Pokročilé metody rozpoznávání řeči

Vyučující: Jan Nouza, ITE

Cíl předmětu: Seznámit se s modernějšími metodami rozpoznávání řeči – s využitím modulové stavebnice HTK (Hidden Model Markov Toolkit) a prakticky si vyzkoušet celý návrh systému pro rozpoznávání řeči

Forma výuky:
a) Přednášky
b) cvičení – formou domácích úloh
c) malý závěrečný projekt

Předchozí znalosti: Nutnost předchozího absolvování PZR

Literatura:

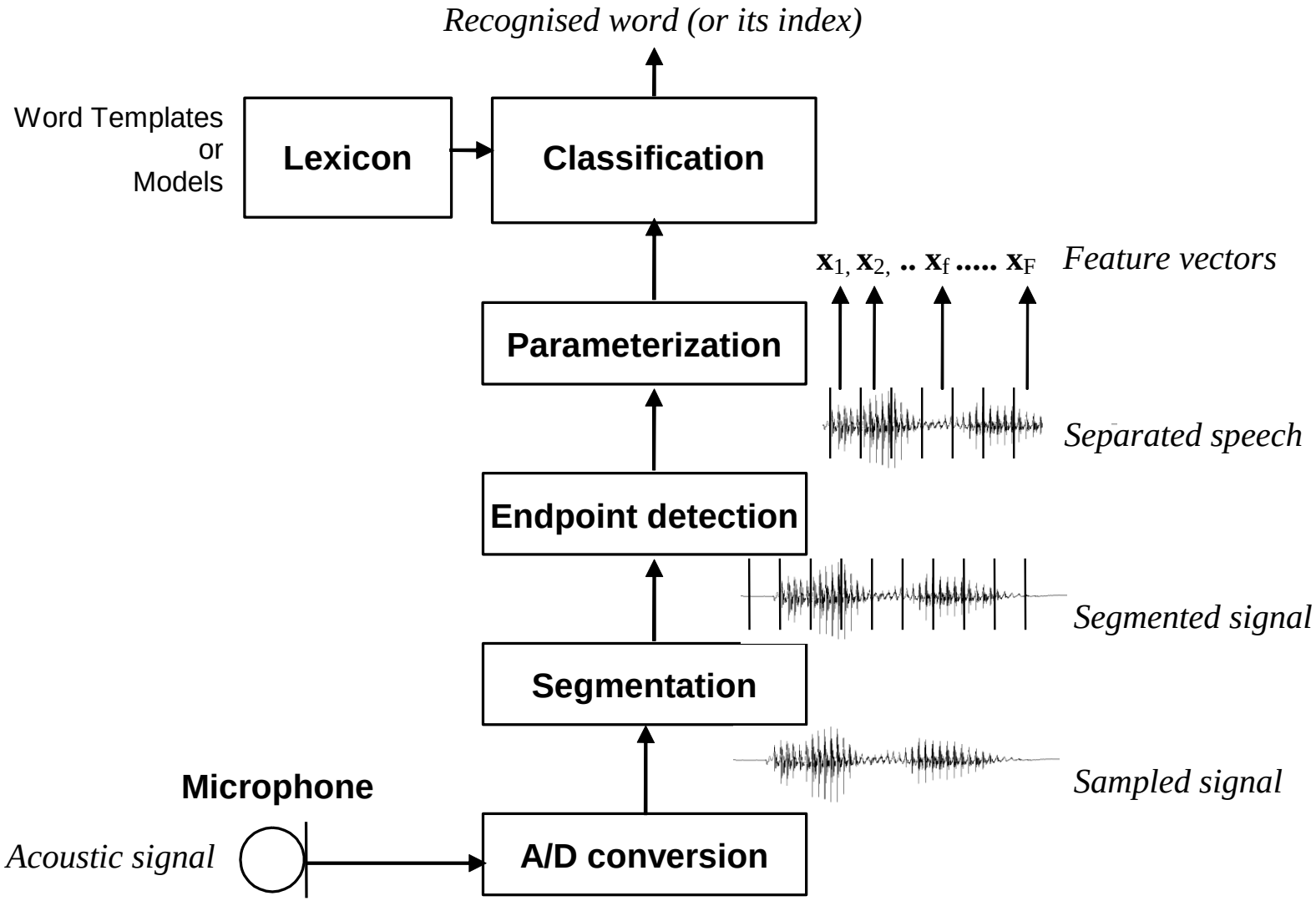
I. Nouza J. (editor): Počítačové zpracování řeči. TUL Liberec 2009.
II. Psutka J., Müller L., Matoušek J., Radová V.: Mluvíme s počítačem česky. Academia Praha, 2006
III. HTK book http://htk.eng.cam.ac.uk/
IV. Přednášky a materiály k předmětu budou na stránce eLearningu

