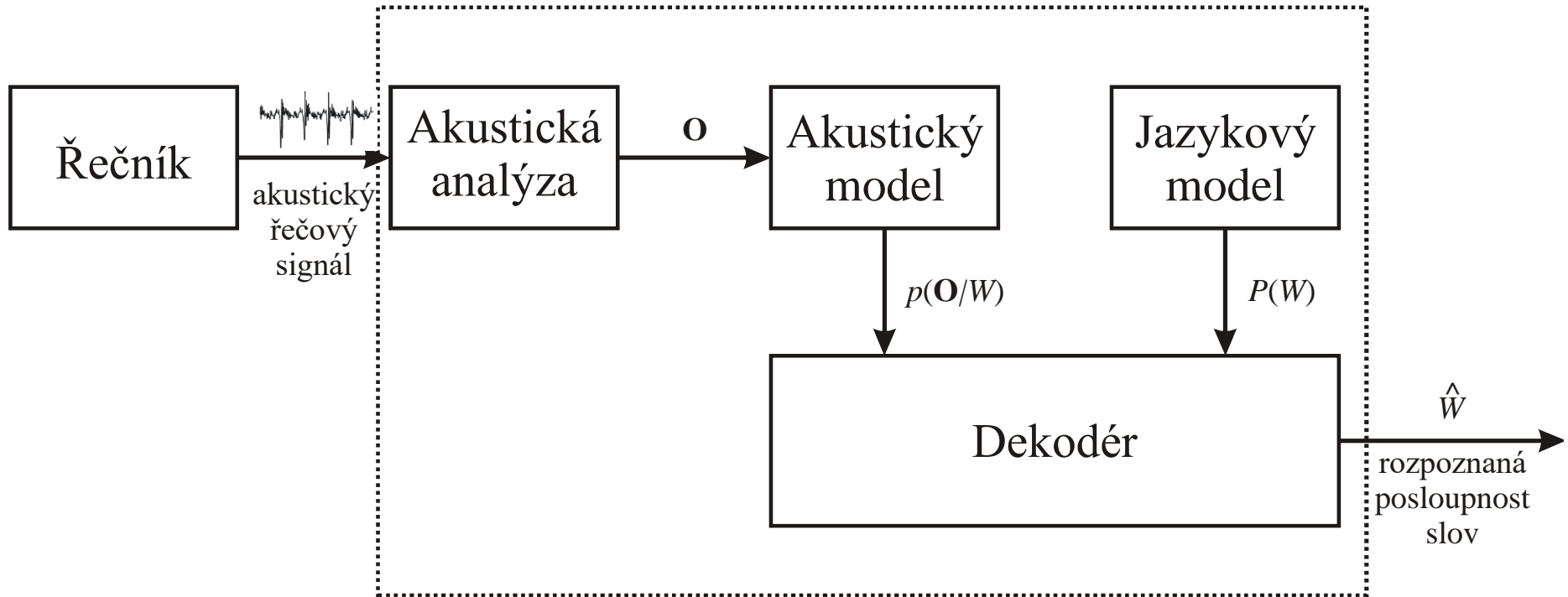


Úvod do praxe stínového řečníka

Automatické rozpoznávání řeči

System rozpoznávání řeči



$$\hat{W} = \arg \max_W P(W | \mathbf{O}) = \arg \max_W p(\mathbf{O} | W) P(W)$$

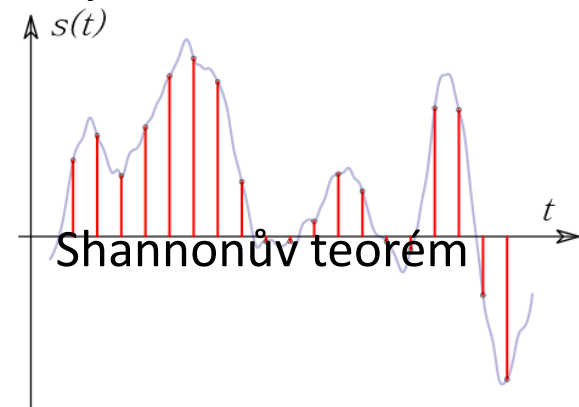
Akustická analýza

- potlačit vysokou informační redundanci řečového signálu z pohledu přenosu slovní informace
 - výška a barva hlasu, hlasitost promluvy (identifikace řečníka)
 - prozodie - přízvuk, intonace, frázování... (syntéza řeči)
 - emocionální stav řečníka (porozumění)
- snížit datový tok digitalizovaného řečového signálu (PCM)

- 🔊 ➤ 8000 Hz – staré telefony
- 🔊 ➤ 16000 Hz – nové telefony
- 🔊 ➤ 44100 Hz – CD
- 🔊 ➤ 48000 Hz a více – profesionální

