

# Obsah

Předmluva .....	15
<b>1. Úvod .....</b>	<b>17</b>
1.1 Řečové technologie v praxi .....	17
1.2 Stručný pohled do historie řečových technologií .....	18
1.3 Poznámky k obsahu knihy .....	20
<b>2. Vytváření a vlastnosti mluvené řeči .....</b>	<b>21</b>
2.1 Úvodní poznámky .....	21
2.2 Proces vytváření řeči .....	22
2.2.1 Proces vytváření řeči člověkem .....	22
2.2.1.1 Dechové ústrojí .....	23
2.2.1.2 Hlasové ústrojí .....	24
2.2.1.3 Artikulační ústrojí .....	26
2.2.2 Modelování procesu vytváření řeči .....	28
2.3 Reprezentace řeči .....	32
2.3.1 Akustická úroveň .....	33
2.3.2 Fonetická úroveň .....	36
2.3.3 Fonologická úroveň .....	37
2.3.4 Řečové jednotky a jejich vlastnosti .....	37
2.3.5 Fonetické inventáře .....	38
2.3.5.1 Mezinárodní fonetická abeceda IPA .....	39
2.3.5.2 SAMPA .....	40
2.3.5.3 Ostatní abecedy .....	41
2.3.5.4 Fonetické abecedy pro češtinu .....	41
2.4 České řečové jednotky .....	44
2.4.1 Samohlásky .....	45
2.4.1.1 Artikulační vlastnosti samohlásek .....	47
2.4.1.2 Akustické vlastnosti samohlásek .....	48
2.4.2 Souhlásky .....	50
2.4.2.1 Artikulační vlastnosti souhlásek .....	50
2.4.2.2 Akustické vlastnosti souhlásek .....	54
2.4.3 Slabiky .....	59
2.4.4 Základní řečové jednotky a jejich použití v oblastech počítačového zpracování řeči .....	62
2.5 Prozodické vlastnosti řeči .....	63
2.5.1 Reprezentace prozodických jevů .....	64
2.5.2 Časování .....	66
2.5.2.1 Členění souvislé řeči .....	66
2.5.2.2 Tempo .....	66
2.5.2.3 Pauzy .....	67
2.5.2.4 Rytmus a přízvuk .....	68
2.5.3 Intonace .....	72

2.5.3.1	Kadence větých úseků koncových	74
2.5.3.2	Kadence větých úseků nekonečných	74
2.5.4	Intenzita	75
2.5.5	Mikroprozodie	75
2.6	Varianty řeči	75
2.7	Informační obsah řeči	76
2.7.1	Informační obsah fonetické formy	77
2.7.2	Informační obsah akustické formy	79
2.8	Fonetická transkripce češtiny	80
2.8.1	Automatická fonetická transkripce	82
2.8.2	Symbole používané v pravidlech fonetické transkripce	84
2.8.3	Základní pravidla	86
2.8.4	Spojení samohlásek	86
2.8.5	Spojení souhlásky a samohlásky	87
2.8.6	Výslovnost samohlásek v okolí pauzy	88
2.8.7	Výslovnost souhláskových skupin	88
2.8.7.1	Asimilace znělosti	89
2.8.7.2	Asimilace artikulační	92
2.8.7.3	Slabikotvorné souhlásky	93
2.8.7.4	Zjednodušená výslovnost některých souhláskových skupin	94
2.8.7.5	Výslovnost dvou foneticky stejných souhlásek	95
2.8.8	Pravidla nespisovné výslovnosti	96
2.8.9	Výslovnost slov přejatých	102
2.8.9.1	Samohlásky	103
2.8.9.2	Souhlásky	104
2.9	Závěrečné poznámky	105
	Literatura	105
<b>3.</b>	<b>Slyšení a vnímání zvuku člověkem</b>	<b>108</b>
3.1	Struktura a funkce lidského ucha	108
3.2	Vnímání zvuku	111
	Literatura	122
<b>4.</b>	<b>Analýza řečového signálu</b>	<b>124</b>
4.1	Úvodní poznámky	124
4.2	Kódování tvaru vlny	125
4.2.1	Pulsní kódová modulace (PCM)	125
4.2.2	Další metody kódování tvaru vlny	128
4.3	Zpracování v časové oblasti	130
4.3.1	Krátkodobá energie	133
4.3.2	Krátkodobá funkce středního počtu průchodů signálu nulou	134
4.3.3	Krátkodobá autokorelační funkce	135
