11 Rozpoznávanie rečových signálov

- Riešenie tejto úlohy je veľmi náročné a komplikované, najmä pre redundantnosť akustického signálu a veľké rozdiely pri vyslovení slov jednotlivými ľuďmi.

- Systémy môžeme rozdeliť na dve skupiny: tie, ktoré klasifikujú *izolovane vyslovené slová*, so slovníkom obsahujúcim niekoľko desiatok až stoviek slov a tie, ktoré klasifikujú *slovné spojenia* alebo *celé frázy*, pracujúce so stovkami až tisíckami slov. Ideálna by bola odozva v *reálnom čase*

Informačné a fonetické hľadiská hovorenej reči

- reč človeka je charakteristická *akustickou štruktúrou, lingvistickou štruktúrou a subjektívnym prejavom* osobnosti rečníka. Za najmenšiu jednotku reči možno považovať *fonému.* Počet foném vo svetových jazykoch sa pohybuje od 12 do 60, v českom jazyku ich je 36, v ruskom 40, v anglickom 42. Fonémy sa spájajú do postupnosti, kde ďalšou stavebnou jednotkou je slabika.

- Množstvo informácie v hovorenej reči: človek vysloví 80 až 130 slov za minútu, čo je asi 10 foném za sekundu. Ak entropia jednej fonémy je 1 až 4 bity, dostaneme rýchlosť prenosu informácie maximálne 40 bit/s. Tento výsledok charakterizuje informačný obsah reči v jej lingvistickej štruktúre. Ak budeme sledovať číslicové kódovanie akustického signálu, ktorý je charakterizovaný priebehmi amplitúdy a frekvencie, tak pri dynamike reči 60 dB a viac, pri frekvenčnom spektre 10 kHz (pri sykavkách) si vyžaduje dokonalé kódovanie 200.000 bit/s. Z toho vyplýva obrovská i*nformačná redundancia.*

- Pri automatickom rozpoznávaní reči je predpokladom úspechu urobiť základnú akustickú a fonetickú analýzu, t.j. popísať signál takými príznakmi, ktoré ponesú dostatočnú informáciu a na druhej strane znížia informačnú nadbytočnosť.

- Frekvencia kmitu hlasiviek charakterizuje *základný tón ľudského hlasu*.„ v rozmedzí od 150 do 400 Hz.

Samohlásky

- Okrem základného tónu sa objavuje aj rad vyšších zosilnených tónov, ktoré sa nazývajú formanty. Pre český a slovenský jazyk sú najdôležitejšie dva formanty:

u: $F_1 = 300 \div 500$ Hz, $F_2 = 600 \div 1\,000$ Hz,
o: $F_1 = 500 \div 700$ Hz, $F_2 = 900 \div 1\,200$ Hz,
a: $F_1 = 750 \div 1\,100$ Hz, $F_2 = 1\,100 \div 1\,500$ Hz,
e: $F_1 = 500 \div 700$ Hz, $F_2 = 1\,500 \div 2\,000$ Hz,
i: $F_1 = 300 \div 500$ Hz, $F_2 = 2\,000 \div 3\,000$ Hz.

Spoluhlásky

- Na rozdiel od samohlások si spoluhlásky vyžadujú charakteristický šum, ktorý vzniká prekážkou pri vydychovanom vzduchu.

- Väčšinu spoluhlások možno rozdeliť do párových dvojíc, niektoré dvojicu nemajú.

TABULKA 5.1. Dělení českých souhlásek

Souhlásky		Závěrové (explozívy)				Úžinové (frikativy)				Polozávěrové (afrikáty)	
Párové	neznělé	p	t	ť	k	s	š	f	ch	c	č
	znělé	b	d	ď	g	z	ž	v	h	dz	dž
Nepárové	znělé	m, n, ň				l, j, r, ř					

Akustická analýza a výber príznakov pre klasifikáciu

Základné charakteristiky

- uplatňuje sa metóda krátkodobej analýzy, pričom sa úseky rečového signálu spracovávajú tak, ako keby to boli oddelené krátke zvuky. Takéto mikrosegmenty sú reprezentované časovým úsekom 10 až 30 ms. Výsledkom je príznak alebo súbor príznakov charakterizujúci mikrosegment. Základné príznaky, ktoré nesú informáciu o energetických a kmitočtových zmenách akustického signálu, sa pri rozpoznávaní izolovaných slov používajú ako príznaky priamo pre klasifikáciu.