lidský hlas – do 10000 Hz

lidský sluch – do 20000 Hz

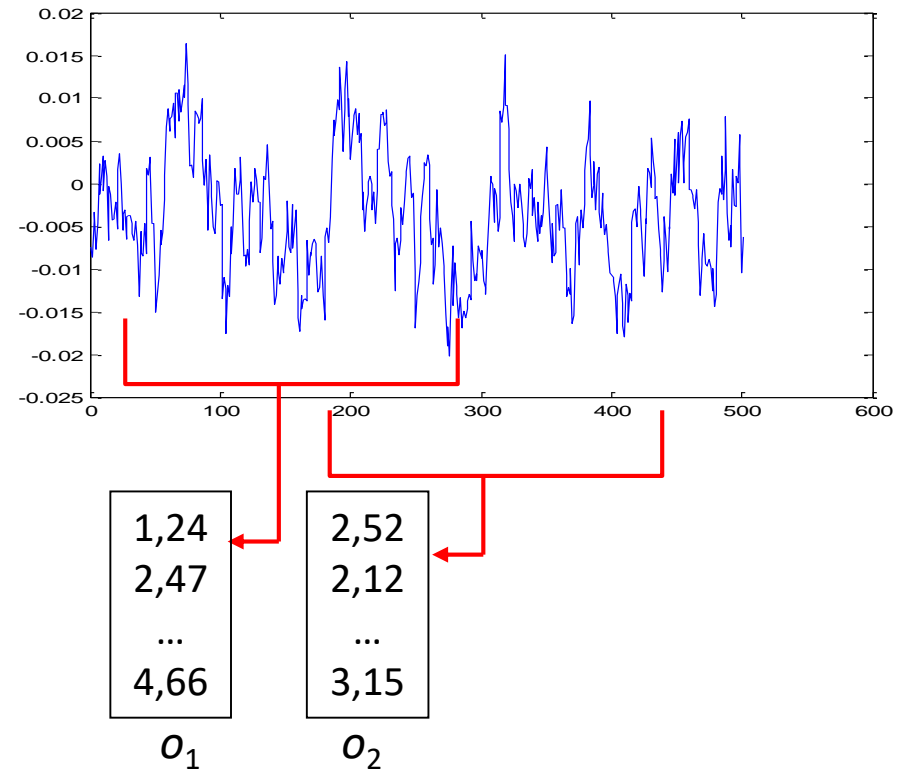


Akustická analýza

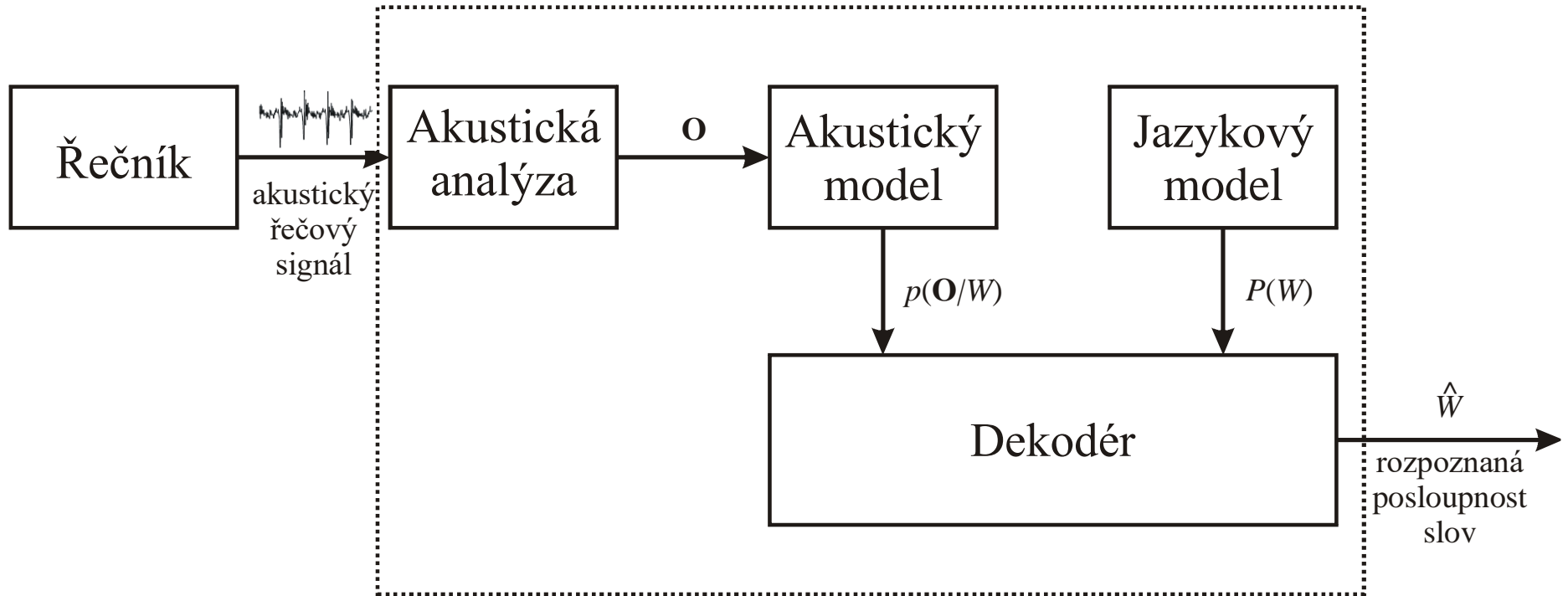
- extrahovat z řeči takové příznaky (parametry), které:
 - co nejvíce potlačí charakteristiky řečníka a prostředí
 - co nejlépe zachovají informaci o slovním obsahu promluvy
 - přiměřeně sníží objem zpracovávaných dat
- metody založené na tzv. krátkodobé analýze signálu
 - řečový signál je v krátkém časovém úseku (~ 10 ms \rightarrow 100/s) stacionární
 - tyto úseky (mikrosegmenty) lze dobře reprezentovat jedním vektorem příznaků (souborem číselných parametrů)
 - výsledkem např. vektor 12ti příznaků \rightarrow 1200 čísel/s

Akustická analýza

- modelování procesu generování řeči člověkem
 - lineární prediktivní analýza (LPC)
- modelování procesu vnímání řeči člověkem
 - perceptivní lineární predikce (PLP)
 - mel-frekvenční keprální koeficienty (MFCC)
- Fourierova transformace



System rozpoznávání řeči



$$\hat{W} = \arg \max_W P(W | \mathbf{O}) = \arg \max_W p(\mathbf{O} | W) P(W)$$

Akustický model

- pro každou akustickou jednotku určuje pravděpodobnost, se kterou je generována daným vektorem pozorování
- modeluje všechny možné akustické jednotky
 - promluvy
 - věty
 - slova
 - hlásky (fonémy)
- kontextově (ne)závislé fonémy – monofóny, trifóny, pentafóny...

