku a dokáže tak na přítomnost nervu upozornit. Někdy je však omezena hybnost hlasivek i při dodržení celistvosti nervu, což se v průběhu několika měsíců napraví. Jsou-li patrné změny hlasu k horšímu, může pacient podstoupit foniatrické rehabilitace. U komplikovaných zákroků, kdy došlo k výrazným zánětlivým

změnám, na štítné žláze se nachází větší zhoubný nádor nebo se jedná o druhou či další operaci, je riziko i riziko trvalého poškození vlivem nepřehlednosti operačního terénu. Situaci komplikuje zjizvení a fixace nervu k nádoru. Proto je zhruba v 0,5-2 % případů poškození zvrátelného nervu trvalé.

Další komplikací může být poškození příštítných tělísek. Odhaduje se, že při TTE dochází ke snížení hodnoty vápníku v krvi pacienta zhruba ve 25-35 % případů. Pacient pak musí dočasně doplňovat vápník např. ve formě šumivých tablet. Ve 2-5 % dojde k poranění nebo úplnému odstranění příštítných tělísek, což vyústí v v jejich trvalou poruchu funkce. Pacient pak doživotně užívá léky, které poruchu kompenzují. Zde se nicméně nejedná o přímo vazbu na tvorbu hlasu. Možnými komplikacemi, které mohou po zákroku nastat se dále věnuje [61].

3.3.2 Aplikace radioaktivního jódu

S tímto druhem léčby se začalo již v roce 1942. Pacientovi je podán radioaktivní jód ve formě tablet, takže léčba může probíhat ambulantně na oddělení nukleární medicíny. Radiojód se zachytává přímo ve štítné žláze, kde působí, a ozáření zbytku organismu je tak zanedbatelné. Buňky štítné žlázy, které radioaktivní jód zachytily, jsou zničeny zářením. Představuje tak neinvazivní alternativu pro GB chorobu či uzly způsobující hypertyreózu. Po podání radiojodu se typicky v horizontu šesti měsíců utlumí funkce štítné žlázy a může dojít k jejímu zmenšení. Tento proces ale může trvat i několik let. Proto je třeba doživotně navštěvovat lékaře kvůli pravidelným kontrolním vyšetřením. Je-li u pacienta podezření na maligní nález nebo trpí-li pacient endokrinní orbitopatií, nesmí se léčba radiojodem použít. Aplikace je zakázána i u dětí, těhotných žen a dospívajících do 18 let.

4 Cíle práce

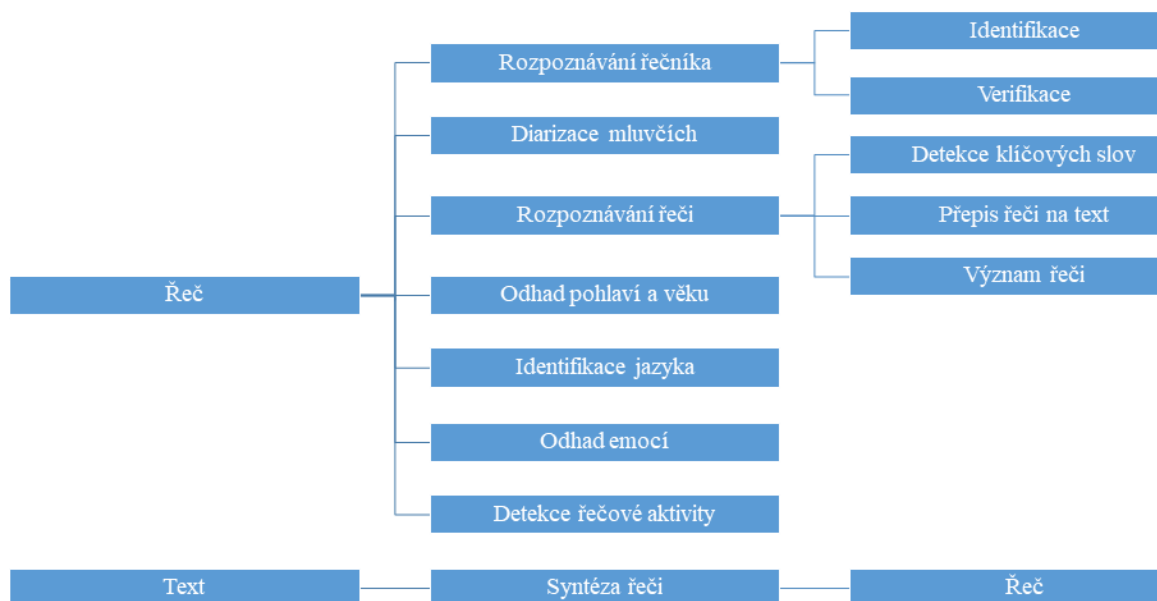
Práce si klade za cíl dosažení teoretických i praktických poznatků v oblasti zkoumání onemocnění štítné žlázy a dopadů chirurgických zákroků při onemocněních štítné žlázy na tvorbu hlasu. Hlavními body, na něž se tato práce zaměřuje, jsou:

- tvorba hlasu, možné poruchy hlasu, jejich příčiny a prevence před vznikem těchto poruch,
- souhrn běžně aplikovaných diagnostických metod, přičemž je kladen důraz na fakt, že většina diagnostických postupů je přímo subjektivně ovlivněna zkušenostmi lékaře,

- onemocnění štítné žlázy v návaznosti na tvorbu hlasu a popis rizik během výkonu operace,
- navržení obecné metodiky pro sběr nahrávek pacientů s hlasovými obtížemi,
- popis klíčových parametrů pro volbu měřicího řetězce – zejména pak mikrofonu,
- ucelený přehled charakteristik řeči a popis extrakce jednotlivých parametrů z řečového signálu,
- hledání objektivních parametrů pro posouzení změn hlasu v důsledku podstoupení operace štítné žlázy,
- vytvoření aplikace, která bude spolupracovat s definovaným měřicím řetězcem a která nabídne lékařům možnost vkládat údaje o pacientech, přehledně pracovat s již uloženými záznamy, exportovat vložené údaje pro další zpracování mimo aplikaci, kalibrovat zařízení dle požadavků nezbytných pro nahrávání, zaznamenávat hlas pacientů, zobrazovat výsledky extrakčního algoritmu a objektivně posuzovat změny hlasu, které u pacienta po chirurgickém zákroku nastaly.

5 Zpracování řeči

Analýza hlasu je v současnosti velmi širokým pojmem. Existuje celá řada aplikací využívající extrakci informací obsažených v řeči, jak znázorňuje schématické dělení na obrázku 5.1. O to složitější je zorientovat se v dané problematice, protože jak uvádí [62], prolíná se zde hned několik vědních oborů, které ani nemusí být na první pohled patrné. Některé obory zde mohou být zastoupeny jen nepatrnou mírou, jiné jsou stěžejní pro kvalitní provedení takovéto analýzy. Je zcela zřejmé, že k porozumění problematice je naprosto nezbytné využití znalostí z matematiky, statistiky, zpracování signálů, informatiky, strojového učení, umělé inteligence, biologie, fonetiky, lingvistiky a v některých případech například i psychologie či medicíny. Soubor nutných znalostí vždy závisí na úhlu pohledu a na způsobu řešení zvoleného problému. V současnosti je například snaha o eliminaci veškerých rušivých vlivů i nepříznivých podmínek a omezení pouze na takovou část informací, která povede k úspěšné klasifikaci. Systémy oproštěné od přebytečných dat nesoucích informaci například o přenosovém kanálu mohou pak lépe a přesněji vyhodnocovat dané hypotézy a jsou navíc i robustnější. Dosažení co nejvyšší robustnosti vyvíjených systémů je aktuálně velkou vědeckou výzvou.



Obrázek 5.1 – Dělení počítačového zpracování řeči

Příkladem využití řečových technologií ve zdravotnických zařízeních může být [63]. Existuje samozřejmě mnoho aplikací jako například dialogové systémy, které běžně využívají hned několik odvětví. Podobně tomu je i v této poměrně netypické úloze diagnostiky poruchy hlasu, kde však nebudou využity jednotlivé funkční celky, ale pouze jejich části. Budou použity tytéž parametry, které se například používají pro detekci řečové aktivity, jen místo toho poslouží pro diagnostické účely.

5.1 Obecná metodika zpracování řečového signálu

Obecně lze práci se zvukovými či řečovými soubory rozdělit na čtyři základní kroky. Nepočítaje převod zvukového signálu na číslicový, který provádí mikrofon a zvuková karta počítače, signál bývá nejprve tzv. předzpracován a následně segmentován. Jednotlivé segmenty se pak parametrizují a podle obdržených parametrů je možné klasifikovat. Zpracováním řeči s využitím neuronových sítí a HMM se zabývá např. [64].



Obrázek 5.2 – Promluva „Vážení spoluobčané“ z projevu V. Havla v roce 1991 [65]

5.1.1 Předzpracování

Předzpracování řeči má v současnosti spíše historický význam. V některých případech však může být užitečné a vést k lepším výsledkům, což je důvodem, proč se stále

ještě v literatuře uvádí. Hlavním přínosem předzpracování řeči je odstranění stejnosměrné složky signálu, která nenesे žádnou užitečnou informaci z pohledu zpracování řeči. Naopak může komplikovat vyhodnocení. Je-li znám celý signál, stejnosměrnou složku lze potom odstranit jejím odečtením od původního signálu. Tento proces se také někdy označuje jako tzv. ustředění signálu, které je dáno vzorcem

$$x'(n) = x(n) - \frac{1}{N} \sum_{k=1}^N x(k), \quad (5.1)$$

- kde: – $x(n)$ je původní signál,
 – $x'(n)$ označuje ustředěný signál po odebrání stejnosměrné složky,
 – N udává celkový počet vzorků.

Problém dosažení rovnoměrného spektra řeči efektivně řeší aplikace tzv. preemfázového filtru, který zvýrazňuje vyšší harmonické složky řeči. Jedná se o jednoduchý derivační filtr typu horní propust. Přenosová funkce filtru je dána vztahem

$$H(z) = 1 - \alpha z^{-1}, \quad (5.2)$$

- kde: – $H(z)$ je přenosová funkce filtru,
 – α představuje koeficient filtru volený z hodnot 0,9 až 1.

Kromě těchto úprav signálu může být doplněn ještě jeden praktický krok. Některé zvukové karty totiž způsobují, že na začátku a konci nahrávky doplní nulovými hodnotami. To následně způsobí problém, budou-li použity metody, které pro práci se signálem využívají dělení či logaritmování. Je proto vhodné doplnit záznam o malý vygenerovaný šum. Hodnoty vzorků nabývají hodnot až 2^{15} , a proto velmi malé hodnoty zpracování nijak neovlivní.

5.1.2 Segmentace

S řečovým signálem se v oblasti zpracování řeči typicky nepracuje v celé jeho délce. Výjimkou jsou pouze některé neuronové sítě využívané ke klasifikaci, kde se na místo zdlouhavého procesu extrakce příznaků obsažených v řeči neuronové sítě předkládají jako trénovací vzory přímo celé (raw) nahrávky, které mohou být uloženy jako audio soubory (kupříkladu s příponou wav). Nicméně při běžném zpracování řeči se segmentace již stala naprostým standardem.

Běžně se tedy signál rozdělí na krátké úseky, které jsou obecně v oblasti zpracování a analýzy zvukových dat nazývány rámce (v angličtině frames). Každý takový rámec se následně zpracovává zvlášť. Hlavním důvodem dělení je fakt, že řeč se může jevit jako

náhodný signál. V případě vymezení dostatečně krátkého úseku řeči lze však signál považovat za kvazy-stacionární a lze tedy aplikovat metody, které by pro celou nahrávku nemohly být použity. Délka rámce se nejčastěji volí v intervalu 10 až 30 ms. Tyto intervalové hodnoty jsou stanoveny v souladu s délkou jednotlivých fonémů, a proto není doporučeno volit hodnoty mimo interval, kdy by mohlo docházet k tomu, že rámec je kratší než daný foném a nedokáže jej tudíž obsáhnout celý.

Tabulka 5.1 – Hodnoty použité při segmentaci v disertační práci

Vzorkovací frekvence	Délka rámce	Překryv rámců
16 kHz	20 ms (320 vzorků)	10 ms (160 vzorků)

Stále je třeba mít na paměti, že signál je diskrétní, a tudíž v každém rámci bude obsažen stejný počet vzorků. Volba délky rámce může být také ovlivněna vzorkovací frekvencí původního signálu. Například bude-li originální audio stopa vzorkována na frekvenci 8 kHz (tj. 8000 vzorků za vteřinu) a délka rámce bude 25 ms, pak každý rámec bude obsahovat právě 200 vzorků původního signálu. Konkrétní hodnoty použité v rámci této práce jsou uvedeny v tabulce 5.1. Přepočet časové délky rámce na vzorky lze snadno realizovat pomocí vzorce

$$L_{VZ} = f_{VZ} \cdot L_t, \quad (5.3)$$

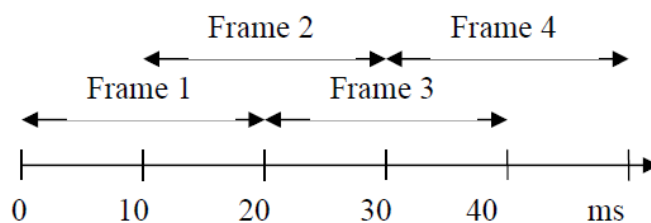
kde: – f_{VZ} označuje vzorkovací frekvenci původního signálu v Hz,

– L_t představuje délku rámce uváděnou v jednotkách času (sekundách),

– L_{VZ} je délka rámce uváděná ve formě počtu vzorků, tedy bezrozměrného čísla.

Při striktním dělení signálu na stejně dlouhé rámce, jenž na sebe navazují (po posledním vzorku jednoho rámce následuje první vzorek následujícího rámce), by ovšem mohlo způsobit, že některé jevy vyskytující se na rozhraní těchto rámců nebude možné identifikovat. Proto se v moderních metodách zpracování řeči často přistupuje k tomu, že se rámce navzájem překrývají. Tak bude zajištěno, že děje a jevy odehrávající se na okraji některého rámce budou celistvěji zachyceny v rámci, se kterým se tento původní překrývá.

Jinak řečeno, vzniká zde duplicita vzorků (příslušnost vždy alespoň do 2 rámců vyjma počátku a konce nahrávky), která vykompenzuje dělení signálu na velmi krátké úseky bez ohledu na jeho obsah. Příklad takového dělení včetně číslování rámců je zachycen na obrázku 5.3.



Obrázek 5.3 – Dělení signálu na rámce délky 20 ms s překryvem 10 ms

5.1.1 Parametrizace

Pod pojmem parametrizace si lze představit určité zjednodušení či substituci původního signálu. Tento krok bezpodmínečně navazuje na předchozí. To znamená, že není parametrizovaný celý původní signál, ale pouze jeho úseky obsažené v rámcích. Pro každý rámeček se vypočtou zvolené parametry. Těmto parametrům se ve zpracování řeči říká příznaky (v angličtině features). To je velká změna oproti dřívějšímu přístupu, kde byl původní signál reprezentován konečným počtem hodnot získaných vzorkováním řečového signálu. Tyto vzorky jsou použity pro výpočet daných příznaků pro každý rámeček a dále se s nimi již nepracuje. Je-li tedy nahrávka segmentována například do padesáti rámečků a je zvolen pouze jeden příznak pro každý rámeček, pak bude nahrávka reprezentována pomocí padesáti hodnot daného příznaku. Tím se významně sníží i výpočetní náročnost potřebná při práci se zvukovým záznamem, který je již reprezentován pouze příznaky. Náročnost výpočtu příznaků je samozřejmě mnohem vyšší.

Je-li příznaků více než jeden, zapisují se do tzv. příznakových vektorů (anglicky feature vectors). Každému rámečku pak odpovídá právě jeden příznakový vektor. Dimenze všech těchto vektorů je stejná, neboť pro každý rámeček jsou vypočítávány hodnoty pro stejné příznaky. Řazení příznaků v příznakových vektorech bývá následující

$$\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_L \end{pmatrix}, \quad (5.4)$$

kde: – \mathbf{p} je příznakový vektor,

– $p_1 \dots p_L$ jsou hodnoty jednotlivých příznaků,

– L představuje dimenzi příznakového vektoru čili počet určovaných příznaků.

Nahrávka jako celek může být tedy zastoupena posloupností příznakových vektorů, které lze v rámci dalšího zpracování spojit do matice, kdy ve sloupcích jsou jednotlivé příznakové vektory tak, jak jdou v nahrávce časově za sebou

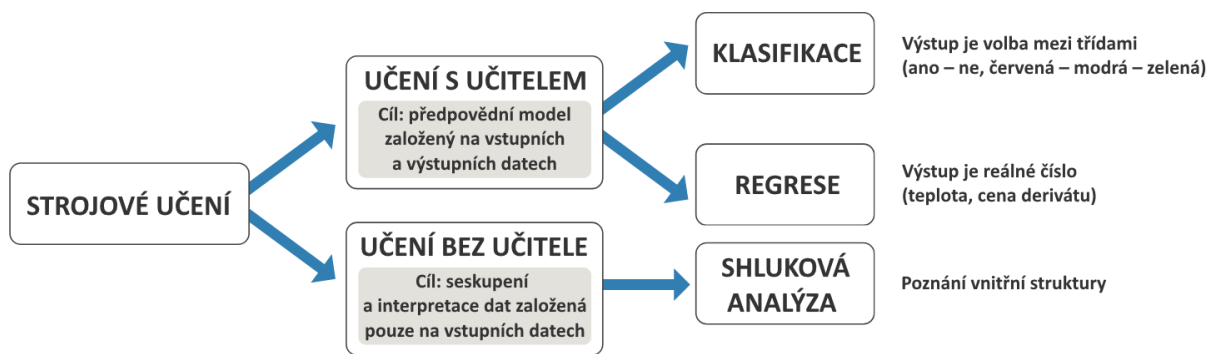
$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{L1} & p_{L2} & \cdots & p_{LK} \end{pmatrix}, \quad (5.5)$$

kde: – \mathbf{P} je příznaková matice charakterizující celý signál po sloupcích,
– K je počet rámců, na které byl analyzovaný signál rozdělen.

Jednotlivé hodnoty příznaků odpovídají charakteru signálu a souvisí tak například s hlasitostí, frekvencí či spektrem řeči. Konkrétním druhům příznaků, jejich popisu a výpočtem ze vzorků jednotlivých rámců je věnována kapitola 4.2.

5.1.2 Klasifikace

Je-li analyzovaná řeč předzpracována, segmentována a parametrizována, přichází na řadu klasifikace. Klasifikace jako taková spadá do oboru strojového učení. Základní rozdělení strojového učení je na obrázku 5.4.



Obrázek 5.4 – Základní dělení strojového učení [66]

Pod pojmem klasifikace se rozumí přiřazení klasifikovaných objektů do jednotlivých tříd. Rozhodnutí o příslušnosti do daných tříd má na starosti klasifikační algoritmus. Objekty se do tříd přiřazují na základě svých atributů, rysů či vlastností. K tomu, aby se klasifikační algoritmus (klasifikátor) naučil správně přiřazovat vstupní data do tříd, slouží množina trénovacích dat někdy též nazývaná trénovací vzory. Tato trénovací data jsou (při učení s učitelem) označena, což znamená, že jejich příslušnost do tříd je již známa. Poté se klasifikační algoritmus verifikuje pomocí testovacích dat, přičemž existuje poučka, která

nabádá k tomu, aby se jako množina testovacích dat nikdy nepoužívaly trénovací vzory. Obecně se rozlišují generativní a diskriminativní klasifikátory [67].

V aplikacích zpracovávajících nejprve izolovaná slova se využívala tehdy přelomová metoda tzv. borcení času (DTW). Později se pro klasifikaci řečových segmentů nejvíce osvědčili tzv. skryté Markovovy modely (HMM) a v poslední době jsou to zejména neuronové sítě (NN).

5.2 Metody a postupy uplatňované při zpracování řeči a extrakci příznaků

Ke zpracování řeči a extrakci příznaků nelze přistupovat bez řádných znalostí teoretického pozadí zpracování signálů. Proto jsou v této kapitole zmíněny některé klíčové myšlenky, techniky a potřebné teoretické znalosti pro lepší pochopení celé problematiky zpracování řeči. Tyto poznatky velmi přehledně prezentuje studijní opora pro zpracování řečových signálů z VUT v Brně [68]. Zbylé postupy využívané pouze pro extrakci určitých příznaků jsou uvedeny přímo u jednotlivých příznaků.

5.2.1 Frekvenční spektrum

Je-li řeč o spektru signálu, jedná se o odlišnou interpretaci původního signálu v časové oblasti. Tento signál je pak transformován do oblasti frekvenční. V zásadě lze každý signál rozdělit na harmonické složky, které jej tvoří. Těchto složek může být teoreticky až nekonečno. Každá harmonická složka signálu je charakterizována svojí frekvencí (případně periodou). Jednotlivé frekvence se pak promítnou do výsledného frekvenčního spektra signálu. Grafická reprezentace se pak taktéž nazývá spektrum. Alternativně se v literatuře používá také název periodogram. K přechodu z časové domény na frekvenční se používá tzv. Fourierova transformace (FT). Vše o FT a nezbytných teoretických souvislostech uvádí [69]. V integrální podobě je FT dána

$$F(i\omega) = \int_{-\infty}^{+\infty} f(t) \cdot e^{-i\omega t} dt, \quad (5.6)$$

kde: – $F(i\omega)$ představuje Fourierův obraz,

– $f(t)$ značí původní funkci před transformací, někdy též nazývanou originál,

– ω je úhlová rychlost udávaná v $\text{rad} \cdot \text{s}^{-1}$

– t označuje čas v s.

Mimo FT existují i jiné druhy transformací. Příkladem může být Laplaceova transformace, která má s FT mnoho společného. Lze dokonce přímo přepočítávat obraz jedné transformace na druhou [70]. Transformace může samozřejmě probíhat i v opačném směru čili z obrazu na originál. V tomto případě se pak hovoří o tzv. inverzní Fourierově transformaci (IFT). IFT je tedy definována následujícím způsobem

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(i\omega) \cdot e^{i\omega t} d\omega. \quad (5.7)$$

Pro úplnost je třeba dodat, že výše uvedená formulace odpovídá spíše matematice. V elektrotechnice a v oborech zabývajících se analýzou a zpracováním libovolných signálů se místo označení obraz užívá název spektrum. V literatuře se tak lze mimo jiné setkat s označením Fourierovo či komplexní spektrum. Transformace originálu na obraz tedy jinak řečeno znamená převod funkce z časové do frekvenční oblasti. Pro jednoduchost se často rovněž ve značení vynechává komplexní jednotka a spektrum signálu se tak v praxi označuje pouze $F(\omega)$. A protože spektrum je komplexní veličina, lze jej podobně jako kterékoliv jiné komplexní číslo zapsat v goniometrickém tvaru, jehož části jsou nazvány podle toho, jaký druh informace nesou. Platí tedy, že

$$F(\omega) = |F(\omega)| \cdot e^{i \arg F(\omega)}, \quad (5.8)$$

- kde: – $F(\omega)$ je Fourierův obraz čili Fourierovo (komplexní) spektrum,
– $|F(\omega)|$ se nazývá amplitudové spektrum,
– $\arg F(\omega)$ nese název fázové spektrum.

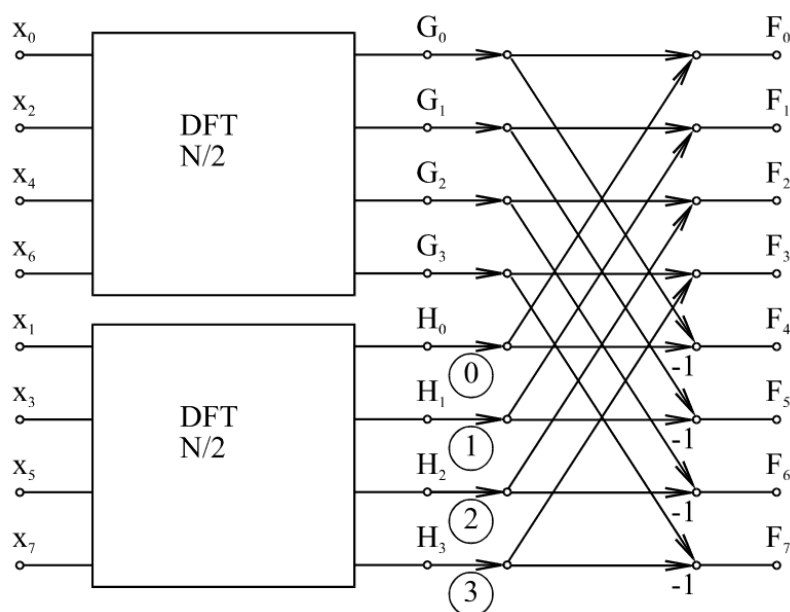
Pro úplnost je vhodné doplnit, že někdy se pracuje s výkonovým spektrem (v angličtině power spectrum). To je dáno jako druhá mocnina amplitudového spektra $|F(\omega)|$. V případě reálně zaznamenaného (vzorkovaného) signálu se však využije Diskrétní Fourierova transformace (DFT). Tu lze také s využitím Eulerova vztahu přepsat do tvaru

$$F(k) = \sum_{n=0}^{N-1} f(n) \cdot e^{-\frac{i2\pi nk}{N}} = \sum_{n=0}^{N-1} f(n) \left[\cos\left(\frac{2\pi nk}{N}\right) - i \sin\left(\frac{2\pi nk}{N}\right) \right], \quad (5.9)$$

- kde: – $F(k)$ je výsledkem N -bodové DFT,

- $f(n)$ představuje původní diskrétní signál před transformací,
- N určuje počet bodů pro výpočet DFT,
- n a k jsou pořadová čísla vzorků $n, k = 0; 1; \dots; N-1$.

Pro zjednodušení výpočetních nároků se pak zavedla zjednodušená varianta, kterou publikoval v roce 1965 Cooley a Tukey [71]. Pro tento algoritmus výpočtu DFT se ustálil název Rychlá Fourierova transformace (FFT). O jejímž využití pojednává např. [72]. Urychlení při použití FFT spočívá v tom, že se zpracovávají výhradně takové počty vzorků, které jsou typicky rovny mocnině 2^N . Hlavními zástupci algoritmů FFT jsou DIT (decimation in time) a DIF (decimation in frequency). Postup výpočtu FFT při použití algoritmu DIT je na znázorněn na obrázku 5.5.



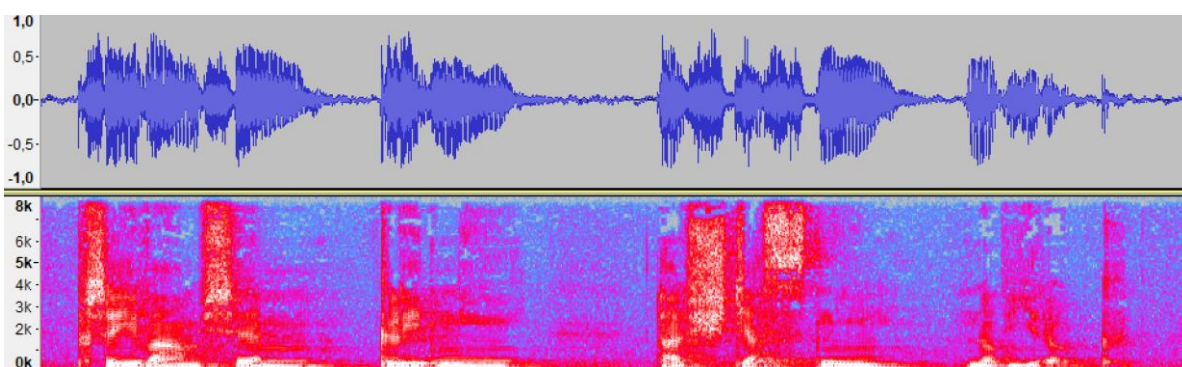
Obrázek 5.5 – Postup výpočtu FFT s využitím algoritmu DIT pro decimování v čase [72]

Základem je rozdělení prvků x na sudé a liché soubory, pro které se pak DFT počítá zvlášť. Obrazy sudých prvků jsou označeny G u lichých je značení H , přičemž výsledek FFT jsou pak hodnoty F . Jednotlivé varianty blíže popisuje např. [73] nebo [74]. Existují i varianty algoritmu FFT, které využívají jiný základ než 2, ty ale nebudou v rámci práce uvažovány. Efektivnost výpočtu při použití FFT z hlediska snížení počtu komplexních operací sčítání (PKS) a násobení (PKN) demonstruje tabulka 5.2. Při vzorkovací frekvenci 16 kHz a délce rámce 20 ms je počet vzorků v rámci roven 320 (použito v práci). Při výpočtu se tedy volilo $N=9$, což odpovídá nejbližší vyšší mocnině, tedy 512 prvkům.

Tabulka 5.2 – Porovnání výpočetní náročnosti (počet operací) DFT a FFT [69]

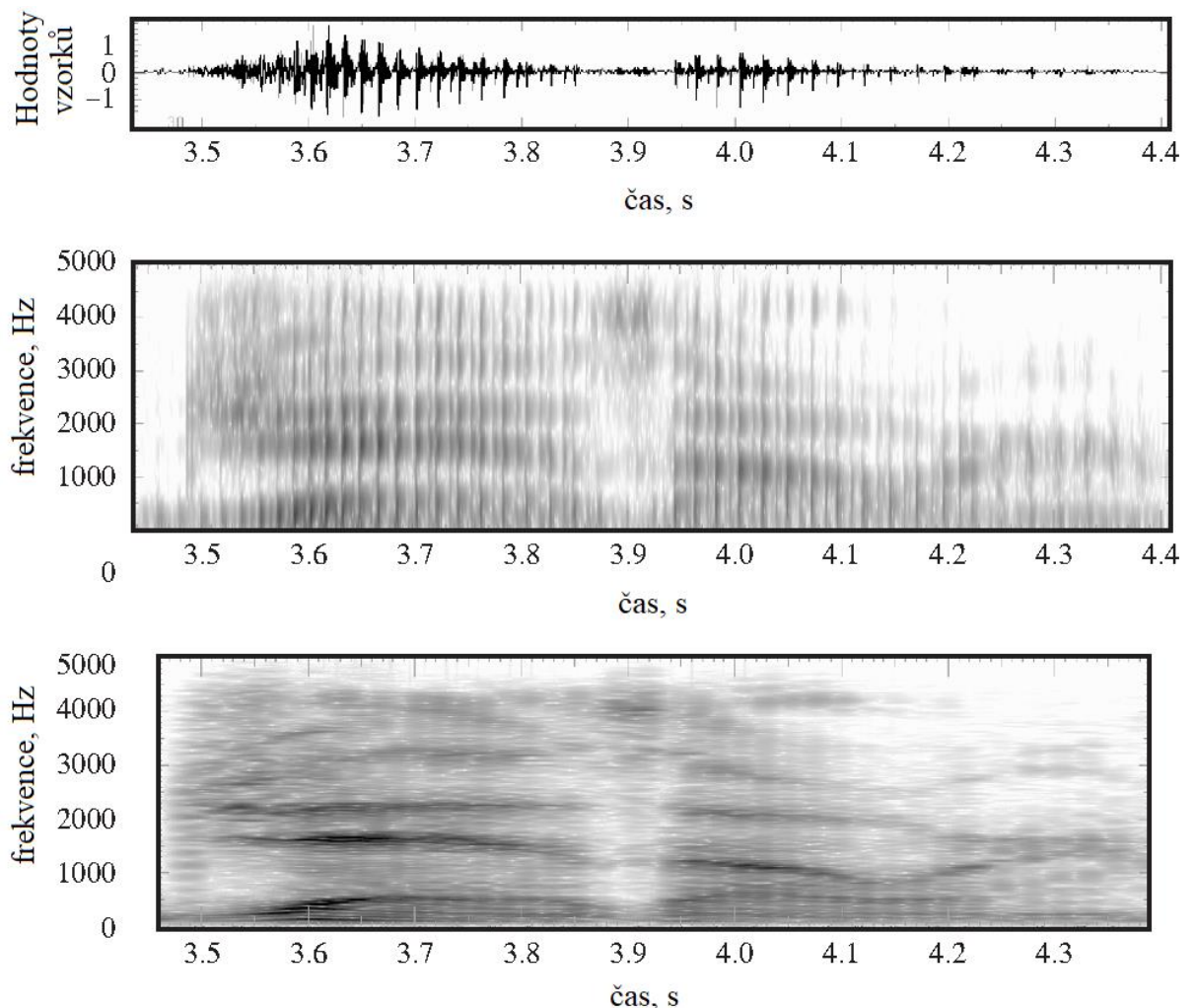
N	2^N	DFT		FFT			
		PKN	PKS	PKN	PKS	úspora PKN	úspora PKS
1	2	4	2	1	2	75,0 %	0,0 %
2	4	16	12	4	8	75,0 %	33,3 %
3	8	64	56	12	24	81,3 %	57,1 %
4	16	256	240	32	64	87,5 %	73,3 %
5	32	1024	992	80	160	92,2 %	83,9 %
6	64	4096	4032	192	384	95,3 %	90,5 %
7	128	16384	16256	448	896	97,3 %	94,5 %
8	256	65536	65280	1024	2048	98,4 %	96,9 %
9	512	262144	261632	2304	4608	99,1 %	98,2 %
10	1024	1048576	1047552	5120	10240	99,5 %	99,0 %

Spektrum (periodogram) má v praxi svá omezení. Hlavním nedostatkem je statická reprezentace spektra. Ve skutečnosti se však charakter analyzovaného signálu typicky v čase mění. Řeč je toho názorným příkladem. Při tvorbě každé hlásky se přestavuje artikulační trakt a moduluje tak zvuk hlasivek do výsledné podoby. Kromě změn frekvence při tvorbě řeči mohou z okolí působit různé druhy šumů a ruchů, které jsou také proměnlivé. V nahrávce jsou pak všechny tyto jevy zachyceny. Místo statického periodogramu se tak více využívá tzv. spektrogram. Při tvorbě spektrogramu se na horizontální osu vynáší čas, zatímco na vertikální ose je frekvence. Barevná škála pak slouží k rozlišení úrovně signálu (energie). Náznorná ukázka jest uvedena na obrázku 5.6. Jedná se o krátký úsek vyčleněný z nahrávky novoročního projevu (2004) tehdejšího prezidenta České republiky Václava Klause. Záznam je pořízen v programu Audacity, kterýžto umožňuje zobrazení, jak signálu samotného, tak i jeho spektrogramu. Program byl rovněž v dřívějších fázích práce pro svoji jednoduchost použit pro záznam.



Obrázek 5.6 – Spektrogram promluvy: „Přeji vám všem klidný a šťastný Nový rok“

Bohužel i spektrogram má jednu podstatnou nevýhodu. Volbou parametrů spektrální analýzy lze totiž docílit buď přesného frekvenčního nebo časového rozlišení. Princip výpočtu je takový, že se nejprve určí tzv. spektrální hustota výkonu (PSD). K výpočtu PSD se používá DFT, která neumožňuje dosáhnout přesnosti v obou parametrech. Proto se často volí kompromis. V této souvislosti se v literatuře rovněž hovoří o použití tzv. vlnkové transformace jako alternativě k DFT. Té se pro analýzu nestacionárních signálů podrobněji věnuje práce [75]. Spektrogramy se tedy často dělí podle zvolené přesnosti na širokopásmové čili krátkodobé (short-term), u kterých je vyšší přesnost v čase, a úzkopásmové čili dlouhodobé (long-term), u kterých je přesněji zakreslena frekvence. Srovnání obou variant spektrogramů je na obrázku 5.7.



Obrázek 5.7 – Původní signál, širokopásmový a úzkopásmový spektrogram [76]

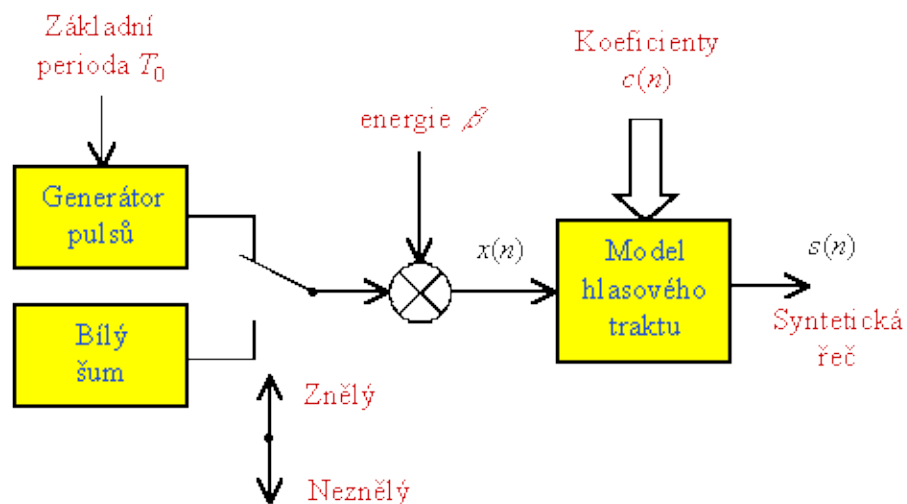
5.2.2 Kepstrální analýza

Kvefrenční kepstrum navazuje na spektrální analýzu. Proto i názvosloví vychází z pojmů používaných při analýze spektra signálu. Pomocí přesmyček tak vznikly jednotlivé termíny, jak v angličtině, tak i v češtině. Obě varianty jsou uvedeny v tabulce 5.3.

Tabulka 5.3 – Pojmy užívané v kepstrální analýze [68]

Spektrální analýza	Kepstrální analýza
Spektrum (Spectrum)	Kepstrum (Cepstrum)
Frekvence (Frequency)	Kvefrence (Quefrequency)
Filtr (Filter)	Liftr (Lifter)

Pro objasnění významu kepstra je třeba se vrátit k teorii zdroje a filtru, která je uvedena v úvodní kapitole. Hlasové ústrojí (hlasivky) má úlohu zdroje, jenž produkuje periodické kmity při tvorbě znělých úseků řeči a šum pro neznělé části. Ve skutečnosti není ani znělý úsek čistě periodický, ale obsahuje i šumový charakter, a to stejné platí i opačně. Vždy se jedná o souběh obou jevů, jak uvádí [77]. Artikulační ústrojí coby filtr svým nastavením upravuje výstup do výsledné podoby. Artikulační ústrojí lze tedy chápat jako filtr s časově proměnlivou přenosovou funkcí. Přičemž parametry tohoto filtru (poloha všech orgánů artikulačního ústrojí) udávají, jak bude upraven signál ze zdroje.



Obrázek 5.8 – Teorie zdroje a filtru: model používaný u syntetizátorů řeči [77]

Informace o zdroji i filtru jsou provázané. V úlohách zpracování řeči však může být žádoucí umět od sebe tyto části oddělit a pracovat s nimi samostatně. Vhodným prostředkem pro tento účel je právě zavedení kepstra. Změny parametrů v filtru lze při zpracování řeči zanedbat, neboť se pracuje s rámci délky 10-30 ms. V takto krátkém úseku

lze tedy předpokládat, že parametry jsou konstantní. Platí tedy, že výsledná řeč je dána konvolucí budícího signálu s impulsní charakteristikou filtru

$$s(n) = x(n) * g(n), \quad (5.10)$$

- kde: – $s(n)$ značí výsledný řečový signál,
 – $x(n)$ je budící signál ze zdroje
 – $g(n)$ představuje impulsní charakteristiku filtru.

Při výpočtu spektra daného signálu pak samozřejmě platí, že konvoluce v časové oblasti se převede na součin v oblasti frekvenční

$$S(k) = X(k) \cdot G(k), \quad (5.11)$$

- kde: – $S(k), X(k), G(k)$ jsou spektra původních signálů vypočtená pomocí DFT (FTT).

Pro získání kepsra se pak nadále pracuje pouze s reálnou částí spekter. Použije se pouze modul (absolutní hodnota) čili amplitudová spektra. Aby od sebe mohly být odděleny části zdroje a filtru zavádí se do vztahu ještě nelinearita v podobě přirozeného logaritmu. Po aplikaci absolutních hodnot a přirozeného logaritmu pak rovnice 5.11 přejde do tvaru

$$\ln[|S(k)|] = \ln[|X(k)| \cdot |G(k)|] = \ln[|X(k)|] + \ln[|G(k)|], \quad (5.12)$$

- kde: – $|S(k)|, |X(k)|$ a $|G(k)|$ představují amplitudová spektra signálů $s(n), x(n), g(n)$.

Zlogaritmování umožnilo oddělit informace o buzení a hlasovém traktu. Nyní už stačí jen provést IDFT. Díky linearitě zpětné transformace zůstane operace sčítání zachována. Je-li uvažována pouze reálná část, výsledkem je reálné kepsrum, u kterého osa n představuje časovou osu.

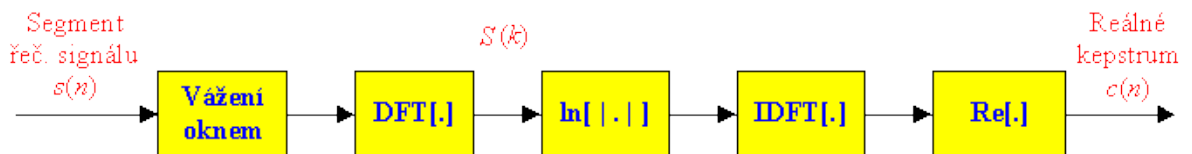
$$c(n) = \text{Re}\left\{\text{IDFT}\left[\ln|S(k)|\right]\right\} = c_x(n) + c_g(n), \quad (5.13)$$

- kde: – $c(n)$ pak označuje reálné kepsrum řečového segmentu $s(n)$,
 – $c_x(n)$ značí kepsrum buzení $x(n)$ (zdroje),
 – $c_g(n)$ představuje kepsrum artikulačního ústrojí $g(n)$ (filtru).

