

Obsah

Seznam použitých symbolů	1
1 Úvod	5
2 Komunikační proces	7
2.1 Řeč – přenos informace	7
2.2 Řečový signál	12
2.2.1 Fonetické segmenty – hlásky	13
2.2.2 Charakteristiky hlásek	15
2.3 Systémy hlasových technologií	17
2.3.1 Analogové a digitální přenosové prostředky	17
2.3.2 Syntéza a rozpoznávání řeči	22
3 Vybrané algoritmy zpracování řeči	26
3.1 Digitalizace a rekonstrukce akustického signálu	26
3.1.1 Vzorkování	26
3.1.2 Lineární PCM	26
3.1.3 Nelineární kvantování – <i>a-law</i> a <i>μ-law</i>	28
3.1.4 Adaptivní kvantování	31
3.1.5 Kvantování s prediktorem	32
3.2 Charakteristiky řeči v časové oblasti	35
3.2.1 Energetické charakteristiky	35
3.2.2 Počet průchodů nulou	36
3.2.3 Energie v pásmech	38
3.2.4 Autokorelační charakteristiky signálů	39
3.2.5 Vzájemná korelace vícekanálových signálů	40
3.3 Spektrální charakteristiky signálů	41
3.3.1 Spektrální analýza na bázi DFT	41
3.3.2 LPC spektrální analýza	43
3.3.3 Ilustrace krátkodobé analýzy řeči pomocí DFT a LPC	46
3.3.4 Dlouhodobá spektrální analýza	46
3.3.5 Kepstrální analýza, kepstrem	48
3.3.6 Koherenční charakteristiky vícekanálových signálů	49
3.3.7 Momentové charakteristiky	54
3.4 Euklidovská vzdálenost	56

4	Akustické vlastnosti řeči	59
4.1	Základní hlasivkový tón a jeho detekce	59
4.1.1	Předzpracování signálu	61
4.1.2	Detekce základní periody v časové oblasti	62
4.1.3	Detekce základní periody ve frekvenční oblasti	64
4.1.4	Postprocessing	65
4.2	Formanty	65
4.2.1	Možnosti výpočtu formantových frekvencí	67
4.2.2	Poznámky k implementaci algoritmů	70
4.2.3	Trasování formantů	72
5	Řeč za přítomnosti šumu	75
5.1	Typy šumů	75
5.1.1	Aditivní šумы	76
5.1.2	Konvoluční šумы	76
5.2	Měření úrovně aditivního šumu v řeči	77
5.2.1	Metody odhadu SNR řeči	79
5.3	Detekce řečové aktivity	83
5.3.1	Deterministické metody detekce řeči	84
5.3.2	Prahování kritériální funkce	87
5.3.3	Kepstrální detektor s adaptivním prahem	91
5.3.4	Stochastické detektory řečové aktivity	92
5.3.5	Vyhazení detekce řečové aktivity	93
6	Zvýrazňování degradované řeči	95
6.1	Základní principy metod redukce šumů	96
6.1.1	Metody pracující v časové oblasti	98
6.1.2	Transformační metody	99
6.2	Vybrané jednovstupové metody	104
6.3	Více vstupové metody	109
6.3.1	Kompenzační systém – redukce aditivních šumů a echa	112
6.3.2	System využívající koherenci	115
6.3.3	Mikrofonní pole s postfiltrací	115
6.4	Hodnocení kvality metod redukce šumu a echa	118
7	Parametrizace řeči	120
7.1	Standardní parametrizace	120
7.1.1	Preemfáze	121
7.1.2	Váhování signálu	121
7.1.3	Energie a logaritmická energie signálu	121