Fonetická abeceda

Hláška	Znak	Příklad	Hláška	Znak	Příklad	Hláška	Znak	Příklad
a	a	máma	h	h	had	p	p	prak
á	A	táta	ch	x	chyba	r	r	rak
au	Y	auto	i	i	pivo	ř (znělé)	R	moře
b	b	bod	í	l	víno	ř (neznělé)	Q	tři
c	c	ocel	j	j	voják	s	s	osel
č	C	oči	k	k	oko	š	S	pošta
d	d	dům	l	l	lod'	t	t	otec
d'	D	děti	m	m	mír	t'	T	kutil
dz	w	leckdo	m	M	nymfa	u	u	rum
dž	W	léčba	n	n	nos	ú (ů)	U	růže
e	e	pes	n	N	banka	v	v	vlak
é	E	lépe	ň	J	laň	z	z	koza
eu	F	eunuch	o	o	bok	ž	Z	žena
f	f	facka	ó	O	jód			
g	g	guma	ou	y	pouto	pauza	#	

Fonetická transkripce

- určuje přepis daného slova do fonetické abecedy
- může vygenerovat více fonetických variant
 - Františka → frantíška, fraňtíška
 - jez → jez, jes
- alternativní výslovnostní varianty
 - osm → osm, osum
 - výjimka → výjimka, vyjímka, výmka
 - zaměstnat → zaměstnat, zaměsнат
 - malý → malý, malej
 - malé → malé, malý

Automatická fonetická transkripce

- produkční (fonologická) pravidla

$$A \rightarrow B / C _ D$$

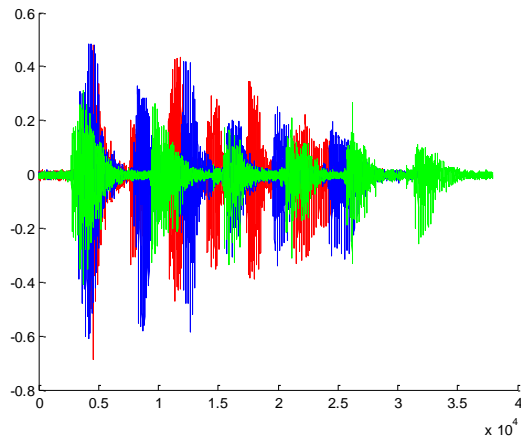
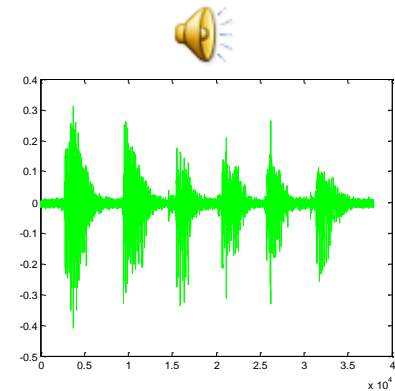
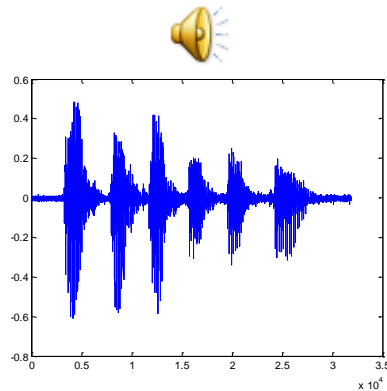
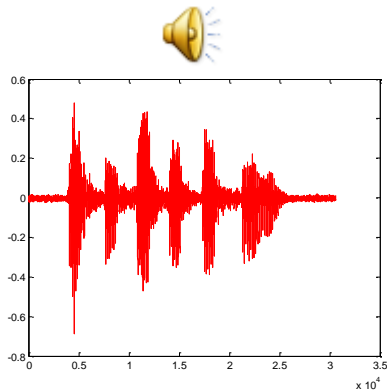
- **JESTLIŽE** řetězci znaků A bezprostředně předchází řetězec znaků C a je bezprostředně následován řetězcem znaků D, **PAK** se A přepíše na symboly B
 - ě → je / [b, p, v, f] _ oběť, opěra, závěr, harfě
 - d → d' / _ [i, í] divák, dítě
 - vz → fs / | _ p vzpomínka
 - zští → ští / _ | francouzští
 - ZPK → -ZPK / _ [NPK, -NPK, |NPK, |JK, |V, |#]

Fonetická transkripce

- slova přejatá
 - romantismus → romantyzmus
 - fotbal → fodbal
 - helium → hélijum
 - junta → chunta
 - Shakespeare → šejkspír
- fonetický slovník výjimek
- u jazyků bez flexe (např. angličtina) se používá expertní fonetický slovník