4.4	Zpracování ve frekvenční oblasti	135
4.4.1	Krátkodobá Fourierova transformace	136
4.4.2	Krátkodobá diskretní Fourierova transformace	138
4.4.3	Pásmová filtrace	141
4.5	Lineární prediktivní analýza	146
4.5.1	Autokorelační metoda výpočtu koeficientů lineární predikce	146
4.5.2	Perceptivní lineární prediktivní analýza	150
4.6	Homomorfní zpracování řeči	156
4.6.1	Základní vlastnosti homomorfních systémů	156
4.6.2	Kepstrální analýza mluvené řeči	159
4.6.3	Kepstrální koeficienty LPC	161
4.6.4	Melovské kepstrální koeficienty	162



4.6.5	Dynamické koeficienty	166
4.7	Lokální míry zkreslení	166
4.8	Určování hraničních bodů promluvy	172
4.9	Fonetická analýza	174
4.9.1	Určování základních fonetických charakteristik	174
4.9.2	Měření základního hlasivkového tónu	175
4.9.2.1	Určování základního tónu pomocí ACF	175
4.9.2.2	Využití zbytkové chyby lineární predikce	177
4.9.3	Určování formantových frekvencí	179
4.9.3.1	Odhad formantových frekvencí ze spektra LPC	180
4.10	Vektorová kvantizace	183
4.10.1	Úvodní poznámky k vektorové kvantizaci	183
4.10.2	Formulace problému	183
4.10.3	Návrh kódové knihy	184
4.10.4	Kvantizér s binárním hierarchickým dělením	186
4.10.5	Praktické použití vektorové kvantizace	190
	Literatura	192
<b>5.</b>	<b>Statistické metody rozpoznávání řeči</b>	<b>195</b>
5.1	Úvodní poznámky	195
5.2	Statistický přístup k rozpoznávání souvislé řeči	197
5.3	Akustické modelování	199
5.3.1	Struktura a parametry skrytého Markovova modelu	200
5.3.1.1	Pravděpodobnosti přechodu a výstupní pravděpodobnosti	201
5.3.1.2	Struktura skrytých Markovových modelů	203
5.3.2	Výpočet pravděpodobnosti generování promluvy modelem	207
5.3.3	Trénování parametrů skrytého Markovova modelu	211
5.3.4	Slučování parametrů a trénování s uvážením fonetického kontextu	219
5.3.4.1	Trifony se sloučenými stavy	219
5.3.4.2	Shlukování řízené daty	221
5.3.4.3	Shlukování založené na fonetických rozhodovacích stromech	222
5.4	Jazykové modelování	225
5.4.1	Stochastický jazykový model	227
5.4.2	Stochastické $n$ -gramové modely	227
5.4.3	Posouzení kvality stochastického jazykového modelu	228
5.4.4	Definice značení	231
5.4.5	Odhady pravděpodobností $n$ -gramových modelů metodou maximální věrohodnosti	232
5.4.6	Goodův–Turingův odhad	236
5.4.7	Metoda odhadu s postupným vynecháváním jednoho jevu	237
5.4.8	Ústupové (backing-off) a interpolační schéma vyhlazování	239
5.4.9	Katzův diskontní model	241
5.4.10	Model s absolutním diskontem	246
5.4.11	Model s lineárním diskontem	249
5.4.12	Wittenův–Bellův model	250
5.4.13	Model Kneserův–Neyův	251
5.4.14	Lineární interpolace	253
5.4.15	Kombinace jazykových znalostních zdrojů metodou maximální entropie	256
5.4.16	Další varianty stochastických jazykových modelů	260
5.4.16.1	Modely založené na třídách slov	261
5.4.16.2	Modely založené na spouštěčích	264
5.4.16.3	Modely s krátkodobou pamětí	265
5.4.16.4	Adaptivní jazykové modely	266

5.4.16.5	Jazykové modely založené na strukturálních vlastnostech jazyka	266
5.4.17	Jazykové modelování češtiny	267
Literatura		270
<b>6.</b>	<b>Dekódovací techniky</b>	<b>274</b>