7.1.4	Lineární DFT a LPC kepstrum	122
7.1.5	Liftering – váhování kepstra	122
7.1.6	Statické, diferenciální a akcelerační parametry	122
7.1.7	Mel-kepstrum	124
7.1.8	PLP kepstrum	126
7.1.9	Maticová reprezentace výpočtu mel-kepstra	128
7.2	Robustní parametrizace	130
7.2.1	Předzpracování řeči s aditivním šumem	131
7.2.2	CMS – odečítání průměrného kepstra	133
7.2.3	RASTA filtrace	134
7.2.4	RASTA-PLP parametrizace	136
7.2.5	Návrh RASTA filtrů pomocí LDA	136
7.3	Speciální parametrizační techniky	138
7.3.1	TRAPs - metoda časových trajektorií	139
7.3.2	Odhad aposteriorních pravděpodobností hlásek pomocí MLP	140
8	Bayesovské metody klasifikace	144
8.1	Bayesův vzorec	144
8.1.1	Diskriminační funkce a klasifikace	146
8.1.2	Jednorozměrný případ klasifikace	146
8.1.3	Příklad jednoduché klasifikace	148
8.1.4	Vícerozměrná klasifikace a Mahalanobisova míra	152
8.1.5	Příklad klasifikace samohlásek	153
8.1.6	Příklad klasifikace frikativ	158
9	Rozpoznávání řeči	161
9.1	Přímé porovnávání, základní metoda DTW	161
9.2	Optimalizace metody DTW	163
9.3	Skryté Markovovy modely, HMM	165
9.3.1	Základní pravděpodobnostní popis shluků	167
9.3.2	Výpočet parametrů funkcí $b()$	169
9.3.3	Pravděpodobnostní popis s více složkami	170
9.3.4	Pravděpodobnostní popis s více proudy	171
9.3.5	Maticový zápis funkcí $b()$	172
9.3.6	Zrychlení výpočtů funkcí $b()$	174
9.3.7	Emitující a neemitující stavy	176
9.3.8	Skládání HMM do vyšších celků	178
9.3.9	Trénování HMM – možné průchody modelem	184
9.3.10	Baum-Welchova reestimace funkcí $b()$	190

9.3.11	Jednoduchý Viterbiho dekodér	192
9.3.12	Token passing dekodér	193
9.3.13	Optimalizace token passing dekodéru	194
9.3.14	Organizace dat v token passing dekodéru	194
9.4	Jazykové modely v HMM	195
9.4.1	Libovolná posloupnost slov ze slovníku	195
9.4.2	Pravděpodobnosti dvojic sousedních slov (bigramů)	197
9.4.3	Pravděpodobnosti dvojic tříd sousedních slov	198
9.4.4	N-gramy	200
9.5	Praktická konstrukce rozpoznávače s HTK	201
9.5.1	Kontextově závislé modely fonémů	202
9.5.2	Příprava dat	204
9.5.3	Volba deskriptorů	204
9.5.4	Volba struktury modelu	204
9.5.5	Globální rozložení jako počáteční stav	205
9.5.6	Vytvoření monofonů klonováním	206
9.5.7	První časové zarovnání modelů a dat	206
9.5.8	Modely ticha	206
9.5.9	Vyřazení chybných položek z databáze	206
9.5.10	Kontextová závislost fonémů	207
9.5.11	Více proudů, více složek	207
9.5.12	Reestimace modelů	208
9.6	Další techniky rozpoznávání	208
9.6.1	Neuronové sítě	208
9.6.2	Rozsáhlé konečné automaty	209
9.6.3	Logické systémy typu Prolog	210
10	Databáze hlasových signálů	211
10.1	Všeobecný přehled řečových databází	211
10.2	Fonetické elementy promluvy	212
10.2.1	Specifikace elementů a jejich inventář	212
10.2.2	Fonetická abeceda IPA	213
10.2.3	Fonetické abecedy pro automatické zpracování češtiny	213
10.3	Obsah databází pro rozpoznávání řeči	217
10.3.1	Foneticky bohatý materiál	217
10.3.2	Aplikačně orientované položky	220
10.4	Záznam řečových signálů	224
10.4.1	Pokrytí variability mluvčích	224
10.4.2	Požadované prostředí	226
10.4.3	Typy nahrávacích zařízení a kanálů	227

10.5	Anotace řečové databáze	230
10.5.1	Ortografická transkripce	230
10.5.2	Fonetická transkripce	230
10.5.3	Anotace neřečových událostí	232
10.5.4	Další informace o nahrávaném signálu	233
10.5.5	Softwarové prostředky pro anotaci	234
10.6	Významné databáze a zdroje	235
10.6.1	Dostupné databáze pro češtinu	235
10.6.2	Nejvýznamnější databáze pro angličtinu	237
10.6.3	AURORA - databáze pro srovnávací experimenty	238
11	Rozpoznávání mluvčího	239
11.1	Formulace úlohy	239
11.2	Digitalizace	240
11.3	Výběr a výpočet příznaků	240
11.4	Modely mluvčího	241
11.4.1	Vektorové kvantování	241
11.4.2	GMM	244
11.5	Problematika rozhodování	244
11.5.1	Obecné poznatky	244
11.5.2	Verifikace a identifikace	247
11.5.3	Systémy závislé na textu	247
11.6	Alternativní řešení	248
12	Syntéza řeči z textu	249
12.1	Formulace úlohy – TTS	249
12.2	Analýza textu	249
12.2.1	Transkripce	250
12.2.2	Fonetizace	250
12.2.3	Prozodie	251
12.3	Syntéza	253
12.3.1	Data pro syntézu	253
12.4	Resyntéza a změna identity mluvčího	258
12.5	Vývoj syntézy řeči	258
Literatura		260
Rejstřík		272