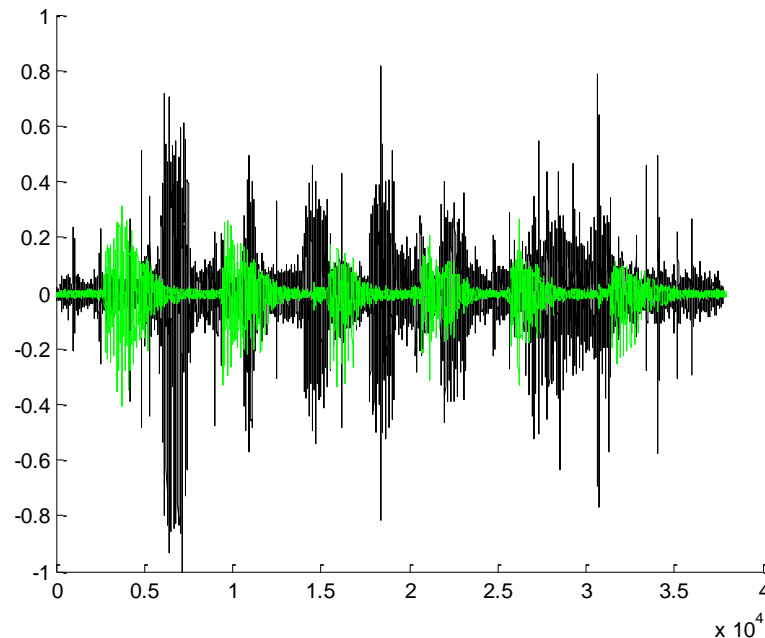
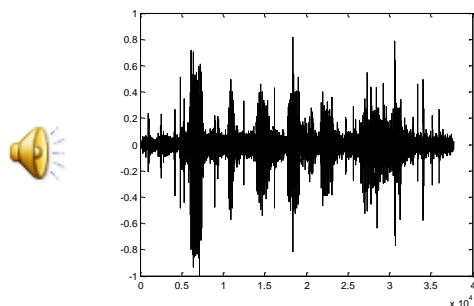
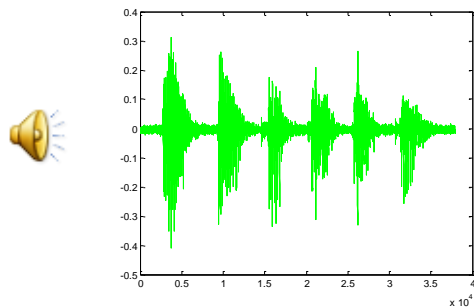
Akustický model

- řečový signál je velmi variabilní
 - tutéž promluvu vysloví každý řečník jinak
 - dokonce stejný řečník vysloví tutéž promluvu pokaždé jinak



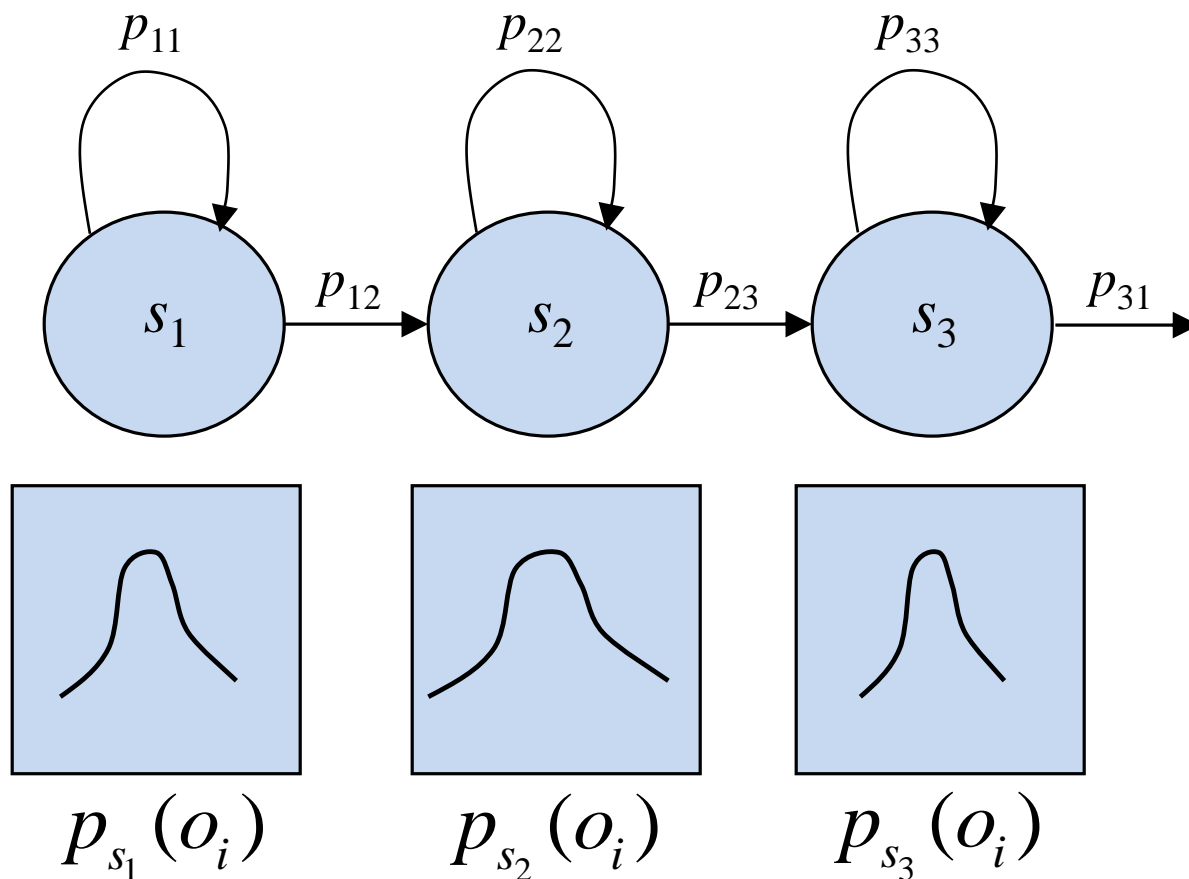
Akustický model

- v řečovém signálu se výrazně projeví jakákoliv změna prostředí (akustika místnosti, rušivé zvuky) nebo přenosového kanálu (jiný mikrofon, řeč přenášená přes telefon)