Obsah předmětu

- Problémy a výzvy moderního počítačového rozpoznávání řeči.
- Fonémově orientované rozpoznávání. Nástroj G2P.
- Principy a metody parametrizace řečového signálu, kepstrum a kepstrální příznaky
- Skryté markovské modely (HMM), nástroj HTK, trénování a testování.
- Trénování fonémových modelů, tvorba trénovací databáze.
- Využití HMM pro rozpoznávání slov a sekvence slov. Viterbiho dekodér.
- Gramatiky. Pravděpodobnostní jazykový model pro rozpoznávání řeči a jeho tvorba.
- Metody dalšího zlepšování úspěšnosti systémů rozpoznávání řeči.
- Neuronové sítě, nástroj Kaldi. E2E systémy.
- Základy rozpoznávání mluvčího

11.-14. týden - Práce na samostatném projektu.

Předchozí znalosti – IWR



Metoda DTW

Princip:

- **framy rozpoznávaného slova jsou přiřazeny framům reference tak, aby vyšla co nejmenší celková vzdálenost**

Výhody DTW:

- **snadná příprava systému** (stačí pouze nahrát referenci pro každé slovo)
- **snadná implementace** i na nevýkonném HW (např. embedded systems)

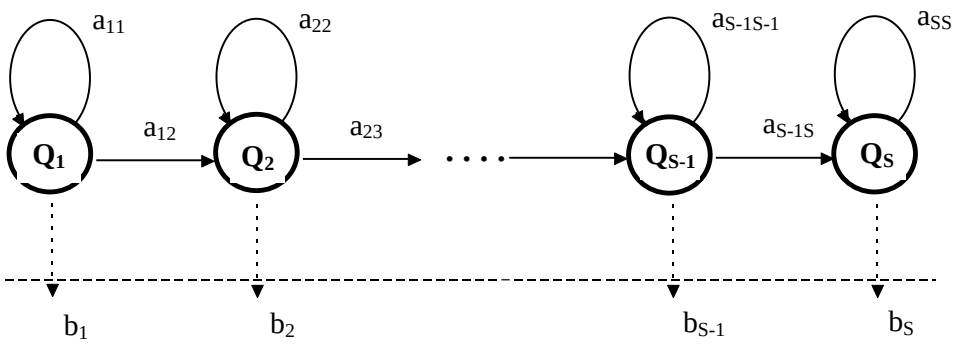
Nevýhody:

- reference jsou **závislé na mluvčím (SD), SI system** vyžaduje **reference od mnoha osob**
- **každé slovo ve slovníku musí být předem nahráno** (nelze přidávat nová slova pouze tím, že bychom je napsali a vložili do seznamu)
- SI systém **s velkým slovníkem** je zákonitě **příliš pomalý** (smyčky pro každé slovo, pro každou jeho referenci, pro každý frame)

Metoda HMM

Princip:

- **každé slovo je reprezentováno jedním stavovým modelem natrénovaným z mnoha nahrávek tohoto slova,**
- **parametry modelu jsou vyjádřeny pravděpodob. rozloženími,**
- **při rozpoznávání jsou framy slova přiřazeny stavům modelu tak, aby vyšla co největší celková pravděpodobnost (Viterbiho algoritmus)**
- **levo-pravý model**



Q_s ... **stavy** (šipky naznačují možné přechody mezi nimi)

a_{ij} .... **přechodová pravděpodobnost** – pravd., že (v aktuálním framu) model přejde ze stavu *i* do stavu *j*

$b_s(\mathbf{x})$...**výstupní pravděpodobnostní rozložení** – funkce určující pravděpodobnost, že příznakový vektor **x** **patří ke stavu *s***

Univerzálnost metod DTW a HMM

Metody nacházejí uplatnění i v jiných oborech než rozpoznávání řeči.
Hodí se tam, kde porovnáváme různě dlouhé sekvence nebo řetězce.