6.1	Definice pojmu dekodování	274
6.2	Výpočetní problém dekodování	275
6.3	Dekodování s akustickým a jazykovým modelem	275
6.3.1	Základní pojmy z teorie grafů	275
6.3.2	Dekodování při dané struktuře akustického a jazykového modelu	278
6.3.3	Definice úlohy dekodování podle kritéria MAP	281
6.3.4	Definice úlohy dekodování podle Viterbiova kritéria	282
6.4	Formulace problému dekodování jako hledání ve stavovém prostoru a jeho řešení	282
6.4.1	Pojem stavového prostoru	282
6.4.1.1	Prohledávání stavového prostoru	283
Seznamy OPEN, CLOSED		285
Prohledávání stavového prostoru do šířky		285
Prohledávání stavového prostoru do hloubky		287
Heuristické prohledávání stavového prostoru		288
Prohledávání stavového prostoru typu BEAM		290
6.4.2	Formulace úlohy dekodování ve stavovém prostoru na úrovni slov	291
6.4.2.1	Ohodnocení vrcholů a hran	293
6.4.3	Formulace úlohy dekodování ve stavovém prostoru na úrovni slov a časového kroku	295
6.4.3.1	Formulace úlohy dekodování s obecným akustickým a jazykovým modelem	295
6.4.3.2	Formulace úlohy dekodování s obecným jazykovým modelem a akustickým modelem založeným na konečném automatu	298
6.4.3.3	Formulace úlohy dekodování s akustickým a jazykovým modelem založeným na konečném automatu	299
6.4.4	Formulace úlohy dekodování ve stavovém prostoru na úrovni stavů akustického modelu slov a časového kroku	300
6.5	Přehled přístupů k řešení úlohy dekodování	301
6.6	Prohledávání stavového prostoru s využitím zásobníkového dekodéru	305
6.6.1	Jednozásobníkový dekodér s heuristickým prohledáváním typu první nejlepší	305
6.6.1.1	Stavový prostor na úrovni slov	306
6.6.1.2	Stavový prostor na úrovni stavů akustického modelu slov a časového kroku	307
6.6.1.3	Dekodování podle kritéria VitMAP	308
6.6.1.4	Dekodování podle kritéria MAP	311
6.6.1.5	Metody urychlení procesu dekodování	314
Aproximace akustického modelu slova		314
Použití lexikálních stromů		317
Prořezávání		319
Prořezávání deaktivací fonémů		320
Pouze jedno vyhodnocení shodných posloupností fonémů v lexikálním stromu		321
Uplatnění jazykového modelu		321
6.6.2	Jednozásobníkový dekodér s obálkovým hledáním	323
6.6.3	Vícezásobníkový start-synchronní dekodér	327
6.6.4	Zhodnocení metod dekodování se zásobníkovým dekodérem	330
6.7	Časově synchronní prohledávání stavového prostoru	330
6.7.1	Viterbiův algoritmus	331
6.7.2	Viterbiův algoritmus pro slovníky s lineární strukturou sítě	335



6.7.3	Metody urychlení procesu dekodování	338
6.7.3.1	Profezávání stromu řešení	339
6.7.3.2	Použití lexikálních stromů	341
6.7.3.3	Časné uplatnění zdrojů znalostí	343
6.7.4	Modelování kontextu	346
6.7.4.1	Vnitroslovní a mezoslovní kontext	346
6.7.4.2	Vliv spodoby znělosti na kontext při modelování promluv v češtině	352
6.7.5	Typy lexikálních stromů a jejich implementace	353
6.7.5.1	Kopie lexikálního stromu podmíněné historií slov	353
6.7.5.2	Reprezentace statickými stromy	356
6.7.5.3	Reentrantní strom se zásobníkem v každém stavu	357
6.7.5.4	Start-synchronní stromy	358
6.7.6	Algoritmus cestování žetonů	361
6.8	Stavově synchronní prohledávání	366