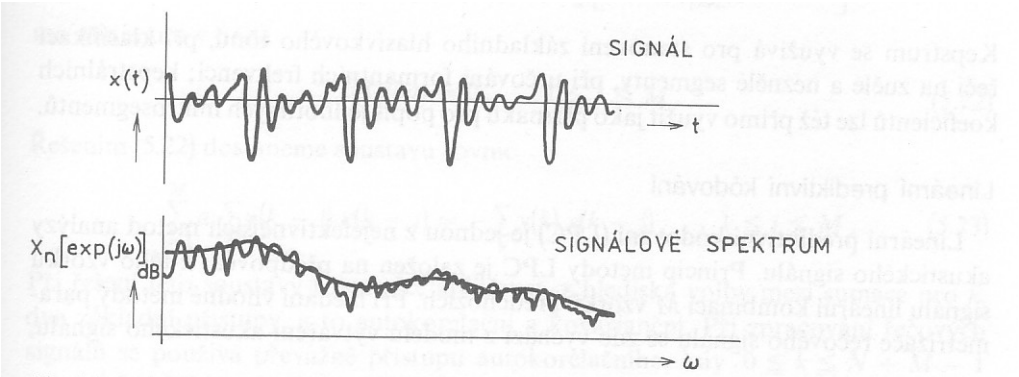
- Niekedy sa ešte spracovávajú základné príznaky v bloku fonetickej analýzy, kde sa pridávajú fonetické charakteristiky (základný hlasivkový tón, frekvencia formantov a ich amplitúdy apod.).

Spracovanie v časovej oblasti

- Používa sa krátkodobá charakteristika, krátkodobá energia, krátkodobá stredná hodnota, krátkodobá funkcia stredného počtu prechodu signálu nulou, krátkodobá autokorelačná funkcia.

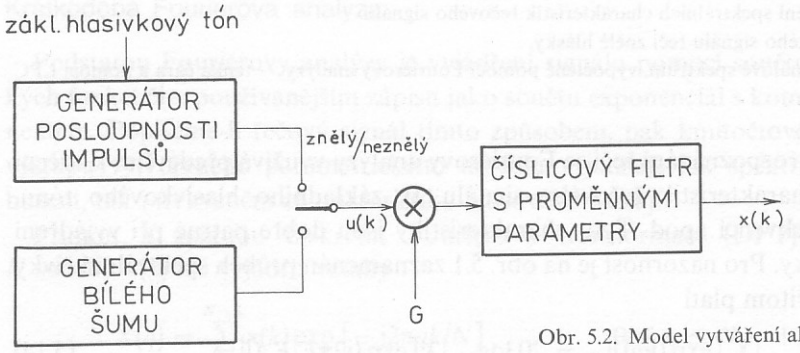
Spracovanie vo frekvenčnej oblasti

- Používa sa krátkodobá Fourierova analýza



Obr. 5.1. Znázornění spektrálních charakteristik řečového signálu
a) průběh akustického signálu řeči znělé hlásky,
b) odpovídající signálové spektrum, vypočtené pomocí Fourierovy analýzy – tenká čára a pomocí LPC – tlustá čára.

- Používa sa tiež kepstrálna analýza a lineárne prediktívne kódovanie (LPC).



Obr. 5.2. Model vytváření akustického signálu.

Rozpoznávanie izolovaných slov

Príznakové metódy klasifikácie

- príklad z ČVUT Praha – 21 povelových slov. Bola urobená časová normalizácia zvukových signálov a slová boli vyjadrené obrazmi konštantnej dimenzie M = 240. Pomocou Karhunenovho-Loeveho rozvoja bola dimenzia príznakového priestoru znížená na N = 8 alebo 16.