Např.

- Varianty DTW se využívají při měření podobnosti psaných slov (třeba nabízení podobných slov při překlepech). Metodě se říká MED (Minimum Edit Distance), nebo Levenshteinova vzdálenost
- Při měření podobnosti vět či textů (odhalování plagiátů, vyhodnocování úspěšnosti systémů rozpoznávání spojité řeči, apod.)
- Porovnávání údajů v databázích s údaji o různé délce
- Porovnávání biologických markerů a sekvencí (např. DNA).

Hláskové HMM – cesta k neomezenému slovníku

Až dosud probírané metody vyžadovaly **nahrání každého slova,** které bylo ve slovníku a **natrénování jeho modelu**. Pro velké slovníky by toto nebylo možné.

Většina současných systémů rozpoznávání řeči používá HMM, které ovšem **modelují jednotky menší než slova.** Jejich počet je omezený a malý.

Nejvhodnějšími jednotkami jsou **fonémy (hlásky).** V mluvené řeči hrají podobnou roli jako písmena v psaném jazyce. (Na rozdíl od grafémů jsou však závislé na řečníkovi a na kontextu.)

Fonémy trvají krátce, lze je modelovat malým počtem stavů – obvykle 3

Fonémy jsou **jazykově závislé**. Ve většině evropských jazyků se jejich počet pohybuje v rozmezí **25 – 65**.

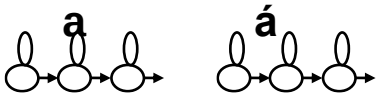
Rozpoznávání s fonémovými HMM

1. Pro **daný jazyk** je třeba definovat **fonémovou sadu**.
2. Pro **každé slovo** ve slovníku musíme určit jeho **fonetický přepis** (sekvenci fonémů).
3. Je nutné nahrát **velké množství záznamů řeči** a pro každou nahrávku vytvořit její **fonetickou transkripci**.
   (Úspěšný systém vyžaduje několik desítek hodin, stovky mluvčích).
4. Je třeba **natrénovat HMM pro všechny fonémy**.
5. Před rozpoznáváním **se sestaví slovní HMM z hláskových modelů**.
6. Pro vlastní rozpoznávání se použije **standardní Viterbiho algoritmus.**