6.8.1	Stavově synchronní prohledávání na úrovni akustického modelu	366
6.8.2	Stavově synchronní prohledávání na úrovni posloupnosti slov	371
6.9	Transducery	373
6.9.1	Polookruh	373
6.9.2	Ohodnocené akceptory a transducery	374
6.9.3	Operace ohodnocených transducerů	376
6.9.4	Ohodnocené transducery při dekodování řeči	380
6.10	Nalezení $N$ nejlepších hypotéz	381
6.10.1	$N$ nejlepších hypotéz vět vyjádřených seznamem	382
6.10.2	$N$ nejlepších hypotéz vyjádřených slovní mřížkou	385
6.10.3	$N$ nejlepších hypotéz při aproximaci slovním párem	389
6.10.4	$N$ nejlepších hypotéz vyjádřených mřížkou při rozšířené rozpoznávací síti	390
6.11	Závěrečné poznámky	394
	Literatura	395
	<b>7. Zvýšení robustnosti systémů rozpoznávání řeči</b>	<b>400</b>
7.1	Úvodní poznámky	400
7.2	Adaptace na hlas řečníka	402
7.2.1	Normalizace délky hlasového traktu	404
7.2.1.1	Warpovací funkce a warpovací faktor	405
7.2.1.2	Metody odhadu warpovacího faktoru	407
7.2.1.3	Rozpoznávání řeči s doprovodnou normalizací délky hlasového traktu	409
7.2.2	Adaptace akustického modelu	410
7.2.2.1	Adaptace založená na MAP odhadu parametrů HMM	410
	Určení hodnot hyperparametrů	415
	Výběr počátečních parametrů HMM	418
7.2.2.2	Odhad parametrů metodou MLLR	419
7.3	Metody kompenzace vlivu prostředí a přenosového kanálu	423
7.3.1	Kompenzace šumu na signálové či příznakové úrovni	424
7.3.1.1	Technika spektrálního odčítání	425
7.3.1.2	Odčítání střední hodnoty kepstra	428
7.3.1.3	Metoda RASTA a její modifikace	430
7.3.2	Kompenzace šumu na úrovni akustických modelů	433
7.3.2.1	Kombinace paralelního modelu	433
7.4	Výběr informativních příznaků	438
7.4.1	Využití časových změn krátkodobého spektra řeči	440
7.4.1.1	Dynamické koeficienty	441
7.4.1.2	TempoRAI PatternS (TRAPS)	441
7.4.2	Dekorelace příznaků	442
7.4.2.1	Diskrétní kosinová transformace	442

7.4.2.2	Normalizace obrazového prostoru	443
7.4.3	Výběr a uspořádání informativních příznaků	444
7.4.3.1	Metoda hlavních komponent (PCA)	445
7.4.3.2	Lineární diskriminační analýza (LDA)	448
7.5	Míry důvěry pro určení spolehlivosti rozpoznávání	449
	Literatura	455
<b>8.</b>	<b>Řečové korpusy</b>	<b>459</b>
8.1	Druhy řečových korpusů	459
8.2	Proces vytváření řečových korpusů	461
8.2.1	Přípravná fáze	461
8.2.1.1	Specifikace řečového obsahu	461
8.2.1.2	Úprava textů pro čtení	462
8.2.1.3	Algoritmus výběru foneticky bohatých a foneticky vyvážených vět	466
	Výpočet skóre pro výběr foneticky bohatých vět	467
	Výpočet skóre pro výběr foneticky vyvážených vět	470
	Poznámky k výběru vět	471
8.2.1.4	Specifikace typu a počtu řečníků	471
8.2.1.5	Specifikace nahrávacích podmínek	473
	Nahrávací prostředí	473
	Způsob nahrávání	473
	Typy signálů	474
8.2.2	Záznamová fáze	474
8.2.3	Dokončovací fáze	477
8.3	Přehled řečových korpusů	484
	Literatura	486
<b>9.</b>	<b>Rozpoznávání řečníka</b>	<b>489</b>
9.1	Úlohy rozpoznávání řečníka	489
9.2	Pracovní režimy systémů rozpoznávání řečníka	492
9.3	Typy systémů rozpoznávání řečníka	493
9.4	Ohodnocení činnosti systémů rozpoznávání řečníka	494
9.4.1	Ohodnocení činnosti systémů verifikace řečníka	495