Změny, které jsou ve spektru pomalé, se projeví u počátečních hodnot keprtra. Proto první část odpovídá právě parametrům artikulačního ústrojí. Na druhou stranu změny, které jsou rychlé a periodické, se promítnou do keprtra špičkou v čase odpovídajícím frekvenci tohoto děje. To platí i pro buzení – viditelná špička v keprtru představuje periodu kmitání hlasivek. Proto lze keprtrum mimo jiné využít i pro detekci základního tónu hlasu. Jednotlivé části keprtra lze na základě dříve zmíněných vztahů určit následovně

$$\begin{aligned} c_x(n) &= \text{Re}\left\{\text{IDFT}\left[\ln|X(k)|\right]\right\} \\ c_g(n) &= \text{Re}\left\{\text{IDFT}\left[\ln|G(k)|\right]\right\} \end{aligned} \quad (5.14)$$

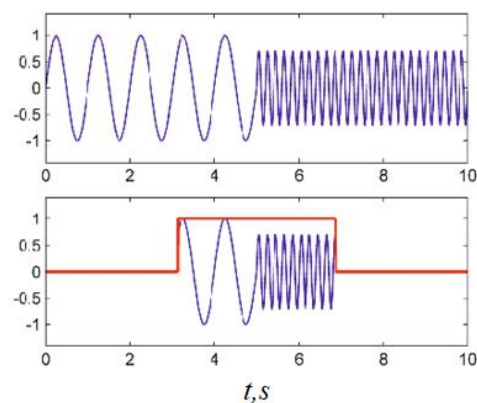
Celý proces výpočtu reálného keprtra rekapituluje obrázek 5.9. Keprtrum je stále poměrně obecný pojem. Existuje totiž několik druhů keprster, které většinou nesou název podle způsobu jejich výpočtu. Patří sem mimo jiné DFT, LPC či Mel-frekvenční keprtrum (MFCC).



Obrázek 5.9 – Schéma výpočtu reálného keprtra segmentu řečového signálu [77]

5.2.3 Aplikace okna

Volba okna je velmi důležitou částí zpracování řečového signálu. Na počátku je třeba zvážit s jakou frekvencí bude řeč vzorkována a na jak dlouhé rámce bude segmentována. Ještě před samotnou parametrizací přichází na řadu mezikrok. Na každý rámeček se totiž aplikuje tzv. okno. Okénkových funkcí existuje celá řada a – Obdélníkové (pravoúhlé), Hannovo, Hammingovo, Blackman-Harrisovo a mnoho dalších.



Obrázek 5.10 – Ukázka aplikace pravoúhlého okna (dole) na původní signál (nahore) [78]

Volba okna závisí i na konkrétní aplikaci, protože okénková funkce může měnit charakter signálu. V případě zpracování řeči lze zvažovat použití Hammingova a obdélníkového neboli pravoúhlého okna. Pravoúhlé okno původní signál v rámci nijak nezmění. Je dáno vztahem

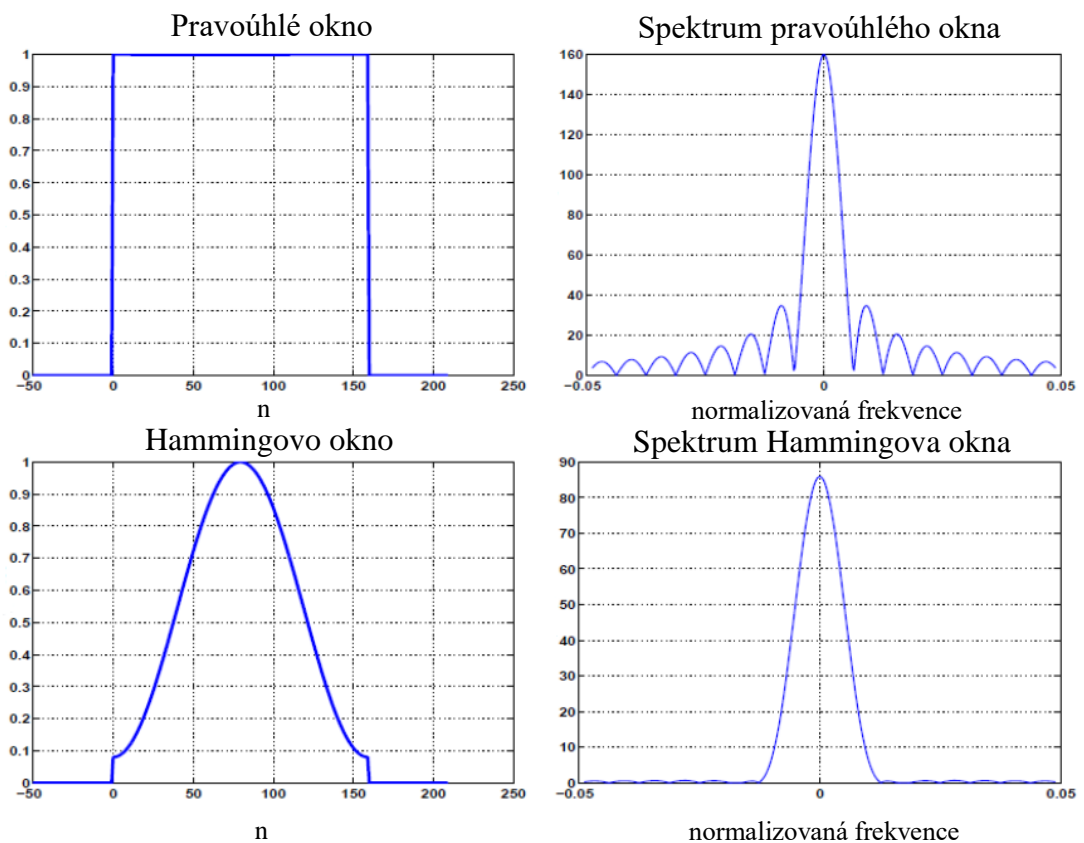
$$w(n) = \begin{cases} 1 & \text{pro } 0 \leq n \leq L_{VZ} - 1 \\ 0 & \text{jinde} \end{cases}, \quad (5.15)$$

kde: – $w(n)$ označuje okénkovou funkci,
 – L_{VZ} značí délku rámce (ve vzorcích).

Při použití Hammingova okna se na rozdíl od pravoúhlého změni charakter signálu tak, že signál bude na okrajích okna utlumen. Hammingovo okno je definováno následovně

$$w(n) = \begin{cases} 0,54 - 0,46 \cos\left(\frac{2\pi \cdot n}{L_{VZ} - 1}\right) & \text{pro } 0 \leq n \leq L - 1 \\ 0 & \text{jinde.} \end{cases} \quad (5.16)$$

Tento jev je patrný i z obrázku 5.11, kde jsou oba typy oken a jejich spektra vyobrazeny.



Obrázek 5.11 – Porovnání frekvenčního spektra obdélníkového a Hammingova okna [68]

Kromě vlivu použitého okna na jednotlivé vzorky signálu se volba okénkové funkce projeví ještě na spektru výsledného signálu, neboť platí že výsledné spektrum bude dáno konvolucí jednotlivých spekter. Čili

$$X(f) = S(f) * W(f), \quad (5.17)$$

- kde: – $X(f)$ je označení pro výsledné spektrum signálu po aplikaci okénkové funkce,
 – $S(f)$ značí spektrum původního signálu (v délce okna) před aplikací okna,
 – $W(f)$ symbolizuje spektrum použité okénkové funkce.

Zatímco obdélníkové (pravoúhlé) okno je mnohem selektivnější, což se projevuje užším centrálním lalokem ve spektru, má také negativní dopad na spektrum signálu, protože obsahuje větší vysokofrekvenční komponenty. Oproti tomu je Hammingovo okno méně selektivnější, avšak nedeformuje výsledné spektrum tak, jako obdélníkové okno.

5.3 Extrakce příznaků z řeči

Proces extrahování příznaků se odehrává během parametrizace již segmentovaného signálu. Pro každý rámeček se vypočtou hodnoty požadovaných příznaků a uloží se do příznakových vektorů. Tento proces probíhá postupně rámeček po rámečku. Výpočet jednotlivých příznaků se liší, přičemž některé mohou přímo souviset s hodnotami vzorků. Příznaky užívané při zpracování řeči lze rozdělit podle charakteru na základní, spektrální, keprální a ostatní.

Obecně se však příznaky dělí ještě na statické, dynamické a akcelerační. Toto platí i pro všechny skupiny předešlého dělení podle charakteru. Zatímco statický příznak vyjadřuje pouze charakteristiku jednotlivých rámečků a jeho hodnota je vždy svázána pouze s příslušným rámečkem a vzorky, které obsahuje, u dynamických a akceleračních příznaků již toto neplatí. Aby mohla být určena hodnota jakéhokoliv dynamického příznaku pro daný rámeček, musí být nejprve známy hodnoty statických příznaků pro sousedící rámečky. Pro dynamické příznaky se v literatuře někdy používá také označení delta. Proces stanovení dynamických příznaků lze popsat vzorcem

$$\Delta p_i(k) = p_i(k+1) - p_i(k-1), \quad (5.18)$$

- kde: – Δp_i označuje i -tý dynamický (delta) příznak,
 – p_i symbolizuje i -tý statický příznak,
 – k je pořadové číslo rámečku takové, že $k \in \{1, 2, \dots, K\}$.

Zcela analogicky lze určit i akcelerační příznaky, také někdy nazývané jako delta-delta, resp. dynamické příznaky druhého řádu. K jejich určení je potřeba nejprve vyčíslit dynamické příznaky okolních rámců. Vztah popisující výpočet akceleračních příznaků je následující

$$\Delta\Delta p_i(k) = \Delta p_i(k+1) - \Delta p_i(k-1), \quad (5.19)$$

- kde: – $\Delta\Delta p_i$ je označením pro i -tý akcelerační (delta-delta) příznak,
 – Δp_i je i -tý dynamický (delta) příznak,
 – p_i značí i -tý statický příznak,
 – k je pořadové číslo rámce takové, že $k \in \{1, 2, \dots, K\}$.

U krajních rámců se delta a delta-delta příznaky musí určit jiným způsobem – úpravou indexů vzorce 4.4 resp. 4.5. Jsou-li uvažovány delta příznaky u prvního rámce se hodnota určí jako

$$\Delta p_i(1) = p_i(2) - p_i(1). \quad (5.20)$$

A podobně tomu bude u rámce posledního

$$\Delta p_i(K) = p_i(K) - p_i(K-1), \quad (5.21)$$

- kde: – K je celkový počet rámců.

Díky zpracování počátečních i koncových hodnot se tak zachovává stejný počet statických, dynamických i akceleračních příznaků, který je současně roven celkovému počtu rámců K . Problematice statických a dynamických příznaků se dále věnuje [79].

5.3.1 Energie hlasu a šumu pozadí

Základním statickým příznakem ve zpracování signálů je energie. Při výpočtu energie s vychází přímo z hodnot vzorků v každém rámcu. Používá se při detekci řeči – rozlišení řečových a neřečových úseků v nahrávce [80]. Další využití může být při rozlišení znělých a neznělých hlásek. Jak již bylo uvedeno v kapitole o hlasivkách, při tvorbě znělých hlásek se hlasivky zapojují, je přítomný základní tón, a tudíž je energie hlasu větší než u neznělých hlásek, které jsou šumového charakteru. I proto může být obtížné neznělé hlásky detekovat, protože rozdíl mezi energií šumu na pozadí a neznělého úseku může být nepatrný. V literatuře se objevuje více variant vztahů pro práci s energií. Krátkodobá energie signálu je dána vztahem

$$E = \sum_{n=0}^{L-1} x^2(n), \quad (5.22)$$

- kde: – E značí energii,
 – L je délka rámce (počet vzorků každého rámce),
 – $x(n)$ představuje jednotlivé vzorky signálu.

Nejčastěji se uvádí střední krátkodobá energie (v literatuře někdy též uváděna jako výkon signálu) získaná podělením krátkodobé energie počtem vzorků

$$E_S = \frac{E}{L} = \frac{1}{L} \sum_{n=0}^{L-1} x^2(n). \quad (5.23)$$

Energie signálu je velmi citlivá na změny rozsahu. Velké změny jsou při výpočtu ještě umocněny. Aby mohlo být docíleno menšího dynamického rozsahu hodnot i při velkých změnách, zavádí se logaritmická energie

$$\log E = \log \left(\sum_{n=0}^{L-1} x^2(n) \right). \quad (5.24)$$

Energie (ať už se jedná o kteroukoliv z uvedených forem) se pro detekci začátku a konce řeči využívá následujícím způsobem. Nejprve je zaznamenán neřečový signál tzn. hluk a šum prostředí, ve kterém bude nahrávka pořízena. Z tohoto záznamu lze určit jakou energii má hluk v dané místnosti apod. Poté při záznamu řeči bude vypočtená energie (ideálně) vyšší. Pro odlišení šumu pozadí a řečových segmentů by měl být rozdíl energie hlasu a energie šumu pozadí co největší. Tomu potom odpovídá volba vhodného rozhodovacího prahu. V praxi však není jednoduché tyto úseky přesně vyčlenit. V běžné řeči často řečník klesá hlasem, dělá záměrné pauzy atd. Krátké pauzy (poklesy energie) lze detekovat i v některých slovech.

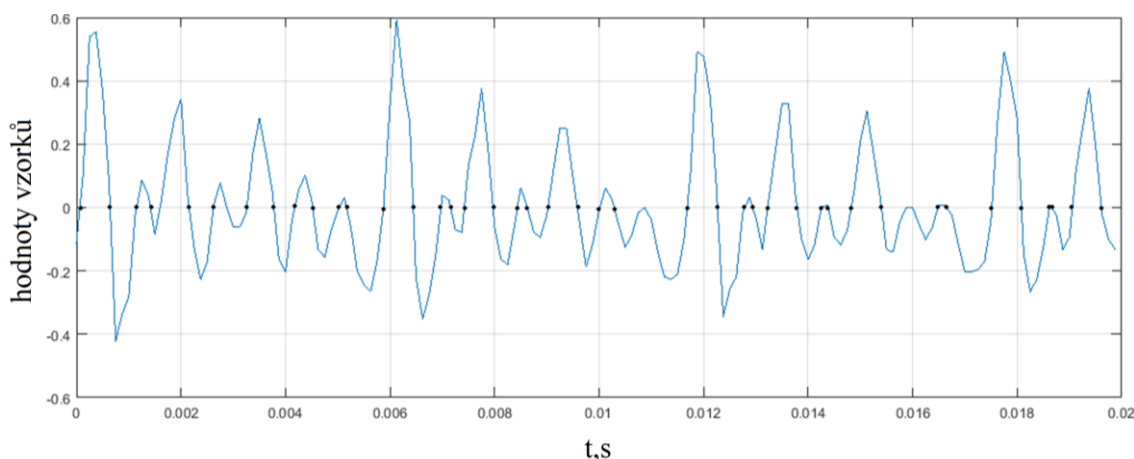
5.3.2 Střední krátkodobá intenzita

Další z veličin, která je příbuzná energii a která může poskytovat řešení v podobě zmenšení dynamického rozsahu, je střední krátkodobá intenzita M . Je dána vzorcem

$$M = \frac{1}{L} \sum_{n=0}^{L-1} |x(n)|, \quad (5.25)$$

5.3.3 Počet průchodů nulou

Tato bezrozměrná veličina se označuje zkratkou ZCR (z anglického zero crossing rate). Je to jednoduchá míra, která souvisí s frekvenčními vlastnostmi analyzovaného signálu. ZCR je číslo, které určuje kolikrát signál prošel x-ovou (časovou) osou a změnil přitom polaritu. Příklad signálu v běžné délce jednoho rámce a označení průchodů nulou pomocí černých teček je na obrázku 5.12.



Obrázek 5.12 – Ukázka ZCR pro rámec délky 20 ms [65]

U periodických signálů by vyšetřování ZCR nemělo smysl, protože počet průchodů nulou bude dán periodou signálu. U řeči však může tento příznak pomoci odlišit v nahrávce segmenty obsahující řeč a ostatní segmenty, ve kterých je pouze ticho či šum na pozadí. Šum typicky nabývá nižších hodnot energie, zatímco ZCR je vysoký, protože signál rychle přechází z jednoho kvadrantu do druhého a naopak. Velká hodnota ZCR platí i pro neznělé úseky řeči, zejména pak sykavky (s, š, z, ž, c, č, dz, dž). U znělých částí řeči je pak počet průchodů signálu časovou osou výrazně nižší. Počet průchodů nulou je matematicky popsán jako

$$ZCR = \frac{1}{2} \sum_{n=1}^{L-1} |\operatorname{sgn} x(n) - \operatorname{sgn} x(n-1)|, \quad (5.26)$$

- kde:
- ZCR je počet průchodů nulou,
 - L je délka rámce (počet vzorků každého rámce),
 - $x(n)$ představuje jednotlivé vzorky signálu,

$$- \operatorname{sgn} x(n) \text{ je znaménková funkce signálu definovaná jako } \operatorname{sgn}(x) = \begin{cases} -1, x < 0 \\ 0, x = 0 \\ 1, x > 0 \end{cases}$$

5.3.4 Spektrální energie

Jak již název napovídá, tato veličina má podobný význam jako klasická energie. Přidává se zde však jedna důležitá informace o příslušnosti do frekvenčního pásma. Pak se hovoří o spektrální energii. Spektrální energie tedy vyjadřuje, jaké je rozložení energie ve frekvenčním spektru zaznamenaného signálu. Obecně lze konstatovat, že největší spektrální energie je rozmístěna u nízkých frekvencí. Z pohledu analýzy spektrální energie může být důležité např. pozorování, jak velká energie je v těchto nižších frekvenčních pásmech, jak je energie rozdělena ve vyšších frekvencích a jaký je poměr mezi jednotlivými pásmy.

Pokud jde o výpočet spektrální energie, nejprve se na vzorky signálu aplikuje v rámci aplikuje Hammingovo okno (o délce rámce L). Dále se provede N -bodová FFT, kde N je nejbližší vyšší mocnina dvou k délce rámce dané počtem vzorků. Tzn. například pro rámec délky 400 vzorků se užije 512bodová FFT. Ještě před samotnou transformací se však chybějící hodnoty musí doplnit nulami. Jedná se o tzv. zero padding. V tomto případě by se tedy přidalo ke vzorkům z rámce 112 nul. Nyní se vezme prvních $\frac{N}{2}$ hodnot a ty se rozdělí do K pásem (kanálů) – v angličtině bin. Tato polovina (v tomto případě prvních 256 hodnot) odpovídá frekvenčnímu rozsahu 0-4 kHz. Typické dělení může být do pásem po 500 Hz – tedy $K = 8$. Což tedy znamená, že každému pásmu přísluší právě 32 komplexních hodnot. Pro určení počátečních hodnot při výpočtu spektrální energie poslouží vztah

$$i_k = (k-1) \frac{N}{2K}, \quad (5.27)$$

- kde:
- i_k je index první hodnoty příslušící ke k -tému pásmu,
 - k označuje pořadové číslo kanálu (pásma) takové, že $k \in \{1, 2, \dots, K\}$,
 - N je počet bodů, ve kterých je počítána FFT.

Obdobným způsobem se určí i indexy posledních hodnot, které patří do daného pásma

$$I_k = k \frac{N}{2K} - 1, \quad (5.28)$$

kde: – I_k je index poslední hodnoty příslušící ke k-tému pásmu.

Poté se pro jednotlivá pásma s využitím rovnic 4.15 a 4.16 určí spektrální energie, která bývá taktéž uváděna v logaritmickém tvaru.

$$\log B_k = \log \left(\sum_{i=i_k}^{I_k} |\hat{x}(i)|^2 \right), \quad (5.29)$$

kde: – B_k symbolizuje spektrální energii k-tého pásma,

– $\hat{x}(i)$ představuje vzorky po aplikaci Hammingova okna a FFT.

5.3.5 Základní tón (frekvence)

Jak již bylo zmíněno v sekci o tvorbě lidského hlasu, hlasivky svým kmitáním upravují proud vzduchu z plic. To má za výsledek generování základního tónu hlasu každého jedince a tento tón je pak dále modifikován artikulačním ústrojím. V úloze určování základního tónu (frekvence) hlasu z nahrávky se jinak řečeno jedná o určení frekvence kmitání hlasivek daného subjektu. Jedná o parametr, který odráží aktuální náladu řečníka a všechny faktory, které na něj působí, ale také udává melodii promluvy na základě obsahu. Využívá se proto mimo jiné v úlohách syntézy řeči (TTS). Malým změnám frekvence (řádově do 10 Hz) po prvním kmitu při vyslovování hlásek se říká „jitter“. Tyto změny charakterizují řečníka [68], a proto je lze využít např. pro hlasovou biometrii.

Pro základní tón (frekvence) je v anglicky psané literatuře uváděn název pitch. Periodu základního tónu lze pak najít pod názvem pitch period. Platí zde klasický vztah mezi frekvencí a periodou

$$T_0 = \frac{1}{f_0}, \quad (5.30)$$

kde: – T_0 je perioda základního tónu udávaná v s,

– f_0 značí základní frekvenci kmitání hlasivek uváděnou v Hz.

Mimo to se v úloze detekce základního tónu využívá ještě perioda základního tónu vyjádřená počtem vzorků. Ta se označuje termínem lag a je dána následovně

$$L_0 = T_0 \cdot f_s, \quad (5.31)$$

- kde: – L_0 je označení pro lag, který se řadí mezi bezrozměrné veličiny,
 – T_0 je perioda základního tónu udávaná v s,
 – f_s představuje vzorkovací frekvenci uváděnou v Hz.

Pro detekci základního tónu řeči existuje celá řada metod. Je zde však mnoho úskalí a omezení, pročež úloha detekce základního tónu rozhodně není triviální. Jedním ze základních přístupů je využití tzv. autokorelační funkce (ACF). ACF se řadí mezi metody pracující se signálem v časové oblasti [81]. Autokorelace značí, že je srovnáván daný signál se sebou samým. Druhý signál je ovšem posunutý v čase o určitý počet vzorků. Tím se stanoví míra podobnosti obou signálů, což je právě ACF definovaná vztahem

$$R(m) = \sum_{n=0}^{L-1-m} s(n) \cdot s(n+m), \quad (5.32)$$

- kde: – $R(m)$ je autokorelační koeficient při posunu signálu o m vzorků,
 – L představuje délku rámce volenou při segmentaci signálu,
 – $s(n)$ symbolizuje původní řečový signál,
 – $s(n+m)$ značí původní řečový signál posunutý o m vzorků.

Protože koeficienty ACF jsou symetrické, lze rovněž vzorec 5.32 uvádět v tomto tvaru

$$R(m) = \sum_{n=m}^{L-1} s(n) \cdot s(n-m). \quad (5.33)$$

Nejvyšší hodnota bude zcela logicky vždy vycházet pro koeficient $R(0)$, kdy je posun signálů $m=0$ vzorků, tudíž se porovnávají dva naprosto stejné signály. Zde také platí, že hodnota $R(0)$ je rovná energii signálu. Rovněž lze dokázat, že je-li analyzovaný signál periodický, pak i ACF bude periodická. Aby se odstranili změny vyvolané kolísáním hlasitosti, mohou být autokorelační koeficienty normovány dle vzorce

$$r(m) = \frac{R(m)}{R(0)}, \quad (5.34)$$

kde: – $r(m)$ značí normovaný autokorelační koeficient při posunu signálu o m vzorků.

Stanovení lagu (tedy periody základního tónu udávané ve vzorcích) je pak možné na základě hledání indexu maxima ACF. Tomu odpovídá rovnice

$$L_0 = \arg \max_m R(m), \quad (5.35)$$

kde: – L_0 je označení pro lag, který je bezrozměrný (udává počet vzorků signálu),

– m představuje posun signálu ve vzorcích a zároveň index maximální hodnoty ACF.

Detekce základního tónu řeči podléhá ještě jednomu problému. Určování výsledné hodnoty je samozřejmě možné pouze v úsecích nahrávky, kde je přítomna řeč. Celou záležitost však ještě komplikuje fakt, že řeč se dělí na znělé a neznělé úseky. Základní tón řeči lze určit pouze pro znělou část. K odhadu znělosti řeči v analyzované nahrávce lze použít pravidlo založené na porovnání hodnoty nalezeného maxima s hodnotou $R(0)$. Pravidlo odhadu znělosti je definováno jako

$$\begin{aligned} R_{\max} < \alpha \cdot R(0) &\Rightarrow \text{neznělý} \\ R_{\max} \geq \alpha \cdot R(0) &\Rightarrow \text{znělý} \end{aligned} \quad (5.36)$$

kde: – R_{\max} je maximální hodnota ACF nalezená pomocí vzorce 5.33,

– α je prahová konstanta, která je volena experimentálně.

Značnou nevýhodou ACF je, že vlivem posunu signálu dochází k neustálému zkracování části, ze které se autokorelační koeficienty počítají. Tuto nevýhodu odstraňuje použití křížové (cross) korelační funkce (CCF). U této metody se totiž pracuje s celým signálem, který musí být znám, a nikoliv pouze s hodnotami jednoho rámce. CCF se určí podle vzorce

$$R_{\text{CCF}}(m) = \sum_{n=z}^{z+L-1} s(n) \cdot s(n-m), \quad (5.37)$$

kde: – $R_{\text{CCF}}(m)$ je cross-korelační koeficient při posunu signálu o m vzorků,

– z je číslo vzorku signálu, přičemž tento vzorek je na počátku zkoumaného rámce,

– L představuje délku rámce volenou při segmentaci signálu,

– $s(n)$ symbolizuje původní řečový signál,

– $s(n-m)$ značí původní řečový signál posunutý o m vzorků.

Protože signál se v čase mění, projevuje se to i na změnách energie, které mohou posunem zkoumané oblasti nastat. Proto je i v případě CCF vhodné použít normalizaci. Jedná se pak o normalizovanou křížovou korelační funkci (NCCF). Normalizace CCF je vyjádřena vztahem

$$R_{\text{NCCF}}(m) = \frac{\sum_{n=z}^{z+L-1} s(n) \cdot s(n-m)}{\sqrt{E_1 \cdot E_2}}, \quad (5.38)$$

- kde: – $R_{\text{NCCF}}(m)$ je koeficient NCCF při posunu signálu o m vzorků,
– E_1 udává hodnotu energie originálního rámce,
– E_2 vyjadřuje hodnotu energie posunutého rámce.

Energie, které jsou potřebné pro výpočet koeficientů NCCF lze určit z následujících rovnic.

$$E_1 = \sum_{n=z}^{z+L-1} s^2(n), \quad (5.39)$$

$$E_2 = \sum_{n=z}^{z+L-1} s^2(n-m), \quad (5.40)$$

Autokorelační metody obecně trpí ještě jedním neduhem. Tím je vliv formantů, které vytváří postranní maxima v autokorelační funkci. Autokorelační metody tato maxima sice potlačují, ne však dostatečně. To může vést k tomu, že základní tón je chybně detekován na základě lagu identifikovaného pomocí právě takového postranního maxima. K odstranění vlivu formantů se tak využívá metoda zvaná Centrální klipování (Center clipping). Jedná se o předzpracování signálu pro autokorelační funkci, aby se tak odstranil vliv postranních maxim. V základu je třeba definovat určitou úroveň c_L . Poté jsou možné dva přístupy

$$c_1[s(n)] = \begin{cases} s(n) - c_L & \text{pro } s(n) > c_L \\ 0 & \text{pro } -c_L \leq s(n) \leq c_L \\ s(n) + c_L & \text{pro } s(n) < -c_L \end{cases}, \quad (5.41)$$

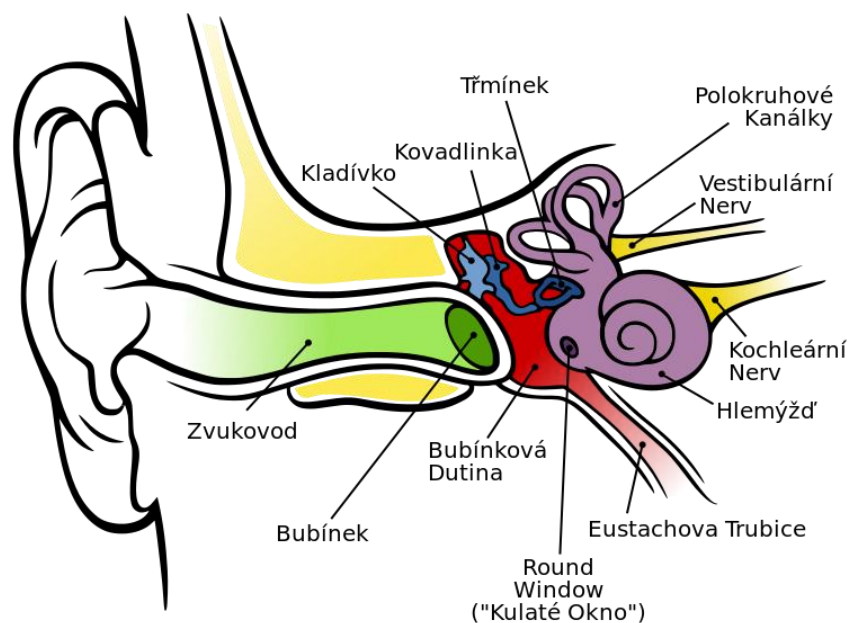
$$c_2[s(n)] = \begin{cases} +1 & \text{pro } s(n) > c_L \\ 0 & \text{pro } -c_L \leq s(n) \leq c_L \\ -1 & \text{pro } s(n) < -c_L \end{cases}, \quad (5.42)$$

5.3.6 Základní akustické veličiny

Definice zvuku a popis jevů s ním souvisejících má své kořeny ve fyzice a samostatném oboru – akustice. Všechny důležité pojmy a souvislosti z oboru akustiky uvádí například [82]. Užívaná terminologie a definice jednotlivých pojmů je taktéž zakotvena v normách. Poslední dodatek k akustickým normám a doplnění terminologie je zpracováno v rámci normy ČSN 01 1600 [83].

Zvuk je vlastně mechanické vlnění vyvolávající ve sluchovém aparátu sluchový vjem. Uvádí se, že pásmo slyšitelných kmitočtů je u lidí v rozmezí 16-20000 Hz, přičemž s narůstajícím věkem se horní hranice snižuje [84]. Tento proces se nazývá presbyakuse. Vlnění na nižších frekvencích se nazývá infrazvuk. Zvuk na frekvenci vyšší nese název ultrazvuk. Referenční rychlost šíření zvukové vlny při teplotě 15 °C a nulové nadmořské výšce je 340 ms⁻¹. Ve vokálním traktu se pak zvuková vlna šíří rychlostí 350 ms⁻¹.

Z pohledu vnímání zvuku je však klíčový zcela jiný parametr. Je jím tzv. akustický tlak. Akustický tlak způsobuje vychýlení membrány lidského sluchového orgánu (bubínku). Dále se šíří ve formě vibrací. V ušním závitku zvaném hlemýžď probíhá frekvenční analýza vnímaného zvuku. Poté se transformuje na elektrické signály, které jsou přenášeny sluchovým nervem (nervus vestibulocochlearis) až do mozku. Sluchový aparát člověka je detailně popsán na obrázku 5.13.



Obrázek 5.13 – Anatomie lidského ucha [85]

Vnímání zvuku je z pohledu frekvenční i dynamické oblasti nelineární. Tuto skutečnost odráží i tzv. sluchové pole. Podobně jako hlasové pole určuje rozsah např. mluvené řeči, sluchové pole označuje oblast od sotva slyšitelných zvuků až po ty, které způsobují bolest. Pro určení subjektivní hladiny hlasitosti se užívá jednotka fón (Ph), přičemž 1 Ph odpovídá při referenční frekvenci 1 kHz právě hodnotě 1 dB. Referenční frekvence je zvolena z důvodu frekvenčně závislé citlivosti sluchu na zvukové podněty. Křivky stejné hlasitosti se nazývají izofóny. Právě ty utváří výsledné sluchové pole, přičemž tzv. nulová izofóna představuje práh slyšitelnosti. Shora je sluchové pole omezeno izofónou na hladině 120 dB. Tato křivka se označuje jako práh bolesti.

Zde je třeba zmínit, že hlasitost je čistě subjektivní veličina. Pro objektivizaci hodnocení hlasitosti zvuku byla podle [86] stanovena jedna ze základních veličin oboru akustiky, a to akustická intenzita zvuku definovaná následovně

$$I = \frac{E}{t} \cdot \frac{1}{S} = \frac{P}{S}, \quad (5.43)$$

- kde: – I je již zmíněná akustická intenzita v jednotkách Watt na metr čtvereční,
 – E vyjadřuje energii vlnění v Joulech,
 – t značí čas v sekundách,
 – P je tzv. akustický výkon udávaný ve Watech daný vztahem 5.45,
 – S představuje velikost plochy (v metrech čtverečních), kterou vlnění prochází.

Pro úplné pochopení souvislostí mezi jednotlivými veličinami je třeba ještě definovat akustický tlak, který lze zapsat ve tvaru

$$p = Z \cdot v, \quad (5.44)$$

- kde: – p je akustický tlak v Pascalech,
 – Z symbolizuje vlnový odpor prostředí a udává se v jednotkách $\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$,
 – v označuje akustickou rychlost v $\text{m} \cdot \text{s}^{-1}$.

Druhým naprosto nezbytným vztahem je vzorec pro určení akustického výkonu na základě znalosti akustického tlaku. Akustický výkon je tedy dán ve tvaru

$$P = p \cdot S \cdot v, \quad (5.45)$$

- kde: – P značí právě akustický výkon ve Watech,
 – p je akustický tlak v Pascalech,

- S je opět plocha v m^2 , kterou prochází vlnění,
- v interpretuje akustickou rychlost udávanou v $\text{m} \cdot \text{s}^{-1}$.

Navíc akustická intenzita je přímo úměrná druhé mocnině akustického tlaku, což lze jednoduše dokázat tak, že se do vzorce 4.18 dosadí nejprve rovnice 4.20 a poté rovněž vztah 4.19, ze kterého je vyjádřena akustická rychlost, přičemž odvození je následující

$$I = \frac{P}{S} = \frac{p \cdot S \cdot v}{S} = p \cdot v = p \cdot \frac{p}{Z} = \frac{p^2}{Z}. \quad (5.46)$$

Jak již bylo zmíněno, vnímání zvuku je nelineární. Pro zvuky s nízkou intenzitou je sluch citlivější, zatímco u hlasitějších zvuků se citlivost významně snižuje, a právě proto se při práci s akustickými veličinami používají logaritmy. Rozdíly v hodnotách akustického tlaku či akustické intenzity mohou být i v rozmezí několika řádů. I proto je aplikace logaritmů vhodná. Mnohem častěji se tedy v literatuře lze setkat s pojmem hladina akustického tlaku (SPL). Mimo to odborná literatura ještě uvádí např. pojmy hladina intenzity či hladina akustického výkonu. Všechny tyto veličiny vychází z logaritmu poměru dané hodnoty k hodnotě referenční a jsou udávány v decibelech. Hladina intenzity se určí jako

$$L = 10 \log \left(\frac{I}{I_0} \right), \quad (5.47)$$

kde: – L označuje hladinu intenzity vypočtenou v dB,

- I vyjadřuje akustickou intenzitu zvuku v jednotkách $\text{W} \cdot \text{m}^{-2}$,
- I_0 představuje intenzitu $10^{-12} \text{W} \cdot \text{m}^{-2}$ odpovídající prahu slyšitelnosti.

Velmi podobně jako hladina intenzity je hladina akustického tlaku (SPL) definována vztahem

$$L_p = 10 \log \left(\frac{p^2}{p_0^2} \right) = 10 \log \left(\frac{p}{p_0} \right)^2 = 20 \log \left(\frac{p}{p_0} \right), \quad (5.48)$$

kde: – L_p je hladina akustického tlaku (SPL) udávaná rovněž v dB,

- p symbolizuje akustický tlak v Pa,
- p_0 je referenční akustický tlak $2 \cdot 10^{-5}$ Pa, který odpovídá prahu slyšitelnosti.

Mezi hladinou intenzity a SPL existuje přímá souvislost. Obě veličiny jsou záměrně definovány tak, aby měly vždy stejné hodnoty (při dodržení jedné podmínky). Jediný rozdíl, který lze mezi nimi najít, je směrové působení. Zatímco akustická intenzita je spojena i se směrem šíření zvuku, akustický tlak nezohledňuje směr, kterým se zvuk šíří, ale pouze popisuje stav prostředí v daném místě. Podmínkou rovnosti obou veličin tedy je, aby plocha, na níž je zjišťována akustická intenzita, byla kolmá na směr šíření zvuku. Potom platí rovnost obou veličin a lze tedy psát

$$L = 10 \log \left(\frac{I}{I_0} \right) = 10 \log \left(\frac{p}{p_0} \right)^2 = 20 \log \left(\frac{p}{p_0} \right) = L_p, \quad (5.49)$$

- kde:
- L označuje hladinu intenzity v dB,
 - L_p je hladina akustického tlaku (SPL) rovněž v dB,
 - I vyjadřuje akustickou intenzitu zvuku ve W,
 - I_0 představuje intenzitu $10^{-12} \text{ W} \cdot \text{m}^{-2}$ odpovídající prahu slyšitelnosti,
 - p značí akustický tlak v Pa,
 - p_0 je referenční akustický tlak $2 \cdot 10^{-5}$ Pa, který odpovídá prahu slyšitelnosti.

Jelikož lidská sluchový orgán je uzpůsobený k tomu, aby mohl přijímat zvuky ze všech směrů využívá se v praxi pro popis intenzity zvuku SPL. Pro větší názornost v tabulce 4.3 uvedeny hodnoty jednotlivých veličin s uvedenými příklady možných prostředí, které odpovídají dané kvantifikaci.

Tabulka 5.4 - Příklady hodnot základních akustických veličin [87]

akustický tlak p [Pa]	akustická intenzita I [Wm^{-2}]	hladina akustického tlaku L [dB]	Prostředí, které odpovídá hodnotám
~ 60	10	130	práh bolesti
2	10^{-2}	100	diskotéka
0,2	10^{-4}	80	rušná ulice
0,02	10^{-6}	60	kancelář
0,002	10^{-8}	40	obývací pokoj
0,0002	10^{-10}	20	ložnice v noci
0,00002	10^{-12}	0	práh slyšitelnosti

5.3.7 Mel-frekvenčních keprální koeficienty

Problematika extrakce Mel-frekvenčních keprálních koeficientů (MFCC) je záměrně zařazena až po představení akustických parametrů. Je to zejména z toho důvodu,

že MFCC byly zavedeny právě proto, že citlivost vnímání zvuku není u lidí stejná na všech frekvencích. Zatímco klasická DFT, používaná pro výpočet kepra, má z hlediska frekvencí všude stejné rozlišení, z předchozí kapitoly je známo, že lidský sluch disponuje větším rozlišením v oblasti nízkých frekvencí a směrem k vyšším frekvencím se pak rozlišení snižuje. Aby mohl být tento jev co nejvíce přenesen i do keprální analýzy, zavádí se příznaky MFCC. Je třeba podotknout, že MFCC se staly v průběhu let nejpoužívanějšími příznaky napříč téměř všemi aplikacemi, kde je třeba zpracování a analýza řeči.

Pro přepočítání frekvence z Hertzů na Mely se používá vzorec

$$f_{Mel} = 2959 \log \left(1 + \frac{f_{Hz}}{700} \right). \quad (5.50)$$

Pro opačný postup se využívá vztahu odvozeného ze vzorce 5.50. Je-li dána nelineární úprava frekvenční osy v Melech, lze frekvenci přepočítat na Hertze následovně

$$f_{Hz} = 700 \left(10^{\frac{f_{Mel}}{2959}} - 1 \right). \quad (5.51)$$

Postup výpočtu MFCC příznaků je takový, že nejprve se provede nelineární úprava frekvenční osy z Hertzů na Mely. Dále se pak na ose označené v Melech lineárně rozmístí filtry (v běžném pojetí frekvencí v Hz to znamená nelineární rozmístění filtrů, jenž odpovídá rozlišovací schopnosti lidského sluchu). Při výpočtu dle definice by se tedy konstruovala banka filtrů a počítala by se energie z jednotlivých filtrů, přičemž by se filtroval vstupní signál v časové oblasti a energie by se počítala jako součet kvadrátů hodnot jednotlivých vzorků. Kvůli nadměrné složitosti výpočtu byl vytvořen zjednodušený postup, jenž byl na univerzitě v Cambridge implementován i do softwaru Hidden Markov Model Toolkit (zkráceně HTK). Tento zjednodušený postup zahrnuje aplikaci DFT, umocnění, vynásobení trojúhelníkovým oknem, a nakonec sečtení získaných hodnot. Takto obdržené hodnoty energie je třeba ještě logaritmovat a na závěr se provede zpětná Fourierova transformace (IFT). Často se IFT nahrazuje diskretní cosinovou transformací (DCT). Mel-frekvenční keprální koeficienty jsou tedy definovány jako

$$c_{mf}(n) = \sum_{k=1}^K \log(E_k) \cdot \cos \left[n(k-0,5) \frac{\pi}{K} \right], \quad (5.52)$$

- kde: – $c_{mf}(n)$ je n -tý Mel-frekvenční keprální koeficient,
- $n = 0, 1, \dots, M$ značí pořadové číslo kalkulovaného keprálního koeficientu,
 - M udává počet zjišťovaných MFCC příznaků,
 - k představuje pořadové číslo filtru z banky filtrů, kde $k = 1, \dots, K$,
 - K stanovuje celkový počet filtrů v bance filtrů,
 - E_k označuje energii na intervalu k -tého filtru.

5.3.8 Rychlost řeči

Jak již název napovídá, tato veličina je jednoduchou mírou pro určení rychlosti řeči daného řečníka. Rychlost, s jakou daný řečník hovoří, je ovlivněna celou řadou faktorů, mezi které patří například jazyk, věk, rétorické schopnosti, charakter člověka, aktuální psychický stav (např. nervozita), ale i fyzický stav (frekvence dýchání, únava apod.), znalost (přípravenost) tématu, omezený čas projevu a mnoho dalších. Mimo jiné se sem řadí i vhodné používání pauzy v souvislé řeči. Rychlost řeči je zkrátka parametrem, který odráží soubor komplexních faktorů, které člověka při jeho orálním projevu ovlivňují.