Slovní modely z hláskových HMM

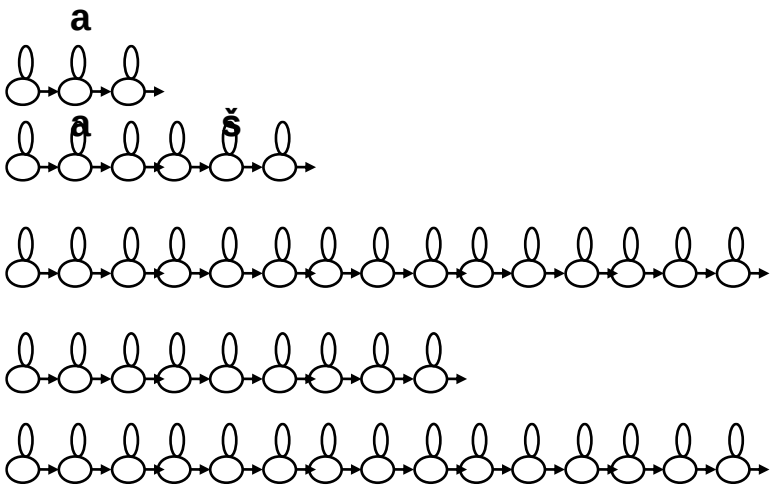
Inventář fonémů
a, á, b, …X … z, ž

Akustické modely



Slovník
a
až
…..
robot
…..
sud
Zürich

Výslovnost
a
aš
…..
robot
…..
sut
ciriX



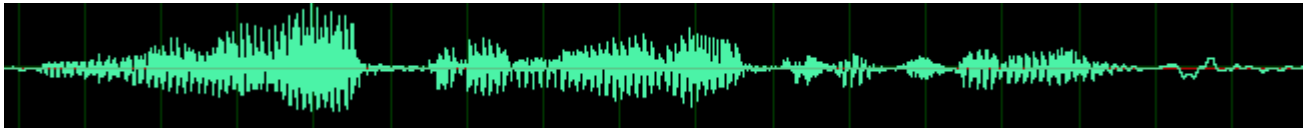
Od izolovaných slov ke spojité řeči

Přirozená plynulá řeč je mnohem složitější pro rozpoznávání

1) V běžné řeči neexistují “oddělovače” (pauzy) mezi slovy. **Jedno slovo navazuje na druhé**. Pauzy vznikají většinou jen při nádechu.
2) V plynulé řeči je **výslovnost slov významně ovlivňována kontextem**. Př. „zavolejnačíslopječesedum“
3) Záznam běžné řeči obsahuje i **„neřečové zvuky“,** např. ticho, nádech a výdech, zvuky rtů, kašel, smích, „ehm“
4) V mnoha situacích má řeč **nespisovný a spontánní charakter,** např. „Ano, ano máte pravdu, ale, ale, když vo tom tak pře přemejšlím …..”
5) **Rychlost řeči** může být velmi různá, závisí na mluvčím, situaci, atd.
6) Pokud nahrávka řeči nevzniká ve studiu, kvalita a srozumitelnost řeči může být dále výrazně ovlivněna **hlukem prostředí**.

Základní problémy spojité řeči

Ukázka záznamu řeči:



Co bylo řečeno a v jakých časových okamžicích?

| - | zavolá | - | premiéra | předseda | - |

| - | zavolá | - | premiér | a | předseda | - |

| - | za | vola | premiéra | předsedá | - |

|-| zvolá | prima | éra | před | sebou | - |

| - | zavolej | premiéra | předsedo | - |

|- | z | vola | prima | tévé | přesedá | - |

Při rozpoznávání spojité řeči není apriori známo:

- Kolik slov bylo řečeno?
- Jaká sekvence slov byla řečena?
- Byla všechna vyřčená slova ze slovníku?
- Byla to skutečně jenom řeč nebo i další zvuky a hluky?
- V jakých časových okamžicích začínala jednotlivá slova?

Řešení úlohy spojité řeči

Principy

1) Sestavit co nejreprezentativnější **slovník** pro danou aplikační oblast
2) Pro každé slovo mít fonetický přepis (sekvenci hlásek),
3) Natrénovat pro všechny hlásky (i hluky) **akustický model** (HMM)
4) Vzít v úvahu statistiky výskytu slov (a jejich kombinací) v dané aplikační oblasti a vytvořit tzv. **jazykový model** (LM – language model)
5) Navrhnout strategii, která pro daný signál najde **nejpravděpodobnější sekvenci slov** s využitím kombinace akustického i jazykového modelu.

Řešíme úlohu nalézt **sekvenci slov** w_1, w_2, …. w_N takovou, že je ze všech možných kombinací slov (a neřečových zvuků) nejpravděpodobnější z hlediska akustického i jazykového.

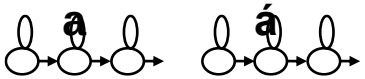
Neznámé je: číslo N, sekvence slov a zvuků, a časy jejich začátků.

