

**TECHNICKÁ UNIVERZITA V KOŠICIACH
FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

Metódy zhlukovania rečových segmentov

Marek Ečegi

Vedúci práce: Ing. Peter Veszlay, PhD.

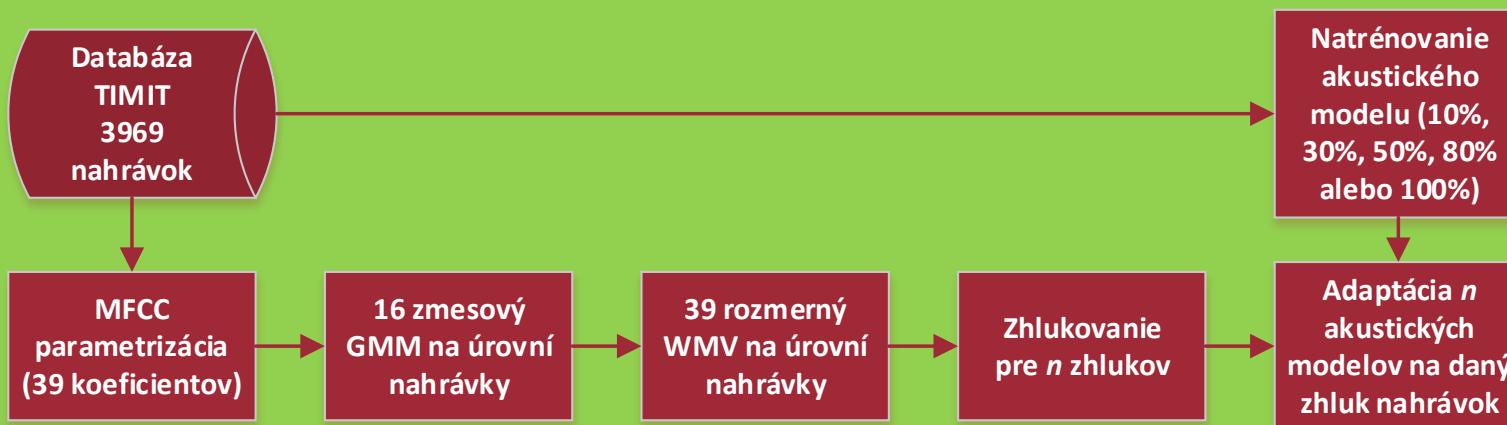
KEMT FEI TU v Košiciach

Hlavné úlohy a ciele - Diplomová práca I. (ZS, 2014/2015):

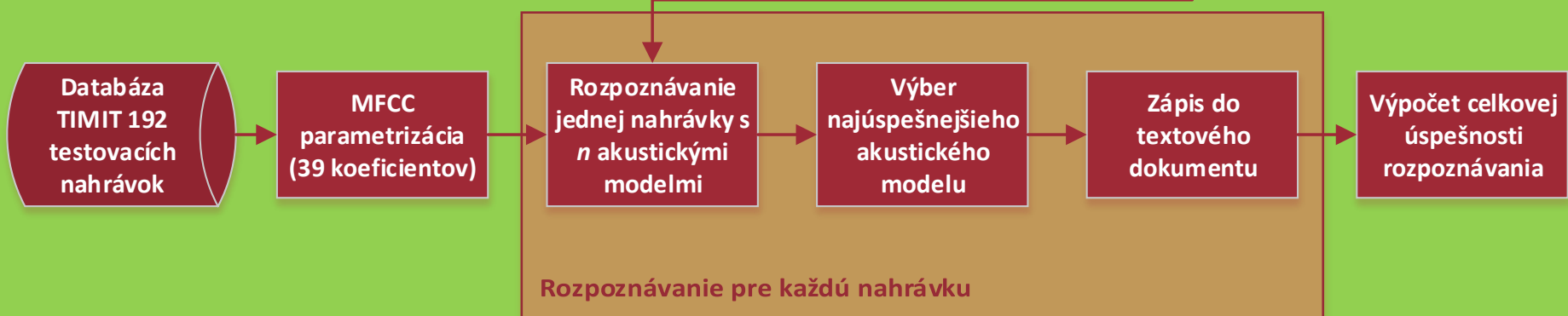
- Návrh systému na využitie vopred vybraných metód zhlukovania v automatickom rozpoznávaní reči (ARR)
- Definovanie a voľba vhodných štatistických reprezentantov na zhlukovanie
- Návrh postupov na vylepšenie presnosti systému ARR adaptáciou akustických modelov s využitím metód zhlukovania
- Implementácia navrhnutých a zvolených postupov v prostredí Matlab
- Iniciálne testovanie navrhnutých postupov v úlohe rozpoznávania reči pomocou databázy TIMIT
- Analýza a vyhodnotenie výsledkov

Bloková schéma navrhnutého systému:

Trénovací proces



Testovací proces



Výber najlepšieho prístupu adaptácie akustického modelu na daný zhluk:

- Testovanie najvhodnejšej adaptácie akustického modelu (v zmysle reestimácie parametrov akustického modelu)
- Testovali sme 2 prístupy:
 - 1. Konvenčná reestimácia parametrov AM (viac iterácií, s prerezávaním, bez prerezávania)
 - 2. Pretrénovanie parametrov AM pomocou špeciálneho režimu jedno-prechodového tréningu (s prerezávaním, bez prerezávania)
- Najvyššia úspešnosť pri Retrain i3

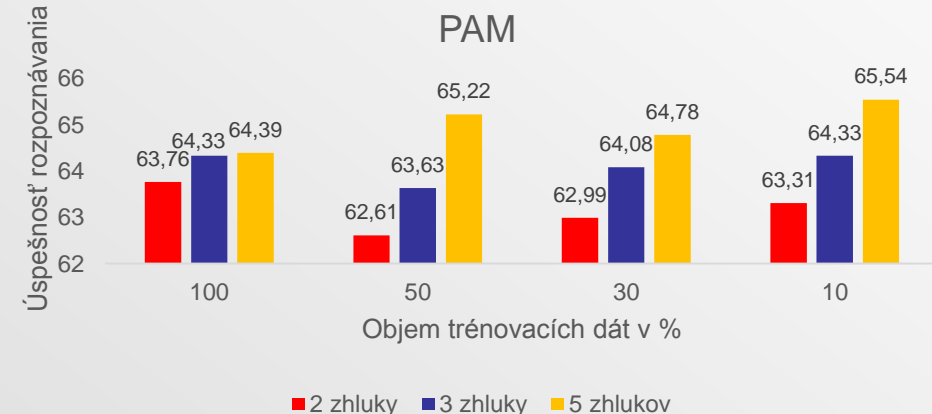
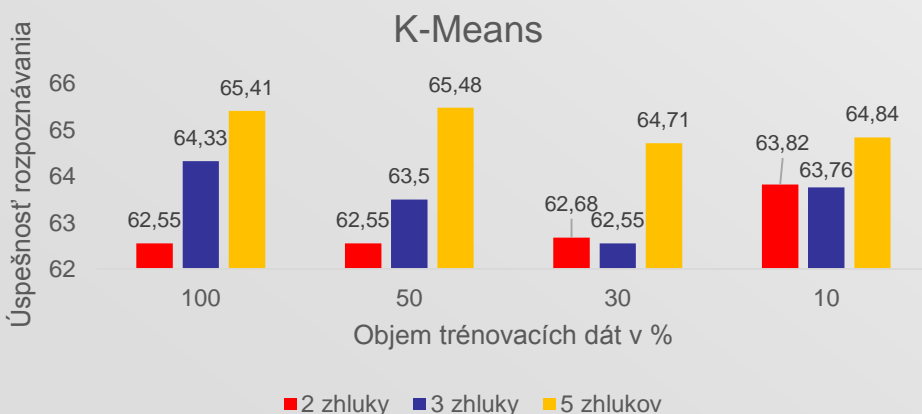
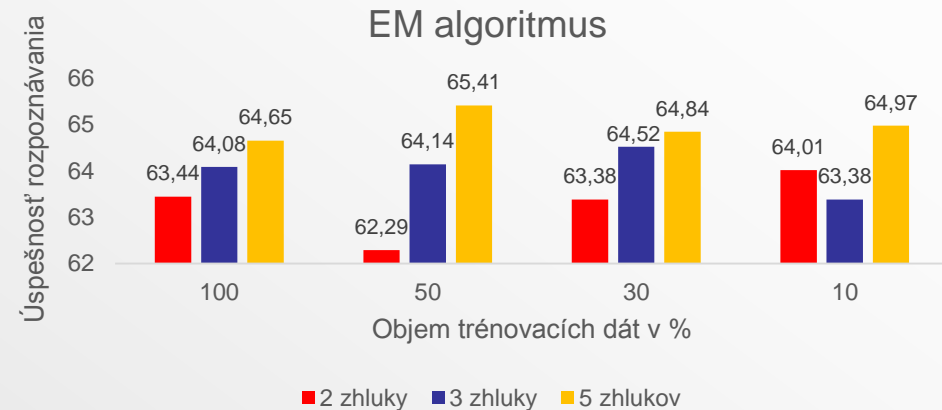
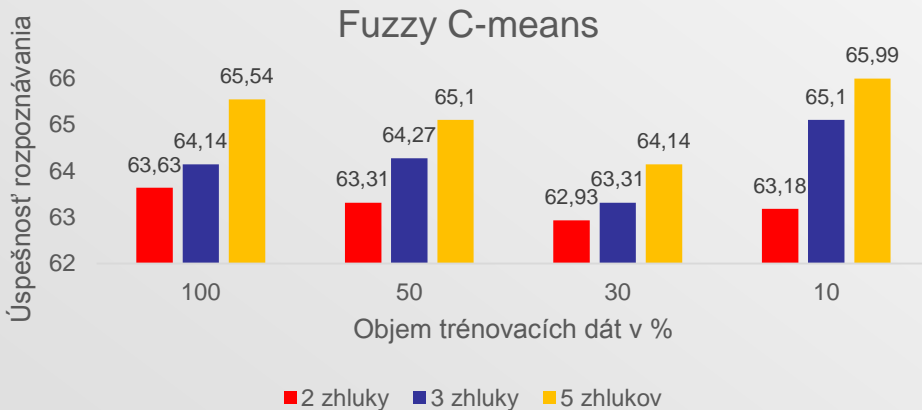
10% Fuzzy C-means