Akustický model

- skryté Markovovy modely (stochastický model)



Akustický model - trénování

Transcriber 1.5.1

File Edit Signal Segmentation Options Help

stopy n- [click] [inhale], [-noise]

- [noise-] [er] nenašel po těch rodičích, zůstal jsem sám [click] [inhale], měl jsem ještě nějaké příbuzné [-noise], vzdálenější [inhale] i bližší příbuzný, který žili [inhale] na různých místech slovenska v bra-, z, zejména tedy v bratislavě, všichni [inhale],
- v průběhu toho dvaadvacátého roku [inhale], ty z příbuzných, který zůstali, doma na slovensku [inhale], všichni zahynuli a zmizeli beze [inhale], beze stopy [inhale] a byla to velká rodina [inhale], [mic] byla to velká rodina mého [inhale], mé, mé matky [inhale],
- její bratři a tak dále a tak dále jako.
- [silence]
- [click] velice málo nás [inhale], nás, nás přežilo tedy [inhale] tohle, z tý, z tý veliké, z tý veliké rodiny.
- [click] [inhale] já jsem v dvaadvacátém roce dostal pokyn [inhale] od [hašomer hacajru]{hašomer hacajru} tedy odejít pryč [inhale] a pravděpodobně tedy přes maďarsko [inhale], kde ještě v té době byla velká relativní svoboda [inhale], relativní svoboda [inhale],
- tedy Židé nemuseli nosit [inhale] hvězdy a tak dále [noise-] [click] [-noise], [er] [hašomer hacajr]{hašomer hacajr}[um] zařídil ilegální přechod [inhale] přes noc [inhale] jednou, poprvé se to nepovedlo [inhale], vrátili jsme se [noise], protože ta spojka [inhale],
- kteřá nás měla čekat na druhé straně hranic [inhale] se nějak nedostavili, tak jsme se vrátili [breath] [click] [inhale] a opět asi ve skupině dvou, tři tedy [inhale] při- [um] d- [er] příslušníků [inhale] rovněž tedy toho [hašomer hacajr]{hašomer hacajr} se nám podruhé povedlo [inhale],


spk2 + spk1

- 1: povedlo překročit hranice [click] [inhale], dostali, někde u seredy se pame- pamatuju na to jméno [inhale],
- 2: kde jste, kde jste překročili [mic].

spk2

- sered jako [inhale], my jsme překročili v noci ilegálně [inhale], tam nás někdo čekal [inhale], tam jsme dostali [mic] [click] maďarské peníze [inhale] a ten člověk, který nás čekal, nebyl to žid jako [inhale] [mic], byl to nějaký maďar, který byl napojen [inhale] na maďarskou složku,
- anebo už na tý [inhale] z [hašomer hacajru]{hašomer hacajru}, který [noise-] už předtím tam byli [inhale] a kteří tuhle cestu

00249_02.trs
00249_002.wav



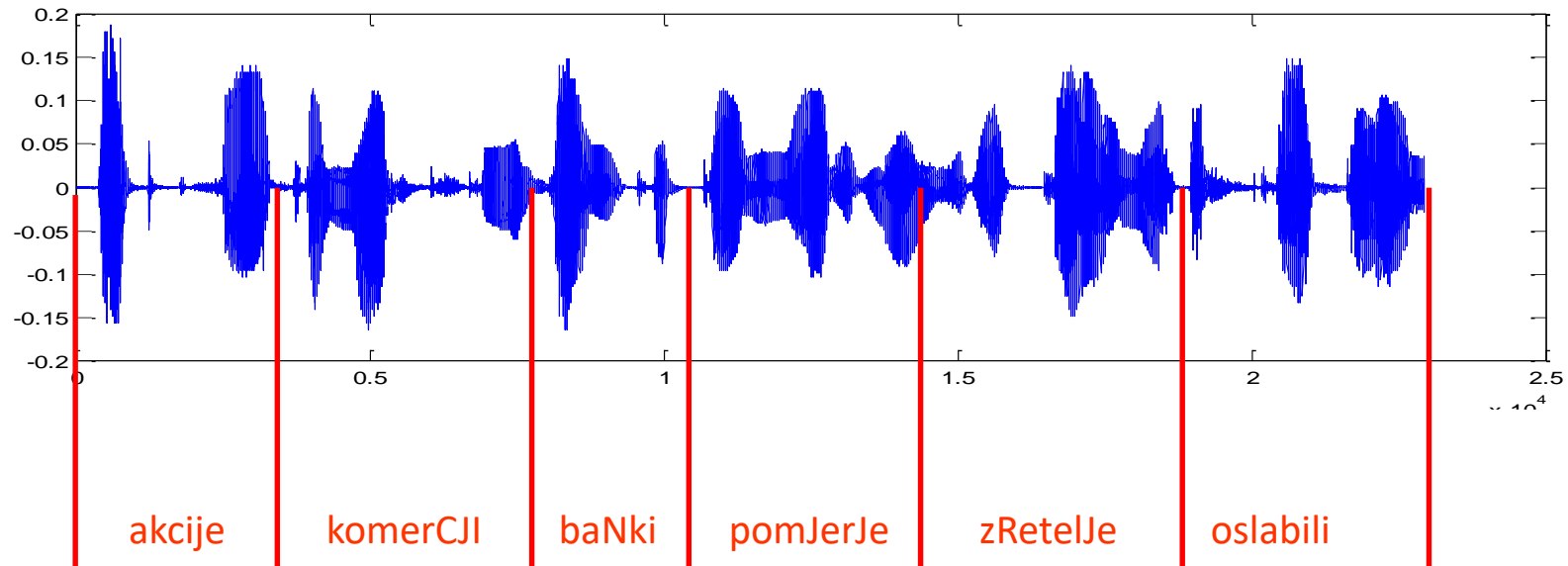
spk1	spk2	sp	spk2	s	spk2	spk2 +...
ty, měl...	[er] doktor... tak já jsem...	da	plný tedy ... kterému v tý ... takže tam ta ...	a	j- já myslím... ale taky ... a vím ...	tu i ta
něl jste vrátil. ...o to [inhale]. ...hacajru][inhale]. ...cim][inhale].	ca	[inhale]. ... [inhale]. ... [inhale]. ... [inhale].	k	... [inhale]. ... [inhale]. ... jako. [unintellig ... [inhale]	

