Rozpoznávanie reči v zjednodušenom anglickom jazyku

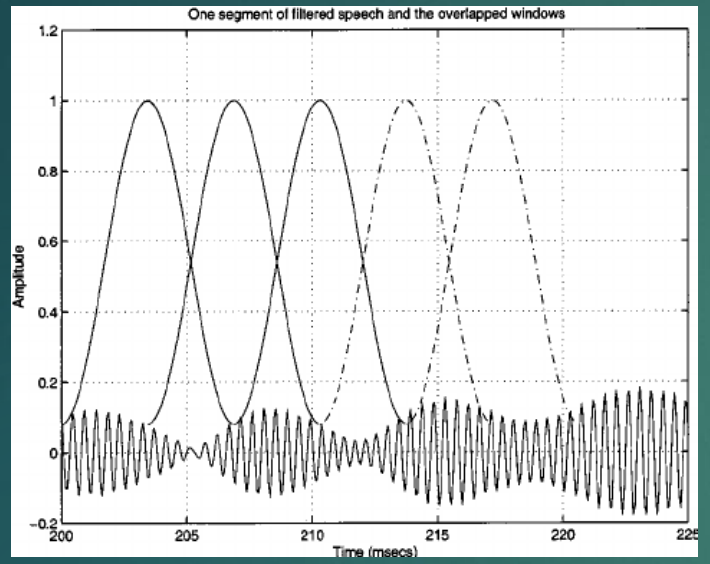
DÁVID ŠUBA

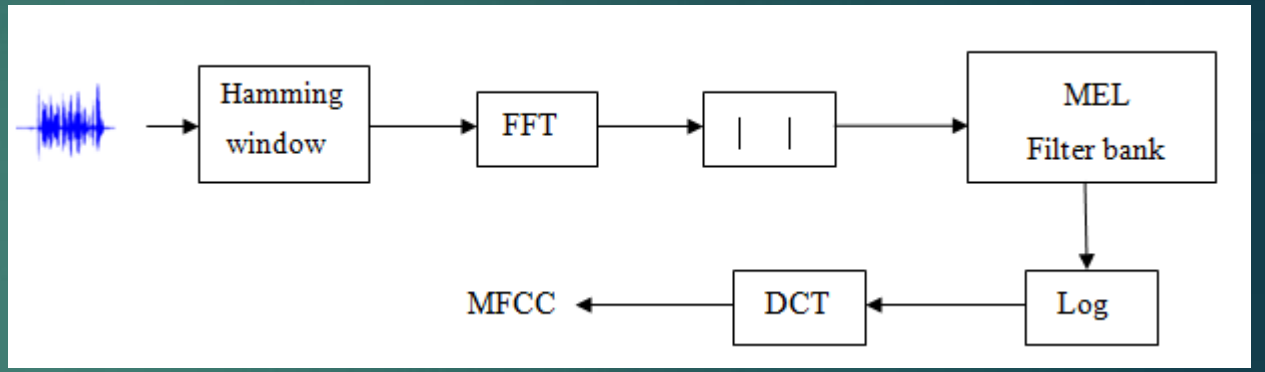
Ciele práce

- Naštudovanie si problematiky rozpoznávania reči pomocou skrytých markovovských modelov
- Pripravenie vlastnej dátovej množiny podľa zadanej slovnej zásoby a natrénovanie systému vytvoreného pomocou HTK toolkit
- Vytvorenie real-time systému rozpoznávajúceho hovorené príkazy