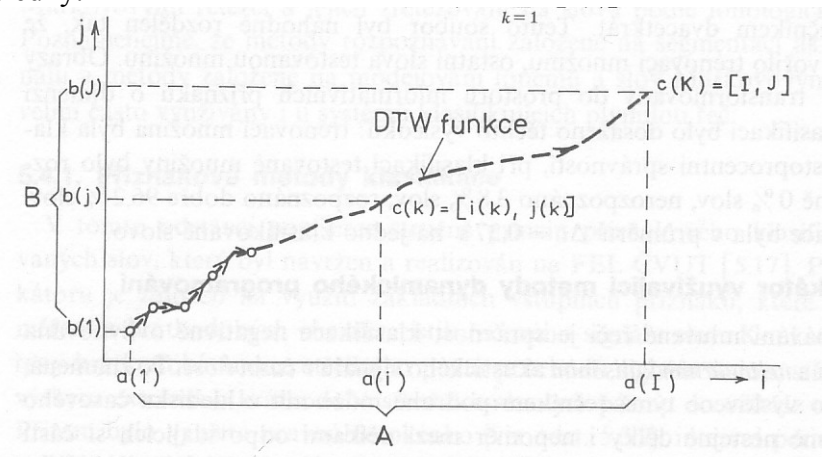
- výsledky

Klasifikátor využívajúci metódy dynamického programovania

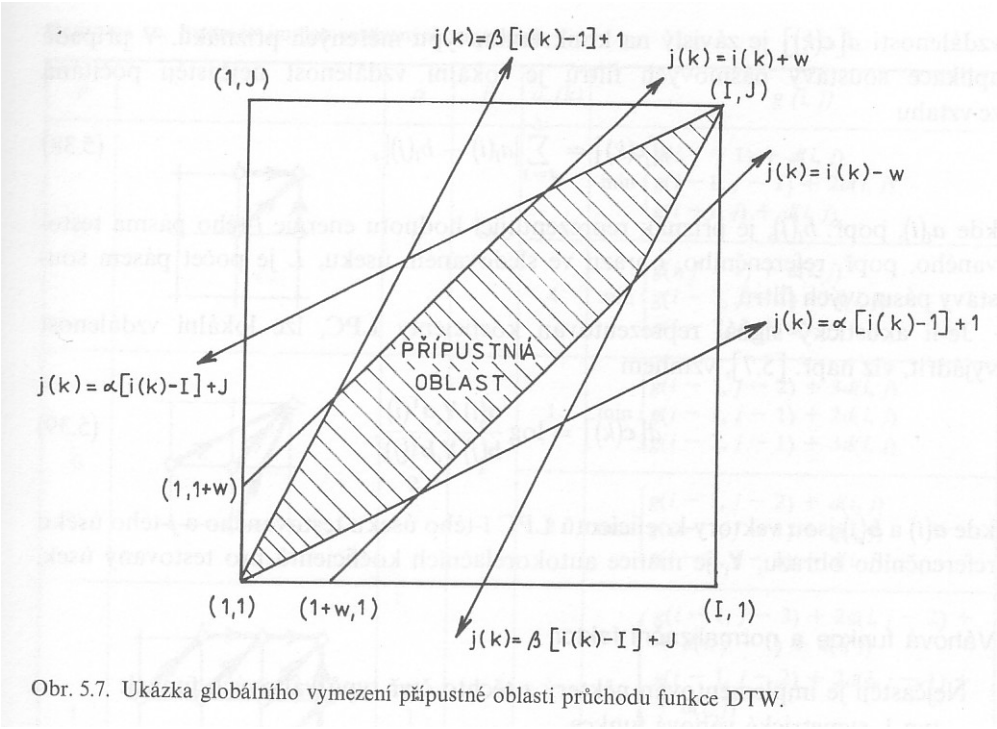
- využíva prístup, ktorý sa nazýva dynamic time warping – DTW. Väčšina komerčných algoritmov pre rozpoznávanie reči pracuje na báze dynamického programovania.

- Dávajú sa do súvisu obrazy dvoch slov, ako postupnosti príznakov. a vymedzuje sa priestor, kadiaľ môže ísť analýza pri hľadaní riešenia.

- Výsledky.



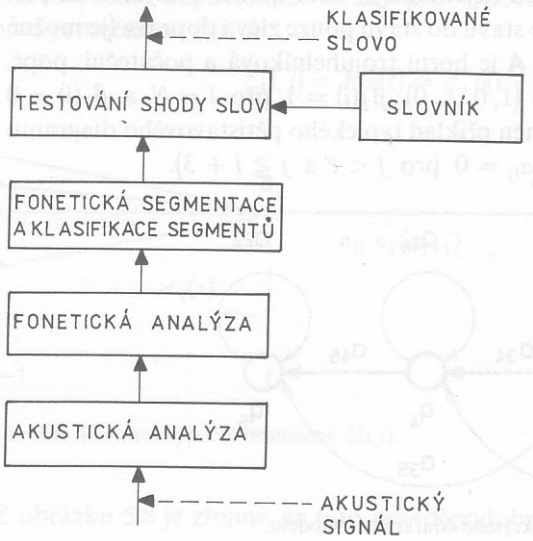
Obr. 5.3. Znázornění průběhu funkce DTW v rovině (i, j).



Obr. 5.7. Ukázka globálního vymezení přípustné oblasti průchodu funkce DTW.

Štrukturálne metódy rozpoznávania

- po bežnej akustickej analýze sa robí fonetická analýza a potom sa signál delí na najkratšie významné úseky – segmenty. Hranice segmentov možno charakterizovať zmenou *spôsobu artikulácie* alebo ostrými zmenami konfigurácie rečového ústrojenstva. Jednotlivým segmentom sa potom priraďujú fonetické prvky (fonémy, slabiky) alebo ich triedy (explozívy, sykavky apod).