2:00 2:30 3:00 3:30 4:00 4:30 5:00

Cursor : 04:35.283

Akustický model - trénování

- 1000 řečníků (600 žen a 400 mužů), 300 hodin řeči

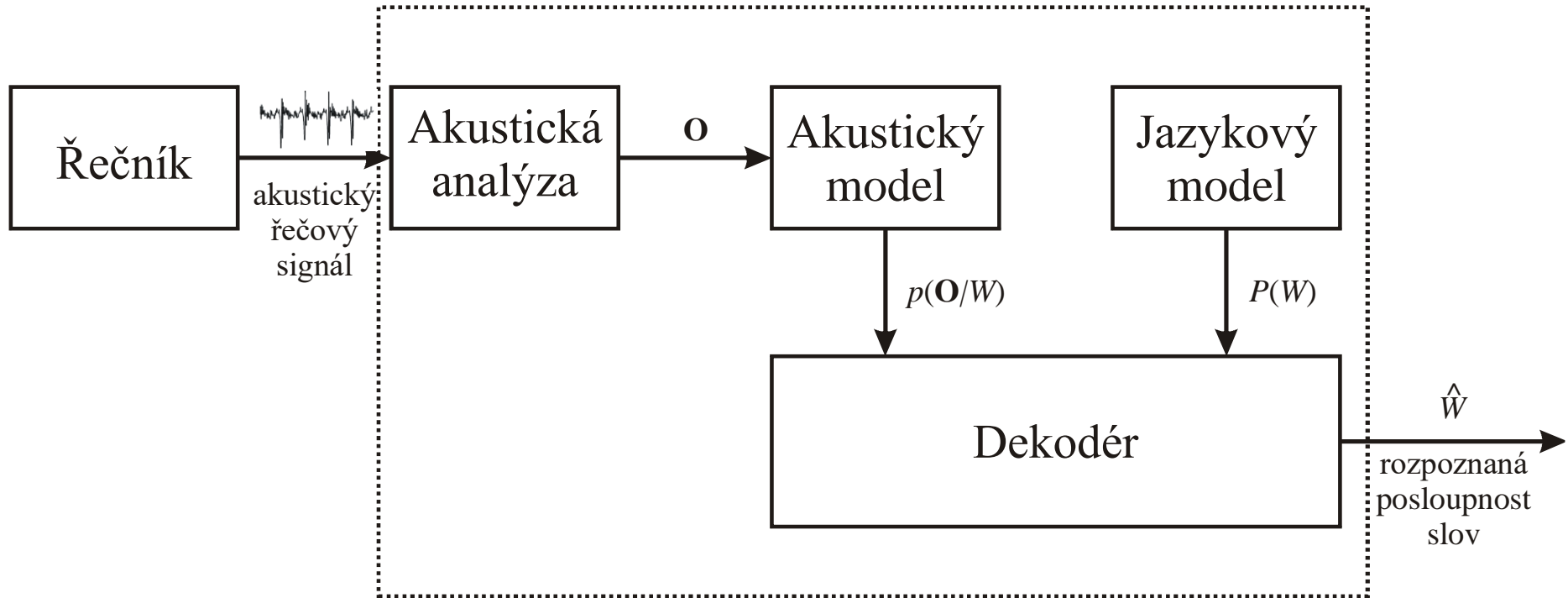


- trifónový akustický model s 50 miliony parametrů

Akustický model - shrnutí

- modeluje fonémy daného jazyka
- využívá statistický přístup (skryté Markovovy modely)
- pro trénování vyžaduje velké množství anotovaných zvukových dat
 - nezávislý na řečnickovi – data od různých řečníků (stovky hodin)
 - závislý na řečnickovi – data od jednoho řečníka (desítky hodin)
- nejlépe funguje za stejných podmínek, za jakých byla pořízena trénovací data (vzorkovací frekvence, mikrofon, akustika místnosti, úroveň hluku, přízvuk, styl řeči...)
- možnost adaptace parametrů na konkrétní přenosový kanál, řečníka apod.