Úvod do rozpoznávania reči

- Konvertovanie signálu na vektory príznakov





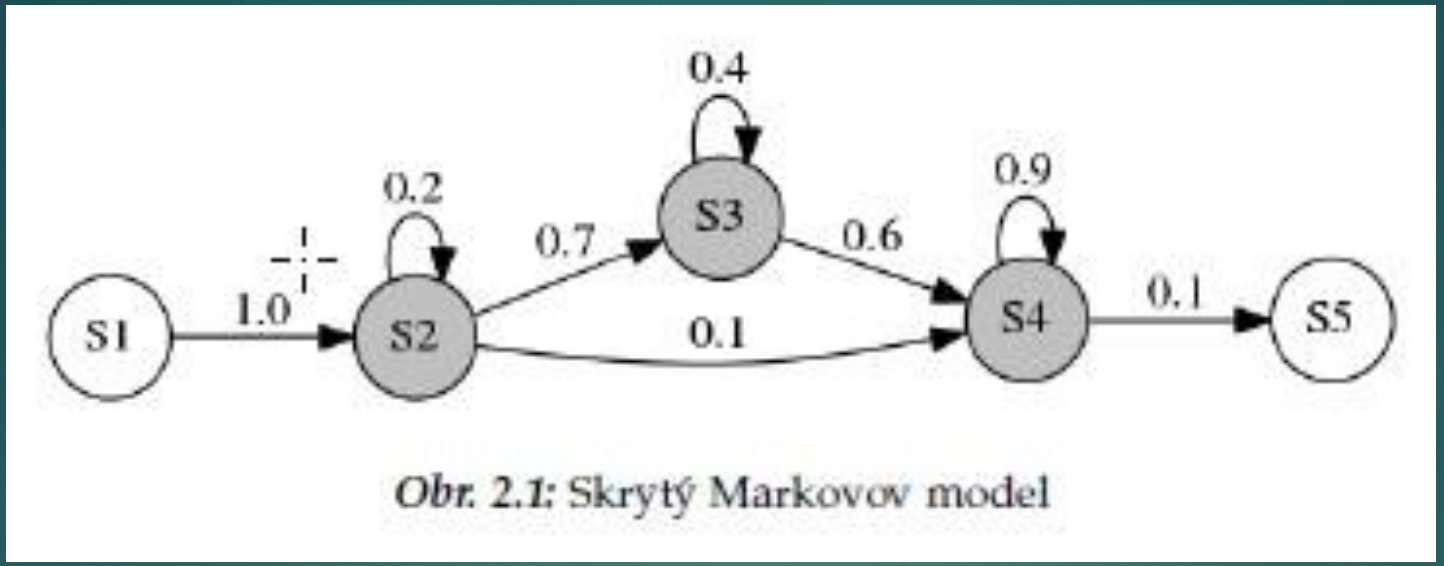
- Reč sa zmení na postupnosť pozorovaní:

$$O = o_1, o_2, \ldots o_t$$

- Rozpoznávanie slova je teda hľadanie:

$\arg\max P(w_i|O)$, w_i - i-te slovo zo slovníka

Skryté markovovské modely



Obr. 2.1: Skrytý Markovov model

- M = (Q, V, π, a, b)
- Konečný automat, ktorý generuje postupnosti symbolov s určitou pravdepodobnosťou
- $P(O, X|M) = a_{12}b_2(o_1)\ a_{22}b_2(o_2)\ a_{23}b_3(o_3)\ldots, \qquad X = (1,2,3)$

Trénovanie a rozpoznávanie

The Baum-Welch algorithm

Initialization:

Pick the best-guess for model parameters (or arbitrary)

Iteration:

1. Forward for each *x*
2. Backward for each *x*
3. Calculate A_{kl}, $E_k(b)$
4. Calculate new a_{kl}, $e_k(b)$
5. Calculate new log-likelihood

Until log-likelihood does not change much

Viterbi Algorithm

- $D(0, \text{START}) = 0$
- **for each** tag t != START **do**: $D(1, t) = -\infty$
- **for** i ← 1 to *N* **do**:
  - **a. for each** tag t^j **do**:

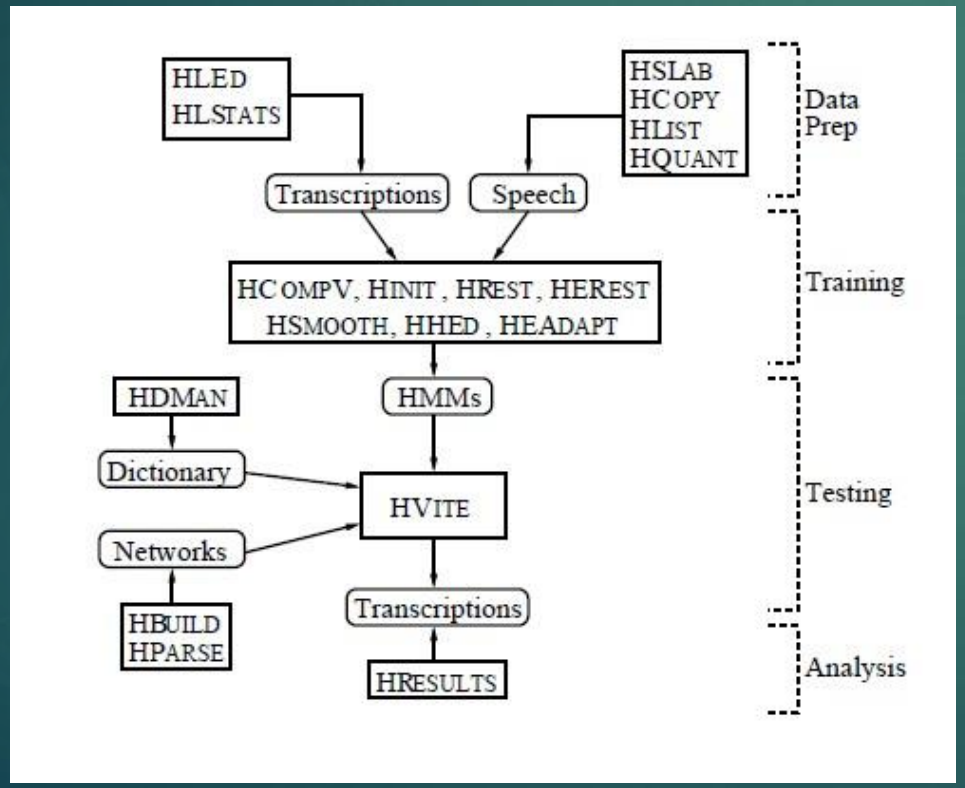
    $D(i, t^j) \leftarrow \mathbf{max}_k\, D(i-1,t^k) b(w_i|t^j) a(t^k \rightarrow t^j)$

    $D(i, t^j) \leftarrow \mathbf{max}_k\, D(i-1,t^k) + logb(w_i|t^j) + loga(t^k \rightarrow t^j)$
- $\log P(W,T) = \mathbf{max}_j\, D(N, t^j)$

where $logb(w_i|t^j) = \log b(w_i|t^j)$ and so forth

HTK Toolkit

- Nástroj slúžiaci na prácu s HMM(Hidden Markov Model)

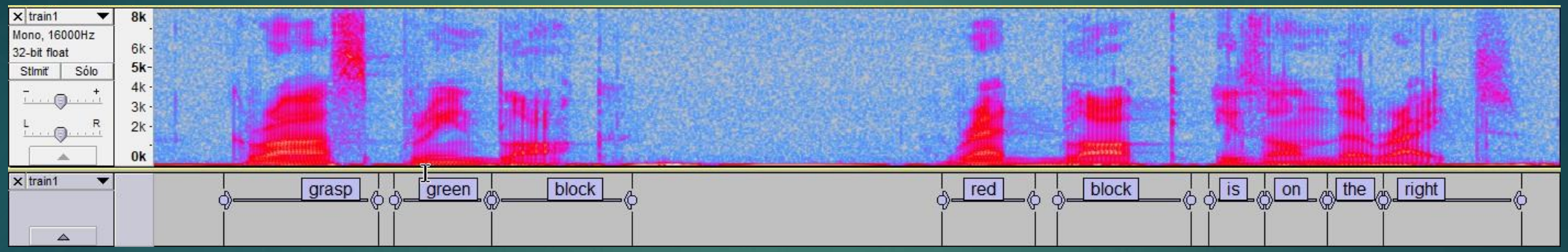



```
~o <VecSize> 39 <MFCC_0_D_A>
~h "proto"
<BeginHMM>
 <NumStates> 5
 <State> 2
    <Mean> 39
      0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
    <Variance> 39
      1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 3
    <Mean> 39
      0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
    <Variance> 39
      1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 4
    <Mean> 39
      0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
    <Variance> 39
      1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <TransP> 5
0.0 1.0 0.0 0.0 0.0
0.0 0.6 0.4 0.0 0.0
0.0 0.0 0.6 0.4 0.0
0.0 0.0 0.0 0.7 0.3
0.0 0.0 0.0 0.0 0.0
<EndHMM>
```