Základní koncepce rozpoznávání spoj. řeči
s využitím slovníku, akustického a jazykového modelu

Fonetický inventář
a, á, b, c, č, X, …… z, ž,

Inventář hluků
ticho, nádech, klik, …

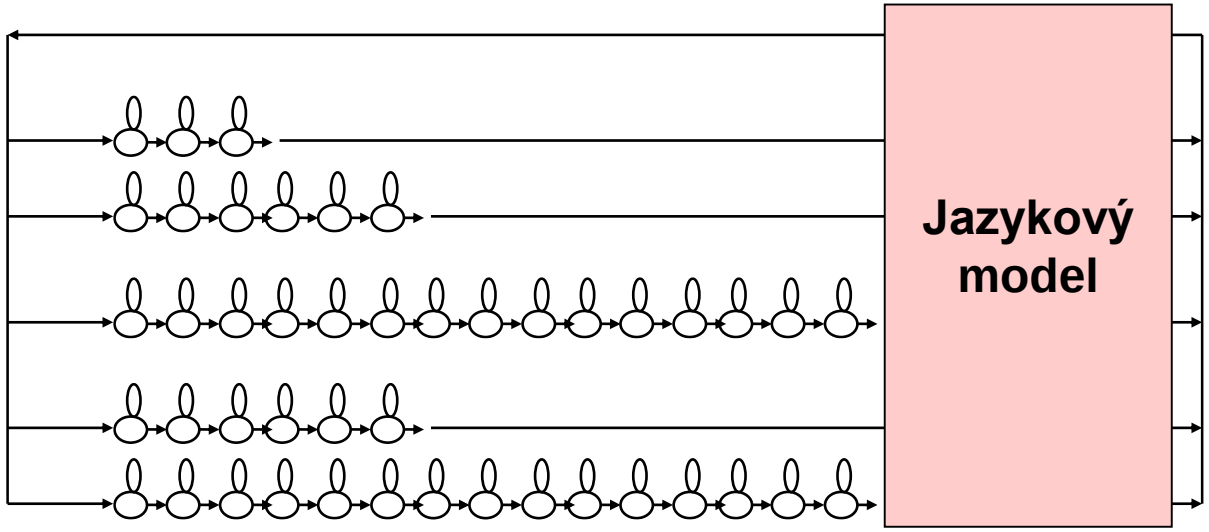
Akustické modely



Lexikon
"ticho"
ať
…..
robot
…
už
…..
Zürich

Výslovnost
-
ať
…..
robot
…..
uš

ciriX



Pozn.: Síť všech modelů slov je nyní "zacyklena", tj. z posledního stavu každého slova vede cesta na počáteční stavy všech slov, a to se započítáním LM.

Základy fonetiky

Jednotky mluvené řeči

Hláska (foném) – základní *stavební* jednotka řeči

- výhody: malý počet, snadný převod z textu na výslovnost,

- nevýhody: stejná hláska má velmi různou akustickou podobu podle kontextu (příklad „abrakadabra“ – 5 variant fonému „a“)

Slabika – základní *vyslovitelná* jednotka řeči

- výhody: slabiky již v sobě obsahují kontext (pro hlásky uvnitř)

- nevýhody: velké množství slabik, automatická segmentace na slabiky není vždy jednoznačná („hrad-ní“, „hra-dní“), schází kontext mezi slabikami, není zahrnuta spodoba

Slovo – základní *významová* jednotka řeči

Promluva – sekvence slov nesoucí sdělení

Fonetika, fonetická abeceda, G2P

Fonetika – zabývá se zvukovou stránkou jazyka (jazyků), detailně zkoumá jak se tvoří a správně vyslovují hlásky.

Fonémy – stavební jednotky pro tvorbu řeči

Grafémy – stavební jednotky pro zápis textu

Fonetická abeceda – soubor znaků používaných pro fonémy při přepisu výslovnosti:

IPA (International Phonetic Alphabet) – používá speciální znaky

PAC (Phonetic Alphabet for Czech) – „inženýrská“ abeceda pro češtinu

Proč při ASR potřebujeme fonetickou abecedu ?

1. **zadávání do slovníku** („měl“ → „mňel“, „CD“→ „cédé“, „5“→ „pjet“)
2. **fonetický přepis nahrávek pro trénování modelů fonémů**

Fonémy v češtině – česká fonetická abeceda

Nouza, J., Psutka, J., Uhlíř, J.: Phonetic Alphabet for Speech Recognition of Czech. In: Radio Engineering, vol. 6, no. 4, December 1997, pp. 16-20.