System rozpoznávání řeči



$$\hat{W} = \arg \max_W P(W | \mathbf{O}) = \arg \max_W p(\mathbf{O} | W) P(W)$$

Jazykový model

- určuje pravděpodobnost, s jakou si řečník přeje vyslovit danou posloupnost slov
- modeluje všechny možné posloupnosti slov
 - promluvy
 - věty
 - n-tice slov
 - n=1 (unigramy) – pravděpodobnost slov bez ohledu na kontext - $P(w_i)$
 - n=2 (bigramy) – pravděpodobnost každého slova je podmíněna slovem bezprostředně předcházejícím - $P(w_i | w_{i-1})$
 - n=3 (trigramy) pravděpodobnost každého slova je podmíněna dvěma slovy bezprostředně předcházejícími - $P(w_i | w_{i-1}, w_{i-2})$

Jazykový model - trénování

dne 13. 10. 1987 bylo usnesením č.j. ORHK – 1895/TČ-80-2006 zahájeno trestní stíhání proti obviněné Marii Šubrové, bytem Vysoké Mýto, Město, Náměstí Přemysla Otakara II. čp. 188.

LS páteře v segmentech L4/5 a L5/S1: spondyloza, osteochondroza L5/S1 s vakuovým fenoménem. Spondylartroza se zúžením laterálních recesů. Nevelký mediální výhřez L4/5, který by při normální šíři neměl mít klinický význam. Drobný hemangiom/8 mm/ v obratlovém těle L5.

Závěr: degenerativní změny na L páteři .Malý výhřez L4/5.

Budka

Jazykový model - trénování

- získání textů
- čištění (nechat jen to, co se má rozpoznávat)
- tokenizace (oddělení rozpoznávacích jednotek)
- normalizace (převod čísel, zkratek, nestandardních slov atd.)
- unifikace (sjednocení synonym, multislova atd.)

dne třináctého desátý tisíc devět set osmdesát sedm bylo usnesením číslo_jednací ORHK - tisíc osm set devadesát pět / TČ - osmdesát - dva tisíce šest zahájeno trestní stíhání proti obviněné Marii Šubrové , bytem Vysoké_Mýto , Město , Náměstí Přemysla_Otakara_II. číslo_popisné sto osmdesát osm .

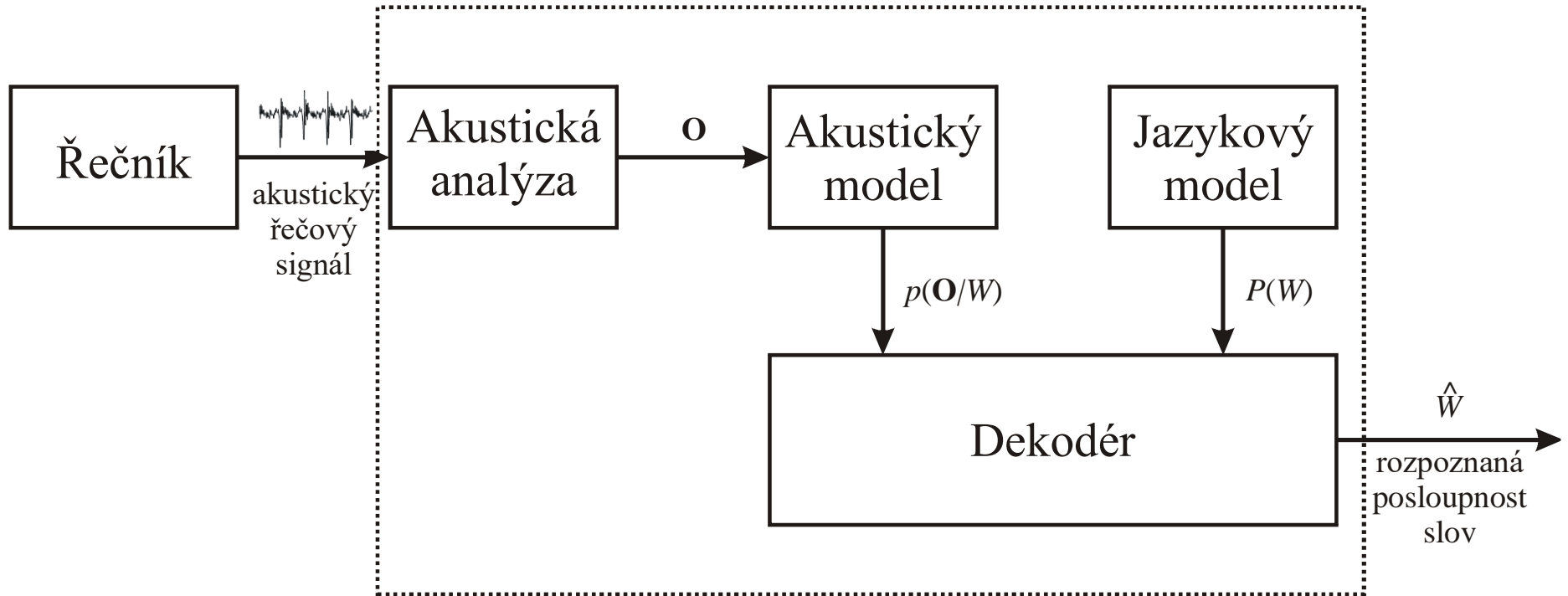
Jazykový model - trénování

- pravděpodobnosti n-gramů z relativního počtu výskytů jednotlivých slovních n-tic
 - nenulová pravděpodobnost pro neviděné n-gramy
- odpískal zakázané uvolnění
 buvol mění
- výslovnostní slovník
 - gynekologie a porodnictví – 100 tisíc slov
 - advokacie – 200 tisíc slov
 - sport – 500 tisíc slov
 - obecný – > 1 milion slov

Jazykový model - shrnutí

- modeluje posloupnosti slov daného jazyka (domény)
- využívá statistický přístup (slovní n-gramy)
- pro trénování vyžaduje velké množství textových dat
 - obecný – desítky GB textu (miliardy slov)
 - omezená doména – stovky MB textu (desítky milionů slov)
- nejlépe funguje na obdobných textech, které byly použity pro trénování (doména, čtená/hovorová řeč, způsob vyjadřování, slovník...)
- možnost adaptace – přidávání slov, n-gramů apod.