Slovná zásoba

```
$color = red | blue | green | yellow;
$verb = put | grasp | move | give-me;
$object = cube | block | cylinder;
$adv = left | right;

(($verb $color $object) | ($color $object is ((in the middle) | (on the $adv))) | (where is $color $object))
```



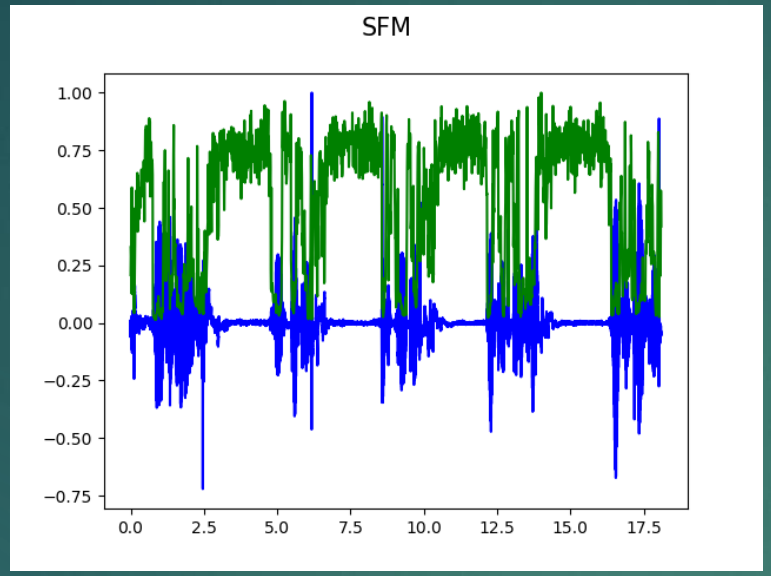
- 13m 36s trénovacích nahrávok od 6 rečníkov, spolu 324 viet