Tato skutečnost se projeví i při čtení textu. V tomto případě budou samozřejmě působit na čtenáře odlišné faktory, než když se bude jednat například o přednášku nebo rozhovor. I tak lze konstatovat, že bude-li několika řečníkům předložen stejný text, nikdy ho nepřečtou stejně rychle. Dokonce i v případě, že bude jeden řečník opakovat stejnou promluvu či větu, nikdy nedojde k naprosté shodě. Přesto jsou-li v maximální možné míře zachovány podmínky v nichž byl záznam pořízen, je pravděpodobné, že vymezený čas bude srovnatelný.

Rychlost řeči, stejně jako základní frekvence či akustické parametry, není počítána pro jednotlivé rámce. Průměrnou rychlost řeči lze stanovit na základě znalosti doby trvání řečového úseku nahrávky a počtu slov v dané promluvě. Není proto třeba využívat žádný algoritmus rozpoznávání řeči, postačí pouhé dosazení do vzorce

$$\bar{v}_s = \frac{W}{t}, \quad (5.53)$$

- kde: – \bar{v}_s představuje průměrnou rychlost řeči udávanou v jednotkách slova $\cdot s^{-1}$,
- W udává celkový počet slov vyřčených v rámci promluvy,
 - t označuje délku trvání promluvy od začátku do konce řeči udávanou v s.

Odlišný přístup, který zároveň poskytuje dynamickou informaci o tom, jak se tempo řeči vyvíjí v čase, představuje akumulovaná rychlost řeči. K určení akumulované rychlosti řeči je již nezbytné využití nástrojů pro automatické rozpoznávání plynulé řeči (ASR). Algoritmus ASR běžně rozpozná slova v přiložené nahrávce – nutno dodat, že v současnosti s přesností nad 90 % (závisí na konkrétní aplikaci). Mimo to může určit i hranice jednotlivých slov. To znamená, že nalezne v nahrávce začátek a konec příslušného slova a zaznamená tyto dva časy. Akumulovanou rychlost řeči tak lze vyhodnotit v kterékoliv části nahrávky. Je vázána vždy pouze ke konkrétnímu slovu a to tak, že

$$v_a^w = \frac{w}{t_w}, \quad (5.54)$$

kde: – v_a^w je akumulovaná rychlost řeči určená pro w -té slovo v jednotkách slova $\cdot s^{-1}$,
– $w = \{1, \dots, W\}$ udává akumulovaný počet slov, přičemž W je celkový počet slov,
– t_w značí akumulovaný čas od počátku nahrávky do konce w -tého slova.

V případě akumulované rychlosti řeči se vlastně jedná o postupné průměrování po vyřčení každého slova. Proto, je-li akumulovaný počet slov roven celkovému počtu slov v nahrávce ($w=W$), pak je i akumulovaný čas roven času celkovému ($t_w = t$) a z toho vyplývá, že akumulovaná rychlost řeči se pak rovná průměrné ($v_a^w = \overline{v_s}$).

5.4 Měření hladiny akustického tlaku (SPL)

Měření SPL nespadá úplně do extrakce příznaků z řeči, i když s ní bezesporu souvisí. Hladina akustického tlaku a její vztah k dalším akustickým veličinám již byly popsány. Ke zjištění hodnoty SPL v praxi jsou však potřeba ještě další znalosti. Před samotným měřením je nutné si uvědomit, že hladina akustického tlaku závisí na vzdálenosti od zdroje zvuku. Pokud tedy bude ve stejném čase měřen např. hlas ze dvou různých vzdáleností, výsledky SPL budou rozdílné, přičemž ve větší vzdálenosti budou hodnoty zcela logicky nižší. Tuto souvislost popisuje zákon o vzdálenosti

$$L_{d_2} = L_{d_1} - 20 \log \left(\frac{d_2}{d_1} \right), \quad (5.55)$$

kde: – d_1 a d_2 představují vzdálenosti, ve kterých je měřeno SPL,

- L_{d1} a L_{d2} jsou hodnoty SPL ve vzdálenostech d_1 a d_2 uváděné v dB.

5.4.1 Časové vážení a průměrování

Dle normy ČSN 01 1600 se zavádí časově průměrovaná hladina akustického tlaku. Ta je definována jako

$$L = 20 \log \left\{ \frac{\left[\frac{1}{T} \int_{t-T}^t p_A^2(\xi) d\xi \right]^{\frac{1}{2}}}{p_0} \right\}, \quad (5.56)$$

- kde:
- L je časově průměrovaná hladina akustického tlaku (SPL),
 - T určuje časový interval průměrování,
 - t označuje čas,
 - p_A symbolizuje akustický tlak,
 - ξ značí pomocnou proměnnou časové integrace,
 - p_0 představuje referenční akustický tlak o hodnotě $20 \mu\text{Pa}$.

Důležitou součástí výše uvedeného vztahu pro časové průměrování je tzv. časově průměrovaná hodnota signálu označována též zkratkou RMS. Je vyjádřena jako

$$RMS = \left[\frac{1}{T} \int_{t-T}^t p_A^2(\xi) d\xi \right]^{\frac{1}{2}}. \quad (5.57)$$

Výše uvedené vztahy samozřejmě platí pouze pro spojitě signály. Je-li uvažován vzorkovaný signál, pak je možné z jednotlivých vzorků akustického tlaku určit RMS následovně

$$RMS = \sqrt{\frac{\sum_{k=1}^K p^2(k)}{K}}, \quad (5.58)$$

- kde: – *RMS* je označení pro časově průměrovanou hodnotu signálu (tlaku),
- *K* označuje celkový počet vzorků záznamu,
 - $p(k)$ představuje mikrofonom zaznamenaný akustický tlak.

Pokud jde o časové vážení hladiny akustického tlaku (SPL), zavádí se obecná forma rovnice v podobě

$$L(t) = 20 \log \left\{ \frac{\left[\frac{1}{\tau} \int_{-\infty}^t p_A^2(\xi) \cdot e^{-\frac{t-\xi}{\tau}} d\xi \right]^{\frac{1}{2}}}{p_0} \right\}, \quad (5.59)$$

- kde: – $L(t)$ je časově průměrovaná hladina akustického tlaku (SPL),
- t označuje čas,
 - τ vyjadřuje exponenciální časovou konstantu,
 - p_A symbolizuje akustický tlak,
 - ξ značí pomocnou proměnnou časové integrace,
 - p_0 je referenční akustický tlak (20 μ Pa).

V praxi se signál časově váží dvěma způsoby – rychle a pomalu. Obě varianty se liší volbou exponenciální časové konstanty. Rychlému vážení F (fast) odpovídá hodnota $\tau = 0,125$. Pomalé vážení S (slow) je spojeno s hodnotou $\tau = 1$.

5.4.2 Filtrace – váhové frekvenční filtry (A, B, C, D a Z)

Protože signál může být při měření negativně ovlivněn šumem, zvláště pak jsou-li fonace tišší, šum se svou intenzitou blíží hlasu v měřené místnosti, je součástí měření SPL i filtrace. Kmitočtové váhové filtry přitom definuje mezinárodní standard IEC. Nejčastěji se využívají filtry s označením A, C a Z, přičemž takto frekvenčně vážené SPL se uvádí v dB společně s typem váhového filtru – např. tedy dB(A) či dB(C).

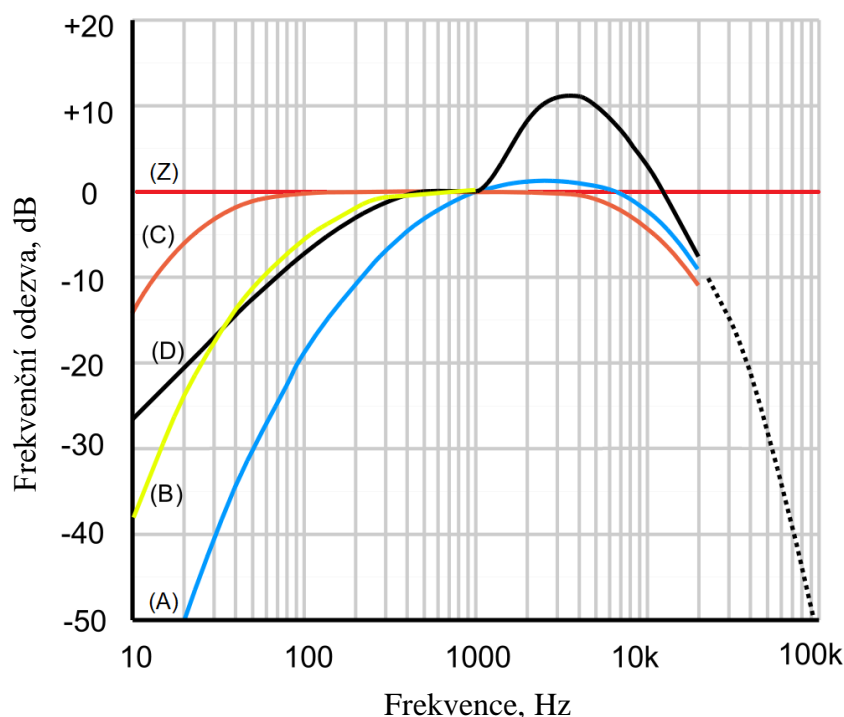
Ze zmíněných filtrů je základní a nejvíce používaný považován filtr typu A. Využití nachází např. u zvukoměrů. Nejvíce zeslabuje nízké frekvence a mírně i ty nad 10 kHz.

Naopak frekvenční odezva je mezi 1-10 kHz téměř rovná, případně jsou některé frekvence zesíleny, což je v souladu s citlivostí lidského sluchu.

Tabulka 5.5 – Hodnoty váhových funkcí A, C a Z dle IEC [88]

Jmenovitý kmitočet	Kmitočtová váhová funkce			Tolerance (dB)	
	dB			Třída	
Hz	A	C	Z	1	2
10	-70,4	-14,3	0	+3,5; -∞	+5,5; -∞
12,5	-63,4	-11,2	0	+3; -∞	+5,5; -∞
16	-56,7	-8,5	0	+2,5; -4,5	+5,5; -∞
20	-50,5	-6,2	0	±2,5	±3,5
25	-44,7	-4,4	0	+2,5; -2	±3,5
31,5	-39,4	-3	0	±2	±3,5
40	-34,6	-2	0	±1,5	±2,5
50	-30,2	-1,3	0	±1,5	±2,5
63	-26,2	-0,8	0	±1,5	±2,5
80	-22,5	-0,5	0	±1,5	±2,5
100	-19,1	-0,3	0	±1,5	±2
125	-16,1	-0,2	0	±1,5	±2
160	-13,4	-0,1	0	±1,5	±2
200	-10,9	0	0	±1,5	±2
250	-8,6	0	0	±1,4	±1,9
315	-6,6	0	0	±1,4	±1,9
400	-4,8	0	0	±1,4	±1,9
500	-3,2	0	0	±1,4	±1,9
630	-1,9	0	0	±1,4	±1,9
800	-0,8	0	0	±1,4	±1,9
1000	0	0	0	±1,1	±1,4
1250	0,6	0	0	±1,4	±1,9
1600	1	-0,1	0	±1,6	±2,6
2000	1,2	-0,2	0	±1,6	±2,6
2500	1,3	-0,3	0	±1,6	±3,1
3150	1,2	-0,5	0	±1,6	±3,1
4000	1	-0,8	0	±1,6	±3,6
5000	0,5	-1,3	0	±2,1	±4,1
6300	-0,1	-2	0	+2,1; -2,6	±5,1
8000	-1,1	-3	0	+2,1; -3,1	±5,6
10000	-2,5	-4,4	0	+2,6; -3,6	+5,6; -∞
12500	-4,3	-6,2	0	+3; -6	+6; -∞
16000	-6,6	-8,5	0	+3,5; -17	+6; -∞
20000	-9,3	-11,2	0	+4; -∞	+6; -∞

Váhový kmitočtový filtr typu C má nejrovnější frekvenční odezvu na intervalu 0,05-4 kHz. Toto frekvenční pásmo je pro záznam a zpracování řeči nejpodstatnější. Filtr typu Z je též označován jako nulový, protože signál nijak nemění. Jeho frekvenční odezva je nulová pro všechny frekvence. Přehled odezev jednotlivých kmitočtových váhových filtrů je znázorněn na obrázku 5.14.



Obrázek 5.14 – Frekvencní charakteristiky váhových kmitočtových filtru

5.5 Určení podobnosti nahrávek

Po úspěšné parametrizaci řečového signálu, přichází na řadu klasifikace. Původní nahrávka je nyní reprezentována sekvencí příznakových vektorů, které mohou být seskupeny do matice, kde počet řádků odpovídá počtu stanovených příznaků a počet sloupců je shodný s počtem rámců, do nichž byla nahrávka segmentována. Zatímco počet příznaků je konstantní, množství rámců se mění v závislosti na délce nahrávky.

5.5.1 Princip borcení časové osy

Za předpokladu, že by byly 2 nahrávky stejně dlouhé, nečinilo by jejich srovnání problém. Stačilo by pak stanovit například Euklidovskou vzdálenost mezi jednotlivými příznakovými vektory a nalezené vzdálenosti sečíst pro všechny tyto vektory. Bohužel toto v praxi při zpracování řeči není možné. Jelikož se tempo různí a člověk při vyslovování dělá různě dlouhé pauzy mezi slovy či větami, pracuje s dechem, melodií apod., nelze očekávat, že budou 2 promluvy stejně dlouhé a přesně časově synchronizované, ačkoliv jsou po obsahové stránce naprosto totožné. Je tedy nezbytné zařadit určitou časovou transformaci, která bude tyto rozdíly kompenzovat.

Obecně metody pro transformaci časové osy předpokládají existenci tzv. reference. Reference, jak již název napovídá, slouží jako vzor, se kterým budou pak další reprezentace porovnávány. Dané algoritmy se pak pokouší rozdíly mezi nahrávkami kompenzovat tak,

že některé rámce zopakují a jiné naopak úplně vynechají. Cílem je, aby byla daná reprezentace co nejpodobnější referenci. Jakmile je určeno, které rámce budou použity pro vzájemné porovnání, určí se hodnota vzdálenosti.

5.5.2 Lineární časová transformace

Jako nejsnazší řešení pro výběr rámců se může jevit tzv. lineární časová transformace (LTW). Tyto rámce jsou voleny na základě lineárního vztahu

$$w(i) = \text{Int} \left[\frac{J-1}{I-1} \cdot (i-1) + 1 + 0,5 \right], \quad (5.60)$$

- kde:
- $w(i)$ je označení pro transformační funkci (cestu),
 - J označuje celkový počet rámců reference,
 - I představuje celkový počet rámců porovnávané reprezentace,
 - $i = \{1, 2, \dots, I\}$ je pořadový index transformační funkce reprezentace.

Využití lineární transformační funkce bohužel nepřináší vždy uspokojivé výsledky, neboť zde není nijak brán zřetel na obsah příznakových vektorů. Hlavní roli zde mají dle zmíněného lineárního vztahu 4.60 pouze počty rámců reference a porovnávané reprezentace. Zde přichází na řadu metoda dynamického borcení časové osy (DTW).

5.5.3 Nelineární časová transformace – dynamické borcení časové osy

Jelikož algoritmus DTW pracuje i s hodnotami z příznakových vektorů, umožňuje volit rámce tak, aby byla reference časově synchronizována s porovnávanou reprezentací. Za zmínku stojí i to, že ačkoliv se metoda hojně používá při zpracování řeči [89], [90], lze ji spolehlivě využít i v celé řadě dalších aplikací. Princip DTW i s příkladem použití z jiné oblasti, než je zpracování řeči, popisuje [91].

Aby mohla být nelineární transformační funkce z pohledu DTW validní, musí splňovat hned několik náležitostí:

- okrajové podmínky,
- monotonicita,
- globální omezení – Itakurovy podmínky spojitosti.

Okrajové podmínky udávají, že transformační cesta musí mít svůj počátek v prvním rámcu reference i dané reprezentace. Obdobně se předpokládá, že cesta bude ukončena posledním rámcem reference i reprezentace. Okrajové podmínky lze pak zapsat ve tvaru

$$w(1) = 1 \wedge w(I) = J. \quad (5.61)$$

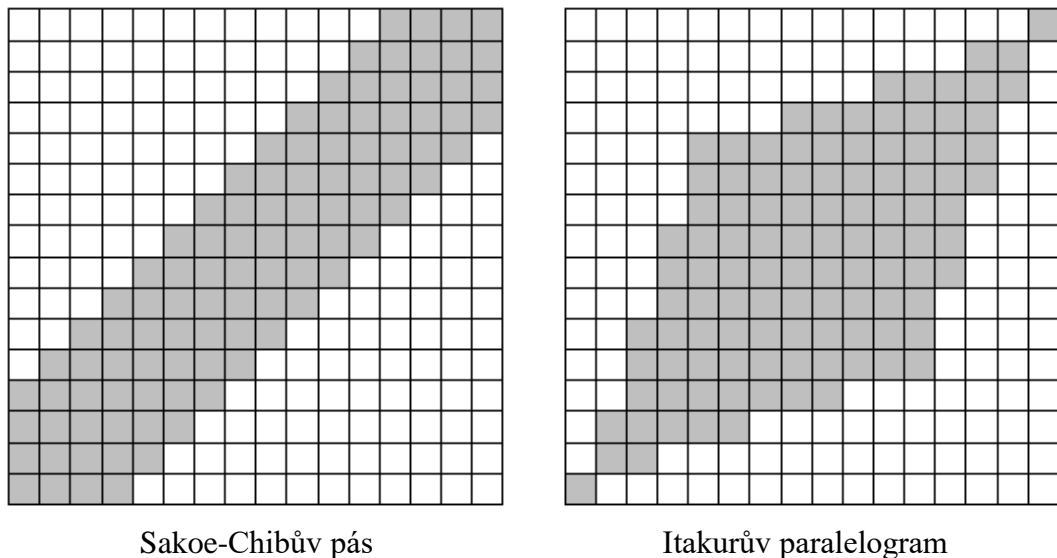
Monotonicita klade důraz na respektování směru toku času. V praxi to znamená, že není možné vracet se k rámcům, které časově předcházejí danému rámcí. Posledním pravidlem, jenž musí být splněno jsou Itakurovy podmínky spojitosti. První část podmínky říká, že jakýkoliv rámeček může být zopakován maximálně jednou. Druhá část popisuje, že se použijí rámečky ve stejném pořadí. Žádný se tedy nevynechá ani opakovaně nepoužije. Poslední část podmínky uvádí, že je možné při konstrukci transformační cesty vynechat rámeček. Do každého bodu se lze tedy dostat ze tří různých směrů. Itakurovy podmínky spojitosti lze pak souhrnně zapsat v následujícím tvaru

$$[w(i) = j \Leftrightarrow w(i-1) = j \wedge w(i-2) \neq j] \vee [w(i-1) = j-1] \vee [w(i-1) = j-2], \quad (5.62)$$

kde: – $j = \{1, 2, \dots, J\}$ značí pořadový index rámečky z pohledu reference,

– $i = \{1, 2, \dots, I\}$ je pořadový index transformační funkce reprezentace.

Splněním Itakurových podmínek je vymezena oblast, v níž mohou ležet nalezené nelineární transformační funkce. Tato oblast se též označuje jako globální omezení. V případě Itakurových podmínek je výsledkem globálních omezení Itakurův paralelogram. Příklady nejpoužívanějších globálních omezení pro DTW jsou na obrázku 5.15.



Obrázek 5.15 – Globální omezení při použití metody DTW [92]

I při zavedení všech zmíněných podmínek a omezení, existuje několik nelineárních transformačních funkcí, které všechny tyto náležitosti splňují. Pro nalezení optimální cesty z pohledu vzdálenosti je nutné nejprve určit tzv. globální vzdálenosti. Cesta s nejnižší globální vzdáleností je pak hledaná nelineární transformační funkce, pro jejíž globální vzdálenost platí

$$D(\mathbf{X}, \mathbf{R}) = \underset{w}{\text{Min}} \sum_{i=1}^I d(\mathbf{x}_i, \mathbf{r}_{w(i)}) \quad (5.63)$$

- kde: – \mathbf{R} je označení pro matici příznakových vektorů \mathbf{r} (reference),
– \mathbf{X} značí matici příznakových vektorů \mathbf{x} (reprezentace),
– $D(\mathbf{X}, \mathbf{R})$ představuje nejmenší globální vzdálenost mezi \mathbf{X} a \mathbf{R} ,
– $w(i)$ je označení pro transformační funkci pro $i = \{1, 2, \dots, I\}$,
– $d(\mathbf{x}_i, \mathbf{r}_{w(i)})$ je lokální vzdálenost mezi příznakovými vektory \mathbf{x} a \mathbf{r} v bodě $[i, w(i)]$.

S využitím Bellemanova principu optimality, jakožto nástroje dynamického programování, lze zefektivnit výpočet DTW. Pro jednoduchost může být počáteční bod transformační funkce $w(1)=1$ označen jako bod A. Obdobně koncový bod splňující vztah $w(I)=J$ bude značen jako bod B. Potom platí, že je-li cesta z bodu A do bodu B optimální a na této cestě se nachází bod K, znamená to, že i cesta z bodu A do bodu K musí být optimální. Díky této strategii se pak optimální cesta hledá rekurzivně. Zavádí se přitom pojem akumulovaná vzdálenost, kterou lze vyjádřit jako

$$A(i, j) = d(\mathbf{x}_i, \mathbf{r}_j) + \text{Min} [A(i-1, j) \cdot q, A(i-1, j-1), A(i-1, j-2)], \quad (5.64)$$

- kde: – $A(i, j)$ představuje akumulovanou vzdálenost v bodě $[i, j]$,
– $q = \{1, \infty\}$ je Itakurův koeficient zohledňující první část podmínky 4.62.

Jakmile algoritmus porovná všechny sady příznakových vektorů, je v samotném závěru určena celková akumulovaná vzdálenost, které je shodná s globální vzdáleností a tady platí, že $A(I, J) = D(\mathbf{X}, \mathbf{R})$. Optimální nelineární transformační funkce se pak hledá od koncového bodu B směrem k počátečnímu bodu A technikou zvanou backtracking. Výsledná globální vzdálenost $D(\mathbf{X}, \mathbf{R})$ hraje klíčovou roli v následné klasifikaci, neboť tato hodnota představuje skóre podobnosti reference a dané reprezentace.

6 Zařízení pro záznam hlasu

Jak již zaznělo dříve, ze záznamu hlasu lze extrahovat celá řada parametrů. Aby to však bylo možné, je nejdříve zapotřebí zvolit způsob, jak hlas subjektu zaznamenat. A k tomu je nutné pořídit odpovídající HW. Nejjednodušším řešením je použití počítače se

zvukovou kartou zabudovaným mikrofonem. Kvalita takto pořízených nahrávek by však v tomto případě byla nedostačující. I v audiometrii a foniatрии platí pro měření závazné normy, a proto musí být volba jednotlivých komponent uskutečněna tak, aby nebyl záznam pacientů negativně ovlivněn použitou technologií. Nejdůležitější částí celého řetězce je nepochybně mikrofon, k němuž se vztahují nejpřísnější požadavky.

6.1 Základní charakteristiky mikrofonu

Volba mikrofonu je podstatnou částí, a proto by neměly být opomenuty některé zásadní znalosti v této oblasti. Samotnému výběru mikrofonu tedy logicky musí předcházet seznámení čtenáře s klíčovými vlastnostmi mikrofonu a jevy, které mohou při nahrávání nastat. Základní povědomí pak zahrnuje například, jaké typy mikrofonů se používají, které charakteristiky jsou uváděny výrobcem v katalogovém listu a jaké informace jsou při výběru relevantní.

6.1.1 Dělení mikrofonů

Mikrofon je zařízení, které slouží jako akusticko-elektrický převodník. To znamená, že převádí signál akustického charakteru (zvuk) na elektrický, který je vhodnější pro další zpracování. Dle [93] existují dva způsoby, kterými je akustický signál snímán. Podle nich se mikrofony dělí na tlakové a gradientní.

Princip funkce tlakových mikrofonů spočívá v působení akustického tlaku vzduchu na pružnou membránu mikrofonu. Výchylka membrány, resp. amplituda jejích kmitů, jakožto i výsledný elektrický signál, jsou pak přímo úměrné intenzitě zvuku a s ním spojeným akustickým tlakem. U tlakových mikrofonů neplatí žádné omezení pro směr přijímaného zvuku, protože se využívají např. v nahrávacích studiích.

Gradientní mikrofony mají oproti tlakovým složitější strukturu, neboť působení akustického signálu na membránu je v tomto případě ze dvou směrů. Na membránu tak působí kromě akustické intenzity ještě její gradient. Jinak řečeno, je zde zohledněn přírůstek akustické intenzity v závislosti na vzdálenosti od zdroje zvuku. To těmto mikrofonům propůjčuje schopnost eliminovat okolní hluk a potlačit sklony ke zpětným vazbám. Jednoduše řečeno, působí-li na mikrofon zvuk ve větší vzdálenosti působení na membránu bude z obou stran srovnatelné, tudíž nebude zvuk zaznamenán. Bude-li zdroj zvuku blízko, pak rozdíl v působení na membránu způsobí výchylku odpovídající tomuto působení, což se pak samozřejmě projeví i na úrovni výstupního elektrického signálu. Výstupní napětí, jenž je přímo úměrné této výchylce, lze vyjádřit jako n -tou derivaci

akustického tlaku podle souřadnice, která je uvažována ve směru šíření akustické vlny. Podkategorií gradientních mikrofonů jsou tzv. rychlostní mikrofony. U těchto zařízení je výchylka membrány přímo úměrná první derivaci akustického tlaku – proto také název rychlostní mikrofony. Úroveň výstupního napětí je tedy přímo úměrná rychlosti vzduchových částic při šíření akustické vlny.

Výslednou kvalitu záznamu i oblast použití, pak ovlivňuje i technologie pro převod zvuku. Dle použité technologie se mikrofony rozdělují na uhlíkové, kondenzátorové, piezoelektrické a elektrodynamické. První a rovněž i nejstarší skupinou jsou mikrofony uhlíkové, které se používají dodnes. Někdy se jim též přezdívá odporové, protože pro svoji funkci využívají princip odporového měniče. Membrána tlakem stlačuje uhlíková zrnka, které mění svůj elektrický odpor, což se pak projeví na velikosti výstupního proudu. Pro svoji jednoduchou konstrukci, ale také značné zkreslení a šum, se využívají převážně tam, kde nemusí být nutně dodržena vysoká kvalita přenášeného zvuku (např. v telefonech). Nutno však podotknout, že uhlíkové mikrofony disponují vysokou citlivostí.

Kondenzátorové neboli elektrostatické mikrofony vyžadují pro svoji funkci elektrostatický převodník. V mikrofonu je obsažen elektrický obvod, jehož součástí je i kondenzátor s proměnlivou kapacitou. Jedna z elektrod tohoto kondenzátoru je pevná, zatímco druhá je tvořena membránou, takže jakákoliv změna výchylky znamená i změnu kapacity. Kondenzátorové mikrofony mohou být jak tlakové, tak i gradientní. Vyznačují se vysokou citlivostí, vyrovnanou frekvenční charakteristikou a téměř až nulovým zkreslením. Řadí se sem i tzv. elektretové mikrofony. Skupina kondenzátorových mikrofonů nachází uplatnění v měřicích přístrojích, ale také v telefonech, počítačích či diktafonech.

Mikrofony pracující na principu piezoelektrického jevu, kdy tlakem či ohybem určitých materiálů vzniká elektrické napětí, se nazývají piezoelektrické. Charakteristickým rysem je nepříliš kvalitní zvuk, relativně velká impedance i citlivost. V dnešní době se používají již jen v omezené míře. V odvětví veřejného ozvučení, kde byly nejčastěji využívány, je nahradily elektrodynamické mikrofony.

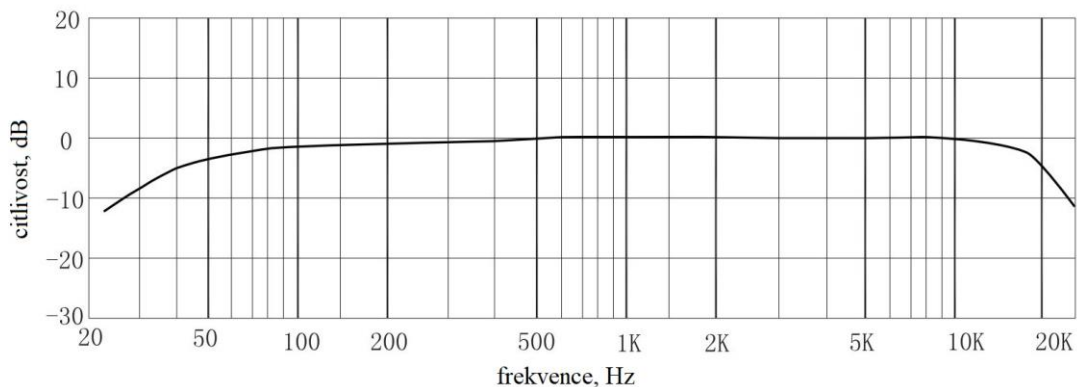
Poslední skupinou jsou již zmíněné elektrodynamické mikrofony. Princip jejich funkce je založen na pohybu cívky v magnetickém poli permanentního magnetu. Pohyb cívky je ovlivněn kmitáním membrány, na kterou je cívka připevněna. Indukuje se v ní tak napětí odpovídající působícímu akustickému tlaku. Výhodou těchto mikrofonů je nízký šum, malé zkreslení i široké frekvenční pásmo. Vyznačují se ale vysokou impedancí. Tyto vlastnosti z nich dělají nejpoužívanější typ mikrofonu pro pódiové ozvučení. Do této

skupiny se rovněž řadí tzv. páskové a elektromagnetické mikrofony, které jsou pouze modifikací výše popsaného principu.

6.1.2 Frekvenční rozsah a frekvenční odezva

Frekvenční odezva neboli frekvenční charakteristika udává, jaké frekvence dokáže mikrofon zaznamenávat. Krom toho vyobrazuje, jestli je elektrický signál ještě zvýrazněn či utlumen a na jakých frekvencích k těmto jevům dochází. Každý mikrofon má tedy pouze určitý výrobcem specifikovaný frekvenční rozsah.

V první řadě je tedy třeba určit, jaký rozsah je vhodný pro záznam řeči. Jinak řečeno, jaká frekvenční informace je v řeči obsažena. Nejhlubší mužský hlas může dosahovat základního tónu 50 Hz. Proto je nezbytné, aby tuto frekvenci již mikrofon dokázal zaznamenat. Je to tedy nejvyšší možná přípustná dolní hranice. Ideální je použití mikrofonu s rozsahem od 10 Hz. Nalezení horní hranice už je obtížnější. Řeč typicky obsahuje různé vysokofrekvenční komponenty. Nejvyšší frekvenční komponenty jsou v řeči přítomny při vyslovování souhlásek zejména pak sykavky „s“, kdy se maximum ve spektru nachází někde v rozmezí 7-8 kHz, ale výjimkou nejsou ani frekvence okolo 10 kHz a vyšší. Nejen proto se ve výsledku doporučuje, aby mikrofon dokázal zachytit stejný rozsah jako lidský sluch. Ideální horní hranice frekvenčního rozsahu by tedy měla být mezi 16-20 kHz.



Obrázek 6.1 – Příklad frekvenční odezvy mikrofonu s ledvinovou směrností [94]

Frekvenční odezva (charakteristika) je výrobcem prezentována v podobě grafu, kde na horizontální osu se vynášejí frekvence a vertikální osa je cejchována v decibelech. Protože intenzita zvuku je závislá na vzdálenosti od zdroje zvuku, bývá často frekvenční charakteristika uváděna pro různé vzdálenosti. Z hlediska vhodnosti volby mikrofonu pro měření a zaznamenávání hlasu je důležitý tvar charakteristiky. Frekvenční odezva správně zvoleného mikrofonu by jednoduše měla být rovná. To totiž znamená, že hlas je pro dané

frekvence realisticky zachycen a nezkrácen. Zachovávají se tak základní vlastnosti zaznamenávaného zvuku (hlasu). Například v USA je zavedeno dle institutu ANSI dělení na mikrofony typu LS1, LS2 a ostatní, které nesplňují zvolené kritérium. Prakticky to znamená, že oba typy mikrofonů LS1 i LS2 musí mít rovnou frekvenční odezvu v celém stanoveném frekvenčním rozsahu, a to s tolerancí 2 dB. Rozdíl mezi nejnižší a nejvyšší úrovní citlivosti mikrofonu v celém rozsahu tuto hodnotu nesmí přesáhnout, aby mohl být mikrofon zařazen do těchto kategorií. Zatímco mikrofony typu LS1 mají frekvenční rozsah 10-8000 Hz, typ LS2 nabízí rozsah 10-20000 Hz.

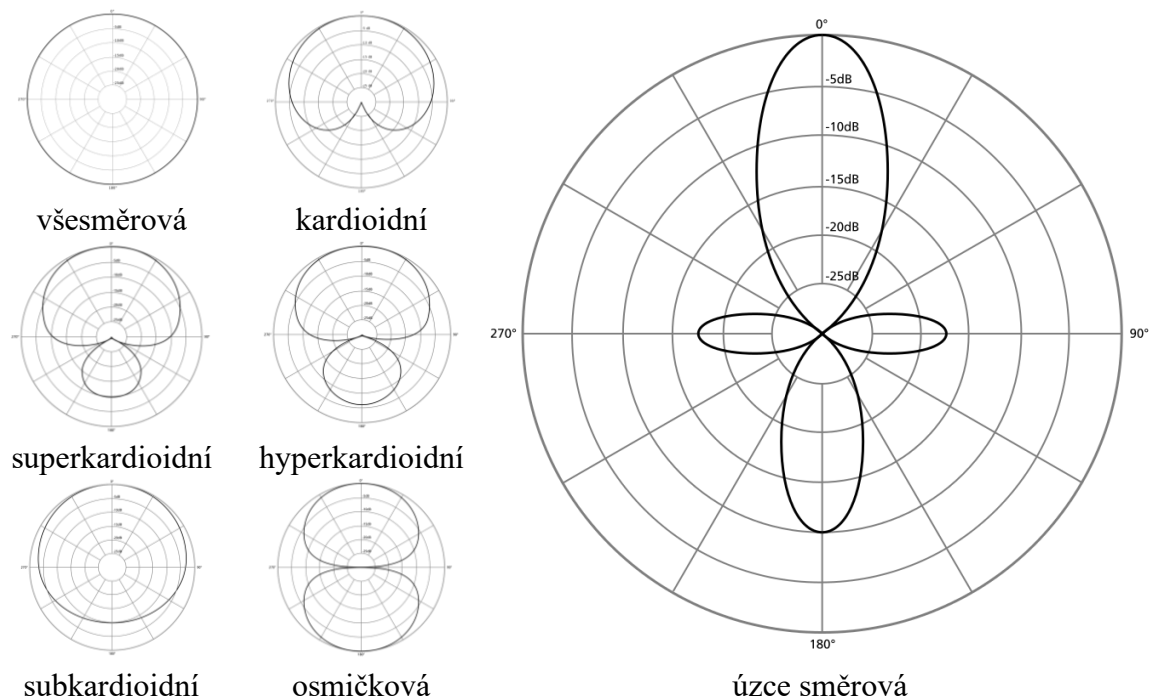
6.1.3 Směrová charakteristika

Kromě frekvencí, které dokáže mikrofon zachytit, je také velmi důležitá tzv. směrovost mikrofonu. Ta je spolu s frekvenční odezvou základní charakteristikou mikrofonu. Směrovost pro daný mikrofon vypovídá o tom, ze kterého směru je zvuk přijímán nejlépe a naopak, který směr je pro použití mikrofonu nevhodný. Obecně lze mikrofony rozdělit na dvě základní skupiny – všesměrové a směrové, přičemž směrové se dále dělí. Graficky se směrovost znázorňuje tzv. směrovou charakteristikou, která je frekvenčně závislá. Svou roli u směrových mikrofonů hraje i vzdálenost od zdroje zvuku. Směrové charakteristiky mikrofonů bývají pojmenovány podle svého tvaru. Nejčastěji bývá směrovost rozdělována do sedmi skupin:

- kulová, všesměrová čili omnidirekcionální (omni-directional),
- ledvinová neboli kardioidní (cardioid),
- superkardioidní (super-cardioid),
- hyperkardioidní (hyper-cardioid),
- subkardioidní (sub-cardioid),
- osmičková neboli bidirekcionální (figure of 8 nebo také bi-directional),
- úzce směrová, někdy také kuželová (shotgun).

Ideální z hlediska směrovosti je všesměrová charakteristika, která je typická pro tlakové mikrofony. Takový mikrofon pak přijímá zvukový signál se stejnou citlivostí nezávisle na směru zdroje zvuku. I všesměrové mikrofony však naráží na své omezení, které spočívá ve frekvenci. Při překročení určitého frekvenčního limitu totiž mikrofon ztrácí svoji všesměrovost a stává se směrovým, jak uvádí [95]. Nastává k tomu, když je vlnová délka srovnatelná s rozměry citlivých částí mikrofonu. Dochází pak k deformaci akustického pole – difrakci zvuku. Proto je frekvenční omezení závislé na rozměrech zvoleného mikrofonu.

Ze směrových mikrofonů jsou pak nejčtenějšími zástupci ty, které mají kardioidní směrovost. Směrová charakteristika je znázorněna v polárních souřadnicích. Zatímco amplituda představuje citlivost mikrofonu v dB, úhel v tomto případě udává polohu zdroje zvuku vzhledem k mikrofonu, přičemž jako výchozí hodnota 0° se vždy bere tzv. akustická osa. Ta bývá většinou totožná s osou geometrickou. Pokud osy stejné nejsou, výrobce musí polohu akustické osy zahrnout do specifikace. V grafu bývají též zaneseny ekvisenzitivní křivky. Jedná se o kružnice spojující body se stejnou citlivostí. V grafu jsou zaneseny pro lepší orientaci a vymezují tak jednotlivé hladiny citlivosti. Běžné je dělení po 5 dB. Více informací o směrovosti shrnuje např. [96]. Grafická podoba směrových charakteristik jednotlivých typů směrovosti včetně detailu úzce směrové charakteristiky je znázorněna na obrázku 6.2.



Obrázek 6.2 – Přehled směrových charakteristik mikrofonů

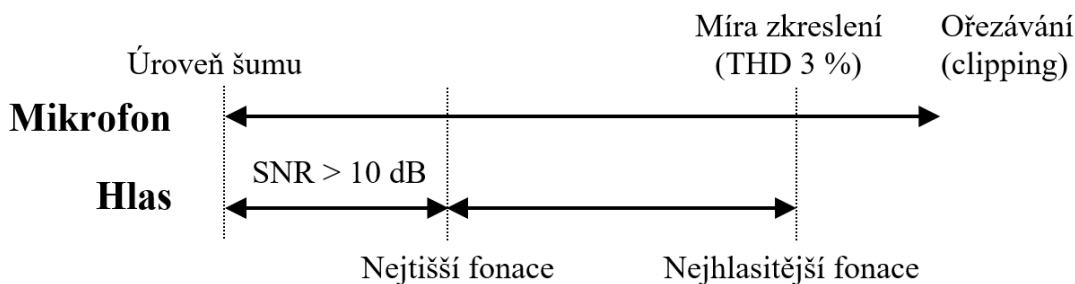
Se směrovými mikrofony se pojí ještě jeden jev, který není radno podceňovat. Řeč je o tzv. „Proximity efektu“ [97]. Ten způsobuje zvýraznění nižších frekvencí. V hudebním průmyslu se někdy jedná o žádoucí a často i využívaný efekt, který způsobí zvýraznění basů. Při záznamu řeči pro účely diagnostiky je však třeba se proximity efektu vyhnout. Tento jev je v rozporu se základní zásadou – zachytit hlas co nejvíce realisticky a nezkrasleně. Proximity efekt je závislý na vzdálenosti. Čím blíže je zdroj zvuku k mikrofonu, tím více jsou zvýrazněny nízké frekvence a zároveň se i posunuje horní

hranice frekvencí, které jsou proximity efektem ovlivněny. Proto též někdy bývají frekvenční i směrové charakteristiky uváděny ve více vzdálenostech od mikrofonu. Řešením, jak se proximity efektu úspěšně vyhnout je použití všesměrových mikrofonů. Ty totiž, jako jediné, tímto neduhem netrpí.

6.1.4 Ostatní parametry

Jedním z dalších parametrů je odstup signálu od šumu (SNR). Pokud by měl šum (teoreticky) nulovou úroveň čili by nebyl vůbec zaznamenán, výsledkem by byl naprosto čistý signál. V praxi je pro přesná měření klíčové, aby SNR byl co možná největší. Čím menší SNR je, tím více je zaznamenávaný signál šumem negativně ovlivněn. Při srovnatelných úrovních signálu a šumu v pozadí pak dochází ke ztrátě informace obsažené v původním signálu. Doporučuje se tak SNR alespoň 10 dB a větší.

Dalším důležitým parametrem je dynamický rozsah. Ten udává rozmezí, v jakém mikrofon generuje výstupní signál. Dolní hranice udává nejnižší možnou zachytitelnou úroveň, která je omezena vlastním šumem mikrofonu. Horní hranice udává maximální hodnotu SPL převedenou na elektrický signál s 3 % zkreslením (THD). Někdy bývá horní hranice též označována jako AOP (Acoustic Overload Point). Při překročení horní hranice pak může docházet k ořezávání signálu. V praxi to například znamená, že špičky signálu, který má být zaznamenán, zcela chybí a jsou nahrazeny konstantní hodnotou odpovídající maximálnímu možnému výstupu mikrofonu. Jedná se o změnu, která je v nahrávce většinou rozeznatelná pouhým sluchem.