System rozpoznávání řeči

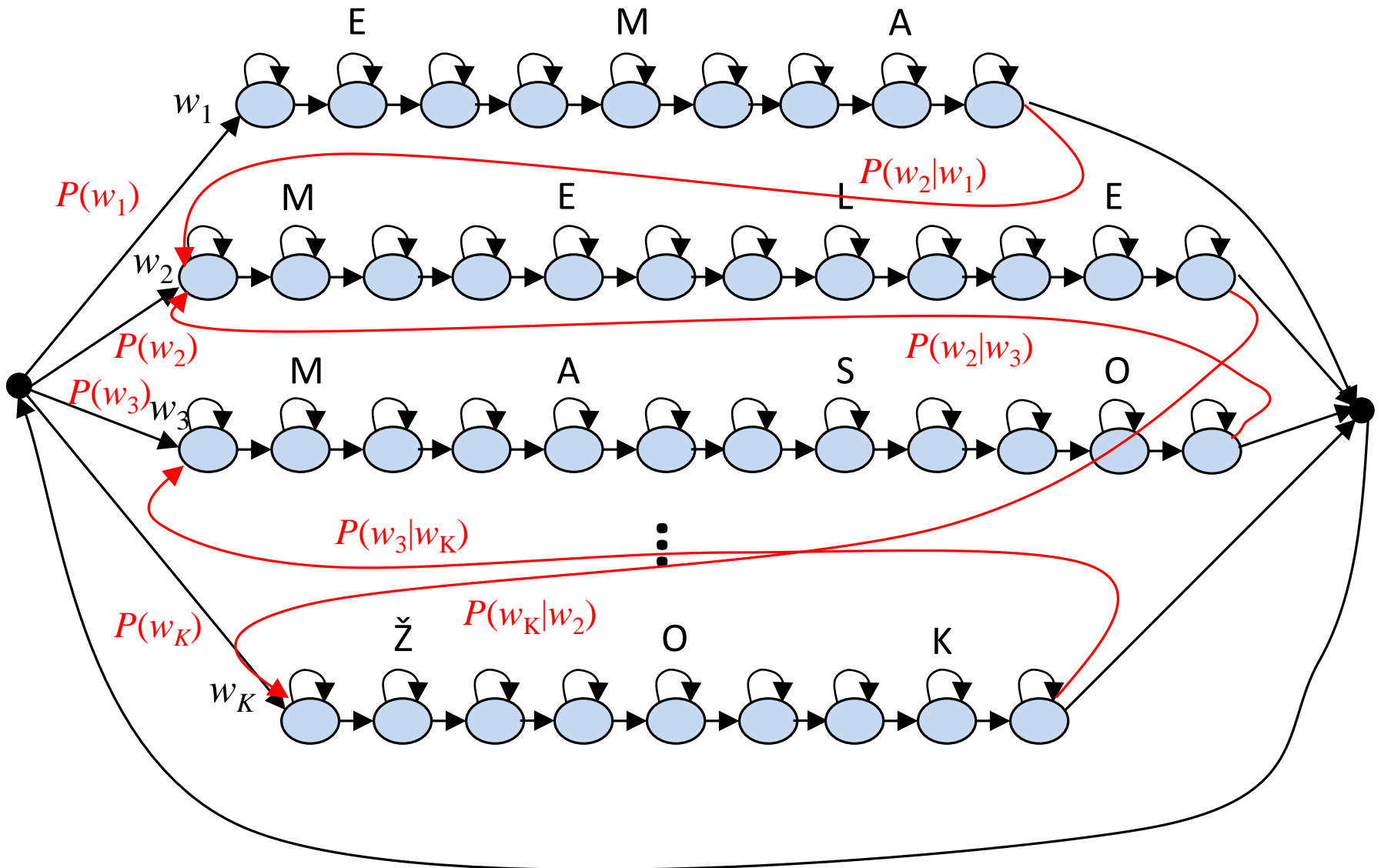


$$\hat{W} = \arg \max_W P(W | \mathbf{O}) = \arg \max_W p(\mathbf{O} | W) P(W)$$

Dekodér

- určuje nejpravděpodobnější posloupnost slov, kterou chtěl řečník vyslovit
 - vstupní vektory pozorování
 - pravděpodobnosti z akustického modelu
 - skryté Markovovy modely fonémů
 - pravděpodobnosti z jazykového modelu
 - slovník s fonetickými transkripcemi
- kompromis mezi přesností a rychlostí

Dekodér



Automatické rozpoznávání řeči - shrnutí

- snaží se převést mluvenou řeč na psaný text
- pracuje s akustickým a jazykovým modelem
- založeno na statistických modelech
- trénuje se na základě zvukových nahrávek a textů
- nejlépe pracuje za obdobných podmínek, za jakých se trénovalo (akustický kanál, jazyková doména)
- může rozpoznat jen slova, která předem zná
- není bezchybné
- má rádo poučeného uživatele