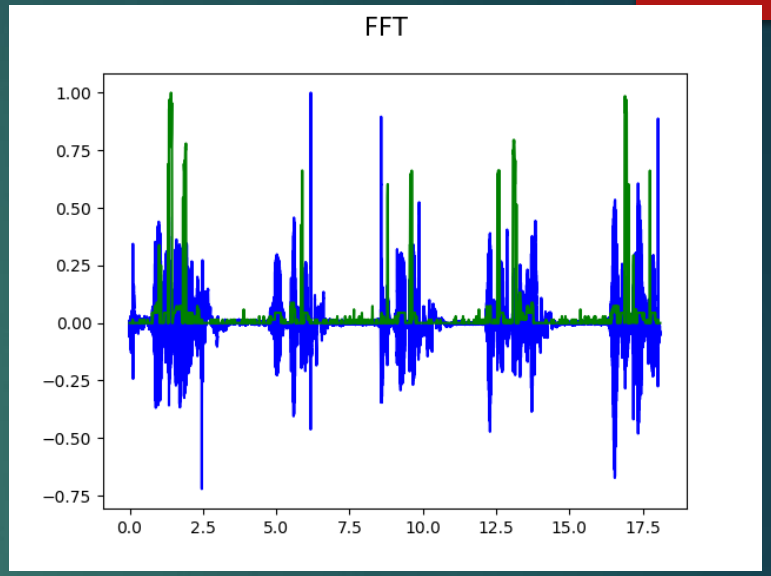
Návrh hmm

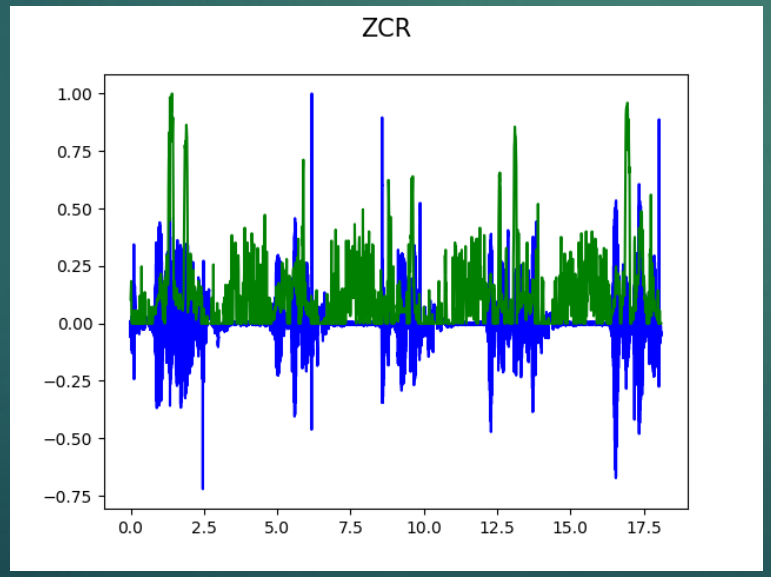
- Aktuálny stav:
  - 5-stavové (z toho 3 generujúce príznaky), ľavo-pravé modely s jedným gausiánom pre celé slová
- Navrhované vylepšenia:
  - Počet stavov podľa foném v slove
  - Postupné rozširovanie počtu gausiánov v stavoch a testovanie úspešnosti
  - Navrhnúť modely pre trifóny a spájať ich do slov

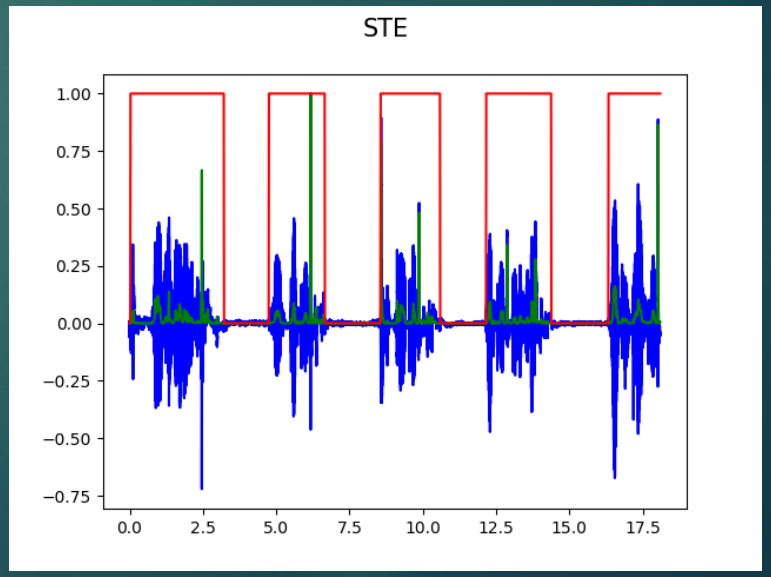
- SENT: %Correct=94.95 [H=188, S=10, N=198]
  WORD: %Corr=98.48, Acc=98.48 [H=780, D=0, S=12, I=0, N=792]

VAD


SFM
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


FFT
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


ZCR
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


STE
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

Ďakujem za pozornosť!