Číslo	Foném vyjádřený českými hláskami	Foném dle PAC	Příklad	Číslo	Foném vyjádřený českými hláskami	Foném dle PAC	Příklad
1	„a“	a	táta	21	„m“	m	máma
2	„á“	á	táta	22	„m“	M	tramvaj
3	„b“	b	bába	23	„n“	n	víno
4	„c“	c	ocel	24	„ň“	N	banka
5	„dz“	C	leckde	25	„ň“	ň	koně
6	„č“	č	čichá	26	„o“	o	kolo
7	„dž“	Č	rádža	27	„ó“	ó	óda
8	„d“	d	jeden	28	„p“	p	pupen
9	„dˇ“	ď	dělat	29	„r“	r	bere
10	„e“	e	lev	30	„ř“	ř	moře
11	„é“	é	méně	31	„ř“	Ř	keř
12	„f“	f	fauna	32	„s“	s	sud
13	„g“	g	guma	33	„š“	š	duše
14	„h“	h	aha	34	„t“	t	dutý
15	„ch“	X	chudý	35	„tˇ“	ť	kutil
16	„i“ nebo „y“	i	bil, byl	36	„u“	u	duše
17	„í“ nebo „ý“	í	vítr, lýko	37	„ú“ nebo „ů“	ú	růže
18	„j“	j	dojat	38	„v“	v	láva
19	„k“	k	kupec	39	„z“	z	koza
20	„l“	l	dělá	40	„ž“	ž	růže
				41	Neutrální samo hláska	E	*)

42 „ticho“ -

Pravidla fonetického přepisu (1)

Slouží k vytvoření (standardního spisovného) fonetického přepisu slova nebo věty.

Jsou využívána v programech označovaných jako G2P (grapheme-to-phoneme)

Pravidla mají podobu

A → B / C _ D **aneb** **CAD → B**

Grafémy **fonémy**

JESTLIŽE grafému A bezprostředně předchází grafém C
a je bezprostředně následován grafémem D
PAK se A přepíše na foném B (platí i pro řetězce)

Samohlásky (SA)	a, á, e, é, i, í, o, ó, u, ú										
Znělé párové souhlásky (ZPS)	b	d	ď	g	z	ž	v	h	dz (C)	dž (Č)	ř
Neznělé párové souhlásky (NPS)	p	t	ť	k	s	š	f	ch (X)	c	č	Ř
Jedinečné souhlásky (znělé) (JS)	m, n, ň, l, j, r,										

Pravidla fonetického přepisu (2)

Základní pravidla:

České *ch* (pozůstatek spřežkového pravopisu) přepisujeme jako [*X*]

***ch* → *X* / _**

České *ů* přepisujeme jako [*ú*]

***ů* → *ú* / _**

Písmeno *w* přepisujeme na [*v*]

***w* → *v* / _**

Písmeno *q* se přepisuje na [*kv*]

***q* → *kv* /_**

Písmeno x se přepisuje na [*ks*]

x → *ks* /_

Samohlásky *y*/*ý* přepisujeme na [*i*/*í*]

***y* → *i* / _**

***ý* → *í* / _**

Pravidla fonetického přepisu (3)

Slabikotvorné j:

Jestliže za *i* následuje jiná samohláska, vloží se mezi *i* a následující samohlásku *j*

i* → *ij* / _ <SA>** ***“marije”, “mariji”, “bijologije”

Pravidla fonetického přepisu (4)

Další pravidla:

Následuje-li *ě* po *b*, *p*, *f*, *v*, přepisuje se na [*je*]

***ě* → *je* / <*b*, *p*, *f*, *v*> _**

Spojení *dě*, *tě*, *ně*, *mě* přepisujeme na [*ďe*], [*ťe*], [*ňe*], [*mňe*]

***dě* → *ďe* / _**

***tě* → *ťe* / _**

***ně* → *ňe* / _**

***ě* → *ňe* / *m*_**

Spojení *di*, *ti*, *ni* přepisujeme na [*ďi*], [*ťi*], [*ňi*]

***d* → *ď* / _<*i*, *í*>**

***t* → *ť* / _<*i*, *í*>**

***n* → *ň* / _<*i*, *í*>**

Pravidla fonetického přepisu (5)