Zhluky	Retrain i1 prunON	Retrain i1 prunOFF	Retrain i2 prunON	Retrain i2 prunOFF	Retrain i3 prunON	Retrain i3 prunOFF	SPR m1 prunON	SPR m1 prunOFF	SPR m3 prunON	SPR m3 prunOFF
2	62,42	62,42	63,82	63,82	64,01	64,01	62,42	62,42	62,42	62,42

- Navrhnutý postup sme testovali pre dve AM: 32 a 128 zmesové
- Adaptovaných bolo celkovo 5 druhov AM: 10, 30, 50, 80 a 100%
- 80% adaptované AM dávajú rovnaké výsledky ako 100% adaptované AM
- Jednotlivé adaptované AM mi boli poskytnuté vedúcim práce
- AM sú založené na báze anglických monofón

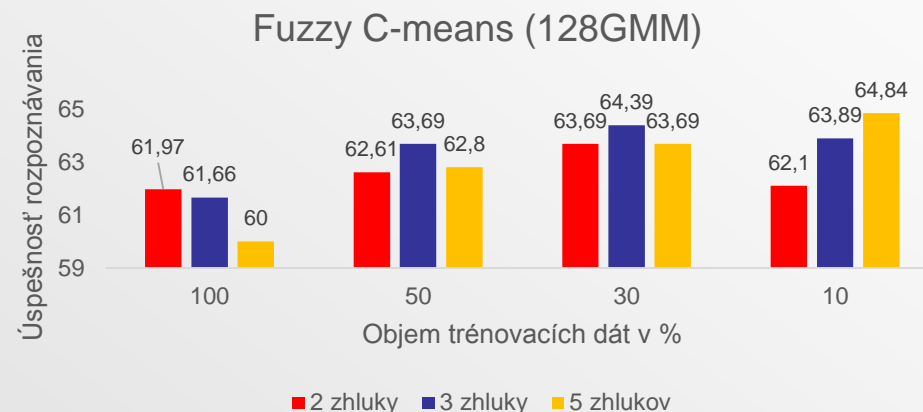
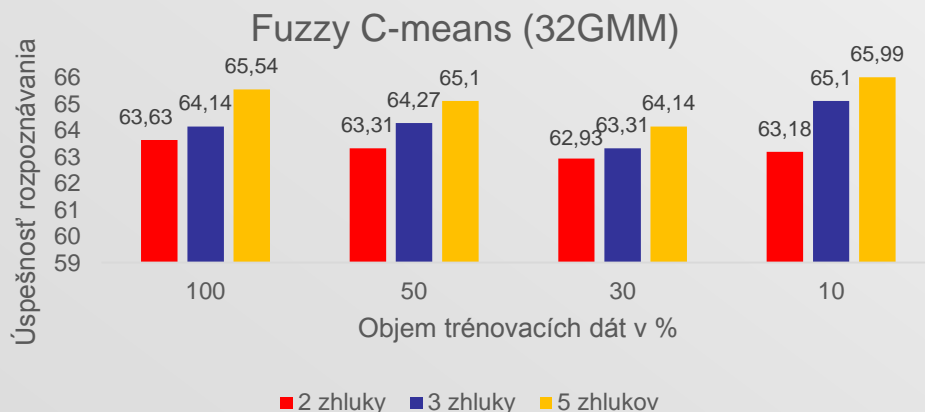
Výsledky testovania 32-zmesového akustického modelu:

- Referenčná hodnota: 61,02% (úspešnosť rozpoznávania so 100% AM)



Porovnanie 32-zmesového a 128-zmesového akustického modelu:

- Najlepšie výsledky dosiahol prístup založený na Fuzzy C-means preto pre porovnanie 32 a 128 zmesového modelu uvádzam výsledky získané týmto prístupom
- Predpokladali sme, že úspešnosť rozpoznávania pre 128-zmesové AM vzrastie, avšak vzrástla iba referenčná hodnota úspešnosti s plne natrénovaným 100% AM