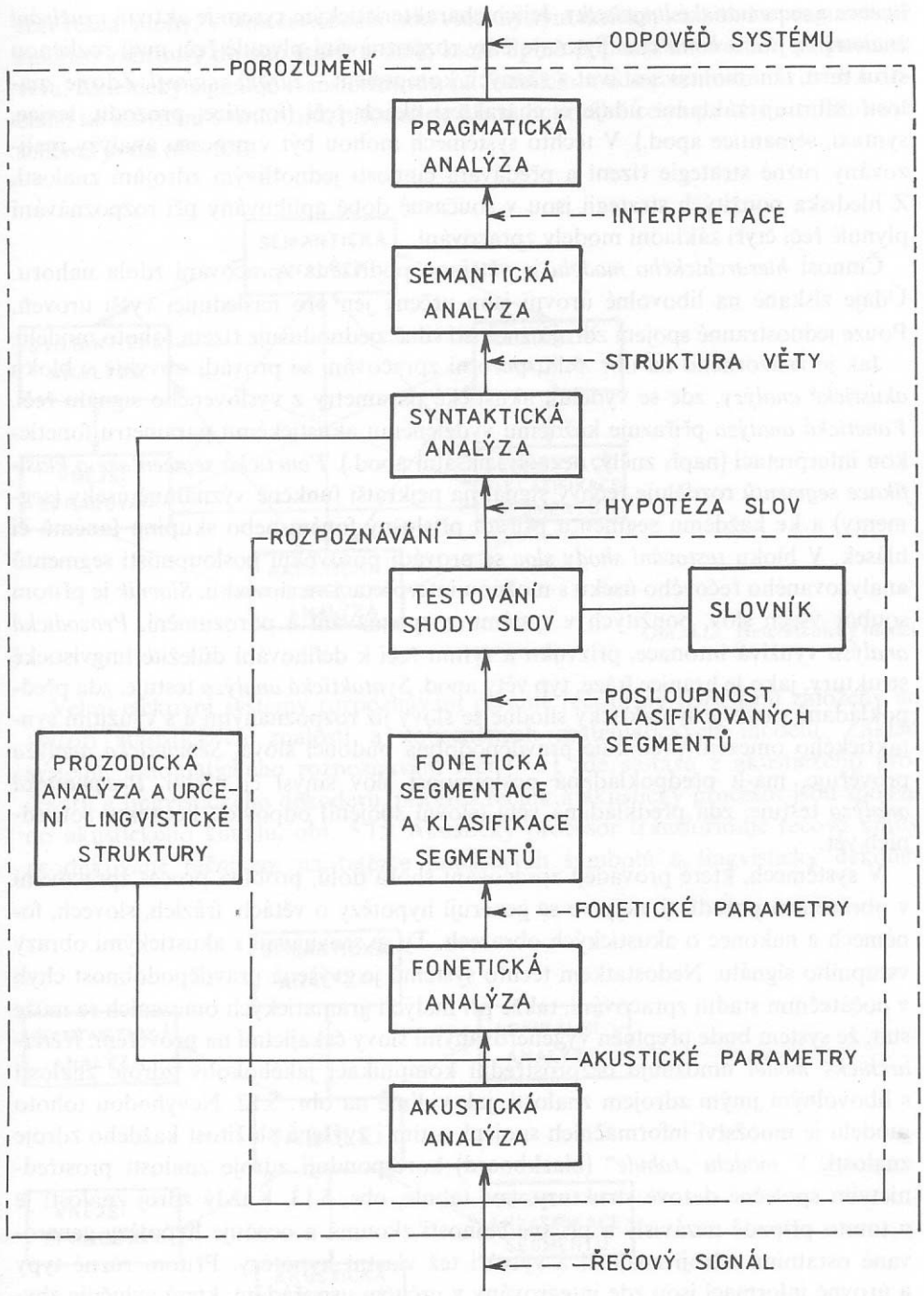


Obr. 5.10. Postupové schéma zpracování akustického signálu při aplikaci strukturálních metod rozpoznávání.