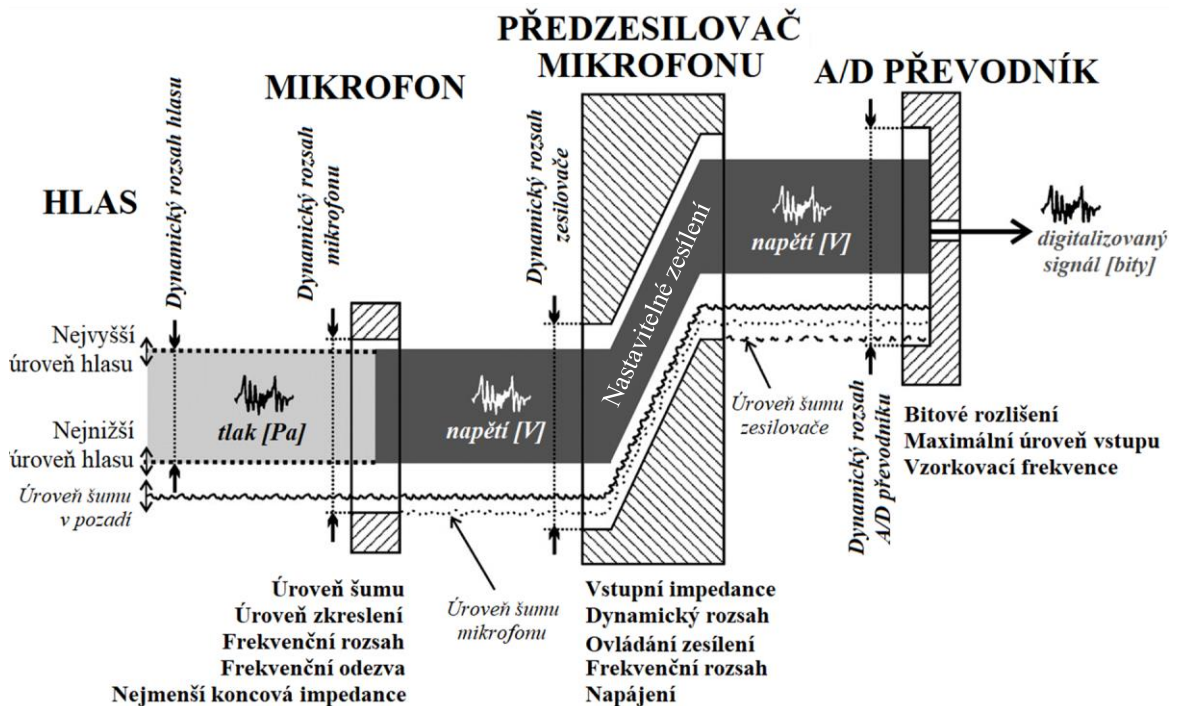


Obrázek 6.3 – Dynamický rozsah mikrofonu

Protože intenzita zvuku s rostoucí vzdáleností klesá, instituce UEP stanovila 2 vzdálenosti doporučené pro používání mikrofonu. V první řadě je doporučena vzdálenost 30 cm od úst, která platí zejména pro stolní a stojanové mikrofony. Alternativou je vzdálenost 5 cm od koutku úst. Ta může být uplatněna u hlavových setů. Obecně platí, že při zdvojnásobení vzdálenosti mezi ústy a mikrofonem dochází k poklesu o 6 dB. Jsou-li uvažovány ony referenční vzdálenosti pak platí, že pokles mezi 5 a 30 cm je přibližně o 15

dB. Proto je dynamický rozsah mikrofону, stejně jako další charakteristiky, závislý na vzdálenosti od zdroje zvuku.

Ve hře jsou i další faktory jako například bitové rozlišení A/D převodníku, použitá vzorkovací frekvence, parametry mikrofonního předzesilovače jako vhodné nastavení zesílení atd. Celý proces převodu akustického tlaku zvukového zdroje na digitální signál pro další zpracování a všechny faktory, které ovlivňují výslednou kvalitu a realistické zachycení signálu, jsou znázorněny na obrázku 6.4.



Obrázek 6.4 – Převod hlasu do digitalizované podoby včetně ovlivňujících faktorů [95]

6.2 Doporučení pro volbu mikrofonu

Dlouhou dobu probíhala analýza hlasu bez jasně stanovených pravidel pro použití mikrofonu. Volbě HW tak často nebyla věnována patřičná pozornost, což mělo za následek nereprodukovatelnost daných měření, resp. nesprávně naměřené hodnoty. Tento nedostatek efektivně řeší článek [95] z roku 2010, který se tak stal hlavní šablonou v tomto ohledu. Další poznatky pak souhrnně prezentuje [48] a přispívá tak k jakési unifikaci, která v této oblasti zcela chyběla. Daná opatření a doporučení tak mají za cíl zajistit významně nezkreslené zachycení SPL, základního tónu či spektra řeči. V souladu s UEP jsou hodnoty uváděny pro vzdálenosti 5 a 30 cm od zdroje zvuku. Stále však platí, že se jedná pouze o doporučení. Jednotlivé hodnoty nemusí být absolutně závazné. Řeč je právě o hodnotách

pro vzdálenost 5 cm, kde situaci komplikuje nejistota ohledně vyzařování zvuku z úst kolem hlavy do stran. Šest hlavních zásad formuluje [48] následovně:

- úroveň šumu mikrofonu (spodní hranice dynamického rozsahu mikrofonu) i šumu v pozadí musí mít od SPL nejnižších fonací odstup alespoň 10 dB,
- pro horní hranici dynamického rozsahu mikrofonu by mělo platit, že je nad úrovní SPL nejhlasitějších zaznamenávaných fonací,
- frekvenční rozsah mikrofonu $\langle f_d; f_h \rangle$ by měl mít takovou dolní hranici, aby byla nižší než základní tón řeči (ideálně menší než 50 Hz – hluboký mužský hlas),
- horní mez frekvenčního rozsahu by měla být nad hodnotou nejvyšší významné ve spektru,
- v celém frekvenčním rozsahu by frekvenční odezva (charakteristika) měla být rovná – s absolutní odchylkou ne větší než 2 dB,
- použití směrových mikrofonů pro určování spektra a měření SPL by mělo být výhradně v takové vzdálenosti, která vylučuje proximity efekt a ve které je frekvenční odezva rovná.

Při dodržení 10 dB odstupů šumu od fonací je SPL ovlivněno o méně než 0,5 dB. Pokud jde o hlavové sety, při jejichž používání je mikrofon umístěn na straně od úst, určité zesílení mezi 5 kHz a horní mezí frekvenčního rozsahu není na škodu, protože vysoké frekvence jsou při vyzařování do stran utlumeny. Doporučení jsou shrnuta v tabulce 6.1.

Tabulka 6.1 – Doporučení pro volbu mikrofonu určeného pro měření hlasu a řeči [48]

Mikrofonní specifikace	Doporučení	Extrémní rozsah		Omezený rozsah	
		30 cm	5 cm	30 cm	5 cm
Hladina šumu	SNR > 10 dB	≤ 25 dB(A)	≤ 40 dB(A)	≤ 40 dB(A)	≤ 55 dB(A)
Maximální SPL (3 % THD)	Nad SPL nejhlasitějších fonací	≥ 130 dB	≥ 145 dB	≥ 120 dB	≥ 135 dB
f_d Dolní mez frekvenčního rozsahu (-2 dB)	Pod nejnižší frekvencí hlasu	≤ 10 Hz	≤ 10 Hz	≤ 50 Hz	≤ 50 Hz
f_h Horní mez frekvenčního rozsahu (-2 dB)	Nad nejvyšší podstatnou frekvencí spektra hlasu	> 16 kHz	> 16 kHz	> 8 kHz	> 8 kHz
Vyrovnanost odezvy mezi f_d a f_h (kromě zesílení nad 5 kHz)	Co nejrovnější	≤ 2 dB	≤ 2 dB	≤ 2 dB	≤ 2 dB
Maximální zesílení mezi 5 kHz a f_h	Žádné	≤ 1 dB	≤ 3 dB	≤ 3 dB	≤ 5 dB

6.3 Kalibrace mikrofonu

Správná volba mikrofonu je pouze prvním krokem. Aby však mohl být mikrofon použit pro diagnostické účely, zejména pak pro měření SPL, musí být mikrofon řádně kalibrován. V případě, že mikrofon kalibrován není, nelze pak měření reprodukovat s jiným mikrofonem a není ani možné výsledky porovnávat s dalšími studiemi. Kalibrace je tedy nezbytným krokem, který je třeba provést ještě před tím, než se bude signál dále zpracovávat. Svě místo tak zaujímá hned po převodu spojitého akustického signálu na vzorkovaný elektrický signál. Celý proces zpracování řeči včetně kalibrace pak zachycuje obrázek 6.5.



Obrázek 6.5 – Kalibrace v řetězci úlohy zpracování řeči

Kalibrování mikrofonu může probíhat několika způsoby. V zásadě jde však o to, aby bylo možné k jednotlivým úrovním vzorkovaného signálu přiřadit odpovídající akustický tlak, resp. SPL. Toho lze docílit tak, že mikrofonem bude zaznamenán zvuk se známým SPL. Tato hodnota pak slouží jako referenční a umožňuje určení přepočtení kalibrační konstanty. Díky této kalibrační konstantě již lze jednotlivým úrovním přiřadit hodnoty akustického tlaku či SPL. Pro generování signálu se známými parametry se používá tzv. kalibrátor. Podle [98] kalibrátor umožňuje nejčastěji generovat harmonický signál na frekvenci 1 kHz a hodnotě SPL 94 dB. Tato hodnota odpovídá akustickému tlaku 1 Pa. Někdy se též jako referenční hodnota pro kalibraci používá 114 dB při téže frekvenci. Kalibrátor je vždy třeba volit tak, aby byl kompatibilní se zvoleným mikrofonem. Podstatné jsou v tomto případě rozměry mikrofonu.

6.3.1 Dvoustupňová (nepřímá) kalibrace

Výše popsaný princip by mohl být označen jako jednostupňová (přímá) kalibrace. Metoda dvoustupňové či dvoufázové kalibrace byla původně publikována v [99] a později i využita v rámci práce [88]. Tento postup zahrnuje použití 2 mikrofonů a kombinuje jejich výhody – realistické zaznamenání hlasu dle dříve zmíněných požadavků a přesné určení SPL. SPL lze měřit i mobilním telefonem s vhodnou aplikací. Podle [100] se však výsledky dostupných aplikací dramaticky různí. Pro přesné měření SPL se v praxi používají pro tento účel vyvinutá zařízení. Řeč je o tzv. zvukoměrech či hlukoměrech. Součástí zvukoměru je velmi přesný měřicí mikrofon, který zaznamenává akustický signál a

přepočítává jej na SPL. I takový zvukoměr však musí být kalibrován. Každý takový přístroj je kalibrován již od výrobce, ale ideální je provést kontrolní kalibraci před každým měřením. Takový zvukoměr tedy dokáže poměrně přesně měřit SPL. Konkrétní přesnost je specifikována výrobcem pro jednotlivé modely a přístroje se podle ní dělí i na třídy v rozmezí 1-5, kde 1. třída označuje ty nejpřesnější.

Mikrofon zvukoměru je dobře uzpůsoben pro měření hluku (SPL), ale pro kvalitní záznam hlasu dle dříve zmíněných kritérií se nehodí. Proto je vhodnější využít hlavový mikrofon připevněný poblíž úst (samozřejmě dle uvedených doporučení). Díky bližší vzdálenosti od úst tak bude větší SNR a hlas bude kvalitněji zachycen.

A nyní již k technice dvoustupňové kalibrace jako takové. V prvním kroku probíhá klasická jednostupňová (přímá) kalibrace zvukoměru. Kalibrátor se nasadí na zvukoměr a při generování známého kalibračního tónu se určí kalibrační konstanta zvukoměru c_Z ze vztahu

$$L_R = 20 \log(c_Z \cdot \text{RMS}_Z), \quad (6.1)$$

kde: – L_R je referenční hodnota 94 dB (kalibrační tón),

- c_Z značí hledanou kalibrační konstantu zvukoměru,
- RMS představuje časově průměrovanou hodnotu zaznamenaného signálu.

Nyní je zvukoměr zkalibrován a umístí se do vzdálenosti 30 cm od zdroje zvuku. Subjekt si nasadí hlavový mikrofon, který bude upevněn ve vzdálenosti 5-10 cm od úst. Není nutné dodržet vzdálenost přesně 5 cm. Za to je ale nezbytné, aby vzdálenost od úst byla stále stejná. Totéž platí o dodržování vzdálenosti od zvukoměru, ta již musí zůstat nezměněna. Pro druhou fázi kalibrace postačí například fonace samohlásky a. Pak je možné s využitím předchozí znalosti konstanty c_Z nalézt i hodnotu konstanty c_M . Přepočet vychází ze vzorce

$$c_M \cdot \text{RMS}_M = c_Z \cdot \text{RMS}_Z \Rightarrow c_M = c_Z \cdot \frac{\text{RMS}_Z}{\text{RMS}_M}, \quad (6.2)$$

kde: – c_M je symbol pro hledanou kalibrační konstantu mikrofonu,

- RMS_M je časově průměrovaná hodnota zaznamenaného signálu mikrofonu,
- c_Z symbolizuje již známou kalibrační konstantu zvukoměru,
- RMS_Z značí časově průměrovanou hodnotu zaznamenaného signálu zvukoměru.

Dvoustupňová (nepřímá) kalibrace je znalostí obou kalibračních konstant ukončena. Při dodržení výše zmíněných vzdálenostních omezení může být hlas zaznamenáván. K záznamu je využit kvalitní zvuk zaznamenaný hlavovým mikrofonem s nižší úrovní šumu, SPL je však přepočítáváno na virtuální vzdálenost 30 cm, tak jak je zobrazována na zvukoměru.

7 Metodika použita pro sběr diagnostických dat

V rámci praktických měření je třeba zvolit metodiku, podle níž se bude postupovat. Musí být nastavena jednotná pravidla, aby byla měření reprodukovatelná a všichni pacienti byly zaznamenáni stejným způsobem. Mimo jiné je třeba dbát i na to, aby do měření nebyla vnesena nežádoucí variabilita, která by ovlivnila výsledky měření. Z toho důvodu je třeba též specifikovat podmínky, které musí být dodrženy, aby se nežádoucí vlivy nepodepsaly na kvalitě pořízené nahrávky. Součástí je nepochybně i volba jednotlivých přístrojů a jejich správné nastavení.

7.1 Měřicí řetězec

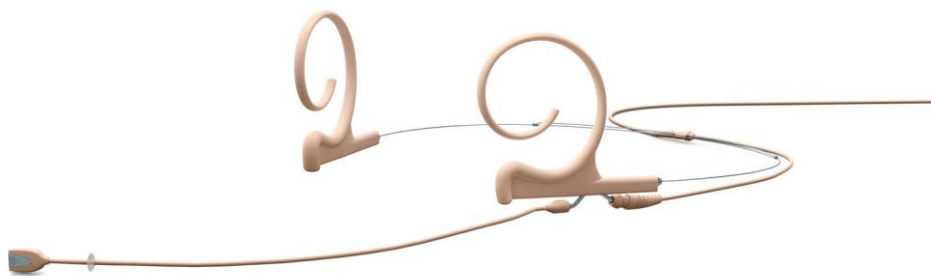
Samotnému návrhu metodiky měření zcela logicky předcházela volba jednotlivých komponent celého měřicího řetězce. Seznam všech použitých komponent (bez kabeláže a redukci) přehledně shrnuje tabulka 7.1.

Tabulka 7.1 – Seznam komponent měřicího řetězce

Mikrofon	DPA d:fine Slim, 110 mm Boom
Zvukoměr	EXTECH 407780A
Kalibrátor	EXTECH 407744
Externí zvuková karta	Focusrite Scarlett 2i2 2nd Generation

7.1.1 Měřicí mikrofon

Aby mohl být hlas pacientů kvalitně a realisticky zaznamenán pro další zpracování, bylo přistoupeno ke dvoufázové kalibraci, ke které jsou potřeba dva mikrofony. Pro záznam hlasu byl použit hlavový mikrofon DPA d:fine Slim, který je na obrázku 7.1.

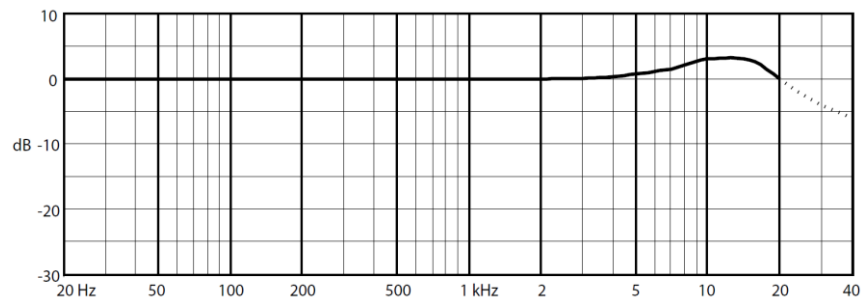


Obrázek 7.1 – Hlavový mikrofon DPA d:fine použitý pro záznam hlasu pacientů [101]

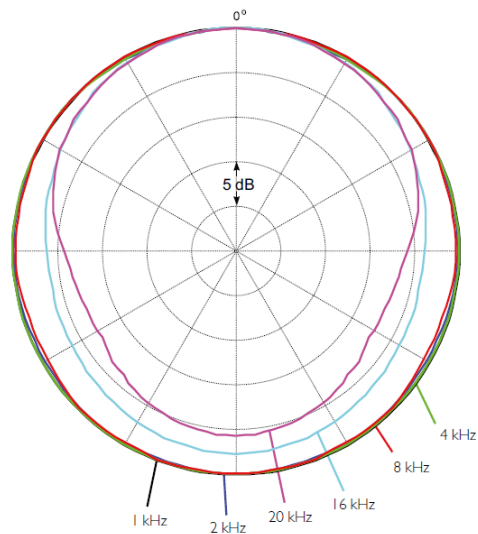
Mikrofon byl zvolen na základě splnění požadavků pro volbu měřicích mikrofonů. Frekvenční odezva mikrofonu je rovná, což je pro pořízení záznamu ideální. Část frekvenčního rozsahu mezi 8-15 kHz navíc disponuje zesílením o velikosti 3 dB, což je u hlavových mikrofonů žádoucí, dle již zmíněných doporučení. Graf frekvenční odezvy mikrofonu je na obrázku 7.2. Mikrofon je všesměrový, aby při měření nedocházelo k proximity efektu. Jeho směrová charakteristika v závislosti na frekvenci je zakreslena na obrázku 7.3. Základní parametry, které byly směrodatné při výběru vhodného mikrofonu, jsou pak vypsány v tabulce 7.2.

Tabulka 7.2 – Klíčové parametry mikrofonu DPA d:fine [101]

Směrová charakteristika	Všesměrová
Frekvenční rozsah	20 Hz - 20 kHz
Dynamický rozsah	97 dB
Úroveň vlastního šumu mikrofonu při 20 μ Pa	26 dB(A), maximálně 28 dB(A)
SNR při kalibrační hodnotě 94 dB (1 Pa)	68 dB(A)
Maximální zaznamatelné SPL bez ořezávání	144 dB



Obrázek 7.2 – Frekvenční charakteristika (odezva) mikrofonu DPA d:fine [101]



Obrázek 7.3 – Směrová charakteristika zvoleného mikrofonu DPA d:figure [101]

7.1.2 Zvukoměr

Pro měření hladiny akustického tlaku (SPL) byl zvolen zvukoměr (hlukoměr) značky EXTECH typ 407780A. Disponuje vyměnitelným 0,5" (12,7 mm) mikrofonem. Konkrétně se jedná o kondenzátorový elektretový mikrofon, který je opatřen krytkou proti větru. Zvukoměr je třídy 2 a vyhovuje standardům IEC 61672-1, 60651, 60804 a ANSI S1.4 2. Má k dispozici celkem 5 měřicích rozsahů po 60 dB. Na displeji může zobrazovat i sloupcový graf právě měřeného SPL. Má též interní paměť pro uchování až 32000 záznamů. Umožňuje volit mezi impulzní, rychlou a pomalou odezvou. Obsahuje AC a DC výstup pro další zpracování. Rovněž má integrované rozhraní pro připojení k počítači přes USB. Umožňuje tak nejen záznam dat, ale s přiloženým SW i vizualizaci v reálném čase. Nejdůležitější parametry jsou sepsány v tabulce 7.3. Podrobnější informace jsou uvedeny v katalogovém listu [102] a manuálu [103].

Tabulka 7.3 – Parametry zvukoměru EXTECH 407780A

Měřicí rozsah (teoretický)	30-130 dB
Měřicí rozsahy	30-90, 40-100, 50-110, 60-120, 70-130 dB
Frekvenční rozsah	31,5 Hz - 8 kHz
Kmitočtové váhové filtry	A nebo C
Přesnost při ref. hodnotě 94 dB, 1 kHz	±1,5 dB



Obrázek 7.4 – Zvukoměr EXTECH 407780A [104]

7.1.3 Kalibrátor

Pro kalibraci byl využit kalibrátor EXTECH 407744, který je kompatibilní se zvoleným zvukoměrem. Zmíněná kompatibilita spočívá v rozměrech mikrofonu, který má být kalibrován. Daný kalibrátor umožňuje kalibraci pouze 0,5" mikrofonů. Zvolený typ

generuje harmonický kalibrační signál o frekvenci 1 kHz a hladině akustického tlaku (SPL) 94 dB, což odpovídá zhruba akustickému tlaku 1 Pa. Přesnost úrovně generovaného signálu je $\pm 0,8$ dB. Uváděná frekvenční přesnost je 5 %. THD je menší než 2 % při referenční hodnotě 94 dB. Kalibrátor je napájen 9 V baterií a má indikační LED diodu pro stav baterie. Po přepnutí přepínače do polohy generování kalibračního tónu, se ze zvukoměru sejme krytka mikrofону proti větru a kalibrátor se přiloží na mikrofon. Zvukoměrem je pak zaznamenána tato referenční hodnota. Další informace o použitém kalibrátoru jsou k dispozici v manuálu [105] a katalogovém listu [106].



Obrázek 7.5 – Kalibrátor EXTECH 407744 [107]

7.1.4 Externí zvuková karta

Protože integrované zvukové karty v počítačích a noteboocích nejsou uzpůsobeny pro přesná měření, byl by záznam, byť pořízen kvalitním mikrofonom, který splňuje všechna doporučení, znehodnocen po připojení k takové kartě. Z toho důvodu se hledala vhodná alternativa, kterou představuje použití externí zvukové karty – v tomto případě Focusrite typ Scarlett 2i2 (druhá generace), jenž je zobrazena na obrázku 7.6. Externí zvuková karta má stejnou roli jako ta integrovaná. Převádí zvuk z mikrofону do počítače (A/D konverze) a umožňuje jej uchovávat na úložném zařízení počítače v digitalizované podobě. Rozdíl spočívá ve větší kontrole nad celým procesem a možností volby vhodné karty dle specifikovaných parametrů.



Obrázek 7.6 – Externí zvuková karta Focusrite Scarlett 2i2 2nd Generation [108]

Připojení karty Scarlett k počítači je řešeno přes USB port. Tato externí karta podporuje vzorkování vstupního signálu s frekvencí 44,1 až 192 kHz. Má dva oddělené vstupy (kanály). Každý ze vstupů je opatřen přepínačem LINE/INST, jehož poloha upravuje vlastnosti daného vstupu. Řeč je zejména o vstupní impedanci, která je při připojení hudebního nástroje (poloha INST) v řádu M Ω . Mírně se ale mění i dynamický rozsah a THD. V poloze LINE je možné použít další zvukařskou techniku nebo mikrofon. I obě tyto varianty mají odlišné vstupní impedance. Oba vstupy umožňují připojit konektory XLR nebo Jack 6,35 mm, přičemž v poloze LINE je jako mikrofon rozpoznáno pouze zařízení připojené přes klasický 3 pinový XLR. Parametry mikrofonního vstupu shrnuje tabulka 7.4.

Tabulka 7.4 – Parametry zvukové karty Focusrite Scarlett 2i2 2nd Generation [109]

Frekvenční rozsah	20 Hz - 20 kHz
Frekvenční odezva	rovná $\pm 0,1$ dB
Dynamický rozsah	111 dB(A)
Nastavitelné lineární zesílení	až 56 dB
Impedance vstupu	3 k Ω

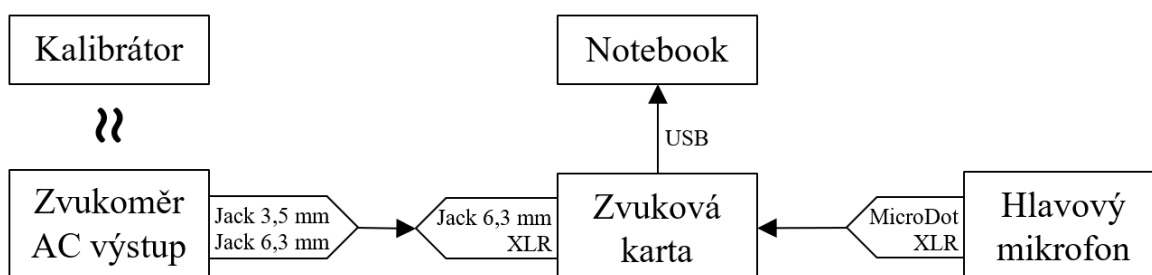
Karta Scarlett 2i2 dále umožňuje nastavovat (pro oba vstupy jednotlivě) zesílení, které je lineární v celém rozsahu. Otočné ovladače zesílení pak disponují ještě barevným podkreslením, které udává, jaká je úroveň signálu. Nesvítlí-li vůbec, pak je signál příliš slabý. Je-li rozsvícena zelená vše je v pořádku. Oranžová pak znamená přiblížení k horní hranici rozsahu. Červená barva značí překročení horní hranice a ořezávání signálu. Toho je samozřejmě nutné se vyvarovat během měření. Je zde i možnost tzv. fantomového napájení vstupů, které je pro funkci zvoleného mikrofonu nezbytné. Je zde i možnost poslechu v reálném čase, k čemuž slouží výstup pro sluchátka s nastavitelným zesílením. Další informace uvádí výrobce [109].

7.1.5 Zapojení a nastavení přístrojů

Aby mohlo vše fungovat, bylo nutné ještě pořídit odpovídající kabeláž. Protože výstupní konektory zvukoměru ani mikrofonu neodpovídají požadovaným, bylo nutné kabeláž doplnit ještě o redukce. Vhodné kabely a redukce byly vybrány po konzultaci s odborníky v oblasti audiotechniky. Zde ještě stojí za zmínku, že čím méně redukcí se pro výsledné zapojení použije, tím více zůstane zachována kvalita přenášeného signálu.

Hlavový mikrofon je opatřen konektorem MicroDot. K němu byla pořízena redukce přímo na požadovaný XLR konektor. AC výstup zvukoměru je kompatibilní pouze s koncovkou Jack 3,5 mm. Protože přímá konverze na XLR není možná, byla pořízena

stereo redukce (označovaná též jako TRS) na Jack 6,3 mm. Propojení s prvním (levým) vstupem externí zvukové karty je pak realizováno pomocí kabelu s konektorem Jack 6,3 mm na jedné straně a XLR na straně druhé. K záznamu a zpracování zvukových dat byl použit notebook Dell Inspiron 15 3000 (3576), který splňuje systémové požadavky pro běh programu Matlab 2018a, ve kterém byl celý experiment navržen. Schéma výsledného zapojení je znázorněno na obrázku 7.7.



Obrázek 7.7 – Schéma zapojení měřicího řetězce včetně použitých redukcí

Zvukoměr je zapojen do prvního vstupu externí zvukové karty (levý kanál), zatímco měřicí hlavový mikrofon je připojen na druhý vstup (pravý kanál). Je tak možné zaznamenávat zvuk z obou kanálů současně a ukládat je do jediného souboru (jako běžná stereo nahrávka). Kalibrátor je použit pouze během kalibrace a není propojen se zvukoměrem ani zbytkem řetězce žádnou kabeláží. Je pouze fyzicky přiložen na mikrofon zvukoměru. Na zvukové kartě je nastavení následující:

- oba přepínače jsou v poloze LINE,
- úroveň zesílení je nastavena tak, aby nebyl otočný ovladač podsvícen červeně,
- tlačítko fantomového napájení (48 V) je stisknuté a svítí červeně.

Nastavení zvukoměru zahrnuje:

- změnu režimu na měření SPL, dB,
- přepnutí měřicího rozsahu na 40-100 dB,
- aplikaci frekvenčního vážení typu C.

7.2 Sběr dat

Po pořízení vybavení bylo třeba dohodnout se na dalším postupu a domluvit se na způsobu získávání dat. Pro možnost měření pacientů, kteří mají absolvovat operaci štítné žlázy nebo trpí jinými hlasovými obtížemi, se Univerzita Pardubice spojila s Pražskou Fakultní nemocnicí Královské Vinohrady (FNKV). Společně obě strany došli k návrhu vhodné metodiky. FNKV pro tento výzkum vyčlenila tým a umožnila nahrávání pacientů v tiché komoře audiometrie na své ORL klinice.

7.2.1 Rozbor metodiky

Pro dosažení co nejvyšší kvality nahrávek je nutné snížit nepotřebnou variabilitu na minimum. Proto bylo každé nahrávání prováděno na stejném místě, se stejným vybavením a za stejných podmínek. Pro odstranění vlivu obsahu promluvy byl pak využit standardní text „Jak v Kocourkově chytali zajíce“, který je v příloze B. Tento text byl použit zejména z důvodu, že stejný text využívá i dříve zmíněná aplikace pro vyšetření hlasového pole [42]. Dále bylo po odborných konzultacích rozhodnuto, že standardní text bude strukturalizován do 3 částí. Důvodů je hned několik. Patří mezi ně snadnější detekce pauzy mezi jednotlivými částmi, což může přispět k časovému zarovnání zvukových stop či oddělené analýze pouze části projevu. Rozdělení má i psychologický efekt na dotyčného pacienta. Určitě lépe se mu budou číst kratší úseky, mezi kterými bude mít dostatek času na nadechnutí a plynulé čtení další části, než kdyby dostal pacient blok textu. Navíc je zde předpoklad, že pacient může být stresován tím, jak celé vyšetření dopadne, na základě toho, jak dobře text přečte. Proto je snaha umělým vytvořením pauzy nutit pacienta, aby se soustředil pouze na čtení.

Jako místo byla vybrána tichá komora FNKV. Jelikož nahrávací zařízení splňuje podmínky a doporučení stanovené v [99] a [48], je díky všem těmto opatřením zcela vyloučena variabilita mikrofону i obsahu promluvy. Vliv prostředí je rovněž minimalizován a analýza tak může být soustředěna pouze na řečové parametry. Variabilita, která však přirozeně zůstává, je mezi jednotlivými řečníky. Dále je projev ovlivněn právě zdravotním stavem pacienta a samozřejmě i emočním rozpoložením.

Byly zvažovány různé varianty s ohledem na dostupnost pacientů, možnost spojit nahrávání s běžnými prohlídkami a určení počtu měření jako takových. Po konzultaci s odborníky v oblasti ORL byly stanoveny termíny nahrávání s dělením do 3 fází:

- fáze 0 – pacient přichází na kliniku se svými obtížemi, zatím nepodstoupil žádnou léčbu,
- fáze 1 – záznam se provádí 2-3 dny po operaci, hned jak je to možné,
- fáze 2 – poslední záznam přichází na řady v průběhu běžné kontroly (typicky 2 týdny po zákroku).

K třífázovému dělení bylo přistoupeno ze 3 základních důvodů. Je zde možnost spojit nahrávání s běžnými návštěvami u lékaře, prohlídkami po operacích, při vytahování stehů apod. Díky 3 fázím je možné sledovat vývoj hlasu před invazivním zákrokem, bezprostředně po něm a po dalším čase na rekonvalescenci. Nabízí se tedy možnost použít první nahrávku pro diagnostické účely a nahrávku z první a druhé fáze pro sledování

průběhu léčení pacienta, přičemž pokud při zákroku nastanou komplikace, bude k dispozici srovnání. Poslední důvod je čistě praktický. Tři nahrávací intervaly jsou hraniční pro realizovatelnost. Při více fázích by bylo obtížné pacienty přesvědčit, aby pokračovali ve výzkumu např. 6 měsíců poté, kdy už budou zcela uzdraveni.

Jako příklad může být uvedena modelová situace, kdy pacient přišel na vyšetření s hlasovými obtížemi (fáze 0). V rámci běžného vyšetření ho bude čekat i záznam hlasu. Celý proces pro nahrávání hlasu pacienta krok po kroku je následující:

- informace o pacientovi jsou zaneseny do systému,
- po vytvoření profilu pacienta se provede dvouступňová kalibrace, přičemž kalibrační signály zůstávají uloženy pro další zpracování,
- po kalibraci následuje čtení standardního textu „Jak v Kocourkově chytali zajíce“,
- celý záznam je společně s kalibračními signály uložen pro další analýzu.

Celý tento cyklus s rozdílem pouze v prvním bodě, kdy se informace o stávajícím pacientovi již převezmou z databáze, se poté znovu zopakuje při pacientově druhé a třetí návštěvě (fáze 1 a 2). Aby mohl být celý proces reprodukovatelný a použitelný v širším měřítku (např. i v jiných nemocnicích), všechny potřebné nástroje byly zapracovány do diagnostické aplikace.

8 Aplikace Voice disorder diagnostician

Jedním z cílů práce je vytvoření SW, který bude použitelný v rámci ORL vyšetření a který bude poskytovat lékařům další informace nejen ohledně diagnostiky. Užitečné může být i zobrazení výsledků v pozdějších fázích (průběh léčby). Hlavními požadavky pro tento SW je samozřejmě uživatelsky přívětivé grafické prostředí, jednoduché ovládání a minimum znalostí potřebných pro správnou činnost.

Vznikla tak aplikace s názvem Voice Disorder Diagnostician (VDD), která všechny tyto předpoklady splňuje. Aplikace byla vytvořena v prostředí Matlab a je dostupná ve formě samostatného instalátoru. Instalace se provádí běžným způsobem. Po spuštění instalátoru uživatel zvolí cestu k adresáři, do něž má být SW nainstalován. Ačkoliv je program vytvořen v prostředí Matlab, jedná se o tzv. standalone (samostatnou) aplikaci. To znamená, že jádro programu Matlab je již integrováno v instalátoru a tento SW ani licence k němu již nemusí být na cílovém zařízení.

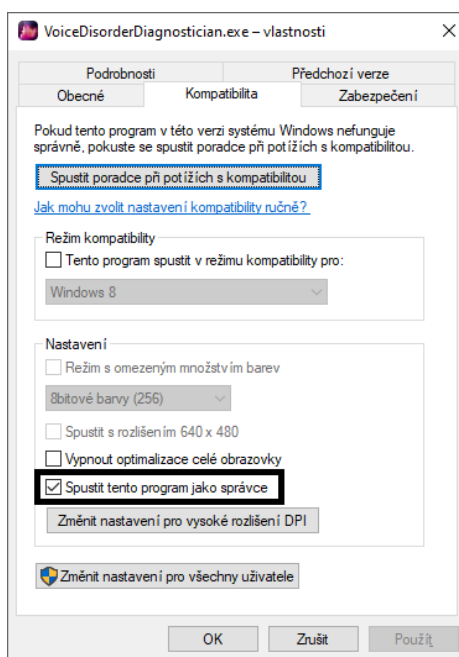
Při prvním spuštění se vytvoří předdefinovaná adresářová struktura jako na obrázku 8.1. Protože každému měření předchází kalibrace ukládají se i tato data do složky

kalibracni_data. Pro uchování veškerých pořízených záznamů ze všech fází pacientova nahrávání slouží adresář *nahravky*. Databáze pacientů se ukládá do složky *udaje_o_pacientech* a jako dočasné úložiště pro odstraněné soubory pak slouží poslední adresář s názvem *_deleted*.

```
C:\Program Files\Voice Disorder Diagnostician\application>tree /F
Folder PATH listing for volume SYSTEM
C:.
  icon.ico
  Launch.bat
  Manuál.pdf
  splash.png
  VoiceDisorderDiagnostician.exe
  kalibracni_data
  nahravky
  udaje_o_pacientech
  _deleted
```

Obrázek 8.1 – Adresářová struktura aplikace zobrazena přes příkazovou řádku

VDD lze spustit pouze v prostředí operačního systému Windows. Mac a Unixové systémy nejsou podporovány. Pro správnou funkci aplikace je nezbytné programu VDD přidělit práva správce, aby bylo možné zapisovat data na disk. To je možné po zobrazení vlastností spouštěcího souboru aplikace *VoiceDisorderDiagnostician.exe*. Náhled nastavení je na obrázku 8.2. Přesto je vhodné aplikaci spouštět přes soubor *Launch.bat* z důvodu kompatibility. Aplikace byla testována v prostředí Windows 10, které je pro používání doporučeno.

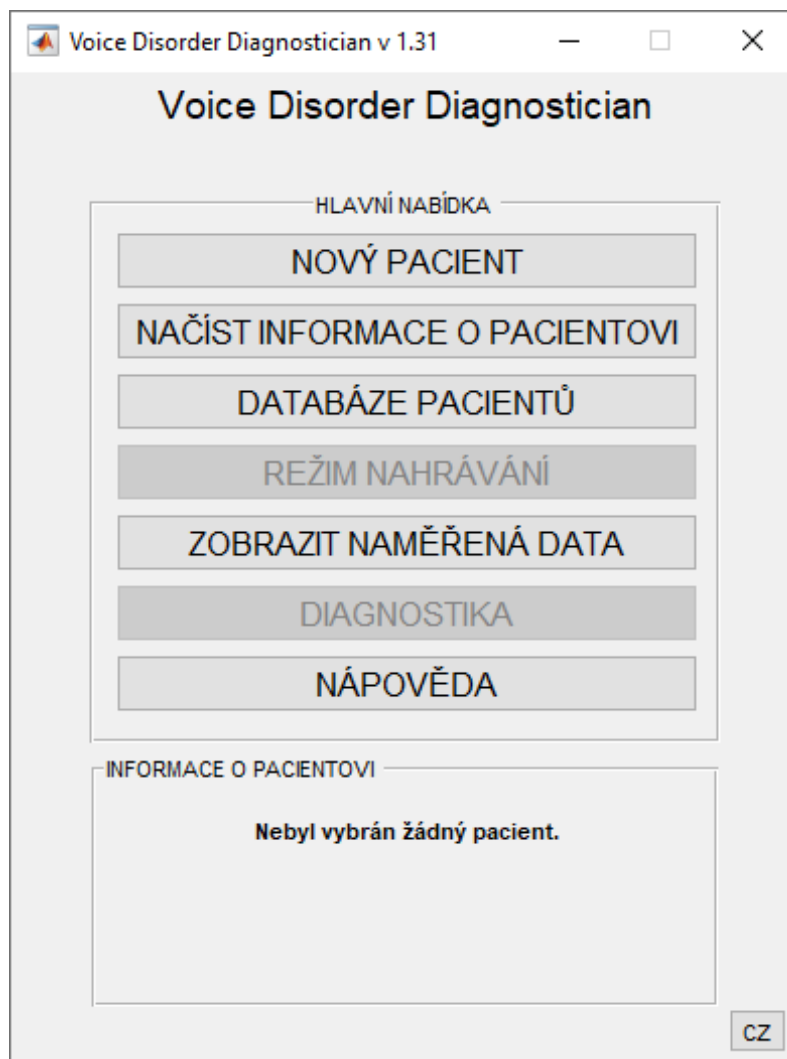


Obrázek 8.2 – Nastavení práv správce pro VDD je vyžadováno pro správnou funkčnost

Systémové požadavky VDD jsou shodné jako pro nástroj Matlab 2018a, což je zmíněno i během instalace. Nyní přichází na řadu popis samotné aplikace včetně všech jejích režimů a obrazovek.

8.1 Režimy pro sběr a analýzu dat

Při spuštění aplikace Voice Disorder Diagnostician se uživateli zobrazí hlavní obrazovka (menu), jejíž ukázka je na obrázku 8.3.



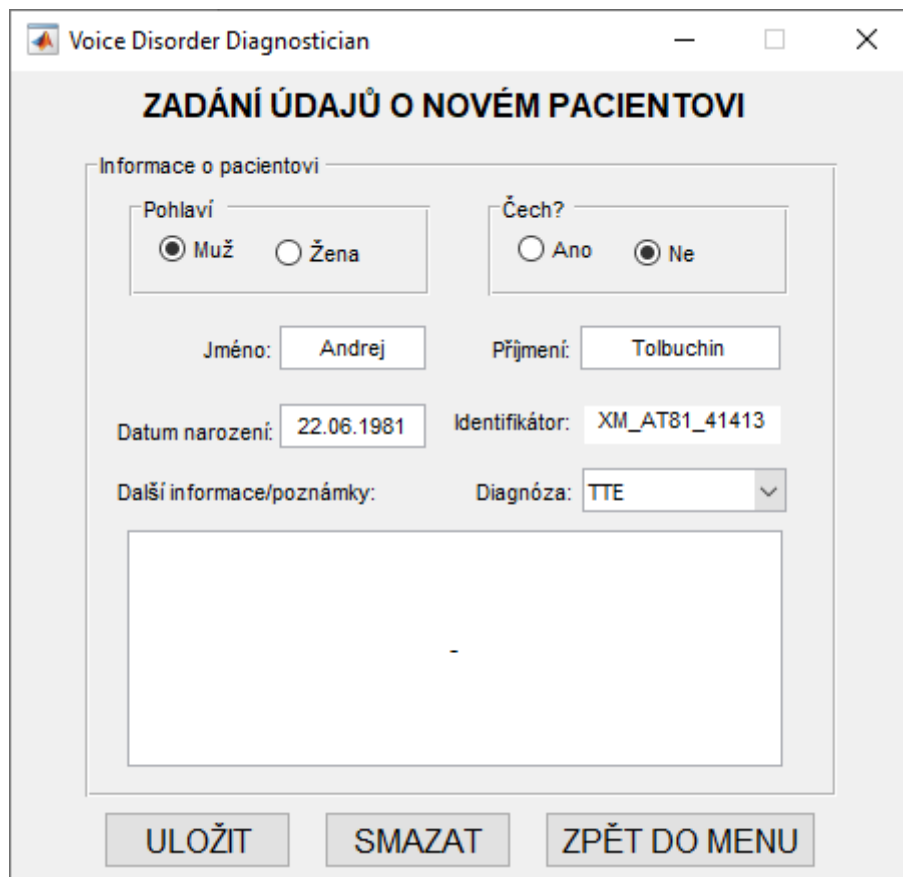
Obrázek 8.3 – Hlavní nabídka aplikace Voice Disorder Diagnostician

Ta je rozčleněna do dvou základních částí: hlavní nabídka a informace o pacientovi. V hlavní nabídce může uživatel zvolit, ve kterém režimu bude chtít s aplikací pracovat. Aplikace umožňuje zaevidovat nového pacienta, načíst informace o již uloženém pacientovi, prohlédnout databázi pacientů, pořídit nahrávky, zobrazit již naměřená data, spustit diagnostiku nebo zobrazit nápovědu. Kromě těchto režimů, je také možné přepínat jazyk celé aplikace tlačítkem v pravém dolním rohu. VDD aktuálně podporuje češtinu a

angličtinu. Verze aplikace je uvedena v titulku okna. Pokud byl vybrán nějaký pacient a aplikace načetla údaje o tomto pacientovi z databáze, pak se náhled zobrazí v sekci informace o pacientovi. Ta je při spuštění samozřejmě prázdná. Načtení údajů je nezbytné pro přístup do režimů nahrávání a diagnostiky, které jsou jinak nedostupné. Logicky nyní následuje popis jednotlivých režimů.