- $65,99\% - 61,02\% = 4,97\%$ nárast úspešnosti rozpoznávania s využitím metódy Fuzzy C-means

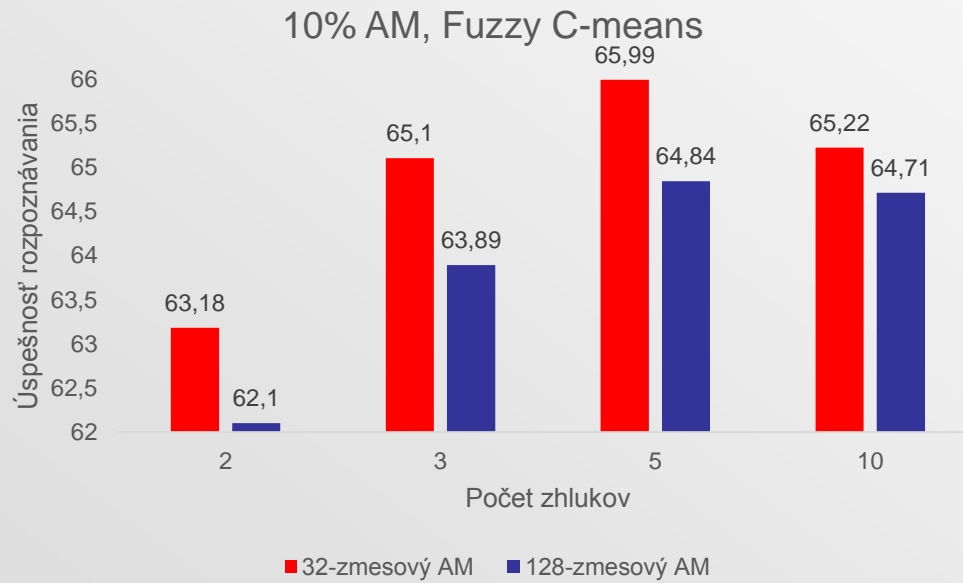
Príklad výsledného textového dokumentu:

- V každom riadku je uvedený:
- 1. názov nahrávky
- 2. najvyššia úspešnosť dosiahnutá pri rozpoznávaní danej nahrávky
- 3. názov adaptovaného AM s ktorým bola dosiahnutá najvyššia úspešnosť
- Na konci celého textového dokumentu je výsledná úspešnosť rozpoznávania

```
dr8_mpam0_si1189 0.2857 adapted_timit_10p_32gm_idx_FUZZY_cluster_2_18_retrain_i3_prunON.mmf
dr8_mpam0_si1819 0.5556 adapted_timit_10p_32gm_idx_FUZZY_cluster_2_18_retrain_i3_prunON.mmf
dr8_mpam0_si1961 0.6667 adapted_timit_10p_32gm_idx_FUZZY_cluster_4_18_retrain_i3_prunON.mmf
dr8_mpam0_sx109 0.9 adapted_timit_10p_32gm_idx_FUZZY_cluster_1_30_retrain_i3_prunON.mmf
dr8_mpam0_sx19 1 adapted_timit_10p_32gm_idx_FUZZY_cluster_1_30_retrain_i3_prunON.mmf
dr8_mpam0_sx199 1 adapted_timit_10p_32gm_idx_FUZZY_cluster_1_30_retrain_i3_prunON.mmf
dr8_mpam0_sx289 1 adapted_timit_10p_32gm_idx_FUZZY_cluster_1_30_retrain_i3_prunON.mmf
dr8_mpam0_sx379 1 adapted_timit_10p_32gm_idx_FUZZY_cluster_1_30_retrain_i3_prunON.mmf
65.99
```


Vplyv zvyšovania počtu zhlukov na rozpoznávanie reči:

- Testovanie iba pre jeden typ adaptovaného AM – 10% pomocou Fuzzy C-means
- Testovalo sa aj pre 20 zhlukov, ale v tomto prípade už nastali problémy pri rozpoznávaní



Zoznam úloh na LS 2014/2015:

- Použitie slovenskej rečovej databázy
- Otestovať GMM supervektory namiesto WMV
- Dokončiť zhlukovanie na báze GMM modelov
- Využitie akustických modelov závislých na pohlaví (gender dependent) v zhlukovaní
- Písanie diplomovej práce

Ďakujem za vašu pozornosť .

- S. Bandyopadhyay, S. Saha. *Unsupervised Classification*, Springer-Verlag Berlin Heidelberg 2013
- Marina Meila. *Clustering – A tutorial overview*, University of Washington, MLSS Taipei 2006
- Apsingekar V. R. – Efficient speaker identification using speaker model clustering, New Mexico State University 2008
- The HTK Book (for HTK Version 3.4)
- <http://www.mathworks.com/>