9.4.2	Ohodnocení činnosti systémů identifikace řečníka v uzavřené množině	498
9.4.3	Ohodnocení činnosti systémů identifikace řečníka v otevřené množině	498
9.4.4	Volba verifikačního prahu	500
9.5	Charakteristiky řeči používané při rozpoznávání řečníka	501
9.6	Základní přístupy k rozpoznávání řečníka	505
9.6.1	Metody využívající vzorové reprezentace	506
9.6.1.1	Rozpoznávání na základě časových funkcí příznakových vektorů	507
9.6.1.2	Rozpoznávání s využitím vektorové kvantizace	513
9.6.2	Metody využívající pravděpodobnostních modelů	514
9.6.2.1	Rozpoznávání řečníka na základě směsi Gaussových hustotních funkcí	521
9.6.2.2	Rozpoznávání řečníka s využitím skrytých Markovových modelů	523
9.7	Závěrečné poznámky	524
	Literatura	526
<b>10.</b>	<b>Syntéza řeči</b>	<b>529</b>
10.1	Základní přístupy	530
10.1.1	Teorie zdroje a filtru	531
10.2	Historie syntézy řeči	532
10.2.1	Mechanické syntetizéry	533
10.2.2	Elektronické syntetizéry	533
10.2.3	Digitální syntetizéry	535
10.2.4	Historie české syntézy řeči	536



10.3 Formantová syntéza	537
10.3.1 Pravidla	538
10.3.2 Vytváření řeči formantovým syntetizérem	539
10.3.3 Klattův formantový syntetizér	543
10.3.4 Shrnutí	546
10.4 Konkatenáční syntéza	547
10.4.1 Vlastnosti konkatenáční syntézy řeči	548
10.4.2 Princip konkatenáční syntézy	550
10.4.3 Volba řečových jednotek	553
10.4.4 Inventář řečových jednotek	558
10.4.4.1 Řečový korpus	558
10.4.4.2 Segmentace	560
Automatická segmentace s využitím skrytých Markovových modelů	560
Automatická segmentace porovnáváním se syntetickou řečí s využitím techniky DTW	563
10.4.5 Korpusově orientovaná syntéza	563
10.4.5.1 Kontextově orientované shlukování	564
10.4.5.2 Shlukování HMM založené na binárních fonetických rozhodovacích stromech	565
10.4.5.3 Syntéza výběrem jednotek	568
10.4.6 Metody prozodických a spektrálních modifikací	574
10.4.6.1 Prostá konkatenace	575
10.4.6.2 Lineární prediktivní syntéza	576
Koeficienty filtru $1/A_p(z)$	576
Koeficienty PARCOR	577
Páry čárového spektra LSP	578
Syntéza řeči	579
Kvalita syntetické řeči	580
Vylepšení LP syntézy	581
10.4.6.3 PSOLA	582
Obecné schéma metody PSOLA	583
Hlasivkové pulsy	586
Určení hlasivkových pulsů v syntetické řeči	589
TD-PSOLA	593
FD-PSOLA	598
LP-PSOLA	599
MBROLA	599
10.4.6.4 Harmonický a šumový model vytváření řeči	601
Princip HNM	601
Analýza	603
Syntéza	607
Kvalita syntetické řeči	610
10.4.7 Výhody a nevýhody konkatenáční syntézy	611
10.5 Artikulační syntéza	613
10.6 Syntéza řeči z textu	614
10.6.1 Zpracování přirozeného jazyka	615
10.6.1.1 Morfologicko-syntaktická analýza	616
10.6.1.2 Fonetická transkripce	618
10.6.2 Generování prozodie	619
10.6.2.1 Generování intonace	619
Fujisakiho model	622
Intonační model Tilt	624

10.6.2.2	Generování intenzity	626
10.6.2.3	Generování časování	626
10.6.3	Hodnocení kvality syntetické řeči	627
10.6.3.1	Testy srozumitelnosti	627
10.6.3.2	Testy přirozenosti	630
10.6.4	Aplikace syntézy řeči	631
10.6.5	Příklady řečových syntetizérů	633
10.6.5.1	Zahraniční systémy	633
10.6.5.2	České systémy	634
Literatura		635