8.1.1 Zadávání údajů o novém pacientovi

Přichází-li pacient na vyšetření poprvé, je třeba nejdříve zadat některé identifikační údaje do systému. K tomu slouží hned první volba z hlavní nabídky. Otevře se okno pro zadání údajů o novém pacientovi, kde doktor vyplní příslušná pole. Náhled této obrazovky se zadáním fiktivního pacienta je na obrázku 8.4.



The screenshot shows a window titled "Voice Disorder Diagnostician" with a subtitle "ZADÁNÍ ÚDAJŮ O NOVÉM PACIENTOVI". The form is titled "Informace o pacientovi" and contains the following fields and controls:

- Pohlaví:** Radio buttons for "Muž" (selected) and "Žena".
- Čech?:** Radio buttons for "Ano" and "Ne" (selected).
- Jméno:** Text input field containing "Andrej".
- Příjmení:** Text input field containing "Tolbuchin".
- Datum narození:** Text input field containing "22.06.1981".
- Identifikátor:** Text input field containing "XM_AT81_41413".
- Další informace/poznámky:** A large empty text area.
- Diagnóza:** A dropdown menu with "TTE" selected.

At the bottom of the form are three buttons: "ULOŽIT", "SMAZAT", and "ZPĚT DO MENU".

Obrázek 8.4 – Hlavní nabídka aplikace Voice Disorder Diagnostician

Povinná jsou textová pole pro jméno, příjmení a datum narození. Dále lékař uvede, jestli se jedná o muže či ženu a zda je pacient rodilý mluvčí či nikoliv. Uvedení skutečnosti, zda se jedná o rodilého mluvčího je klíčové z důvodu zjišťování rychlosti řeči. Pokud se nejedná o rodilého mluvčího, velmi pravděpodobně nebude pacient správně vyslovovat některá slova standardního textu a běžné algoritmy ASR pak nebudou schopny správně

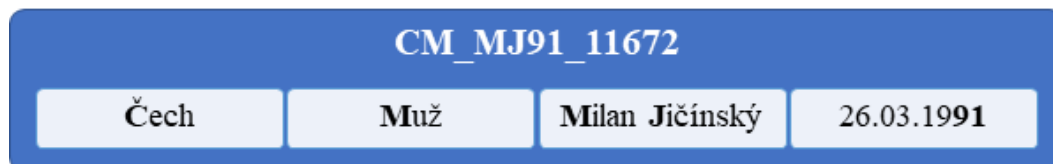
detekovat tato slova a měřit tak rychlost řeči. Dále je ještě nutné z rolovacího menu vybrat diagnózu, která je stanovena na základě předchozího běžného vyšetření. Vyplnění dalších poznámek či upřesnění diagnózy je volitelné a slouží spíše informativně pro lékaře.

Na základě všech zadaných údajů vygeneruje algoritmus unikátní identifikační kód pacienta, který se zobrazí v needitovatelném poli s názvem identifikátor. Tento identifikační kód je pak aplikací využíván ve všech režimech. Jedná se o třináctimístný kód ve tvaru MP_IIRR_DUUUU, kde M je označení pro rodilého mluvčího a může nabývat hodnot Č – čech nebo X – není rodilý mluvčí. P představuje pohlaví a je zastoupeno buď písmenem M – muž nebo Z – žena. II jsou iniciály daného pacienta v pořadí jméno a příjmení. RR jsou zástupné znaky pro poslední dvě číslice roku narození. D představuje diagnózu zvolenou z rozbalovacího menu a dané číslo odpovídá jejímu pořadí v nabídce. Čtveřice U na konci je pak unikátní kód vygenerovaný algoritmem na základě dříve zadaných údajů. Výpočet unikátního kódu probíhá dle vztahu

$$unikatniKod = \sum_n j(n) + \sum_k p(k) - r + d - 10m, \quad (8.1)$$

- kde:
- j představuje křestní jméno, jehož písmena jsou převedena na čísla dle ASCII,
 - n je počet písmen křestního jména,
 - p označuje příjmení, jehož písmena jsou převedena na čísla dle ASCII,
 - k vyjadřuje počet písmen příjmení,
 - r symbolizuje rok, kdy se pacient narodil,
 - d značí pořadové číslo dnu v měsíci, kdy se pacient narodil,
 - m je pořadové číslo měsíce narození.

Přidání unikátní části identifikátoru bylo zařazeno zejména z důvodu, aby nedocházelo ke shodě identifikátorů jiných pacientů. Přesto, kdyby se tak někdy stalo, bude identifikátor rozšířen ještě o datum, kdy byl pacient zaevidován do databáze. To znamená, že shoda identifikátorů je vyloučena. Jako příklad může posloužit obrázek 8.5.



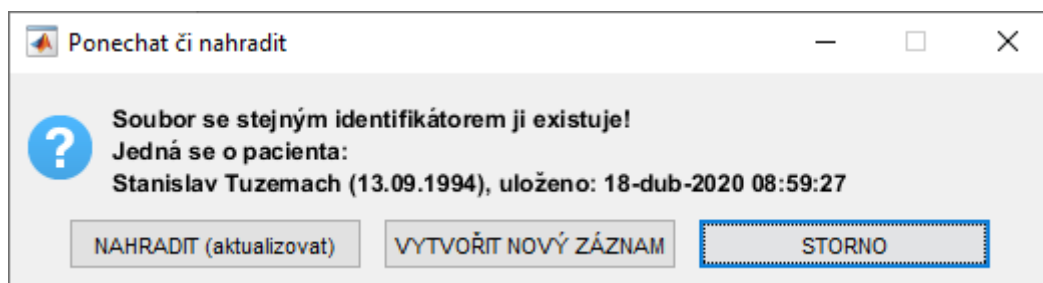
Obrázek 8.5 – Ukázka vygenerovaného identifikátoru na základě údajů o pacientovi

Identifikátor je uveden nahoře a je vygenerován na základě údajů o autorovi práce, které jsou uvedeny níže. Jednotlivé části údajů, které se přímo promítnou do výsledného

identifikátoru, jsou pro přehlednost uvedeny tučně. Unikátní kód tvoří poslední čtyři znaky identifikátoru.

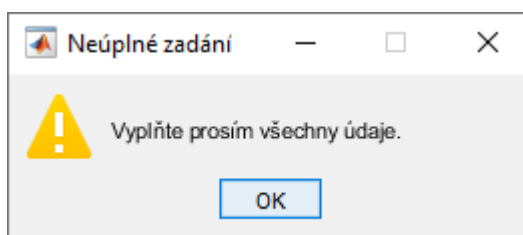
Po vyplnění všech potřebných údajů již nic nebrání tomu, aby mohl být pacient přidán do databáze pacientů. To se provede stiskem tlačítka uložit. Proběhne-li uložení v pořádku tlačítko zezelená a text se změní na „uloženo“.

Pokud by při pokusu o ukládání byla nalezena shoda s již evidovaným pacientem (identifikátor by již existoval), je lékař vyzván, aby rozhodl o dalším postupu. V takovém případě se zobrazí dialogové okno jako je na obrázku 8.6. Aby se mohl uživatel relevantně rozhodnout ve chvíli, kdy je na shodu upozorněn, jsou mu předloženy základní údaje, o již uloženém pacientovi – tj. jméno, příjmení, datum narození a datum, kdy byl pacient do systému přidán. V případě, že se jednalo pouze o aktualizaci dat již známého pacienta (např. změna nepovinného komentáře), zvolí lékař možnost nahradit (aktualizovat). Pokud se jedná o jiného pacienta, přesto však došlo ke shodě identifikátorů, pak je třeba zvolit možnost vytvořit nový záznam, čímž se identifikátor rozšíří o aktuální datum. Tedy např. do tohoto formátu CM_MJ91_11672_17-01-20. Tím zůstane zachován původní záznam a může být přidán o nový pacient. Stiskem tlačítka storno se naopak uživatel vrátí zpět a může upravit zadané údaje.



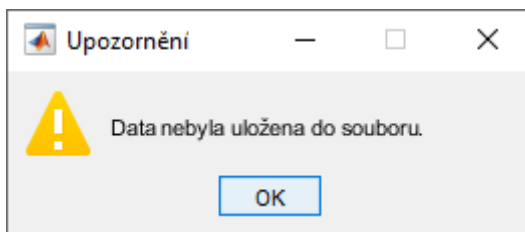
Obrázek 8.6 – Dialogové okno zobrazované při shodě identifikátoru

Údaje jsou ukládány do souboru s příponou vdd do adresáře *udaje_o_pacientech*. Jako název slouží identifikátor pacienta. Pokud by při ukládání nebyly uvedeny všechny vyžadované údaje, každý pokus o uložení pak automaticky skončí hlášením na obrázku 8.7.



Obrázek 8.7 – Upozornění na chybějící informace při vyplňování údajů

Tlačítko smazat pak slouží k vymazání údajů ze všech zadaných polí a své využití nalezne například při zadávání více pacientů. Uživatel pak nemusí ručně mazat nebo přepisovat jednotlivá pole. Tlačítkem „zpět do menu“ se uživatel může kdykoliv vrátit zpět na hlavní obrazovku. Pokud však před návratem do hlavní nabídky nebyly údaje uloženy, je na to uživatel upozorněn dialogovým oknem na obrázku 8.8.



Obrázek 8.8 – Upozornění, že údaje o pacientovi nebyly uloženy do databáze

Jestliže uložení údajů proběhlo úspěšně, daný pacient se stává aktivním, což znamená, že se při návratu do hlavní obrazovky:

- v sekci informace o pacientovi zobrazí výpis uložených údajů,
- zpřístupní se režim nahrávání a kalibrace,
- rovněž se zpřístupní režim diagnostiky.

Aktivním může být v danou chvíli vždy právě jeden pacient, u kterého je možné pořizovat nahrávky či jej diagnostikovat. Aktivním pacientem je vždy poslední pacient, který byl v tomto režimu uložen. Změnu aktivního pacienta je možné provést rovněž v následujícím režimu.

8.1.2 Import dat z databáze pacientů

Tento režim slouží převážně k tomu, aby bylo možné měnit aktivního pacienta, který je zobrazován na hlavní obrazovce. Rozložení okna je velmi podobné, jak je tomu při vytváření záznamu o novém pacientovi. Rozdíl je v tom, že místo poznámek je zde seznam uložených záznamů a na pravé straně je náhled do jednotlivých souborů. Grafické rozvržení okna je znázorněno na obrázku 8.9.

Režim načítání údajů tak slouží primárně k tomu, aby bylo možné přehledně vyhledávat a filtrovat ve velkém počtu záznamů. Pro filtraci výsledků slouží volba pohlaví a rodilého mluvčího. V sekci nalezené záznamy se pak zobrazují jen ty soubory, které odpovídají volbě těchto dvou kritérií. Pokud je i tak záznamů mnoho, je možné ještě vyhledávat konkrétního pacienta, a to zadáním jeho jména, příjmení a data narození. Aby bylo možné ověřit, že je skutečně vybrán správný subjekt (v seznamu jsou pacienti viditelní pouze pomocí svých identifikátorů), je po zvolení souboru ze seznamu na levé straně

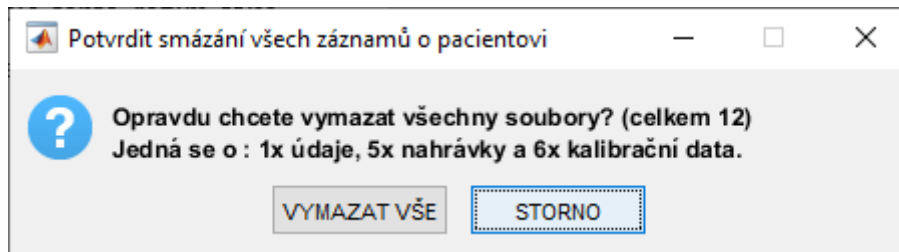
zobrazen náhled na straně pravé. Poté stačí potvrdit výběr tlačítkem načíst. To následně zezelená a nápis se změní na „načteno“. Daný záznam se pak stane aktivním pacientem. Pak už se stačí jen vrátit do hlavní nabídky tlačítkem zpět do menu a pokračovat například nahráváním. Při chybném zadání lze použít reset. Při stisku se smaže obsah všech editovatelných polí.

Obrázek 8.9 – Režim načítání údajů o pacientech v aplikaci VDD

Jelikož při používání mohou být vytvořeny i záznamy, které jsou z nějakého důvodu nerelevantní (ať už špatné zadání, vyřazení pacienta apod.), umožňuje tento režim také záznamy odstraňovat. Opět stačí vybrat záznam podle identifikátoru a stisknutím tlačítka odstranit se vyvolá dialogové okno, kde je uživatel vyzván k potvrzení smazání všech souborů, které s daným pacientem souvisí. Tím je myšlen záznam o pacientovi, všechny jeho nahrávky i data pořízená v rámci kalibrace. Ukázka potvrzení smazání všech souborů je na obrázku 8.10.

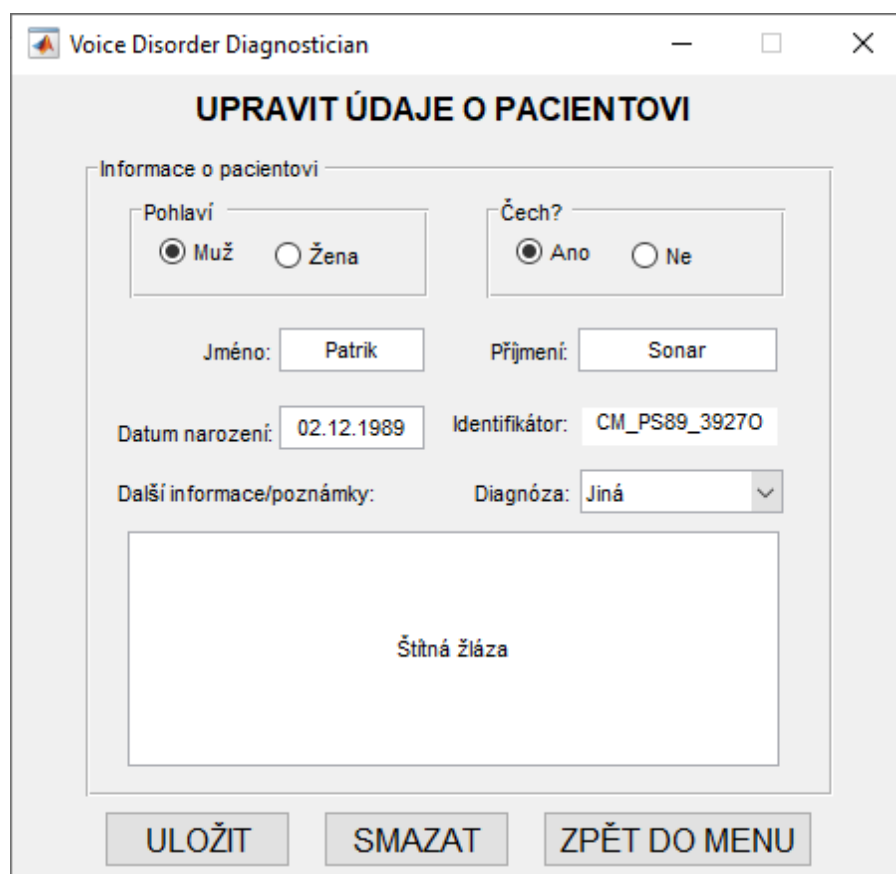
Po odstranění záznamu tlačítko také změní barvu na zelenou. Společně s tím se přepíše nápis na „odstraněno“. V případě, že by byl odstraněn aktivní pacient, pak žádný pacient po návratu do hlavní nabídky aktivní není. Ačkoliv je řeč o odstraňování, ve skutečnosti nejsou soubory mazány, ale jsou pouze přesunuty do adresáře *_deleted*.

Soubory v tomto adresáři již nejsou v aplikaci VDD viditelné. K tomuto způsobu bylo přistoupeno zejména proto, že pokud by došlo k nechtěnému smazání, je možné soubory ještě ručně obnovit zkopírováním do příslušných adresářů. Pokud nebudou do 30 dnů soubory ručně obnoveny, budou automaticky smazány z disku (tentokrát již definitivně).



Obrázek 8.10 – Dialogové okno zobrazované při shodě identifikátoru

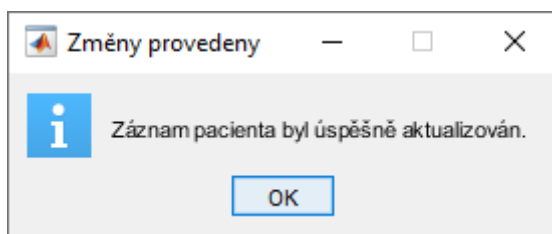
Kromě mazání je zde ještě možnost změnit již zadané údaje u pacienta, který je vybrán v seznamu vlevo. K tomu slouží tlačítko upravit. Po jeho stisknutí se otevře okno pro zadávání nového pacienta v režimu upravování, který je na obrázku 8.11.



Obrázek 8.11 – Obrazovka upravování záznamů pacienta hned po stisku tlačítka upravit

To znamená, že všechna pole budou s předvyplněnými údaji tak, jak byly informace uloženy. Úprava může být kdykoliv zrušena návratem do menu. V tomto režimu lze změnit všechny údaje. Protože však jakákoliv změna (kromě doplnění komentáře) vede

k pozměnění identifikátoru pacienta, budou touto změnou ovlivněny všechny soubory daného pacienta. Jinak řečeno, všechny relevantní souboru budou přejmenovány, aby odpovídali novému identifikátoru a byly tak dohledatelné. Jakmile se změny uloží, uživatel je automaticky navrácen do hlavního menu. Jako zpětná vazba slouží hlášení na obrázku 8.12.



Obrázek 8.12 – Potvrzení zobrazené při automatickém návratu do hlavního menu

8.1.3 Databáze pacientů

Vytvoření nového záznamu, úpravy a odstraňování byly popsány v předchozích režimech. Režim databáze pacientů byl přidán na přání FNKV zejména kvůli přehlednosti a možnosti spravovat získaná data i mimo aplikaci. Neumožňuje žádné úpravy, slouží pouze k zobrazení všech evidovaných pacientů ve formě tabulky. Lékař tak nemusí dělat dvojí evidenci a ručně přepisovat do tabulky, kteří pacienti byli již evidováni. Navíc je zde vidět, kdo již má v systému nahrávku svého hlasu. Ukázka databáze, kde jsou samozřejmě použity smyšlené údaje, je na obrázku 8.13.

DATABÁZE PACIENTŮ											
Filtrování: Pohlaví: - Rodilý mluvčí: -											*PN - Počet nahrávek
	Identifikátor	Příjmení	Jméno	M/Ž	Datum narození	Stát	Diagnóza	Poznámky	PN		
1	XZ_LB54_31432	Becherová	Laura	Z	1954.07.16	mimo ČR	Jiná	Chrapot	-	0	
2	XM_PB79_11212	Blackhead	Peter	M	1979.11.10	mimo ČR	Zdravý subjekt	-	-	0	
3	CM_MD80_11450	Duhový	Miroslav	M	1980.10.14	ČR	Zdravý subjekt	-	-	0	
4	XM_JD86_11598	Džusík	Jiří	M	1986.12.17	mimo ČR	Zdravý subjekt	-	-	1	
5	CZ_ŽF95_41739	Faustová	Žaneta	Z	1995.05.13	ČR	TTE	-	-	0	
6	CZ_MJ84_31631	Jalovicová	Monika	Z	1984.08.22	ČR	HTE	-	-	0	
7	XM_JJ91_11806	Jägermeister	Johny	M	1991.03.26	mimo ČR	Zdravý subjekt	Zcela zdravý	-	5	
8	CM_RK90_3832U	Kufr	Radim	M	1990.01.31	ČR	Jiná	Chrapot	-	0	
9	CM_RL67_12012	Louškáček	Radovan	M	1967.11.08	ČR	Zdravý subjekt	-	-	0	
10	CM_MM86_31312	Maďar	Miroslav	M	1986.12.17	ČR	HTE	-	-	1	
11	XM_PM77_11182	Messi	Patricio	M	1977.10.19	mimo ČR	Zdravý subjekt	-	-	0	
12	CM_LM55_11451	Mravůrek	Libor	M	1955.12.27	ČR	Zdravý subjekt	-	-	0	
13	XM_MR71_41335	Rockonut	Martin	M	1971.07.04	mimo ČR	TTE	-	-	9	
14	CM_KR88_11062	Rvačka	Kuba	M	1988.03.23	ČR	Zdravý subjekt	-	-	0	
15	CM_PS89_49270	Sonar	Patrik	M	1989.12.02	ČR	TTE	-	-	0	
16	CM_MS79_41434	Sousedů	Martin	M	1979.11.10	ČR	TTE	-	-	0	
17	CM_PS96_31364	Studený	Patrik	M	1996.04.01	ČR	Jiná	Chrapot	-	0	
18	XM_AT68_31473	Toibuchin	Andrej	M	1986.02.27	mimo ČR	Jiná	Chrapot	-	0	
19	CM_ST94_41611	Tuzemach	Stanislav	M	1994.09.13	ČR	TTE	-	-	0	
20	CZ_BV93_32019	Valášová	Berenika	Z	1993.02.08	ČR	Jiná	Chrapot	-	2	
21	CZ_HV94_3980A	Vanová	Hana	Z	1994.06.11	ČR	HTE	-	-	0	
22	XM_PV81_1895Í	Vitr	Peter	M	1981.09.01	mimo ČR	Zdravý subjekt	-	-	0	
23	CM_JR81_31635	Řejzek	Jonáš	M	1991.04.12	ČR	HTE	-	-	0	
24	CM_JŠ92_11121	Štova	Jan	M	1992.12.17	ČR	Zdravý subjekt	-	-	0	

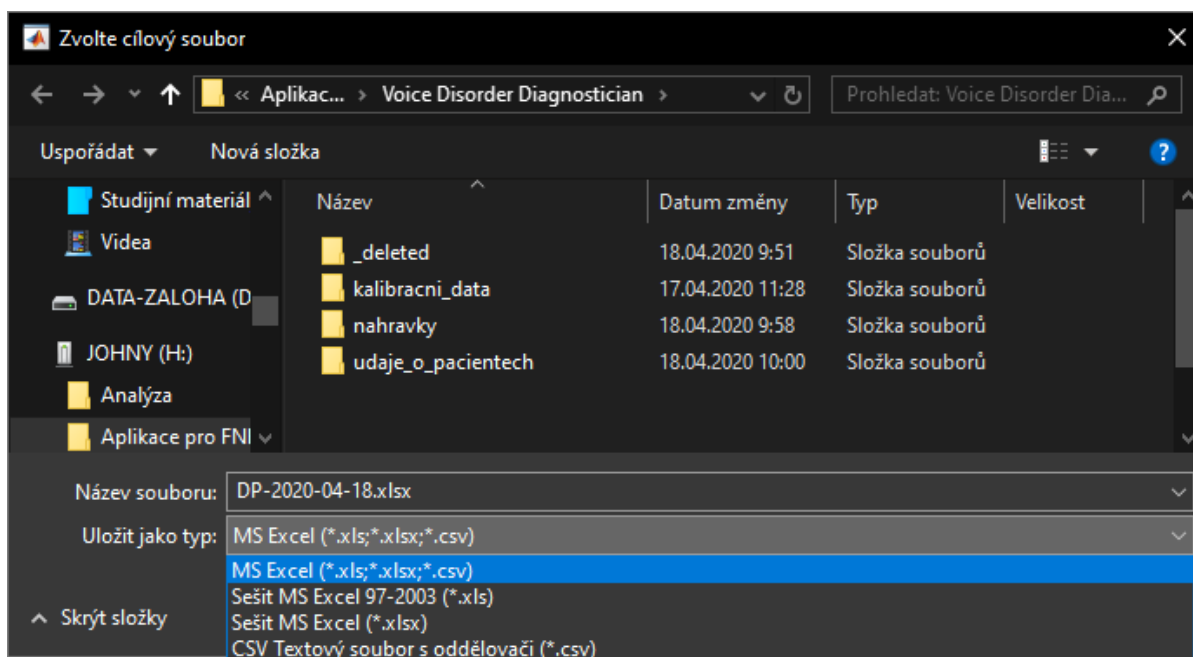
Tipy: Je možné myší označit pouze část tabulky a exportovat pouze tuto část. Chcete-li buňky odznačit, provedete tak stiskem CTRL+levé tlačítko myši.

EXPORTOVAT TABULKU ZPĚT DO MENU

Obrázek 8.13 – Ukázka obrazovky databáze pacientů s řazením příjmení podle abecedy

U jednotlivých sloupců jsou k dispozici tlačítka umožňující řazení tabulky vzestupně či sestupně s ohledem na daný sloupec, což je vidět i na náhledu, kde je databáze řazena podle abecedy s ohledem na sloupec příjmení. Kromě toho je zde možnost filtrování. Filtry se vybírají vlevo nad tabulkou a umožňují filtrovat dané pohlaví či rodilé mluvčí. Kombinací filtrů a řazení lze tak docílit požadovaných výsledků.

Jakmile je tabulka řazena v požadovaném pořadí, je možnost exportovat ji do souboru stiskem tlačítka exportovat tabulku. Otevře se okno pro ukládání jako na obrázku 8.14. Pro export tabulky jsou podporovány pouze formáty xls,xlsx a csv tabulkového editoru Microsoft Excel. Kromě exportu celé tabulky nebo její vyfiltrované části, je možné vybrat pouze její část. Stisknutím levého tlačítka myši a tažením lze vybrat oblast několika buněk. Tlačítka poté změni podobu na exportovat výběr. Po stisku se opět objeví okno z obrázku 8.14, ale exportována bude pouze vybraná část (exportují se vždy celé řádky). Odznačení již označených buněk lze provádět stiskem klávesy CTRL a levého tlačítka myši na příslušných buňkách, což je uvedeno i v nápovědě v levém dolním rohu databázového režimu.

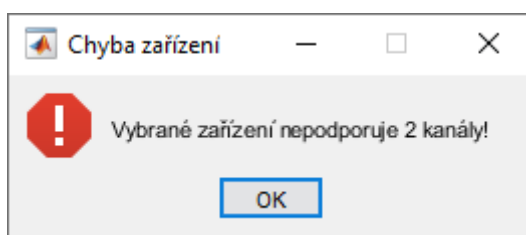


Obrázek 8.14 – Export databáze do souboru a podporované typy souborů

8.1.4 Režim nahrávání

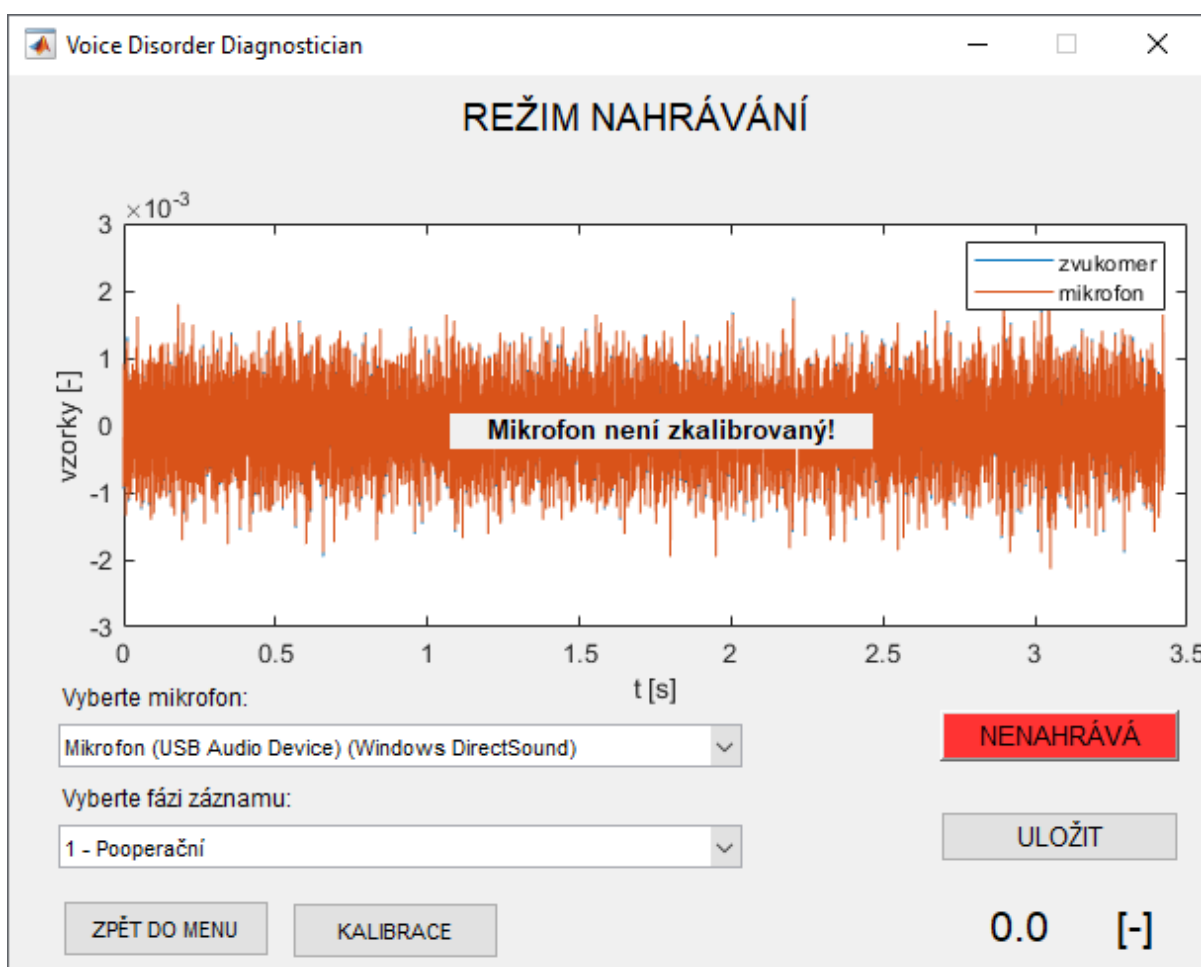
Je-li zvolen aktivní pacient, zpřístupní se v hlavní nabídce režim nahrávání. Ten slouží k tomu, aby se pro daného pacienta pořídily zvukové nahrávky, které pak poslouží jako zdroj informací pro další zpracování a diagnostiku. Při spuštění se otevře nové okno, kde je nejprve nutné zvolit záznamové zařízení, které podporuje 2 kanály. V opačném

případě je uživatel upozorněn na nevhodně zvolené zařízení, stejně jako na obrázku 8.15.



Obrázek 8.15 – Hláška upozorňující na zvolení nevhodného zvukového zařízení

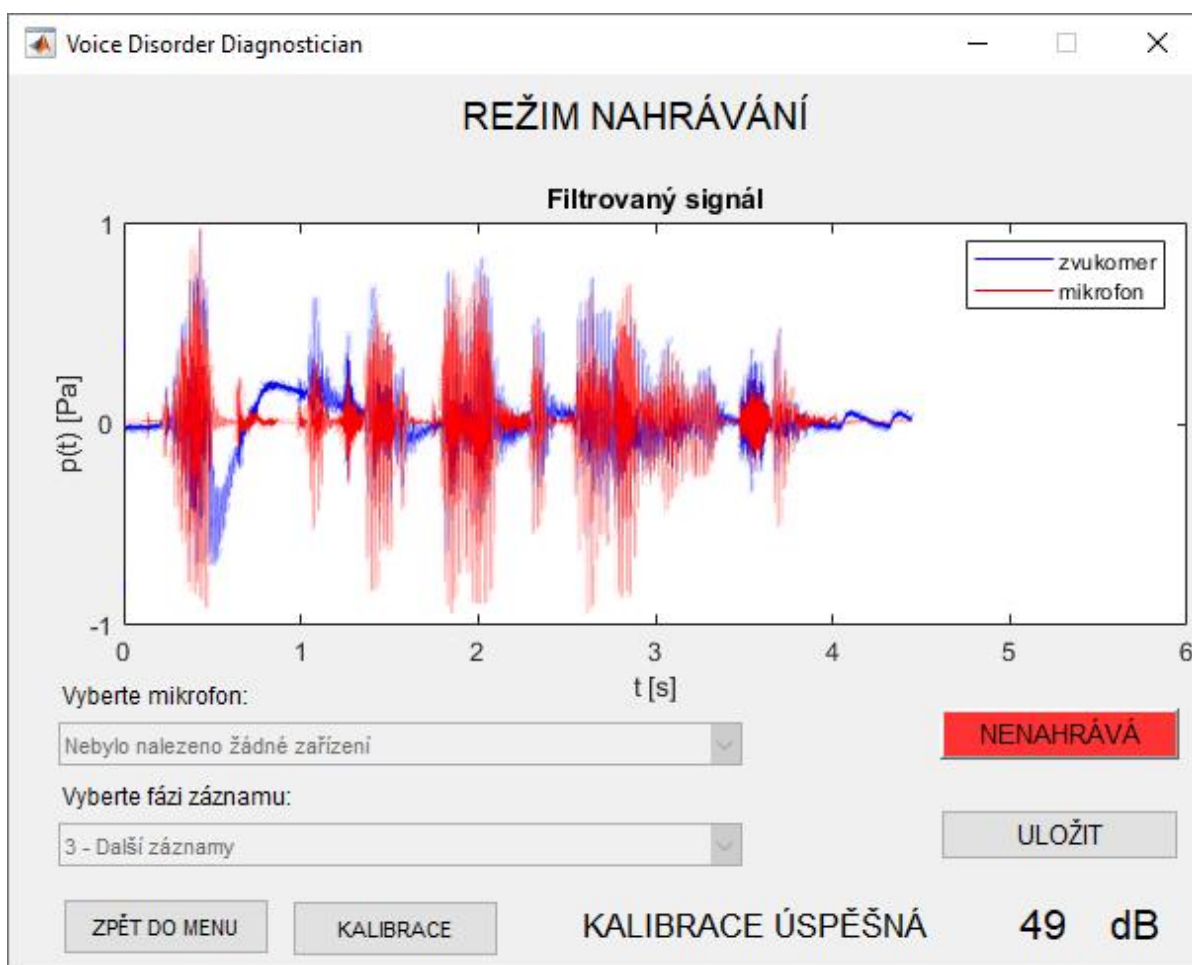
S danými komponentami měřicího řetězce by nahrávacím zařízením byla v tomto případě zvuková karta Focusrite. Dále je nutné vybrat z rolovacího menu fázi nahrávání pacienta, jak bylo zmíněno již dříve. Ve zkratce lze tedy označit fázi 0 jako předoperační, fázi 1 jako pooperační (2-3 dny po zákroku) a fázi 2 jako nahrávání s odstupem (typicky 2 týdny po zákroku). Aplikace nabízí ještě fázi 3 (další záznamy), kterou je možné využít při dalším nahrávání, pokud by to bylo třeba. Po volbě obou povinných údajů se v grafu zobrazí průběh signálu v reálném čase. Uživatel je však zároveň upozorněn na skutečnost, že zařízení, které zvolil není kalibrované. Tato fáze je viditelná rovněž na obrázku 8.16.



Obrázek 8.16 – Režim nahrávání po volbě nahrávacího zařízení (před kalibrací)

V pravém dolním rohu se zobrazují aktuálně naměřené hodnoty SPL v dB. Před kalibrací však jednotlivé vzorky signálu nenesou informaci o akustickém tlaku, a proto je ukazatel zatím vynulován. Tlačítko nahrávání a ukládání hraje roli až po samotném kalibrování. Než bude tedy uživatel moci pokračovat, je nejprve třeba stisknout tlačítko kalibrace, čímž se otevře nové okno stejnojmenného režimu.

Po kalibrování zařízení již z grafu zmizí informační hlášení o neprovedené kalibraci. Dále je možné si všimnout toho, že místo vzorků je již díky kalibračním konstantám zakreslován průběh akustického tlaku. Jednoduchým vizuálním vodítkem a potvrzením, že kalibrace proběhla úspěšně, je vzájemné překrývání signálů z obou kanálů (zvukoměru a mikrofonu). V pravém dolním rohu se nyní již zobrazují hodnoty SPL. Nahrávací zařízení ani fáze pacienta již nemohou být po kalibraci změněny. Proto bylo nutné, vybrat je ještě před kalibrací. Tlačítko pro nahrávání slouží zároveň i jako indikátor, který nabývá dvou stavů. Je-li nestisknuté, svítí červeně s titulkem nenahrává. Pokud je stisknuté, barva je zelená a nápis nahrává upozorňuje uživatele na zaznamenávání zvukového signálu. Náhled tohoto režimu po kalibraci je na obrázku 8.17.

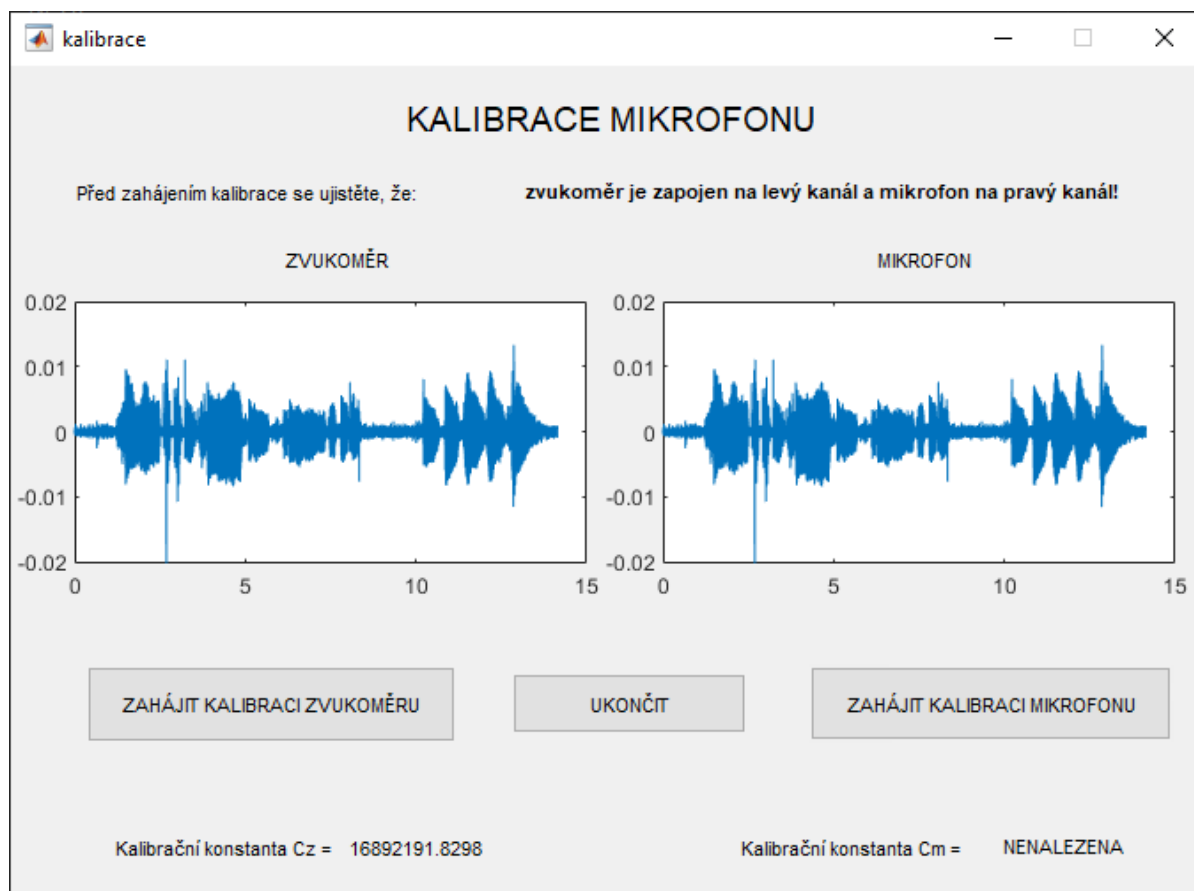


Obrázek 8.17 – Režim nahrávání po úspěšné kalibraci

Nyní na řadu přichází samotné nahrávání. Lékař vyzve pacienta, aby se připravil. Poté stiskne nyní červené tlačítko nenahrává, které se vzápětí rozsvítí zeleně. Pacient začne číst standardní text. Po dočtení posledního slova lékař opětovným stiskem, nahrávání ukončí. Pokud je s výsledkem spokojen a nebylo nutné nahrávání z nějakého důvodu přerušovat, je možné nahrávku uchovat stiskem tlačítka uložit. Ukládání probíhá automaticky do formátu wav, přičemž se zaznamenávají oba dva kanály (zvukoměr i mikrofon). Název souboru je pak tvořen kombinací identifikátoru pacienta a fáze jeho nahrávání. V případě, že by chtěl například pacient začít znovu od začátku, podmínky měření by byly nějakým způsobem narušeny apod., stačí opětovným stiskem tlačítka nenahrává spustit nové nahrávání. Předchozí zaznamenaný signál se automaticky odstraní z paměti, pokud již nebyl uložen.