Pravidla spodoby znělosti

Označíme ¬ ZPS jako neznělý protějšek ke znělé souhlásce ZPS,
tj ¬ *b* = *p*, ¬ *d* = *t*, ¬ *d'* = *t'*, ¬ *g* = *k*, ¬ *v* = *f*, ¬ *z* = *s*,

¬ *ž* = *š*, ¬ *h* = *ch*, ¬ *C* = *c*, ¬ *Č* = *č*, ¬ *Ř* = *ř*.

podobně ¬ NPS je znělý protějšek k neznělé souhlásce NPS

ZPS1 → ¬ ZPS1 / _ < - , NPS, ZPS2 - > **„hrad“ → „hrat“, „vůz“ → „vús“,**
„Radka“ → „ratka“, „drozd“, → „drost“,

NPS1 → ¬ NPS1 / _ ZPS **„kresba → „krezba“, „kdo“ → „gdo“,**
„leckde“ → „leCgde“,

U písmene ř rozhoduje o jeho výslovnosti znělost hlásky před i za, takže např. „keř → „keŘ“, „břicho → „břiXo“, „tři → „tŘi“, „přímo → „pŘímo“, „pařba → „pařba“,

Pravidla fonetického přepisu (6)

Pravidla spodoby artikulační

Jestliže souhláska *n* stojí před *k* nebo *g*, spodobuje se na [*N*] (nosové n)

n* → *N* / _ < *k*, *g* >** ***"baNka", "goNg"

Jestliže souhláska *m* stojí před *f* nebo *v (retozubé hlásky)*, spodobuje se v [*M*]

m* → *M* / _ < *f*, *v* >** ***"traMvaj", "niMfa", "trijuMf", "aMfóra", "eMvej"

Jestliže souhlásky *s*, *z* stojí před *t* nebo *d*, mění se následovně

***ts* → *c* / _**	**pět set → *"pjecet",***
***tš* → *č* / _**	**větší → *"vječí",***
***ds* → *c* / _**	***„beskickí",***
***dš* → *č* / _**	***"mlačí",***
***dz* → *C* / _**	***"poCemí", „leCgde", „kamikaCe"***
***dž* → *Č* / _**	***"Čorč", "loČije", "ČuNgle"***

Další pravidla v článku (na e-learningu)

Dana Nejedlová: Transkripce psaného českého textu do fonetické podoby. 2009

Příklady přepisu vět

Spolek byl založen devatenáctého listopadu roku devatenáct set třicet dva.
spoleg_bil_založen_devatenáctého_listopadu_roku_devatenácetŘicedva

Sejdeme se v naší restauraci ve čtvrt na sedm večer.
sejdeme_se_v_naší_restauraci_ve_čtvrt_na_sedum_večer

Kdy dnes odjíždí poslední vlak nebo autobus z Liberce do Pardubic.
gdi_dnes_odjížďí_posledňí_vlak_nebo_autobuz_z_liberce_do_pardubic

Na konferenci senátor rovněž kritizoval současné právní prostředí.
na_konferenci_senátor_rovňeš_kritizoval_současné_právňí_prostŘeďí

Výkon brankáře znamenal pro hokejové družstvo dobré umístění v tabulce.
víkon_braNkáře_znamenal_pro_hokejové_drušstvo_dobré_umísťeňí_f_tabulce

Dnes bude oblačno až polojasno, místy možno očekávat přeháňky.
dnez_bude_oblačno_aš_polojasno_místi_možno_očekávat_pŘeháňki

Najdeš to ve zdrojovém kódu HTML, stačí hledat řetězec mp3
najdeš_to_ve_zdrojovém_kódu_hEtEmElE_stačí_hledat_řeťezec_empétŘi

Úkol do příště

1. Vytvořit si program G2P, který z textové podoby (slova nebo věty, tj. sekvence slov) vygeneruje její výslovnost

   Účel programu:

   a) bude vám sloužit k tvorbě výslovnostního slovníku
   b) pomůže vám vygenerovat fonetický přepis trénovacích dat, které budete tvořit příště