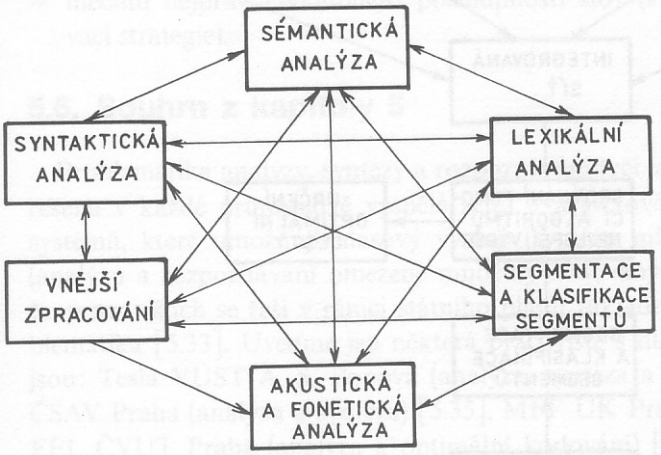
- segmentácia sa robí dvoma spôsobmi : v jednej sa skúmajú mikrosegmenty a priraďujú sa do segmentov. Druhá skupina sleduje prekročenie prahovej hodnoty niektorého parametra rečového signálu.

- Výsledky.

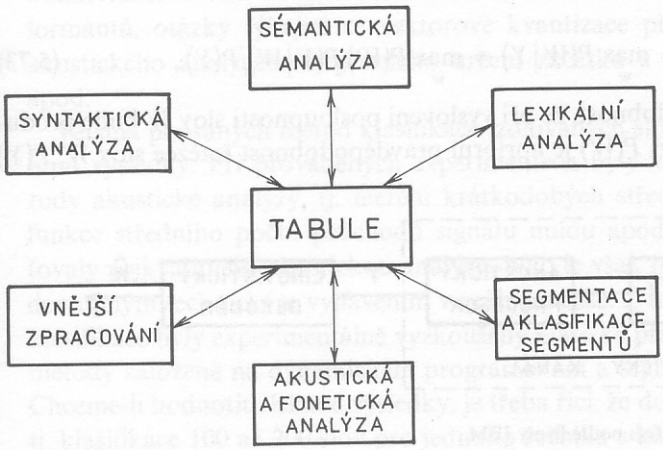
Systémy pre rozpoznávanie a porozumenie plynulej reči



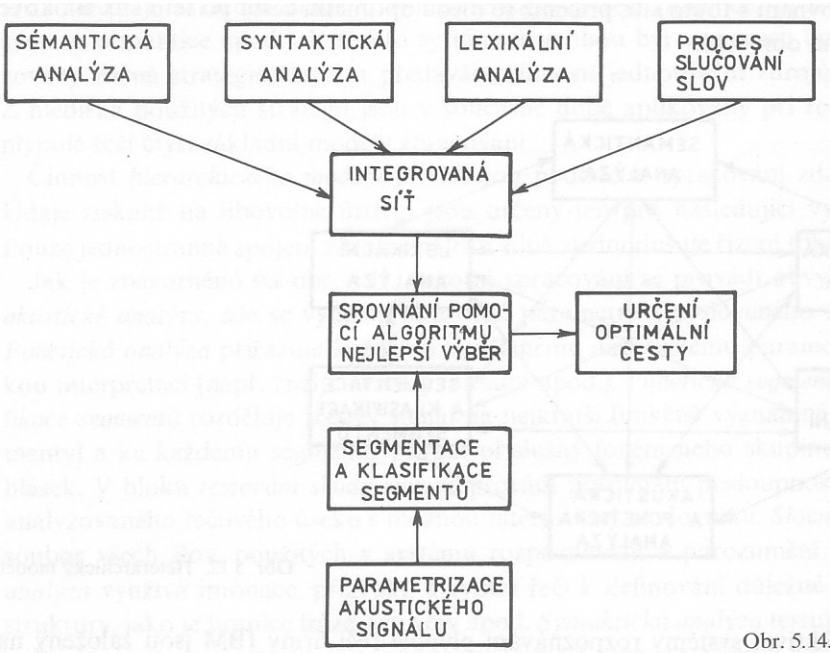
Obr. 5.11. Hierarchický model.



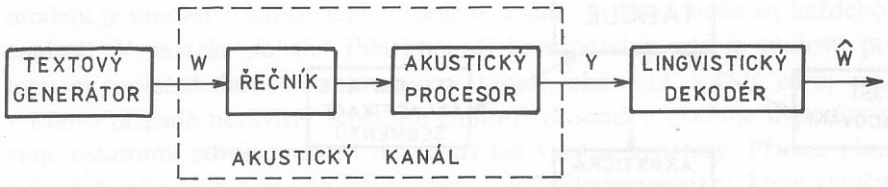
Obr. 5.12. Heterarchický model.



Obr. 5.13. Model „tabule"



Obr. 5.14. Síťový model.



Obr. 5.15. Systém rozpoznávání řeči podle firmy IBM.