8.1.5 Kalibrace

Režim kalibrace je přístupný pouze z obrazovky nahrávání a vyžaduje tedy aktivního pacienta. Z režimu nahrávání je převzato záznamové zařízení a fáze nahrávání pacienta. Náhled kalibračního módu je vyobrazen na obrázku 8.18.



Obrázek 8.18 – Režim kalibrace umožňuje v reálném čase náhled signálů z obou kanálů

Obrazovka je rozdělena na dvě části. Dle schématu zapojení je signál ze zvukoměru (vstup 1) zobrazován vlevo a signál z mikrofonu (vstup 2) vpravo. Uživatel je upozorněn, aby se ujistil, že zařízení jsou připojena požadovaným způsobem. Prohození vstupů by totiž znamenalo nesprávnou kalibraci.

Dvoustupňová kalibrace je implementována formou dvou tlačítek. První krok je zahájení kalibrace zvukoměru (fáze 1). Na zvukoměr se přiloží kalibrátor, který generuje harmonický kalibrační tón 94 dB při frekvenci 1 kHz. Poté stiskem tlačítka zahájí kalibraci zvukoměru započne záznam harmonického signálu. Opětovným stiskem se záznam ukončí. Pro nalezení kalibrační konstanty zvukoměru c_z , postačí několikvteřinový záznam. Celý signál se uloží ve formátu datového souboru Matlab (s příponou mat). Jako název opět poslouží identifikátor pacienta a fáze nahrávání, se kterou je kalibrace spojena. Nalezená konstanta se ihned zobrazí v aplikaci. Následně může být kalibrátor sejmuto ze zvukoměru a na řadu přichází fáze 2. Ta započne stiskem tlačítka zahájí kalibraci mikrofonu. Nyní zvukoměr i mikrofon zaznamenají stejný signál, například fonace samohlásky anebo ku příkladu přečtení nadpisu standardního textu. Ukončení probíhá opětovným stiskem. Do druhého souboru se uloží záznamy z obou kanálů. Pro každé nahrávání tak existují dva kalibrační datové soubory (první a druhý stupeň kalibrace). Při následném přepočítání se zobrazí i kalibrační konstanta mikrofonu c_m . Po nalezení obou konstant se uživatel může vrátit zpět do režimu nahrávání kliknutím na tlačítko ukončit.

8.1.6 Analýza zaznamenaných dat

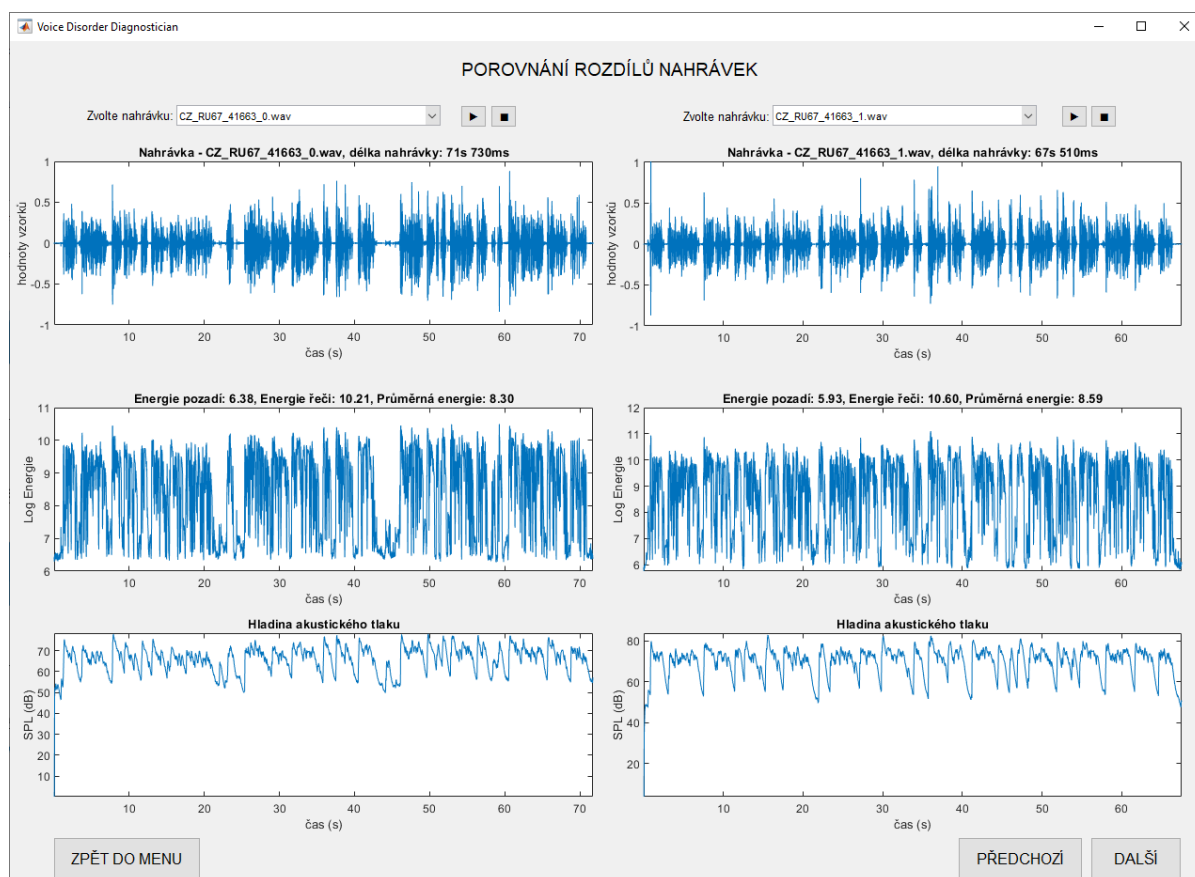
Režim analýzy zaznamenaných dat sloužil zejména při vývoji aplikace k verifikaci implementovaných algoritmů a vizualizaci obdržených příznaků. Nicméně i ze strany FNKV byl mód analýzy vyžadován, a proto zůstal i ve výsledné aplikaci. Dle slov vyšetřujících lékařů se sami pacienti ptali, jak vyšetření dopadlo a zda jim lékař může ukázat nějaké výsledky či grafy. I takto tedy režim analýzy mohl posloužit. Lékař měl možnost pacienta seznámit s tím, co všechno je možné z jeho hlasu vyčíst.

Tento režim lze spustit z hlavní nabídky tlačítkem zobrazit naměřená data, přičemž není vyžadováno, aby byla předem načtena data konkrétního pacienta. Analýza zaznamenaných dat slouží primárně k porovnávání rozdílů. Může se jednat o srovnání různých fází nahrávání téhož pacienta, například před zákrokem a po něm, ale také může být porovnáván zdravý subjekt s pacientem trpícím hlasovými obtížemi apod. Z toho důvodu je okno rozděleno na 2 části. Každá část pak představuje sadu grafů reprezentujících příznaky extrahované ze zvolené nahrávky. Každá polovina má v horní

části okna svou rozbalovací nabídku, z níž je třeba zvolit nahrávku, která má být analyzována. Není-li z nabídky vybrána žádná z nahrávek, oblast grafů zůstává prázdná.

Po volbě nahrávky se pak na pozadí spustí extrakční algoritmus a v daném sloupci se po chvíli zobrazí výsledné grafy. Každý graf disponuje základními možnostmi, jak upravovat rozsahy os. Je tedy možné nastavovat u jednotlivých grafů požadované přiblížení (možnost nastavovat separátně horizontální a vertikální), případně stiskem pravého tlačítka vyvolat nabídku, kde je možné vrátit rozsahy os do původního nastavení.

Kromě interakce s grafickými výsledky, má uživatel aplikace možnost zaznamenanou nahrávku také přehrát. Slouží k tomu sada tlačítek vedle rozbalovací nabídky, z níž se volí daná nahrávka. Přehrávání je jednoduše indikováno posuvem svislého ukazatele po časové ose grafu. Uživatel má tak lepší představu o tom, pro která slova či věty byly hodnoty určeny. Tato funkce byla do aplikace přidána na vyžádání lékařů z FNKV, neboť představuje i zpětnou vazbu o kvalitě pořízené nahrávky. Náhled obrazovky při analýze dat s první sadou grafů je k dispozici na obrázku 8.19.



Obrázek 8.19 – Režim analýzy naměřených dat a výsledky extrakčního algoritmu

Na první obrazovce v režimu analýzy naměřených dat je pro každou nahrávku k dispozici trojice grafů. První z nich představuje samotnou nahrávku – hodnoty vzorků

zaznamenaných z mikrofonu, což je běžná interpretace. Tento graf slouží spíše k lepší orientaci v nahrávce. Kromě názvu souboru je zde zobrazena i doba trvání zvoleného záznamu. Druhá v pořadí je energie zaznamenaného signálu. U grafu jsou též uvedeny hodnoty energie pozadí (ticha), řeči a průměrná energie celé nahrávky. Poslední graf pak představuje hladinu akustického tlaku (SPL).

Pro zobrazení dalších charakteristik zaznamenaného řečového signálu je třeba využít navigačních tlačítek ve spodní části obrazovky. Pro přechod z první sady grafů na druhou stačí použít tlačítko „další“. Druhou trojici tvoří charakteristiky související s frekvencí. První v pořadí je základní tón f_0 získaný extrakcí. V druhém obrázku je pak histogram základního tónu, který je pro srovnání dvou nahrávek vhodnější. Na posledním místě jsou pak zobrazeny hodnoty počtu průchodů řečového signálu nulou (ZCR). Na posledním listu jsou uživateli vyobrazeny spektrální a keprální charakteristiky. První graf zobrazuje spektrogram pořízeného záznamu. Po něm následují výsledky keprální analýzy – hodnoty třinácti keprálních koeficientů MFCC. Rovněž je zde přehled distribuce energie v jednotlivých frekvenčních pásmech.

8.1.7 Nápověda

Posledním režimem, který může být z hlavní nabídky vybrán je Nápověda. Spuštěním tohoto režimu se zobrazí uživatelský manuál, jenž slouží k popisu všech funkcí aplikace, ale rovněž se v něm uživatel dočte i jak správně propojit měřicí HW nebo ku příkladu, jak postupovat, aby byla kalibrace úspěšná. Dokument byl průběžně aktualizován během postupného získávání dat a užívání aplikace. Obsahuje tak i řadu praktických tipů, které pomáhaly lékařům z FNKV obsluhovat aplikaci. Uživatelský manuál je též k dispozici jako samostatný soubor ve formátu PDF. Celý dokument je rovněž v příloze C.

8.2 Diagnostika

Hlavní účel, jemuž má aplikace VDD sloužit, je objektivní posouzení změn hlasu po podstoupení zákroku kvůli onemocnění štítné žlázy. Aby aplikace tuto funkci měla, je nejprve důležité zjistit, které z extrahovaných řečových příznaků mění nejvíce svoji hodnotu v čase.

8.2.1 Metodika pro objektivní posouzení změn hlasu

K rozlišení reprezentativních příznaků bylo třeba stanovit vhodnou metodiku. Logickou úvahou lze dospět k závěru, že nejpodstatnější roli budou mít takové příznaky, jejichž hodnota se změnila mezi nultou fází a pooperační fází (fáze 1). To samo o sobě

však nestačí. Je nutné si uvědomit, že ke změnám může docházet i u zdravých osob. Jak bylo vysvětleno dříve, hlas se mění v závislosti na řadě různých faktorů a reflektuje například denní dobu, psychickou pohodu, únavu a mnoho dalších aspektů. Klíčové tedy je v první řadě objektivně určit, jak se hlas a jeho charakteristiky mění u zdravých jedinců, čímž se získají referenční hodnoty pro normální změny charakteristik hlasu. Poté lze srovnávat rozdíly u zdravých a nemocných a zjistit, zda jsou změny v rozsahu běžných hodnot či nikoliv.

Pro zjištění normálních hodnot u zdravých subjektů byla vybrána skupina 15 osob bez hlasových obtíží. Všichni prošli stejným procesem, jako každý z pacientů, aby byly dodrženy stejné podmínky. Za zdůraznění stojí i dodržení stejných časových úseků mezi jednotlivými fázemi nahrávání, tak jak byly definovány pro pořizování nahrávek pacientů. Každý z referenční skupiny tedy v tiché komoře přečetl standardní text celkem třikrát. Poprvé se mohl řečník dostavit kdykoliv dle svých časových možností a nahrávka byla označena jako fáze 0. Další fáze již byly pečlivě plánovány. Fáze 1 tedy po 2-3 dnech, jako by tomu bylo u pacienta po operaci. Poslední nahrávka byla pak pořizována po uplynutí 2 týdnů od fáze 0.

U pacientů byl plán nahrávání stejný. Zúčastnilo se celkem 18 osob. Bohužel pro fázi 2 bylo pořizováno neuspokojivé množství nahrávek. Jelikož se nejednalo o dostatečně reprezentativní vzorek, byly tyto nahrávky z analýzy vyřazeny a výsledky jsou stanoveny pouze s použitím fáze 0 a 1.

Fáze 0 byla použita pro vyhodnocení rozdílů jako referenční nahrávka. To znamená, že nahrávky z fází 1 a 2, respektive jejich extrahované příznaky, byly porovnávány s fází 0. Na základě tohoto porovnání bylo určeno skóre, které představuje odlišnost jednotlivých nahrávek. Stanovení hodnoty skóre (globální vzdálenosti) bylo provedeno metodou DTW.

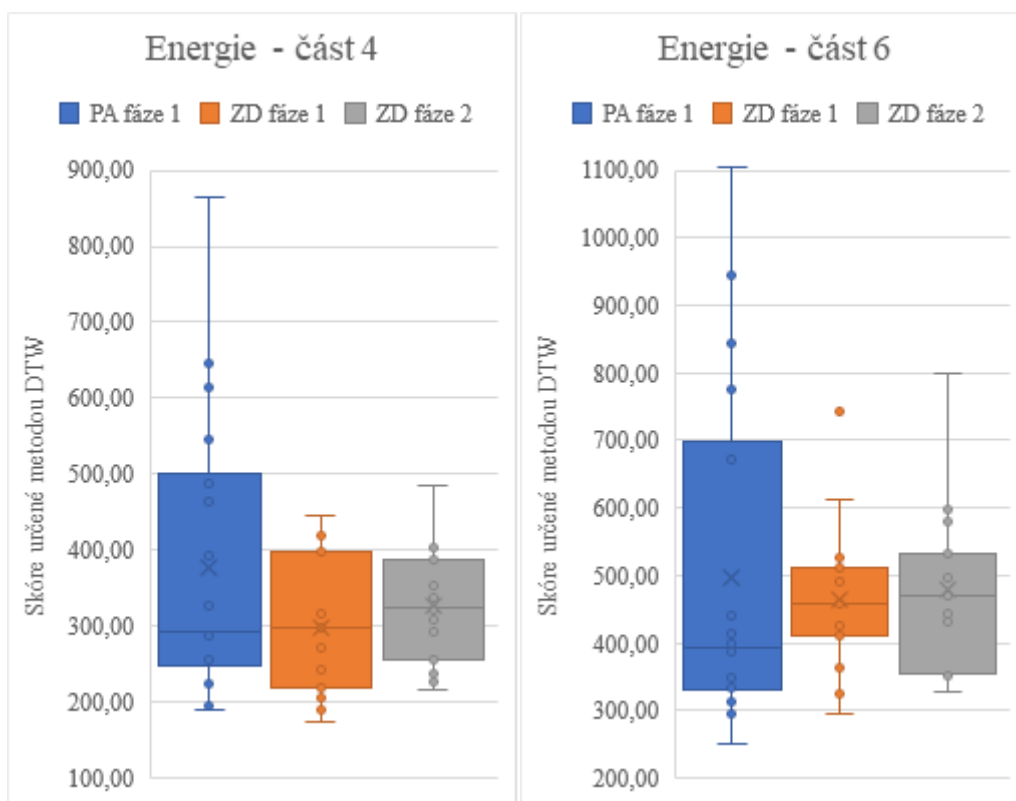
8.2.2 Určení reprezentativních příznaků na základě srovnání DTW skóre

Aby bylo možné určit, které příznaky jsou reprezentativní, bylo skóre určeno pro všechny zvolené příznaky jednotlivě, a to pro celé nahrávky. Výsledky byly z velké části neprůkazné, neboť v rámci celé nahrávky se spousta změn příznaků vykompenzovala a nebylo možné oddělit zdravé od nemocných. K získání detailnějších výsledků ze vstupních dat byla každá nahrávka navíc rozdělena na šest částí, které jsou u všech mluvčích z pohledu obsahu textu totožné. DTW skóre bylo následně vypočítáno pro každou z šesti částí nahrávek samostatně. Tyto části jsou v příloze B odlišeny novým odstavcem. Z první části nahrávek byl rovněž odebrán nadpis textu, protože ne všichni zúčastnění jej přečetli.

Výsledkem analýzy je celkem 210 grafů (35 příznaků pro 6 částí každé nahrávky). Pouze na základě některých úseků nahrávek a konkrétních příznaků lze blíže určit, zda jsou změnu hlasu pacientů v běžném rozmezí. Dále budou prezentovány a diskutovány pouze dostatečně reprezentativní hodnoty, které lze k tomuto účelu využít.

Obecně platí, že čím vyšší je hodnota vypočítaného skóre, tím jsou nahrávky rozdílnější z pohledu sledovaného parametru. Všechny grafy mají v názvu uvedeno, o jaký příznak a část nahrávky se jedná. Každý krabicový graf zobrazuje celkem 3 datové pruhy. První s označením „PA fáze 1“ představuje hodnoty DTW skóre pacientů při porovnání nahrávek z fáze 1 (reprezentace) s fází 0 (reference). Druhý sloupec „ZD fáze 1“ zaznamenává totéž ovšem u zdravých osob. Poslední soubor dat s označením „ZD fáze 2“ ukazuje hodnoty DTW skóre při porovnání fáze 2 (reprezentace) s fází 0 (reference).

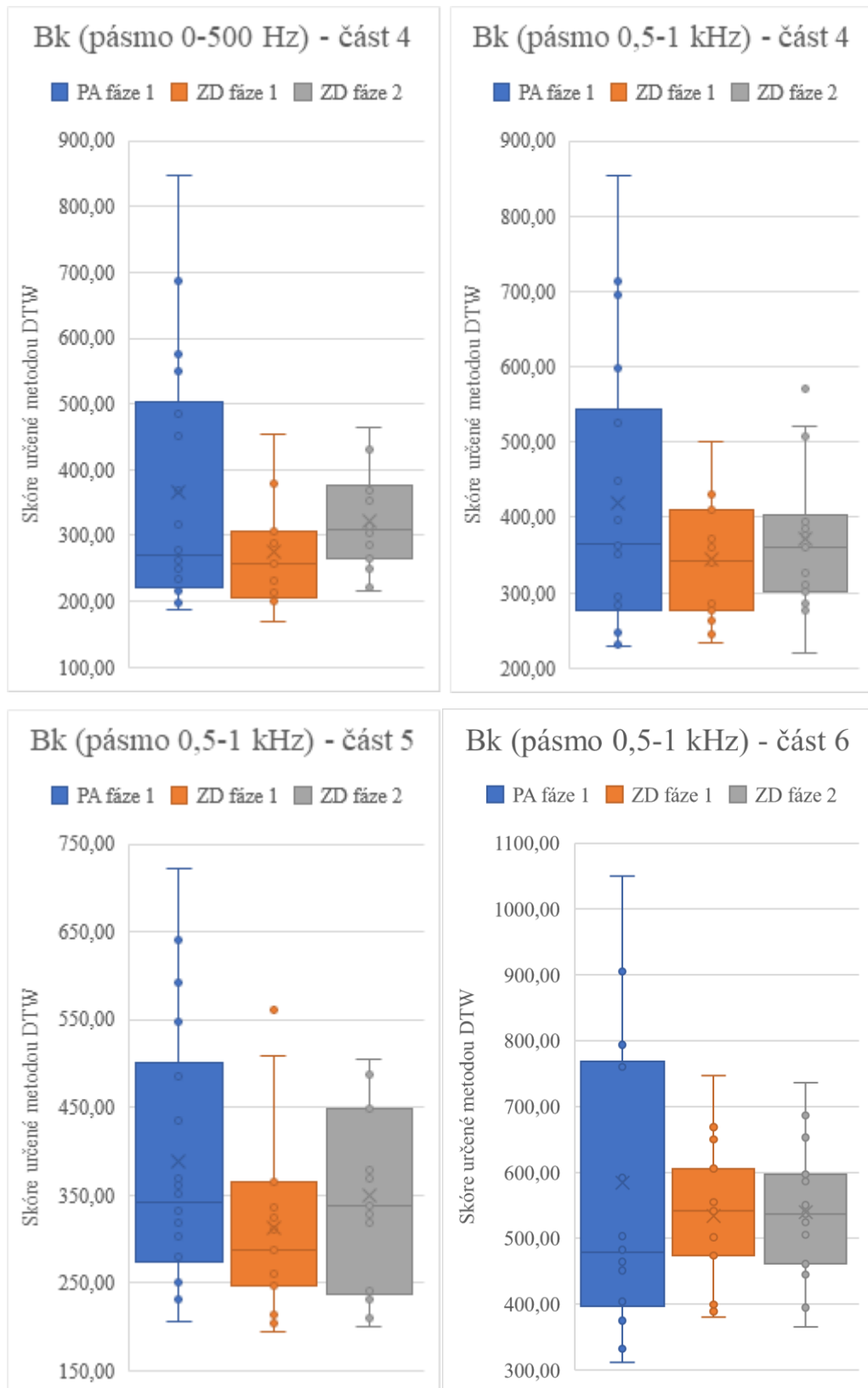
Prvním příznakem umožňujícím odlišit patologické změny hlasu je logaritmická energie. Přímou souvisí s charakterem signálu a jeho úrovněmi. Zvýšené hodnoty se projeví hned ve dvou částech nahrávky. Detail je viditelný na obrázku 8.20. Na druhou stranu dynamické příznaky energie (ΔE a $\Delta \Delta E$) žádné výrazné změny skóre nepřinesly.



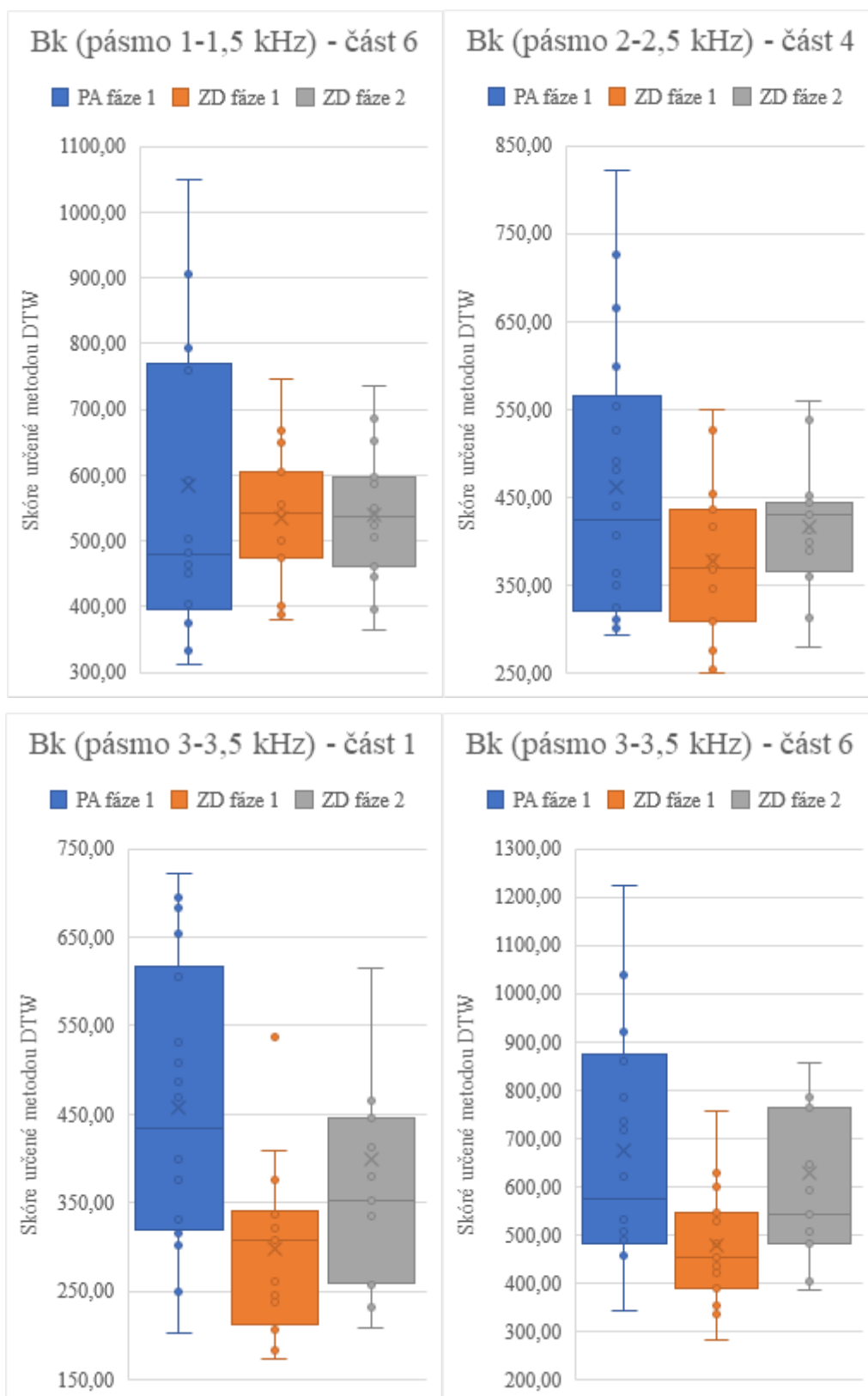
Obrázek 8.20 – Změny logaritmické energie u zdravých a nemocných

Podrobnější informace o rozložení energie ve frekvenčním spektru přináší analýza osmi frekvenčních pásy v rozsahu 0–4 kHz. Zatímco v některých pásmech se rozdíly

příliš neprojevily, v jiných jsou naopak výraznější. Výsledky porovnání spektrální energie B_k shrnují grafy na obrázcích 8.21, 8.22 a 8.23.

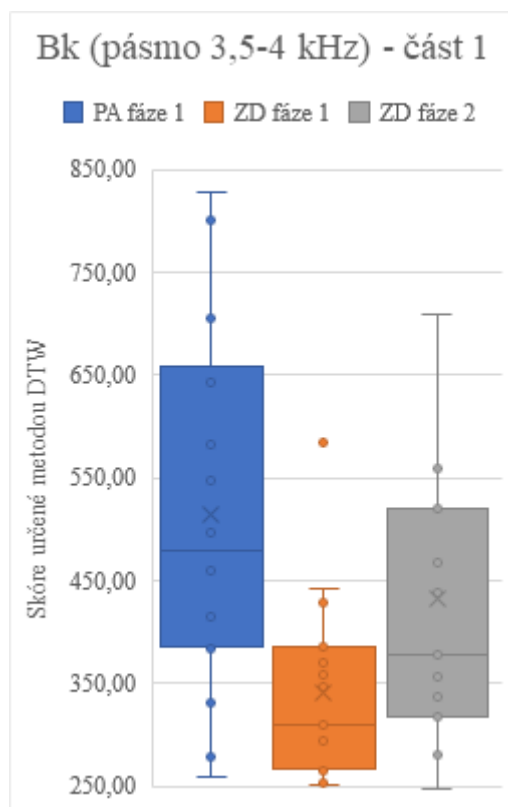


Obrázek 8.21 – Změny spektrální energie v pásmech 0-1 kHz u zdravých a nemocných



Obrázek 8.22 – Změny spektrální energie v pásmech 1-3,5 kHz u zdravých a nemocných

Jelikož se změny logaritmické energie projevily především ve čtvrté a šesté části nahrávek, lze totéž očekávat i u spektrální energie. Odchytky ve skóre se nicméně projevily i mimo tyto úseky. Dyn. příznaky ΔB_k nepřinesly signifikantní výsledky, podobně jako ΔE .

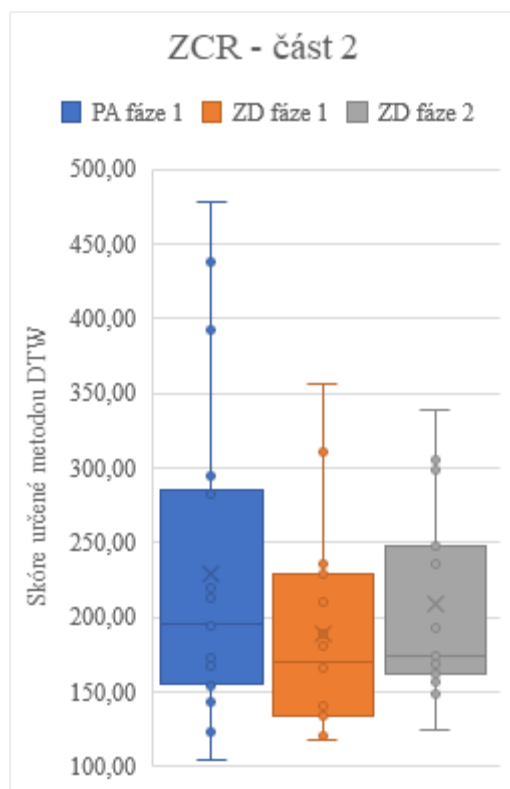


Obrázek 8.23 – Změny spektrální energie v pásmu 3,5-4 kHz u zdravých a nemocných

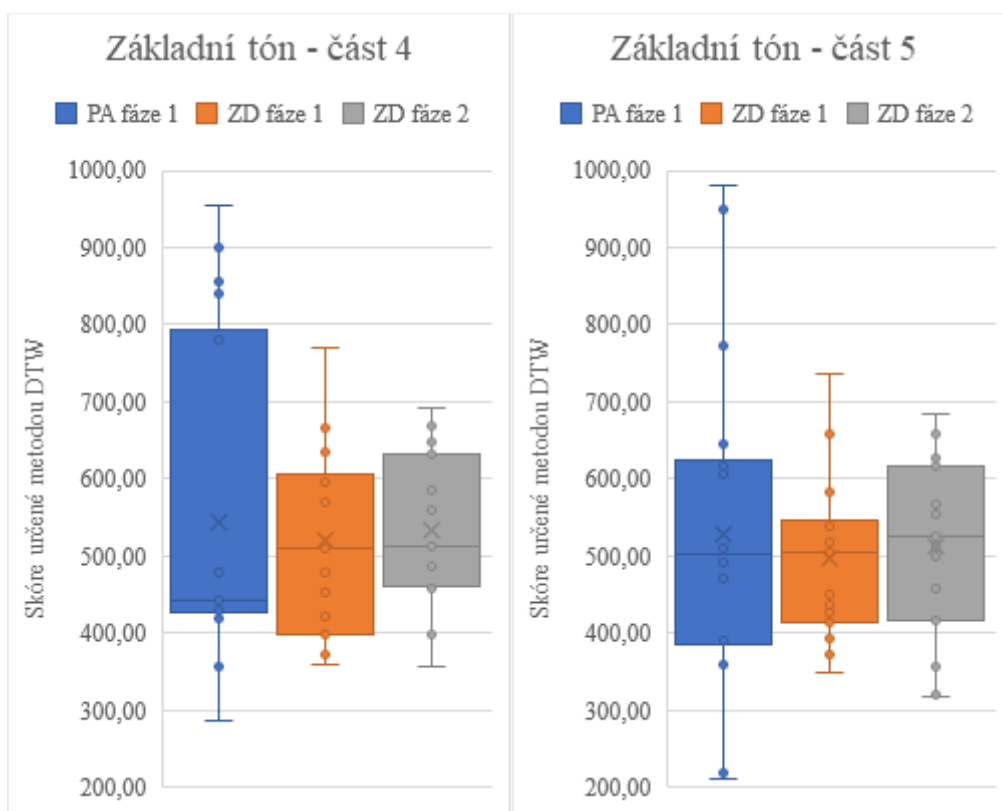
Dalším z příznaků, u kterých se po zákroku u pacienta objevily hodnoty, které přesahují změny hlasu u zdravých osob, je počet průchodů nulou. Vzhledem k tomu, že tento parametr je citlivý na rozlišení znělých a neznělých hlásek, může to znamenat, že došlo ke zvýšení šumového charakteru hlasu. Fonace znělých hlásek tak nemusí být úplně „čistá“, jako by tomu bylo u zdravého jedince. Výsledky shrnuje obrázek 8.24.

Základní frekvence (základní tón) je parametrem, který rovněž vykazuje vyšší hodnoty skóre, než je obvyklé. Jedná se o veličinu, která se již v oblasti diagnostiky poruch hlasu využívá při měření hlasového profilu (VRP). Základní tón je veličina, jež udává, s jakou frekvencí hlasivky kmitají. Opět lze předpokládat, že po invazivním zákroku může dojít ke změně rozsahu generovaných frekvencí. Významnější změny zachycuje obrázek 8.25.

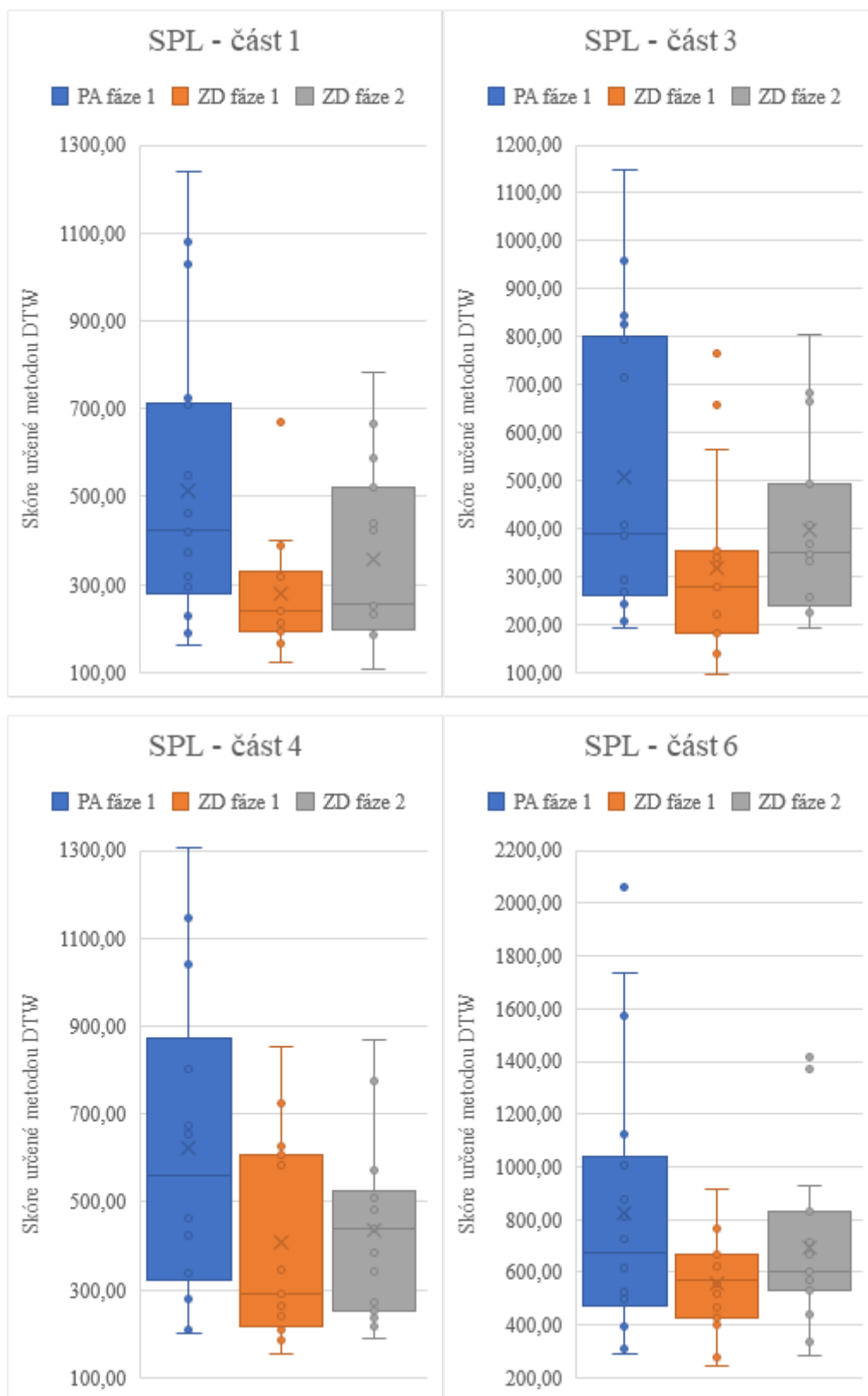
Hlasitost mluveného projevu reflektuje hladina akustického tlaku (SPL) udávaná v decibelech. Je druhým stěžejním faktorem pro vyhodnocení VRP. Zde se projevil významné změny téměř ve všech úsecích nahrávky. Může to znamenat, že bolest po zákroku či další vlivy vyvolané operací neumožňují pacientovi mluvit na jinak komfortní úrovni hlasitosti řeči. SPL se ukazuje jako nejvýznamnější parametr pro odlišení změn hlasu u pacientů. Přehled skóre určeného pro SPL je na obrázku 8.26.



Obrázek 8.24 – Změny ZCR u zdravých a nemocných

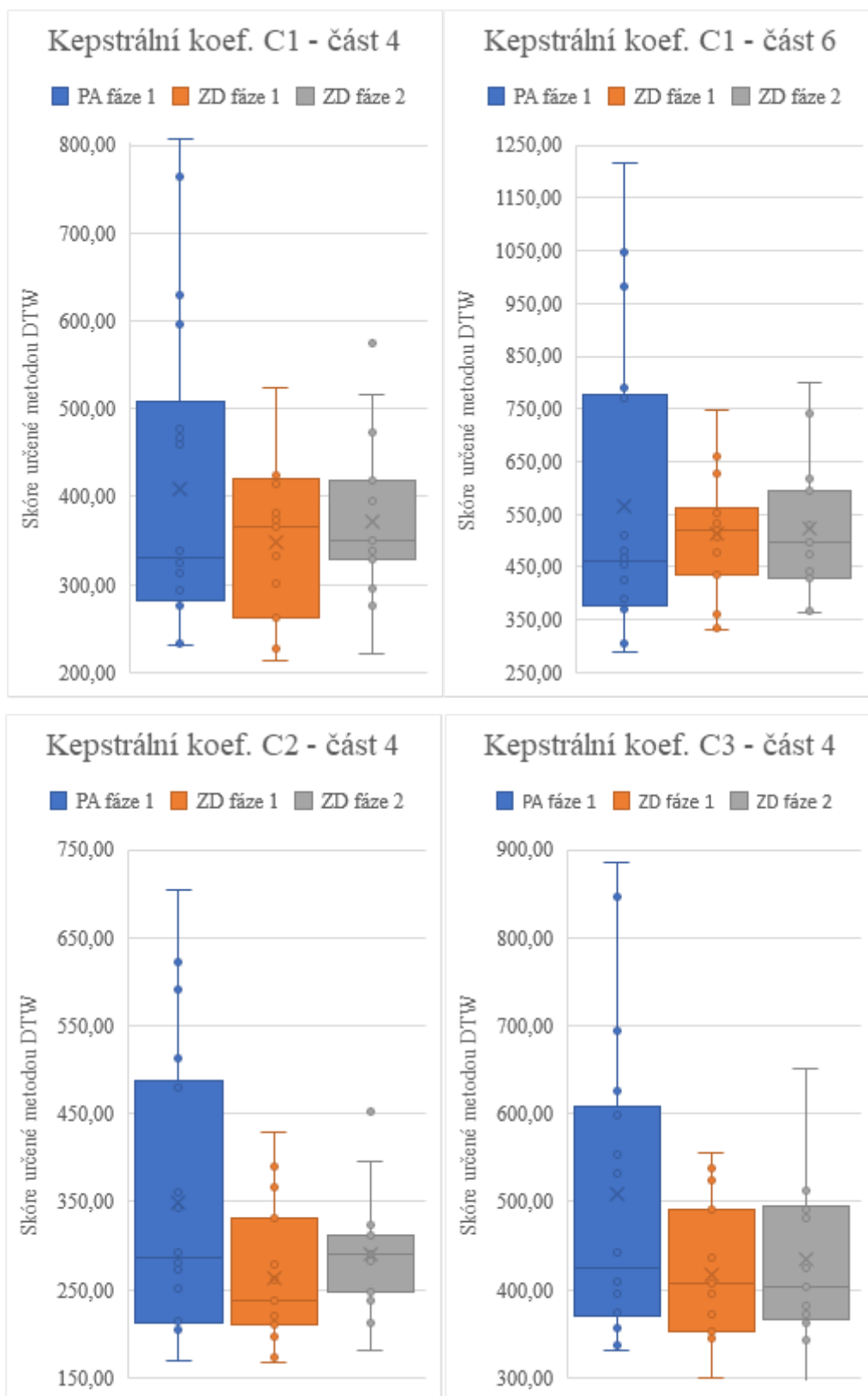


Obrázek 8.25 – Změny základního tónu u zdravých a nemocných

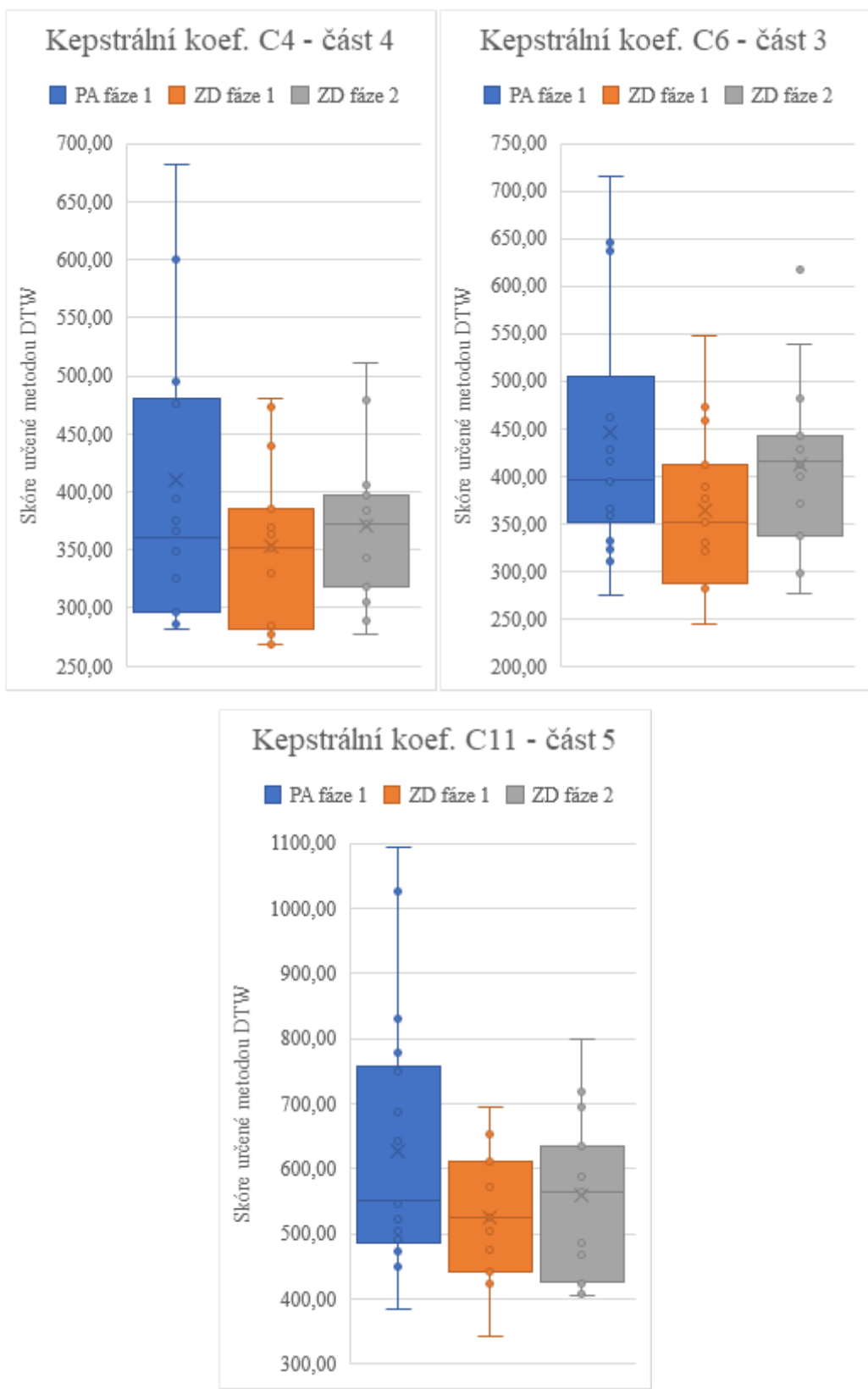


Obrázek 8.26 – Změny hladiny akustického tlaku u zdravých a nemocných

Další sada příznaků, jenž se často používá při zpracování řeči jsou koeficienty MFCC. Z celkem třinácti stanovených MFCC, se změny projevíly opět pouze u některých. Přehled skóre pro keprální analýzu ukazuje obrázek 8.27 a 8.28.