<b>11. Hlasové dialogové systémy</b>		<b>644</b>
11.1	Přístupy k návrhu hlasového dialogového systému	664
11.2	Lidský faktor v hlasovém dialogu s počítačem	645
11.3	Obecná architektura hlasového dialogového systému	646
11.3.1	Rozpoznávání řeči	647
11.3.1.1	Problémy se spontánní řečí	647
11.3.1.2	Vliv prozodie spontánní řeči	649
11.3.1.3	Využití detekce klíčových slov v hlasovém dialogovém systému	650
11.3.2	Porozumění mluvenému jazyku	651
11.3.2.1	Formální syntaxe	652
11.3.2.2	Reprezentace znalostí	658
Predikátový počet prvního řádu		659
Sémantické sítě		661
Sémantické rámce		662
Scénáře		663
11.3.2.3	Porozumění rozpoznáním promluvám	664
11.3.3	Generování odezvy	668
11.4	Řízení dialogu	669
11.4.1	Vedení dialogu	670
11.4.1.1	Modelování zdrojů znalostí pro vedení dialogu	670
Model prostředí		670
Model interakce systému		670
Model uživatele		671
Model dialogu		671
11.4.1.2	Ověřování správnosti interpretace promluvy uživatele	673
11.4.2	Strategie řízení dialogu	674
11.4.2.1	Dialog s iniciativou systému	674
11.4.2.2	Dialog se smíšenou iniciativou	675
11.4.2.3	Dialog s iniciativou uživatele	676
11.4.3	Typy dialogových systémů	676
11.4.3.1	Dialogový systém s konečným počtem stavů	676
11.4.3.2	Dialogový systém využívající strukturu rámců	678
11.4.3.3	Dialogový systém založený na agentech	681
11.5	Statistický přístup k návrhu a řízení hlasového dialogového systému	683
11.5.1	Základní popis dialogu se statistickým přístupem	683
11.5.2	Porozumění promluvám	685
11.5.2.1	Dekódování sémantických konceptů	688
11.5.2.2	Detekce dialogových akcí	689
11.5.2.3	Řízení dialogu	690
11.5.2.4	Generování odpovědi	691
11.6	Závěrečné poznámky	692
Literatura		693
<b>12. Závěr</b>		<b>697</b>



<b>Dodatky</b> .....	699
A. Fonetické abecedy .....	699
A.1 SAMPA .....	699
A.1.1 SAMPROSA .....	699
A.1.2 Čeština .....	700
A.1.3 Britská angličtina .....	701
A.1.4 Americká angličtina .....	702
A.1.5 Francouzština .....	703
A.1.6 Němčina .....	704
A.1.7 Ruština .....	705
A.1.8 Polština .....	706
A.1.9 Slovenština .....	707
A.2 Mezinárodní fonetická abeceda IPA .....	708
B. Formální popis prozodie .....	709
B.1 Obsah a komunikační funkce .....	709
B.2 Prozodická struktura .....	710
B.2.1 Definice pojmů .....	710
B.2.2 Prozodická frázová gramatika (PFG) .....	712
B.3 Prozodická forma .....	716
B.3.1 Systém syntézy prozodie z textu (TTP) .....	716
B.3.2 Emergující prozodická forma .....	717
C. Diskrétní Fourierova transformace (DFT) .....	718
C.1 Základní vlastnosti DFT .....	718
C.2 Rychlá Fourierova transformace (FFT) .....	720
C.3 Algoritmus výpočtu koeficientů DFT (IDFT) rychlou Fourierovou transformací (metoda decimování ve frekvenci) .....	722
D. Tabulka hodnot modifikované Besselovy funkce 1. druhu 0. řádu .....	724
E. Trénování parametrů skrytých Markovových modelů vložených jednotek .....	725
F. Trénování parametrů skrytých Markovových modelů izolovaných jednotek .....	727
G. Gramatiky .....	732
G.1 Deterministické gramatiky .....	732
G.2 Stochastické gramatiky .....	733
Literatura .....	738
Rejstřík .....	739