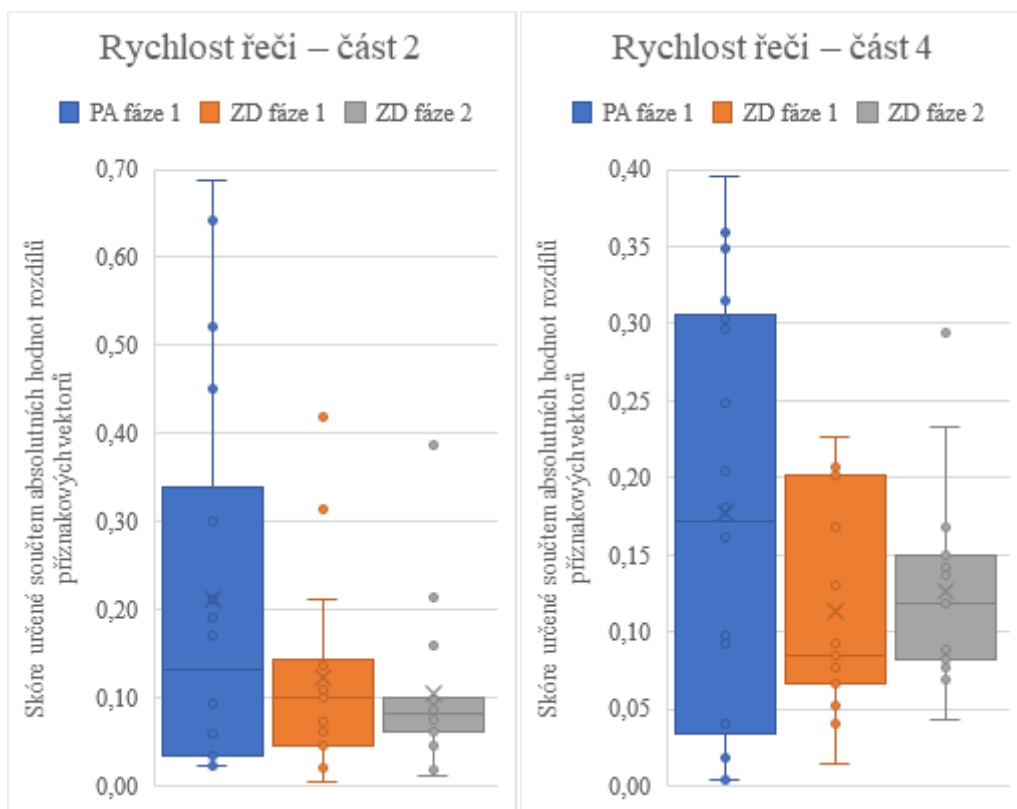
Obrázek 8.27 – Změny koeficientů kepstrální analýzy (C1-C3) u zdravých a nemocných



Obrázek 8.28 – Změny koeficientů kepstrální analýzy (C4-C13) u zdravých a nemocných

Jediným příznakem, pro který nebylo nutné použít DTW je rychlost řeči. Jelikož byly nahrávky rozděleny na stejné úseky obsahující 27, 21, 20, 25, 22 a 35 slov. Pro jednotlivé

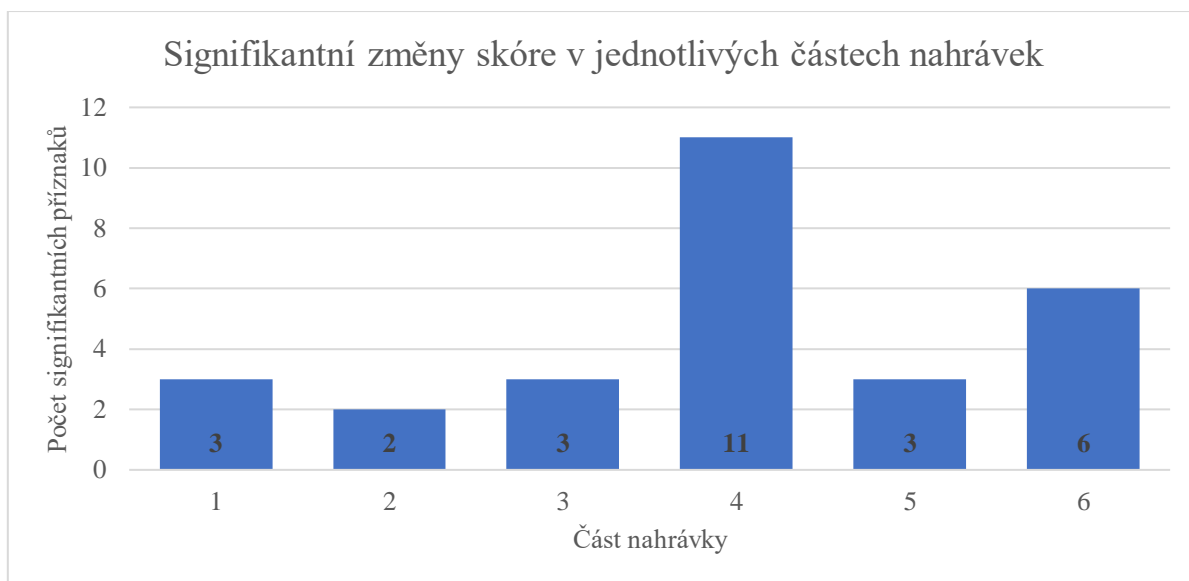
úseky byly určeny rychlosti řeči a vznikly tak příznakové vektory stejné délky. Sečtením absolutních hodnot rozdílů těchto příznakových vektorů pak bylo určeno skóre podobnosti nahrávek z pohledu rychlosti řeči. Toto skóre shrnují grafy na obrázku 8.29.



Obrázek 8.29 – Změny rychlosti řeči u zdravých a nemocných

Jak je ze všech grafů patrné, nikdy nelze přesně odlišit pacienty po zákroku od skupiny zdravých osob. Ale to je zcela logické. Předpokladem je, že pokud operace a následné hojení proběhne hladce, u pacienta by nemělo dojít k žádným výraznějším změnám, a proto není překvapením, že velká část vzorků se pohybuje v rozmezí normálních změn hlasu. Pokud ovšem hodnoty skóre přesahují běžný rozsah zdravých osob, je to dobrý ukazatel toho, že pacienta bude dobré pečlivěji sledovat.

Při pozornějším prostudování výsledků se lze zaměřit i na to, která část nahrávky vlastně přinesla nejvýznamnější výsledky. Jak bylo patrné na dříve uvedených grafech, významnější změny DTW skóre se projevily pouze v některých částech nahrávek. A u každého příznaku byl klíčový jiný úsek. Po sumarizaci počtu signifikantních změn vzhledem k příslušnosti k dané části nahrávky vznikl obrázek 8.30, který ukazuje, že z pohledu sledování změn hlasu pacientů jednoznačně nejvíce tyto změny reflektuje čtvrtý segment nahrávky. Poslední část nahrávky, která je v počtu signifikantních příznaků druhá v pořadí, přitom vykazuje zhruba poloviční počet výraznějších změn. Ostatní úseky pak mají přibližně stejné zastoupení.



Obrázek 8.30 – Nejvýznamnější změny DTW skóre vztažené k částem nahrávek

Jelikož největší odchylky v rychlosti čtení byly zaznamenány ve druhém a čtvrtém úseku nahrávky, mohlo by se zdát, že právě toto bude mít klíčovou roli ve změnách skóre jednotlivých příznaků. Jak ale ukazuje již obrázek 8.30, v druhé části se toto neprojevílo. Pro zamítnutí této hypotézy byla rovněž pro čtvrtou část nahrávek vytvořena sada grafů, kde na ose x bylo vyneseno skóre rychlosti řeči a na osu y bylo postupně zaneseno všech 10 příznaků se signifikantními změnami. Ani u jednoho z deseti grafů se neprokázalo, že s vyšším rozdílem rychlosti čtení by došlo k výrazným změnám skóre daného příznaku. Na druhou stranu se opět jedná o očekávané chování, neboť metoda DTW by totiž měla zajistit, aby se právě rozdílné tempo řeči nepodepsalo na výsledném skóre.

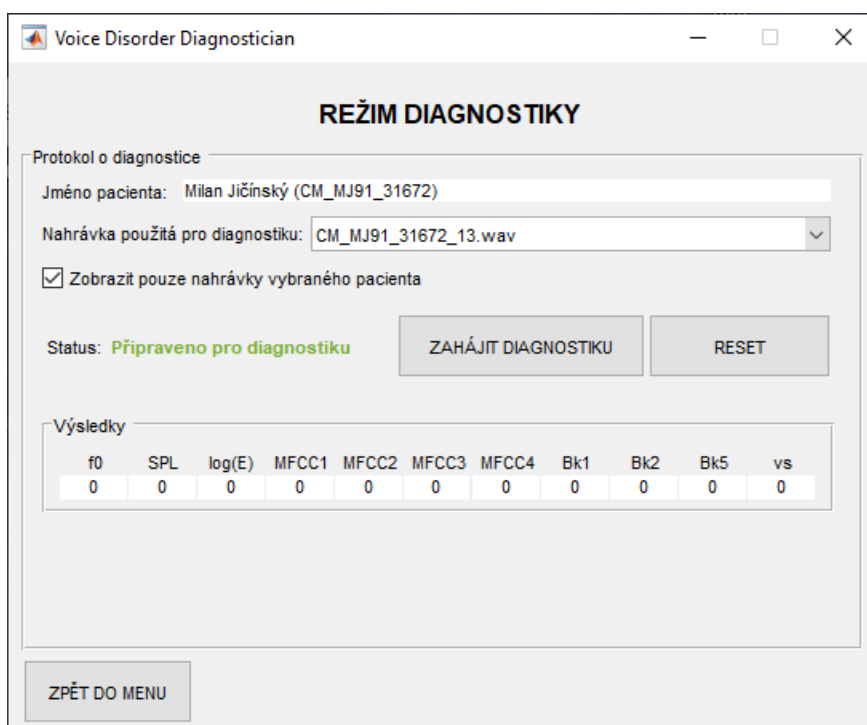
U každého příznaku bylo možné dle získaných dat experimentálně stanovit práh, při jehož překročení lze již předpokládat, že se jedná o vyšší hodnoty než lze očekávat u zdravých osob. Tyto mezní hodnoty jsou samozřejmě závislé na části analyzované nahrávky, a proto je zcela logicky prahová hodnota téhož příznaku rozdílná pro každý z šesti úseků. S danými prahy pak již pracuje aplikace VDD.

8.2.3 Diagnostický režim aplikace VDD

Pro lékaře je bezpochyby nejužitečnější částí celé aplikace VDD diagnostický režim, jenž se v hlavní nabídce skrývá pod tlačítkem diagnostika. Toto tlačítko je ovšem zprvu neaktivní. Aby mohl být spuštěn diagnostický režim, je vyžadováno načtení údajů o pacientovi, který má být diagnostikován. Jakmile je tato podmínka splněna, diagnostika je již dostupná a po stisku příslušného tlačítka se zobrazí nové okno, stejně jako na obrázku 8.31.

V horní části okna se zobrazí jméno pacienta společně s přiřazeným identifikátorem v závorce. Na dalším řádku si může uživatel z rozbalovací nabídky vybrat, kterou z dostupných nahrávek bude chtít použít pro následnou diagnostiku. Mimo to je zde i možnost zrušit omezení výběru nahrávek a zvolit libovolnou nahrávku, což bylo využíváno zejména při testování aplikace, ale tato možnost nakonec ve výsledné aplikaci zůstala.

Hned pod omezením výběru nahrávek se zobrazuje aktuální status. Probíhá-li vše bez problému je aktuální stav zobrazen zeleným písmem. V opačném případě, například pokud chybí referenční nahrávka, se objevuje červený text, který na chybu upozorňuje. Tlačítkem zahájit diagnostiku se spouští extrakční algoritmus, vyhodnocuje se skóre a dochází k porovnání se stanovenými prahovými hodnotami. Pro smazání výsledků a uvedení do výchozího stavu slouží tlačítko reset. Poté může být opět spuštěna diagnostika (například s jinou nahrávkou). Sekce s označením výsledky slouží pro prezentaci hodnot DTW skóre.



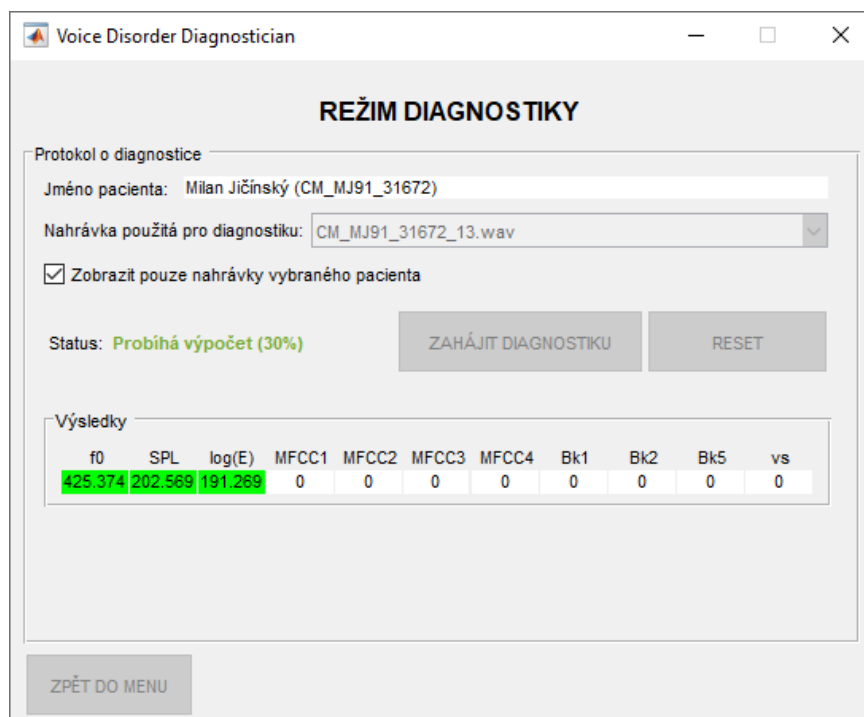
Obrázek 8.31 – Diagnostický režim

Aplikace pracuje právě s 11 vybranými příznaky, k jejichž změnám docházelo ve čtvrté části nahrávek. Rovněž prahové hodnoty jsou tedy vztaženy pouze ke čtvrté části, která je algoritmem analyzována. Tyto experimentálně určené hodnoty přehledně shrnuje tabulka 8.1.

Tabulka 8.1 – Prahové hodnoty příznaků 4. části nahrávek

f_0	SPL	E	$MFCC1$	$MFCC2$	$MFCC3$	$MFCC4$	$Bk1$	$Bk2$	$Bk5$	v_s
800	900	530	590	475	675	550	490	590	580	0,3

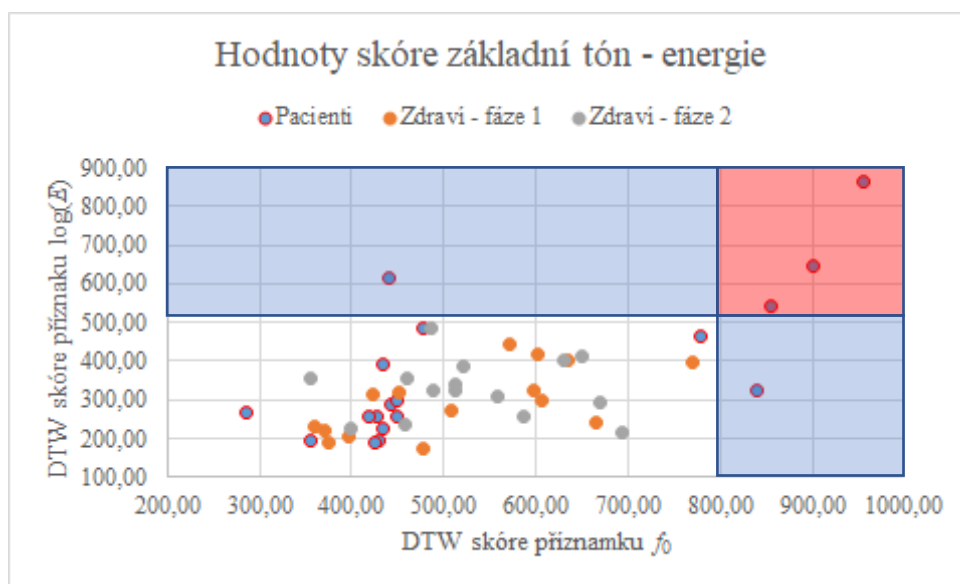
Po spuštění diagnostiky se postup výpočtu průběžně aktualizuje v rámci statusu, kde je uživatel informován, kolik procent výpočtu již bylo provedeno. Ukázka výpočtu je na obrázku 8.32. Nepřekračuje-li hodnota DTW skóre stanovený práh, je buňka s číselnou hodnotou vybarvena zeleně. Naopak příznak, jehož skóre přesáhlo prahovou hodnotu, je zvýrazněn červenou barvou.



Obrázek 8.32 – Nejvýznamnější změny DTW skóre vztažené k částem nahrávek

Po dokončení výpočtu je uživatel informován o tom, kolik příznaků překročilo stanovený práh. Obecně lze předpokládat, že přesahuje-li skóre některého příznaku danou mez, dochází ke změnám hlasu, které jsou mimo běžné hodnoty. V takovém případě je zvýšená pravděpodobnost, že u pacienta dochází k nežádoucímu vývoji v oblasti tvorby hlasu v důsledku prodělaného zákroku. Jedná-li se však o překročení mezních hodnot hned několika příznaků, je to pro lékaře jednoznačná indikace k tomu, aby přistoupil k pečlivějšímu sledování zdravotního stavu pacienta a průběhu léčení. Ve zjednodušené podobě tuto situaci demonstuje obrázek 8.33, který je vyhotoven pro dvojici příznaků energie a základního tónu. Bílá oblast představuje normální změny hlasu. Modré úseky označují překročení prahové hodnoty jednoho příznaku a nabádají ke zvýšené pozornosti ze strany lékaře. Červená oblast pak poukazuje na to, že zvýšené skóre je u obou příznaků a je zde tedy podstatně vyšší riziko. Pokud si bude pacient následně u svého lékaře stěžovat,

že svůj hlas vnímá jinak než dříve, může tento údaj posloužit k podložení či vyvrácení tohoto tvrzení.



Obrázek 8.33 – Oblasti s vyšším rizikem negativních změn hlasu

8.3 Možné směry dalšího rozvoje tématu

Ačkoliv práce jako taková představuje uzavřený celek z pohledu zkoumání změn hlasu u pacientů po operaci štítné žlázy, lze nastínit jakým směrem by se mohl ubírat výzkum, který by na získané znalosti navázal. V první řadě by jistě stálo za úvahu nasazení ASR modelu natrénovaného na co největším objemu dat. Díky přesnému rozpoznání slov, včetně jejich hranic (časové značky začátku a konce daného slova), by bylo možné velmi přesně oddělovat například jednotlivé věty či samostatná slova. Bylo by tak možné se v následné analýze zaměřit na kratší řečové segmenty a zkoumat vývoj jednotlivých příznaků podrobněji než pouze v rámci celé nahrávky či jejích pevně daných šesti částech. Segmentace by navíc probíhala plně automaticky. Obdobně by mohl být ASR model natrénován na rozpoznání zvuku otočení stránky, který je v nahrávkách též přítomen. Dalším zajímavým jevem pro výzkum je práce s dechem. Opět by bylo možné s vhodným modelem zaznamenávat časy, kdy se řečník nadechuje nebo kolik slov bylo na jeden nádech vyřčeno. Je pravděpodobné, že i tyto parametry by mohly být základem ovlivněny. V samotném závěru práce bylo nastíněno, že zvýšení hodnoty skóre u dvojice příznaků již může signalizovat nepříznivý vývoj z pohledu změn hlasu. Aplikace však nebere v potaz případné vazby mezi příznaky. Zobrazuje zkrátka počet příznaků, které překročily stanovenou mez. Dalším logickým krokem by tedy mohla být analýza vícedimenzionálního prostoru tvořeného jednotlivými příznaky za pomoci analýzy hlavních komponent (PCA).

Závěr

Základem pro pochopení celé problematiky je porozumění procesu tvorby hlasu, jemuž byla věnována úvodní kapitola práce. V podobném duchu byla sestavena i úvodní část práce čítající několik kapitol, jenž se zaměřují zejména na potřebnou teorii z oblasti medicíny. Bezprostředně po představení principu tvorby hlasu následoval rovněž souhrnný popis nejen možných poruch tvorby hlasu ale i prevence před jejich vznikem. Pro seznámení se se současnými postupy uplatňovanými na ORL klinikách byly rozebrány běžné i méně časté diagnostické metody. V závěru medicínské části byl pak podrobněji představen problém, na který se soustředil další výzkum – onemocnění štítné žlázy, rizika při operaci a spojitost s tvorbou hlasu. V rámci technického pozadí byly představeny parametry (příznaky) řeči, které se využívají i v jiných aplikacích jako je ASR apod. Byl vytvořen a popsán unikátní postup, jehož cílem bylo minimalizovat variabilitu všech měření tak, aby se bylo možné zaměřit pouze na změny hlasu daného řečníka. Nebyl opomenut ani HW potřebný pro realizaci všech měření včetně zdůraznění klíčových charakteristik jednotlivých přístrojů pro tuto konkrétní aplikaci.

Vyvinutý algoritmus pak umožňuje daný HW kalibrovat a provádět záznam hlasu. Sběr nahrávek bohužel neprobíhal dle očekávání. Zdaleka ne všichni pacienti byly ochotni se do výzkumu zapojit. V období pandemie byly rovněž všechny činnosti v nemocnici omezeny na minimum a sběr dat tak nedostal tolik prostoru. Byly rovněž vyřazeny nahrávky, kde byla k dispozici pouze fáze 0 či 1. Stejným způsobem byly z reprezentativního vzorku vyjmuty nahrávky se špatnou kvalitou či nesprávně provedenou kalibrací. Fáze 2 nakonec bohužel nebyla u pacientů vůbec vyhodnocována, neboť existuje pouze minimální počet záznamů. Z pořízených nahrávek, které byly vybrány pro další zpracování, je následně možné extrahovat jednotlivé příznaky, které slouží pro porovnání vybrané dvojice nahrávek.

Jelikož analýza celých nahrávek nebyla příliš průkazná, přistoupilo se k dělení na kratší segmenty. V pořadí čtvrtá část ze šesti obsahovala největší počet signifikantních změn, a proto se podrobnější výzkum zaměřil právě na ni. Z nahrávek zdravých osob byl stanoven normální rozsah změn hlasu z pohledu jednotlivých parametrů. Z naměřených dat byly experimentálně určeny mezní hodnoty, které umožňují identifikovat takové změny hlasu, jenž přesahují běžný rozsah u zdravého člověka. Jakmile algoritmus určí, že bylo překročeno 2 a více prahových hodnot, lékař by měl pacienta pečlivě prohlédnout a sledovat jeho stav. Pro zjednodušení celého procesu byla pak vytvořena aplikace Voice

Disorder Diagnostician, která všechny tyto poznatky integruje. Umožňuje tak snadnou správu dat o pacientech, záznam hlasu dle stanovených požadavků, manuální kontrolu výsledků i provedení diagnostiky. Všechny tyto funkce jsou opatřeny uživatelsky přívětivým grafickým rozhraním. SW byl postupně aktualizován na základě zpětné vazby od FNKV až do své finální verze tak, aby co nejvíce vyhovoval lékařům, pro které byl vyvinut. Všechny cíle disertační práce tak byly úspěšně splněny.

Literatura

- [1] O tvorbě a produkci lidského hlasu. In: *Technický týdeník* [online]. Praha: Business Media CZ, 2019 [cit. 2019-11-02]. Dostupné z: https://www.technickytydenik.cz/rubriky/medicinska-technika/o-tvorbe-a-produkci-lidskeho-hlasu_37540.html
- [2] Kalendář akcí: Světový den hlasu. In: *Lékárna AGEL* [online]. Ostrava: Agel, 2013 [cit. 2019-11-23]. Dostupné z: <https://lekarna.agel.cz/akce-a-novinky/kalendar-akci/svetovy-den-hlasu.html>
- [3] Forthcoming INTERSPEECH Conferences. In: *Welcome to ISCA Web* [online]. Baixas: International Speech Communication Association, c2009-2019 [cit. 2019-11-23]. Dostupné z: <https://www.isca-speech.org/iscaweb/index.php/conferences/interspeech>
- [4] VAN MERSBERGEN, Miriam, Julie OSTREM a Ingo R. TITZE. Preparation of the Speech-LanguagePathologist Specializing in Voice. In: *Journal of Voice*. 2001, , s. 237-250. ISSN 08921997. Dostupné z: doi:10.1016/S0892-1997(01)00024-8
- [5] Voice anatomy & physiology. In: *The voice foundation* [online]. Philadelphia: The Voice Foundation &vert, 2017 [cit. 2019-11-03]. Dostupné z: <https://voicefoundation.org/health-science/voice-disorders/anatomy-physiology-of-voice-production/>
- [6] KNEIFLOVÁ, Eliška. *Hlasivky: neviditelný, ale důležitý orgán*. Teplice, 2018. Seminární práce. Gymnázium Teplice. Vedoucí práce Pavla Bergmanová.
- [7] LUNGOVÁ, Vlasta. Stavba a funkce hlasového ústrojí. In: *E-learningová podpora mezioborové integrace výuky tématu vědomí na UP Olomouc* [online]. Olomouc: Univerzita Palackého v Olomouci, 2010 [cit. 2019-11-02]. Dostupné z: <http://pfyziolmysl.upol.cz/?p=2661>
- [8] KNOTKOVÁ, Simona. Hlasivky: Příznaky a léčba. In: *Symptomy: Encyklopedie nemocí a jejich příznaků* [online]. Brno: Syntex, c2009–2017 [cit. 2019-11-02]. ISSN 2336-6540. Dostupné z: <https://www.symptomy.cz/anatomie/hlasivky>
- [9] FANT, Gunnar. *Acoustic theory of speech production*. 2nd ed. The Hague: Mouton & Co., 1970. ISBN 9027916004.
- [10] ČAPEK, Karel. *Vytvoření webové aplikace pro objektivní analýzu hypokinetické dysartrie ve frameworku Django*. Brno, 2017. Diplomová práce. Vysoké učení

technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Vedoucí práce Zoltán Galáž.

- [11] Vědci vyvíjejí program, který podle hlasu pozná nemoc. In: *E15: Byznys, politika, ekonomika, finance, události* [online]. Praha: CZECH NEWS CENTER, c2001-2019 [cit. 2019-11-23]. Dostupné z: <https://www.e15.cz/magazin/vedci-vyvijejí-program-který-podle-hlasu-pozna-nemoc-906837>
- [12] BIJOTA, Jan. *Aplikace statistické analýzy řeči pacientů s Parkinsonovou nemocí*. Brno, 2016. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií. Vedoucí práce Zoltán Galáž.
- [13] DUHAIME-ROSS, Arielle. From Alzheimer's to ADHD: what doctors can diagnose from your voice alone. In: *The Verge* [online]. New York: Vox Media, 2019 [cit. 2019-11-23]. Dostupné z: <https://www.theverge.com/2014/7/17/5912491/diagnosing-sound-scientists-voice-recordings-Parkinsons-ADHD-Alzheimers>
- [14] RŮŽIČKOVÁ, Lucie. *Hlas pedagoga a práce s hlasem v učitelství profesí*. České Budějovice, 2016. Bakalářská práce. Jihočeská univerzita v Českých Budějovicích, Fakulta Teologická, Katedra pedagogiky. Vedoucí práce Karel Ochozka.
- [15] KAVKOVÁ, Tereza. *Výskyt poruch hlasu u pedagogů na základních školách*. Olomouc, 2017. Diplomová práce. Univerzita Palackého v Olomouci, Fakulta pedagogická, Ústav speciálněpedagogických studií. Vedoucí práce Kateřina Vitásková.
- [16] KŘEMENÁKOVÁ, Pavla. *Prevence hlasových poruch u profesionálních zpěváků*. Praha, 2009. Diplomová práce. Univerzita Karlova v Praze, 3. lékařská fakulta, Otorinolaryngologická klinika FNKV. Vedoucí práce Zina Hánová.
- [17] KUČERA, Martin, Marek FRIČ, Kateřina FRITZLOVÁ a Martin HALÍŘ. *Vokologie I: Funkční diagnostika a léčba hlasových poruch*. 1. vydání. Praha: Institut pro léčbu a výzkum poruch komunikace, 2019. ISBN 9788027062003.
- [18] KUČERA, Martin, Marek FRIČ a Martin HALÍŘ. *Praktický kurz hlasové rehabilitace a reedukace*. Opocno: Martin Kučera, ORL ambulance – centrum hlasových poruch v Rychnově nad Kněžnou, 2010. ISBN 978-80-254-6592-9.
- [19] DRŠATA, Jakub, Viktor CHROBOK, ed. *Foniatrye - hlas*. Havlíčkův Brod: Tobiáš, 2011. Medicína hlavy a krku. ISBN 978-807311-116-8.
- [20] Vocal cord polyp. In: *UCI Head and neck surgery* [online]. Irvine (California): UCI Irvine Health, 2019 [cit. 2019-11-10]. Dostupné z: <https://uciheadandneck.com/clinical-specialties/vocal-cord-polyp/>

- [21] Laryngitida: běžné, ale závažné onemocnění. In: *České noviny* [online]. Praha: Česká tisková kancelář, 2019 [cit. 2019-11-11]. ISSN: 1213-5003. Dostupné z: <https://www.ceskenoviny.cz/zpravy/laryngitida-bezne-ale-zavazne-onemocneni/368469>
- [22] HYBÁŠEK, Ivan a Jan VOKURKA. Eotorinolaryngologie: Multimediální podpora výuky klinických a zdravotnických oborů. *Portál Lékařské fakulty v Hradci Králové* [online]. 2019 [cit. 2019-11-02]. ISSN 1803-280X. Dostupné z: <https://mefanet.lfhk.cuni.cz/clanky.php?aid=18>
- [23] *Základy otorinolaryngologie a foniatrie pro studenty speciální pedagogiky* [online]. Brno: Klinika dětské ORL Lékařské fakulty MU a Fakultní Nemocnice Brno, 2009 [cit. 2019-11-03]. Dostupné z: <https://is.muni.cz/elportal/estud/pedf/js09/orl/web/index.html>
- [24] ZELENÍK, Karol, Pavel KOMÍNEK, Michal ZÁBRODSKÝ a Ondřej JOR. *Příručka pro praxi: Stenózy hrtanu a průdušnice*. 1. vydání. Praha: RETIS GROUP, 2019. Dostupné také z: http://www.otorinolaryngologie.cz/dokumenty/PPP_stenozy_hrtan.pdf
- [25] STANĚK, Petr. Foniatrická péče o pacienty s dočasnou nebo trvalou obrnou hlasivek po operacích na štítné žláze. *Otorinolaryngologie a Foniatrie* [online]. 2015, **64**(3), 168-172 [cit. 2019-11-21]. ISSN ISSN 1805-4528. Dostupné z: <https://www.prolekare.cz/casopisy/otorinolaryngologie-foniatrie/2015-3/foniatricka-pece-o-pacienty-s-docasnou-nebo-trvalou-obrnou-hlasivek-po-operacich-na-stitne-zlaze-55759>
- [26] SELEMENTOVÁ, Zuzana. Chrapot, sípání i úplnou ztrátu hlasu vyřeší rehabilitace. In: *Žena: magazín pro ženy* [online]. Praha: Economia, c1999-2019 [cit. 2019-11-15]. Dostupné z: <https://zena.aktualne.cz/zdravi/ztratili-jste-hlas-poslete-hlasivky-na-rehabilitaci/r~i:article:805438/>
- [27] KLOZAR, Jan. *Speciální otorinolaryngologie*. Praha: Galén, 2005. ISBN 80-246-1125-2.
- [28] HORÁČKOVÁ, Zuzana. Onemocnění hlasivek u dětí. Co se může stát, když dítě často křičí?. In: *MotherClub* [online]. Praha: CzechNetMedia, 2019 [cit. 2019-11-21]. Dostupné z: <https://motherclub.cz/onemocneni-hlasivek-u-deti-co-se-muze-stat-kdyz-dite-casto-krici/>
- [29] MACEK, Petr a Tomáš LAMBERSKÝ. Lad'a Kerndl: Těžká operace. Je na kapačkách!. In: *Aha: online magazín plný žhavých drbů ze showbyznysu* [online]. Praha: CZECH NEWS CENTER, c2001-2019 [cit. 2019-11-17]. Dostupné z:

[https://www.ahaonline.cz/clanek/zhave-drby/4115/lada-kerndl-tezka-operace-je-na-
kapackach.html](https://www.ahaonline.cz/clanek/zhave-drby/4115/lada-kerndl-tezka-operace-je-na-
kapackach.html)

- [30] PAVELEC, Václav. *Rizikové faktory vzniku Reinkeho edému a možnosti jejich ovlivnění*. Plzeň, 2006. Disertační práce. Lékařská fakulta Univerzity Karlovy v Plzni. Vedoucí práce Jaroslav Slípka.
- [31] ZELENÍK, Karol, Pavel KOMÍNEK a Jan CHLUMSKÝ. *Příručka pro praxi: Extraezofageální reflux*. 1. vydání. Praha: Merck, 2013. Dostupné také z: http://www.otorinolaryngologie.cz/dokumenty/PPP_reflux.pdf
- [32] *HPV College* [online]. HPV College, 2019 [cit. 2019-11-18]. Dostupné z: <https://www.hpv-college.cz/>
- [33] ŠPIČKOVÁ, Nikola. *Hodnocení kvality života a hlasových obtíží u pacientů sonemocněním hrtanu*. Pardubice, 2017. Diplomová práce. Univerzita Pardubice, Fakulta zdravotnických studií, Katedra ošetřovatelství. Vedoucí práce Jana Škvrňáková.
- [34] Jednotlivé poruchy hlasu. In: *Nakladatelství Portál* [online]. Praha: PORTÁL, 2019 [cit. 2019-11-19]. Dostupné z: <https://nakladatelstvi.portal.cz/nakladatelstvi/aktuality/79797/jednotlive-poruchy-hlasu>
- [35] HAVLÍK, Radan. *Příručka pro praxi: Hyperkinetická dysfonie*. 1. vydání. Praha: Merck, 2013. Dostupné také z: http://www.otorinolaryngologie.cz/dokumenty/PPP_dysfonie.pdf
- [36] VELDOVÁ, Zuzana. Možnosti léčby hlasových profesionálů v ambulantní praxi. *Interní medicína pro praxi* [online]. 2005, 7(11), 496-498 [cit. 2019-11-19]. ISSN 1803-5256. Dostupné z: <https://www.internimedicina.cz/pdfs/int/2005/11/07.pdf>
- [37] JEDELSKÁ, Eva. Zaměřeno na hlasivky: Proč alkohol mění váš hlas?. In: *Proženy* [online]. Praha: Seznam.cz, c1996-2019 [cit. 2019-11-19]. Dostupné z: <https://www.prozeny.cz/clanek/zamereno-na-hlasivky-proc-alkohol-meni-vas-hlas-16794>
- [38] NOVÁK, Alexej. *Foniatric a pedaudiologie II: poruchy hlasu u dětí a dospělých - základy anatomie a fyziologie hlasu, diagnostika, léčba, reedukace a rehabilitace poruch hlasu*. 2. přepracované vydání. Praha: Alexej Novák, 2000. ISBN 9788023863246.
- [39] VOHLÍDKOVÁ, Monika, Pavel NOVÝ a Jiří PEŠTA. Srovnání akustických a optických metod vyšetření hlasu u pacientů s patologickými nálezy na hlasivkách. In: *4. Česko-slovenský foniatrický kongres* [online]. Ostrava, 2006 [cit. 2019-11-05]. Dostupné z: <http://www.kiv.zcu.cz/~novyp/publ/foniatr06.pdf>

- [40] BLOOTHOOFT, Gerrit a Peter PABON. Vocal Registers Revisited. In: *Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)*. Budapešť: ISCA, 1999, s. 423-426. Dostupné také z: https://www.isca-speech.org/archive/archive_papers/eurospeech_1999/e99_0423.pdf
- [41] FRIČ, Marek. *Objektivní a psychoakustické aspekty hodnocení lidského hlasu*. Praha, 2013. Disertační práce. České vysoké učení technické v Praze, Fakulta elektrotechnická, Katedra radioelektroniky. Vedoucí práce Václav Syrový.
- [42] NOVÝ, Pavel a Jiří PEŠTA. *Fonetogram 4.5 – LITE: Program pro vyšetření a analýzu hlasového pole* [online]. Plzeň [cit. 2019-11-02]. Dostupné z: <http://www.kiv.zcu.cz/~novyp/foniatry/phonetogram.pdf>
- [43] FRIČ, Marek, Zdeněk OTČENÁŠEK a Václav SYROVÝ. Akustika hlasu. In: *Sborník abstraktů a příspěvků 2. symposia Umělecký hlas*. Praha: Zvukové studio HAMU a Hlasové centrum Praha, 2010, s. 53-65. ISBN 978-80-7331-170-4.
- [44] PEŠTA, Jiří, Pavel NOVÝ a Zdeněk KASL. Pooperační objektivní posouzení hlasu. In: *XV. celostátní foniatrické dny Evy Sedláčkové: 2. česko-slovenský foniatrický kongres*. Plzeň, 2004. Dostupné také z: <http://www.kiv.zcu.cz/~novyp/publ/foniatr04.pdf>
- [45] STOMEIO, Francesco, Vittorio RISPOLI, Mariachiara SENSI, Antonio PASTORE, Nicola MALAGUTTI a Stefano PELUCCHI. Subtotal arytenoidectomy for the treatment of laryngeal stridor in multiple system atrophy: phonatory and swallowing results. In: *Brazilian Journal of Otorhinolaryngology*. 2016, , s. 116-120. ISSN 18088694. Dostupné z: doi:10.1016/j.bjorl.2015.03.016
- [46] ŠVEC, Jan G., Mojmír LEJSKA a Jana FROSTOVÁ. Česká verze dotazníku Voice Handicap Index pro kvantitativní hodnocení hlasových potíží vnímaných pacientem. *Otorinolaryngologie a foniatrie* [online]. 2009, **58**(3), 132-139 [cit. 2019-11-07]. ISSN 1805-4528. Dostupné z: <https://www.prolekare.cz/casopisy/otorinolaryngologie-foniatrie/2009-3/ceska-verze-dotazniku-voice-handicap-index-pro-quantitativni-hodnoceni-hlasovych-potizi-vnimanych-pacientem-7605>
- [47] K čemu jsou laryngoskopy?. In: *Oceán zdraví* [online]. Makawiel Publishing [cit. 2019-11-24]. Dostupné z: <https://oceanzdravi.cz/k-cemu-jsou-laryngoskopy/>
- [48] ŠVEC, Jan G. *Základní a aplikovaný výzkum tvorby lidského hlasu*. Olomouc, 2019. Habilitační práce. Univerzita Palackého v Olomouci, Přírodovědecká fakulta.
- [49] *The visible voice: HreS ENDOCAM 5562 - Laryngoscopic diagnosis system*. Knittlingen. Dostupné také z: <https://itarmed.org/files/product/636087637636412271-Diagnosticheskaya-sistema-dlya-foniatirii-HreS-ENDOCAM-5562-Richard-Wolf.pdf>

- [50] FIALOVÁ, Kateřina. *Hlasová hygiena v mateřských školách*. Brno, 2006. Bakalářská práce. Masarykova univerzita, Fakulta pedagogická, Katedra speciální pedagogiky. Vedoucí práce Ilona Bytešnicková.
- [51] Hlasová hygiena. In: *Nové služby* [online]. c2011-2019 [cit. 2019-11-24]. Dostupné z: <http://novesluzby.cz/zdravi-a-zdrava-vyziva/hlasova-hygiena>
- [52] FRITZLOVÁ, Kateřina. Hlasové poruchy. In: *Klinická logopedie* [online]. Žamberk: dStudio, 2013 [cit. 2019-11-25]. Dostupné z: http://www.logopedie-fritzlova.cz/?page_id=101
- [53] JISKRA, Jan. *Poruchy štítné žlázy: praktický přehled nejen pro laickou veřejnost* [online]. 1. vydání. Praha: Mladá fronta, 2011 [cit. 2020-04-30]. Lékař a pacient. ISBN 978-80-204-2456-3. Dostupné z: https://www.nfozp.cz/download/Poruchy_stitne_zlazy_dvoustrany.pdf
- [54] KADLČÁKOVÁ, Eva. Včasná operace štítné žlázy může zachránit život. In: *Český rozhlas* [online]. Praha: Český rozhlas, c1997-2020 [cit. 2020-04-30]. Dostupné z: <https://budejovice.rozhlas.cz/vcasna-operace-stitne-zlazy-muze-zachranit-zivot-7058095>
- [55] Těhotenství a porucha štítné žlázy. In: *Naseporodnice.cz: těhotenství, výběr porodnice, rodina a péče o dítě* [online]. Praha: naseporodnice.cz, c2008-2020 [cit. 2020-05-01]. Dostupné z: <https://www.naseporodnice.cz/tehotenstvi-a-porucha-stitne-zlazy.php>
- [56] Proč je důležitá štítná žláza. In: *Oblastní nemocnice Příbram* [online]. Příbram: Internet PB & Smartim, c2003-2020 [cit. 2020-04-30]. Dostupné z: <http://www.nemocnicepribram.cz/clanky/detail/proc-je-dulezita-stitna-zlaza>
- [57] NEČASOVÁ, Pavlína. *Onemocnění štítné žlázy a možnosti prevence tohoto onemocnění*. Brno, 2019. Dostupné také z: https://is.muni.cz/th/s37ur/Onemocneni_stitne_zlazy_a_moznosti_prevence_tohoto_onemocneni.pdf. Bakalářská práce. Katedra tělesné výchovy a výchovy ke zdraví. Vedoucí práce Jitka Slaná Reissmannová.
- [58] PŘÍHODOVÁ, Anna. Struma, vole. In: *Vitalia: chytře na život* [online]. Praha: Internet Info, c2009–2020 [cit. 2020-05-02]. ISSN 1802-8012. Dostupné z: <https://www.vitalia.cz/clanky/struma-vole/>
- [59] HTE – TTE (operace štítné žlázy). In: *Kroměřížská nemocnice* [online]. Kroměříž: Raab Computer, 2015 [cit. 2020-04-30]. Dostupné z: <https://www.nem-km.cz/140-hte-tte-operace-stitne-zlazy.html>

- [60] Chirurgie štítné žlázy. In: *Zdravotnictví a medicína* [online]. Praha: Mladá fronta, 2020 [cit. 2020-05-03]. Dostupné z: <https://zdravi.euro.cz/clanek/postgradualni-medicina/chirurgie-stitne-zlazy-148149>
- [61] NOVÁKOVÁ, Michaela. *Pooperační komplikace u pacientů s onemocněním štítné žlázy*. Pardubice, 2015. Dostupné také z: https://dk.upce.cz/bitstream/handle/10195/60807/NovakovaM_PooperacniKomplikace_AD_2015.pdf?sequence=1&isAllowed=y. Bakalářská práce. Univerzita Pardubice, Fakulta zdravotnických studií. Vedoucí práce Alena Dvořáčková.
- [62] WAIBEL, Alex. a Kai-Fu. LEE. *Readings in speech recognition*. 1. vydání. San Mateo, Calif.: Morgan Kaufmann Publishers, 1990. ISBN 978-1558601246.
- [63] CHALUPNÍČEK, Kamil. *Rozpoznávání diktované řeči pro medicínské aplikace*. Brno, 2004. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce Pavel Matějka.
- [64] THIRION, Jan Willem Frederik. *Recurrent neural network-enhanced HMM speech recognition systems*. Pretoria, 2011. Disertační práce. University of Pretoria, Faculty of Engineering.
- [65] JIČÍNSKÝ, Milan a Jaroslav MAREK. New Year's Day Speeches of Czech Presidents: Phonetic Analysis and Text Analysis. In: *Computer Information Systems and Industrial Management*. Cham: Springer International Publishing, 2017, s. 110-121. Lecture Notes in Computer Science. ISBN 978-3-319-59104-9. Dostupné z: doi:10.1007/978-3-319-59105-6_10
- [66] JIRKOVSKÝ, Jaroslav. Strojové učení a prostředí MATLAB. In: *Humusoft: ve službách vědy a techniky* [online]. Praha: HUMUSOFT, c1991-2019 [cit. 2019-11-29]. Dostupné z: <https://www.humusoft.cz/blog/20180817-strojove-uceni/>
- [67] SILOVSKÝ, Jan. *Generativní a diskriminativní klasifikátory v úlohách textové nezávislého rozpoznávání a diarizace mluvcích*. Liberec, 2011. Disertační práce. Technická Univerzita v Liberci, Fakulta mechatroniky, informatiky a mezioborových studií, Ústav informačních technologií a elektroniky. Vedoucí práce Jan Nouza.
- [68] ČERNOCKÝ, Jan. *Zpracování řečových signálů: studijní opora*. Brno, 2006. Dostupné také z: <https://docplayer.cz/12788883-Http-www-fit-vutbr-cz-cernocky.html>
- [69] BROKEŠOVÁ, Johana. *Teoretické základy*. Praha, 2008. Dostupné také z: <http://geo.mff.cuni.cz/studium/Brokesova-FourierovaSpektralniAnalyza.pdf>
- [70] JIČÍNSKÝ, Milan. *Integrální transformace v teorii automatického řízení*. Pardubice, 2015. Diplomová práce. Univerzita Pardubice. Vedoucí práce Libor Kupka.

- [71] COOLEY, J. W. a J. W. TUKEY. Cooley, J. W., and J. W. Tukey. "An algorithm for machine calculation of complex Fourier series. In: *Mathematics of computing*. New York: IEEE Press, 1965, s. 223-227.
- [72] TŮMA, Jiří. *Zpracování signálů získaných z mechanických systémů užitím FFT*. 1. vydání. Štramberk: Sdělovací technika, 1997. ISBN 80-901-9361-7.
- [73] OPPENHEIM, Alan V. *Digital Signal Processing*. 1st edition. New Jersey: Pearson, 1975. ISBN 978-0132146357.
- [74] ČÍŽEK, Václav. *Diskrétní Fourierova transformace a její použití*. 1. vydání. Praha: Státní nakladatelství technické literatury, 1981.
- [75] MAREŠ, Jiří. *Časově-frekvenční analýza nestacionárních signálů*. Liberec, 2009. Disertační práce. Technická univerzita v Liberci, Fakulta mechatroniky, informatiky a mezioborových studií. Vedoucí práce Ivan Jaksch.
- [76] QUATIERI, Thomas F. *Discrete-time speech signal processing: principles and practice*. 1. vydání. Upper Saddle River, N.J.: Prentice-Hall, 2002. ISBN 01-324-2942-X.
- [77] VONDRA, Martin. Kepstrální analýza řečového signálu. In: *Elektrorevue: Časopis pro elektrotechniku* [online]. Brno: ISES (International Science and Engineering Society), 2013 [cit. 2020-03-07]. Dostupné z: <http://www.elektrorevue.cz/clanky/01048/index.html>
- [78] MÜLLER, Meinard. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. 1. vydání. Cham: Springer, 2015. ISBN 978-3-319-21944-8.
- [79] NOUZA, Jan. On the speech feature selection problem: Are dynamic features more important than the static ones?. In: *Proc. of EUROSPEECH'95*. Madrid: ISCA, 1995, s. 919-922. Dostupné také z: https://www.isca-speech.org/archive_v0/archive_papers/eurospeech_1995/e95_0919.pdf
- [80] GERVEN, Stefan Van a Fei XIE. A comparative study of speech detection methods. In: *Fifth European Conference on Speech Communication and Technology*. Rhodes: ISCA, 1997, s. 1095-1098. ISSN 1018-4074.
- [81] MATUŠTÍK, Daniel. *Určování základního hlasového tónu*. Brno, 2013. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce Milan Sigmund.
- [82] VÁGNEROVÁ, Monika. *Základy akustiky: příručka pro začátečníky*. Praha, 2013. Dostupné také z: <http://www.greif.cz/download/its075-zaklady-akustiky-prirucka-pro-zacatecniky.pdf>

- [83] ČSN 01 1600. *Akustika: Terminologie*. 1. vydání. Praha: Český normalizační institut, 2003.
- [84] *CzechEncy: Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita, c2012–2018 [cit. 2019-12-19]. Dostupné z: <https://www.czechency.org/>
- [85] Research: Committee/Areas of interest/Open-access policy/Request for Information on Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001 [cit. 2020-01-19]. Dostupné z: https://meta.wikimedia.org/wiki/Research:Committee/Areas_of_interest/Open-access_policy/Request_for_Information_on_Public_Access_to_Peer-Reviewed_Scholarly_Publications_Resulting_From_Federally_Funded_Research
- [86] VŠETIČKA, Martin a Jaroslav VŠETIČKA. Hlasitost a intenzita zvuku: Základní definice. In: *Encyklopedie fyziky* [online]. Praha: WebArchiv, c2006-2020 [cit. 2020-01-20]. Dostupné z: <http://fyzika.jreichl.com/main.article/view/208-zakladni-definice>
- [87] KAŇKA, Jan. Intenzita zvuku. In: *TZB-info: stavebnictví, úspory energií, technická zařízení budov* [online]. Praha: Topinfo, c2001-2020 [cit. 2020-01-20]. ISSN 1801-4399. Dostupné z: <https://stavba.tzb-info.cz/akustika-staveb/216-intenzita-zvuku-decibel>
- [88] ŠRÁMKOVÁ, Hana. *Vyšetření dynamického rozsahu akustického tlaku u lidského hlasu*. Olomouc, 2010. Diplomová práce. Univerzita Palackého v Olomouci. Vedoucí práce Jan G. Švec.
- [89] MUDA, Lindasalwa, Mumtaj BEGAM a Irraivan ELAMVAZUTHI. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *JOURNAL OF COMPUTING* [online]. 2010, 2(3), 138-143 [cit. 2022-05-11]. ISSN 2151-9617. Dostupné z: [doi:https://doi.org/10.48550/arXiv.1003.4083](https://doi.org/10.48550/arXiv.1003.4083)
- [90] AMIN, Talal Bin a Iftekhar MAHMOOD. Speech Recognition using Dynamic Time Warping. In: *2008 2nd International Conference on Advances in Space Technologies*. Islamabad: IEEE, 2008, s. 74-79. ISBN 978-1-4244-3299-8. Dostupné z: [doi:10.1109/ICAST.2008.4747690](https://doi.org/10.1109/ICAST.2008.4747690)
- [91] PORTILLA, Ricardo, Brenner HEINTZ a Denny LEE. Understanding Dynamic Time Warping: Part 1 of our Using Dynamic Time Warping and MLflow to Detect Sales Trends Series. In: *Databricks: Data Lakehouse Architecture and AI Company* [online]. San Francisco: Databricks, 2022 [cit. 2022-05-05]. Dostupné z: <https://databricks.com/blog/2019/04/30/understanding-dynamic-time-warping.html>

- [92] CHEN, Qian, Guyu HU, Fanglin GU a Peng XIANG. Learning optimal warping window size of DTW for time series classification. In: *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)* [online]. Montreal: IEEE, 2012, s. 1272-1277 [cit. 2022-05-08]. ISBN 978-1-4673-0382-8. Dostupné z: doi:10.1109/ISSPA.2012.6310488
- [93] ŠKOPEK, Pavel. Techbox: mikrofon změni nejen váš hlas v elektrinu. In: *Mobilenet.cz* [online]. Praha: 24net, 2020 [cit. 2020-03-14]. Dostupné z: <https://mobilenet.cz/clanky/techbox-mikrofon-zmeni-nejen-vas-hlas-v-elektrinu-14057>
- [94] MOLINA, Adam. What to look for in a microphone: Get the right mic the first time. In: *SoundGuys* [online]. SoundGuys, 2020 [cit. 2020-03-20]. Dostupné z: <https://www.soundguys.com/what-to-look-for-in-a-microphone-26035/>
- [95] ŠVEC, Jan G. a Svante GRANQVIST. Guidelines for Selecting Microphones for Human Voice Production Research. In: *American Journal of Speech-Language Pathology*. Vol. 19. American Speech-Language-Hearing Association, 2010, , s. 356-368. ISSN 1058-0360. Dostupné z: doi:10.1044/1058-0360(2010/09-0091)
- [96] JIRÁSEK, Ondřej. Jak vybrat správný mikrofon do studia. Jakou směrovou charakteristiku?: Díl 3. Směrovost. In: *Disk* [online]. Boskovice: DISK Multimedia, 2020 [cit. 2020-03-20]. Dostupné z: <https://magazin.disk.cz/cs/jak-vybrat-spravny-mikrofon-do-studia-jakou-smerovou-charakteristiku>
- [97] VLACHÝ, Václav. Mikrofony. In: *Muzikus: hudební portál* [online]. Praha: MUZIKUS, 2013 [cit. 2020-03-20]. Dostupné z: <http://www.muzikus.cz/pro-muzikanty-clanky/Mikrofony~23~unor~2013/>
- [98] Why do I need to calibrate my sound level meter?. In: *Pulsar instruments* [online]. Filey: Pulsar instruments, 2020 [cit. 2020-03-23]. Dostupné z: <https://pulsarinstruments.com/en/post/pulsar-recalibration-services>
- [99] G. ŠVEC, Jan, Peter S. POPOLO a Ingo R. TITZE. Measurement of vocal doses in speech: experimental procedure and signal processing. *Logopedics Phoniatrics Vocology*. 2009, **28**(4), 181-192. ISSN 1401-5439. Dostupné z: doi:10.1080/14015430310018892
- [100] NOVÁK, Jan a Radovan ZADRAŽIL. Měření hluku mobilem. In: *SVMTech* [online]. Praha: 11 Studio [cit. 2020-03-23]. Dostupné z: <https://www.svmtech.cz/cz/news/mereni-hluku-mobilem>
- [101] *Get closer to DPA microphones: General program*. Lillerød, 2016. Dostupné také z: <https://www.dpamicrophones.com/dpa/media/dpa-brochures/dpa-catalog-2016.pdf>

- [102] FLIR SYSTEMS. *407780A: Integrating Sound Level Meter*. Waltham, c2013-2014. Dostupné také z: http://www.extech.com/products/resources/407780A_DS-en.pdf
- [103] FLIR SYSTEMS. *Integrating Sound Level Meter and Datalogger - Model 407780A: User Manual*. Waltham, c2014-2017. Dostupné také z: http://www.extech.com/products/resources/407780A_UM-en.pdf
- [104] Hlukoměr Extech datový záznamník 407780A. In: *Conrad Electronic: velkoobchod s moderní elektronikou a technikou* [online]. Praha: Conrad Electronic, 2020 [cit. 2020-03-27]. Dostupné z: <https://velkoobchod.conrad.cz/hlukomer-extech-datovy-zaznamnik-407780a-30-130-db-kalibrovano-dle-bez-certifikatu.k1385163>
- [105] FLIR SYSTEMS. *Model 407744 and 407766 Sound Level Calibrators: User manual*. Wilsonville, 2017. Dostupné také z: http://www.extech.com/products/resources/407744_407766_UM-en.pdf
- [106] EXTECH INSTRUMENTS CORPORATION. *Sound Level Calibrators*. Waltham, c2008-2011. Dostupné také z: http://www.extech.com/products/resources/407744_407766_DS-en.pdf
- [107] Extech 407744 94dB Sound Calibrator for 0.5" and 1" Microphones. In: *Amazon* [online]. Seattle: Amazon, c1996-2020 [cit. 2020-03-27]. Dostupné z: <https://www.amazon.com/Extech-407744-Sound-Calibrator-Microphones/dp/B001AFFVR6?th=1>
- [108] Focusrite Scarlett 2i2 USB Audio Interface (2nd Generation). In: *B&H: Photo Video Digital Cameras, Photography, Camcoders* [online]. New York: B & H Foto & Electronics Corp., c2000-2020 [cit. 2020-03-28]. Dostupné z: https://www.bhphotovideo.com/c/product/1251655-REG/focusrite_scarlett_2i2_2nd_gen_scarlett_2i2_usb_audio.html
- [109] FOCUSRITE AUDIO ENGINEERING. *Scarlett 2i2: User guide*. High Wycombe, 2016. Dostupné také z: https://www.bhphotovideo.com/lit_files/374166.pdf

Seznam publikací autora

- 1) JIČÍNSKÝ, M. a J. MAREK. New Year's Day Speeches of Czech Presidents: Phonetic Analysis and Text Analysis. In: *Computer Information Systems and Industrial Management*. Cham: Springer International Publishing, 2017, 2017-05-17, s. 110-121. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-59105-6_10. ISBN 978-3-319-59104-9. Dostupné také z: http://link.springer.com/10.1007/978-3-319-59105-6_10.

- 2) JIČÍNSKÝ, M. a J. MAREK. Clustering analysis of phonetic and text feature vectors. In: *2017 IEEE 14th International Scientific Conference on Informatics*. IEEE, 2017, 2017, s. 146-151. DOI: 10.1109/INFORMATICS.2017.8327237. ISBN 978-1-5386-0888-3. Dostupné také z: <http://ieeexplore.ieee.org/document/8327237/>.
- 3) HONC, Daniel a Milan JIČÍNSKÝ. Analytic Model Predictive Controller in Simple Symbolic Form. In: *Innovation, Engineering and Entrepreneurship*. Cham: Springer International Publishing, 2019, 2019-06-03, s. 82-88. Lecture Notes in Electrical Engineering. DOI: 10.1007/978-3-319-91334-6_12. ISBN 978-3-319-91333-9. Dostupné také z: http://link.springer.com/10.1007/978-3-319-91334-6_12
- 4) JIČÍNSKÝ, Milan a Jan MAREŠ. Measurable Changes of Voice After Voice Disorder Treatment. In: *Intelligent Systems Applications in Software Engineering*. Cham: Springer International Publishing, 2019, 2019-09-20, s. 295-305. Advances in Intelligent Systems and Computing. DOI: 10.1007/978-3-030-30329-7_27. ISBN 978-3-030-30328-0. Dostupné také z: http://link.springer.com/10.1007/978-3-030-30329-7_27
- 5) JIČÍNSKÝ, Milan, Jan MAREŠ, Ludmila VEREŠPEJOVÁ a Martin CHOVANEC. Speech Processing in Diagnosis of Vocal Cords Diseases. In: *Applied Electronics 2019*. Plzeň: Západočeská univerzita v Plzni, 2019, s. 75-78. ISBN 978-80-261-0812-2. ISSN 1805-9597. Dostupné také z: [https://dspace5.zcu.cz/bitstream/11025/35519/1/Jic%
c3%adnsk%c3%bd.pdf](https://dspace5.zcu.cz/bitstream/11025/35519/1/Jic%c3%adnsk%c3%bd.pdf)
- 6) JIČÍNSKÝ, Milan a Jan MAREŠ. Software tool for voice disorder diagnostics. *Transactions of the VŠB - Technical University of Ostrava, Mechanical Series* [online]. Ostrava: VŠB-Technical University of Ostrava, 2019, (1), 11-18 [cit. 2019-11-03]. DOI: 10.22223/tr. ISSN 1804-0993. Dostupné z: <https://dspace.vsb.cz/bitstream/handle/10084/138520/2051.pdf>

Příloha A – Dotazník VHI

Index hlasového postižení – Voice Handicap Index (VHI), česká sjednocená verze.

Jméno a příjmení: _____ Datum narození: _____

Dnešní datum: _____

Níže uvedené výroky vyjadřují běžně používané popisy vlivu hlasu a jeho postižení na lidský život.

Zakroužkujte odpověď, která nejlépe odpovídá Vaší situaci.

Stupeň Vašich hlasových potíží dnes: Žádné Mírné Střední Těžké

	Nikdy	Téměř nikdy	Někdy	Téměř vždy	Vždy
P1. Stává se, že mi během mluvení dochází dech.	0	1	2	3	4
P2. Zvuk mého hlasu se v průběhu dne mění.	0	1	2	3	4
P3. Stává se, že se mě lidé ptají, co to mám s hlasem.	0	1	2	3	4
P4. Můj hlas zní skřípavě a vyprahle.	0	1	2	3	4
P5. Abych ze sebe vydal/a hlas, musím vynaložit úsilí.	0	1	2	3	4
P6. Stává se, že dopředu nevím, jak můj hlas bude znít, když promluvíím.	0	1	2	3	4
P7. Když mluvím, snažím se měnit svůj hlas, aby zněl jinak.	0	1	2	3	4
P8. Mluvení mě stojí hodně úsilí.	0	1	2	3	4
P9. Večer je můj hlas znatelně horší než ráno.	0	1	2	3	4
P10. Stává se, že mi během mluvení hlas z ničeho nic vypoví službu.	0	1	2	3	4
F1. Stává se, že můj hlas lidé špatně slyší.	0	1	2	3	4
F2. V hlučném prostředí mi lidé špatně rozumějí.	0	1	2	3	4
F3. Moje rodina mě špatně slyší, když na ně doma zavolám.	0	1	2	3	4
F4. Kvůli potížím s hlasem používám telefon méně často než bych chtěl/a.	0	1	2	3	4
F5. Kvůli potížím s hlasem se raději vyhýbám situacím, kde bych měl/a mluvit ve skupině.	0	1	2	3	4
F6. Kvůli potížím s hlasem mluvím méně často s přáteli, sousedy či rodinou.	0	1	2	3	4
F7. Lidé mě při rozhovoru často žádají, abych jim něco zopakoval/a.	0	1	2	3	4
F8. Potíže s hlasem mě omezují v osobním a společenském životě.	0	1	2	3	4
F9. Kvůli potížím s hlasem se cítím vyloučen/a, když si ostatní povídají.	0	1	2	3	4
F10. Mé potíže s hlasem mají nepříznivý dopad na mé výdělky.	0	1	2	3	4

	Nikdy	Téměř nikdy	Někdy	Téměř vždy	Vždy
E1. Kvůli potížím s hlasem jsem nervózní, když mám s někým mluvit.	0	1	2	3	4
E2. Zdá se mi, že můj hlas je lidem nepříjemný.	0	1	2	3	4
E3. Zdá se mi, že ostatní mé potíže s hlasem nechápu.	0	1	2	3	4
E4. Potíže s hlasem mi způsobují rozladění/rozčilení/nespokojenost.	0	1	2	3	4
E5. Kvůli potížím s hlasem jsem méně podnikavý/á, společenský/á.	0	1	2	3	4
E6. Kvůli potížím s hlasem se cítím znevýhodněn/a, hendikepován/a.	0	1	2	3	4
E7. Rozčiluje mě, když mě lidé žádají, abych opakoval/a, co jsem říkal/a.	0	1	2	3	4
E8. Cítím se trapně, když mě lidé žádají, abych opakoval/a, co jsem říkal/a.	0	1	2	3	4
E9. Kvůli potížím s hlasem se cítím neschopný/á.	0	1	2	3	4
E10. Stydím se za své potíže s hlasem.	0	1	2	3	4

Vyhodnocení: P hodnota _____, F hodnota _____, E hodnota _____. Celková hodnota ____.

Příloha B – Standardní text používaný při analýze řeči pacientů

Jak v Kocourkově chytali zajíce

První část

Čtyři kocourkovští sousedé se domluvili, že půjdou v noci na zajíce. Ve dne si vyhledali pelech, v kterém zajíc každou noc spával. V noci šli na lov.

Jeden soused nesl lucernu se svíčkou. Druhý nesl buben s paličkou. Třetí nesl kropicí konev s vodou a čtvrtý široký pytel.

Druhá část

Přišli do lesa. Potichu se přikradli až k zaječímu pelechu pod nízkým smrčkem. Opatrně se rozestavili kolem něho a začali.

První soused zamával rozsvícenou lucernou několikrát nad smrčkem. Zajíc se probudil. Jak byl tak ještě rozespálý, pomyslíl si: „No, no, to se nějak začíná blýskat!“

Závěrečná část

Vtom druhý soused udeřil paličkou několikrát na buben. Zajíc rozumoval: „Aha, už bije hrom!“ Potom ten třetí soused rozespálého zajíce rychle pokropil.

Zajíc se otřepal a řekl si: „Už prší! To abych se honem někam schoval!“ Zatím čtvrtý soused nastavil ke smrčku rozevřený pytel. Zajíc uviděl pěknou skrýši u samého smrčku. Vyskočil, hop, už byl v pytli.

Příloha C – Uživatelský manuál pro sběr dat v aplikaci VDD

Zapojení

- Zvukovou kartu Focusrite na zadní straně propojit červeným kablíkem přes USB s počítačem,
- i** Propojení s počítačem je nutné před spuštěním aplikace,



- Zvukoměr se zapojí do konektoru vlevo,
 - Mikrofon se zapojí do konektoru vpravo,
 - Oba přepínače LINE/INST musí být v poloze LINE,
 - Tlačítko 48V musí být zmáčknuté a červeně svítit,
 - Zesílení by během nahrávání mělo být nastaveno tak, aby svítilo **zeleně**, **NIKOLIV červeně**,
- i** Během kalibrace/nahrávání se již nesmí zesílení měnit.

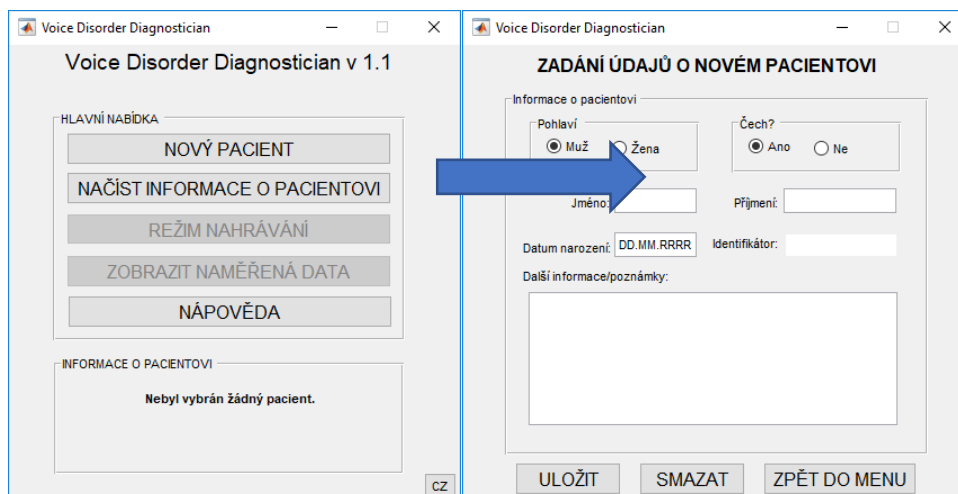
Nastavení softwaru Voice Disorder Diagnostician

Po spuštění se zobrazí hlavní nabídka. Je třeba vybrat pacienta:

Nový pacient – umožňuje zadat do systému údaje o novém pacientovi.

Každému pacientovi bude přiřazen unikátní identifikátor např.: *CM_MJ91_1676*.

C – Čech (X – cizinec), M – muž (Z – žena), MJ91 – iniciály a rok narození, 1676 – unikátní kód



Načíst informace o pacientovi – načtení údajů již existujícího pacienta ze souboru.

Pacienty lze vyhledat zadáním údajů (stejně jako při vytváření záznamu o novém pacientovi). S každým vyplněným údajem se aktualizuje seznam nalezených souborů z databáze pacientů. Přičemž vlevo je seznam souborů a vpravo informace o zvoleném souboru.

NAČTENÍ ÚDAJŮ O PACIENTOVI

Informace o pacientovi

Pohlaví: Muž Žena

Čech?: Ano Ne

Jméno: Příjmení:

Datum narození: Identifikátor:

Nalezené záznamy:

- CM_AL87_700.vdd
- CM_MJ91_1676.vdd
- CM_UK66_1314.vdd

Rychlý náhled

Jméno: **Milan Jičínský**

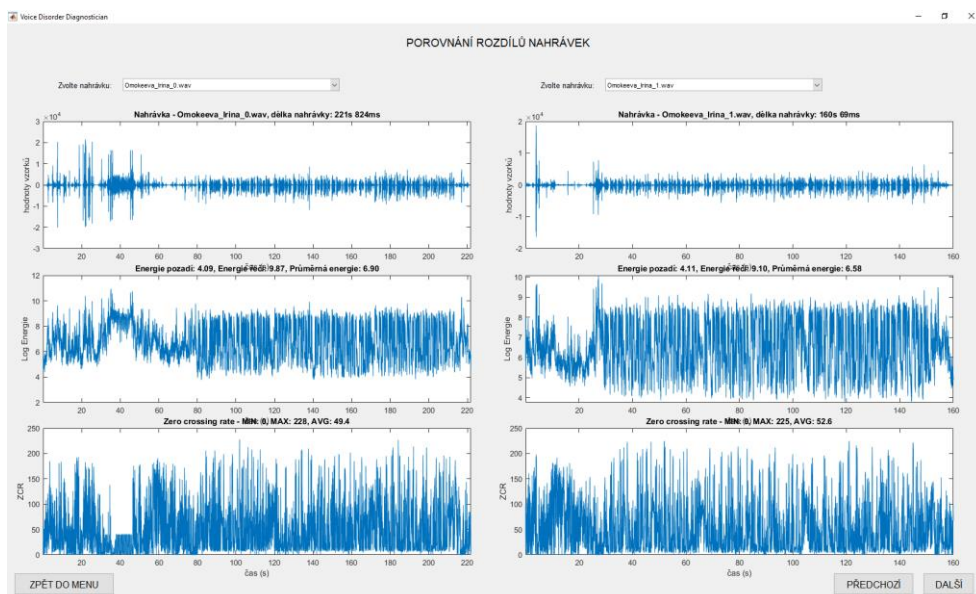
Datum narození: **26.03.1991**

NAČÍST SMAZAT ZPĚT DO MENU

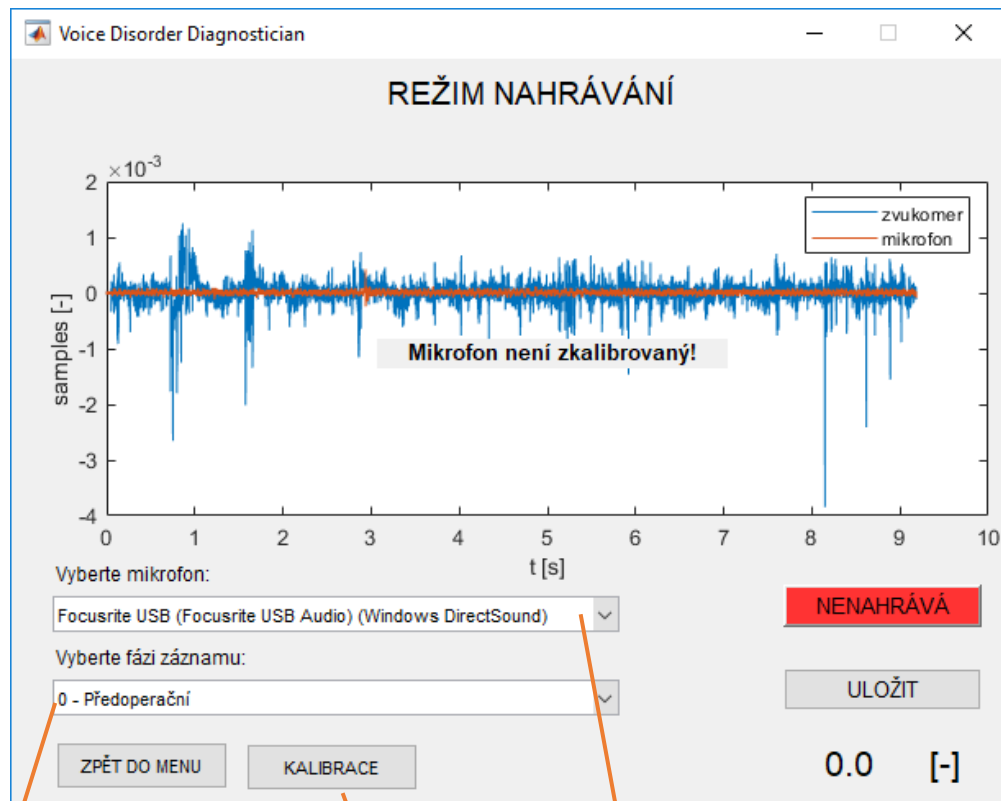
Zobrazení naměřených dat:

Tento režim slouží k zobrazení dříve zaznamenaných nahrávek a jejich klíčových parametrů.

- Nejprve **vyberte obě** nahrávky z rolovacích nabídek nahoře (mezi výběry chvíli počkejte, data se zpracovávají),
- Poté můžete zobrazit charakteristiky na dalších stranách,
- K navigaci používejte tlačítek „Předchozí“ a „Další“.



Režim nahrávání (před kalibrací) – přístupný po načtení dat o pacientovi.



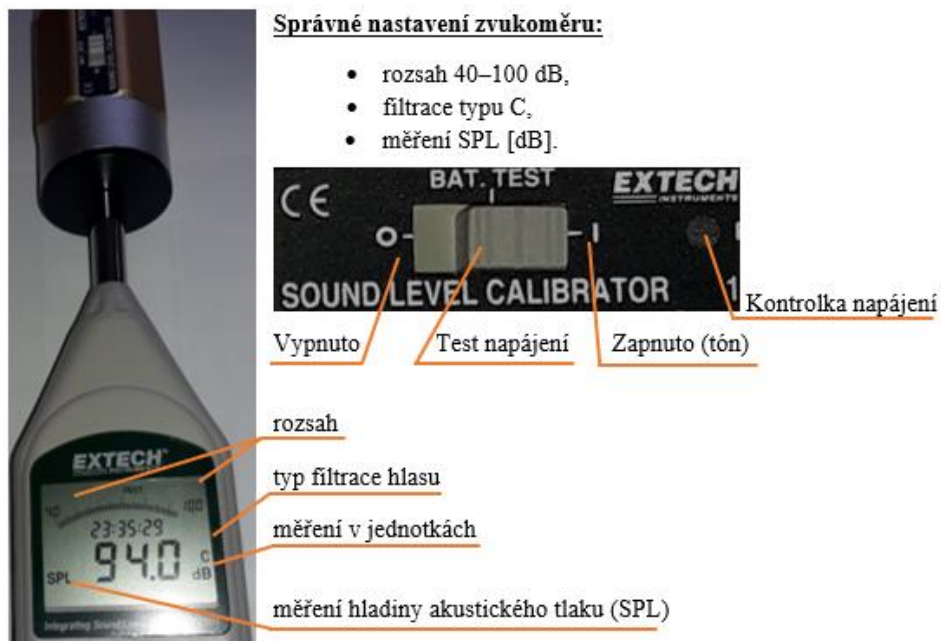
Výběr fáze záznamu

Režim kalibrace

Volba mikrofonu

Postup při nahrávání:

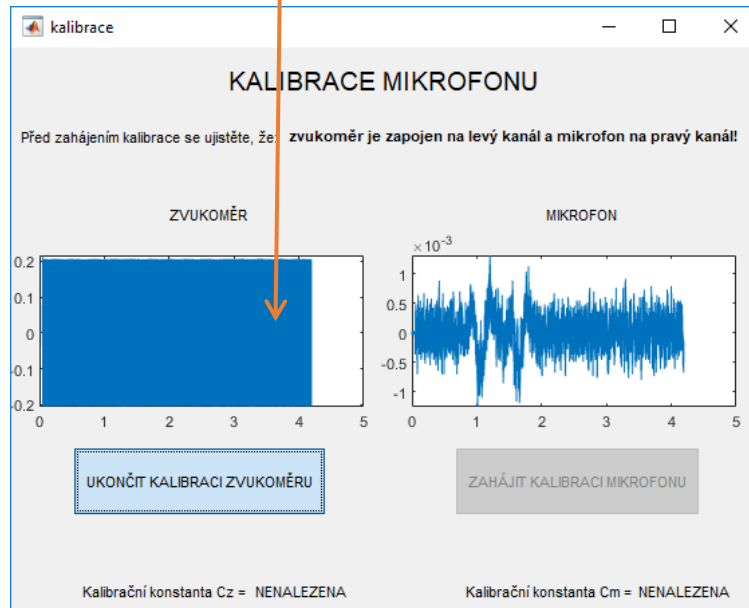
- Vyberte mikrofon **Focusrite USB**, při výběru správného zařízení se začne zobrazovat zvukový signál a hlášení, že mikrofon není kalibrován,
- Vyberte fázi (číslo) záznamu,
- Poté se přepněte do režimu kalibrace.



Režim kalibrace – přístupný z režimu nahrávání.

Kalibrace zvukoměru

- nejprve přiložte kalibrátor na zvukoměr a zapněte generování kalibračního tónu,
- po chvíli stiskněte **Zahájit kalibraci zvukoměru**,
- v levé části **musí být vidět rovný téměř obdélníkový průběh** (po celou dobu záznamu),
- po několika vteřinách stiskněte **Ukončit kalibraci zvukoměru**,
- odstraňte kalibrátor a dejte na zvukoměr ochrannou pěnovou krytku

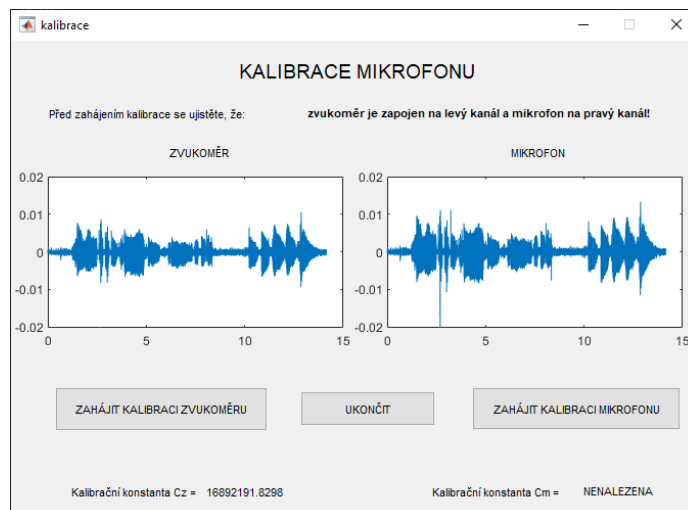


Kalibrace mikrofonu

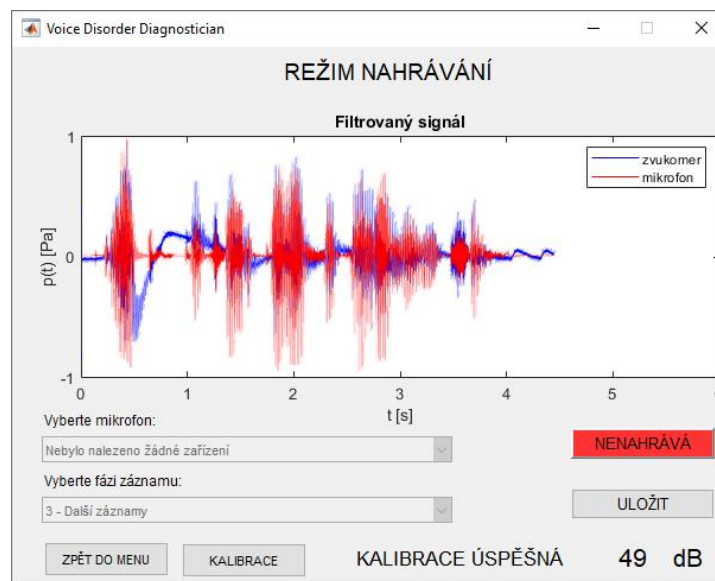
i Vzdálenost mezi mikrofonem a zvukoměrem by měla být 30 cm.

Je třeba dohlížet na to, aby pacient tuto vzdálenost dodržoval až do konce nahrávání.

- Kalibraci zahájíte stiskem tlačítka **Zahájit kalibraci mikrofonu**,
- po krátké promluvě ji ukončíte tlačítkem **Ukončit kalibraci mikrofonu**,
- **oba grafy při kalibraci mikrofonu by si měly být velmi podobné, téměř stejné**



Režim nahrávání (po kalibraci)



- Nyní již není možné změnit fázi nahrávání pacienta ani zařízení,
- stisknutím červeného tlačítka **nenahrává** se spustí nahrávání,
- pro ukončení záznamu stiskněte zelené tlačítko **nahrává**,
- stiskem **uložit** se nahrávka uloží ve formátu .WAV a uživatel se vrátí do hlavní nabídky,
- při návratu z hlavní nabídky je mikrofon stále kalibrovaný pro další nahrávání.

Režim diagnostiky – přístupný z hlavní nabídky až po volbě pacienta

- Nejprve zkontrolujte jméno pacienta a zvolte **zobrazit pouze nahrávky pacienta**,
- poté zvolte nahrávku, která bude použita pro diagnostiku (lze vybrat pouze fázi 1 a 2),
- analýza se spustí tlačítkem **zahájit diagnostiku**,
- zelené pole – parametr je v normě, červené pole – byla překročena prahová hodnota,
- závěr: pokud jsou označeny 2 a více polí červeně, dochází k abnormálním změnám.

