Automatické hodnocení anglické výslovnosti nerodilých mluvčích

Diplomová práca

Peter Gazdík

30. augusta 2019

Automatické hodnotenie výslovnosti

Druhy chýb

- **segmentálne chyby**,
- prozodické chyby.

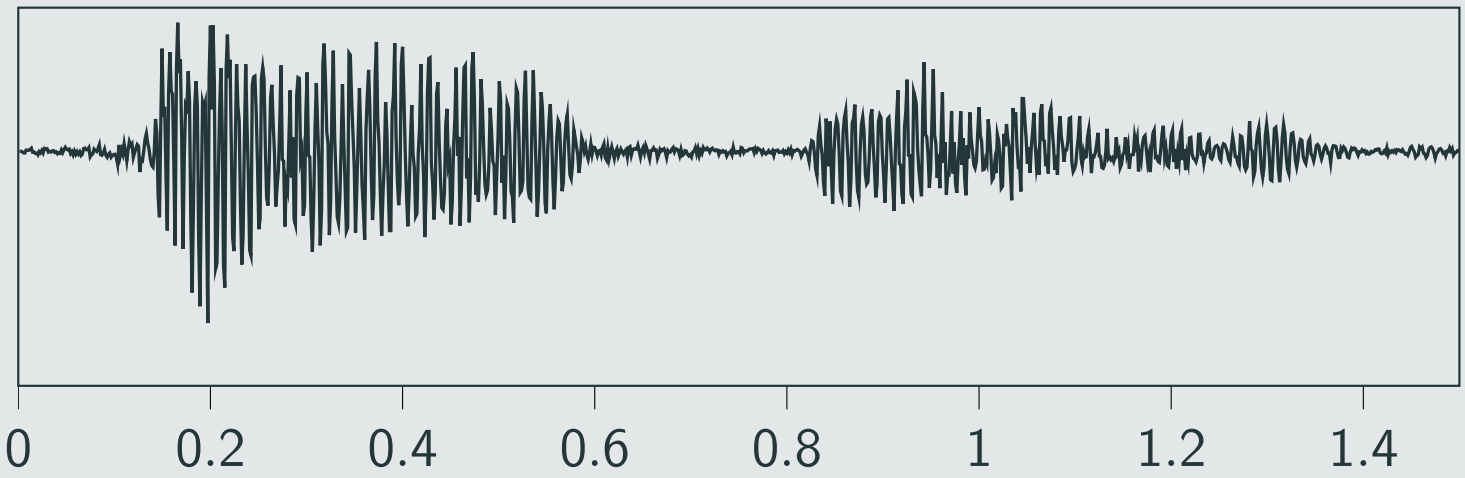
Úrovne hodnotenia výslovnosti

- **detekcia**,
- diagnostika.

Hodnotenie segmentálnych chýb

Text	Fonémový prepis
hello world	həlˈəʊ wˈɜːld

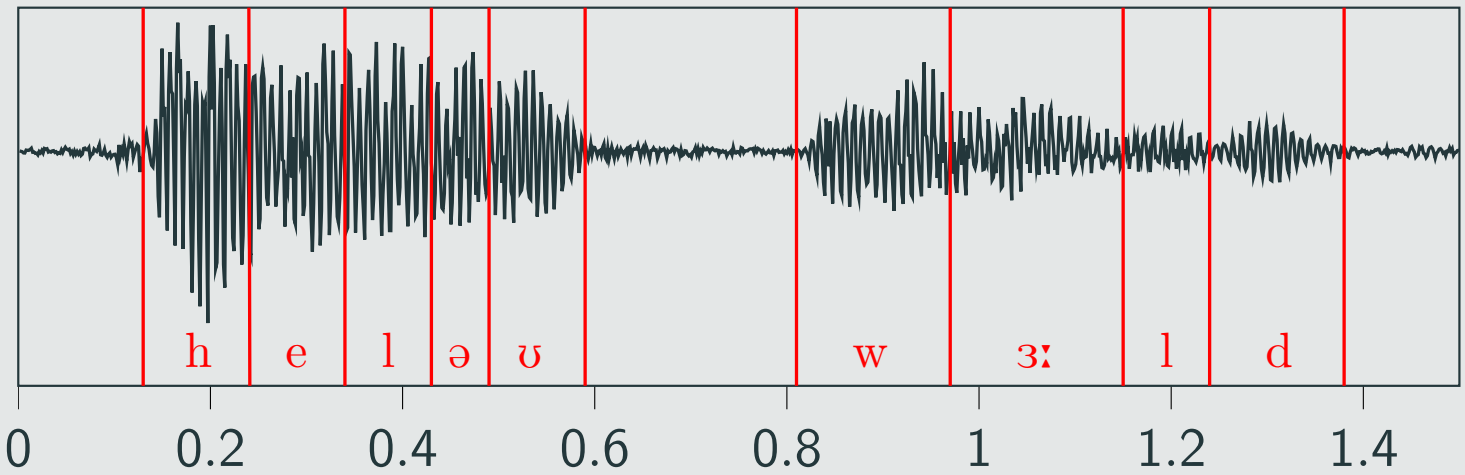
Nahrávka



Hodnotenie segmentálnych chýb

Text	Fonémový prepis
hello world	heləʊ wɜːld

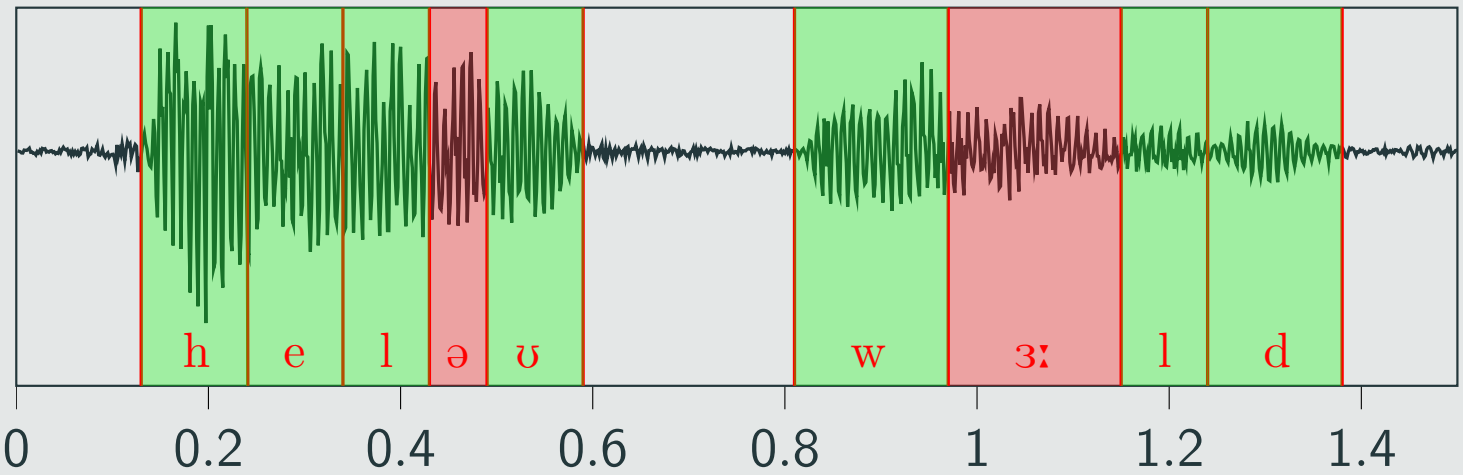
Nahrávka



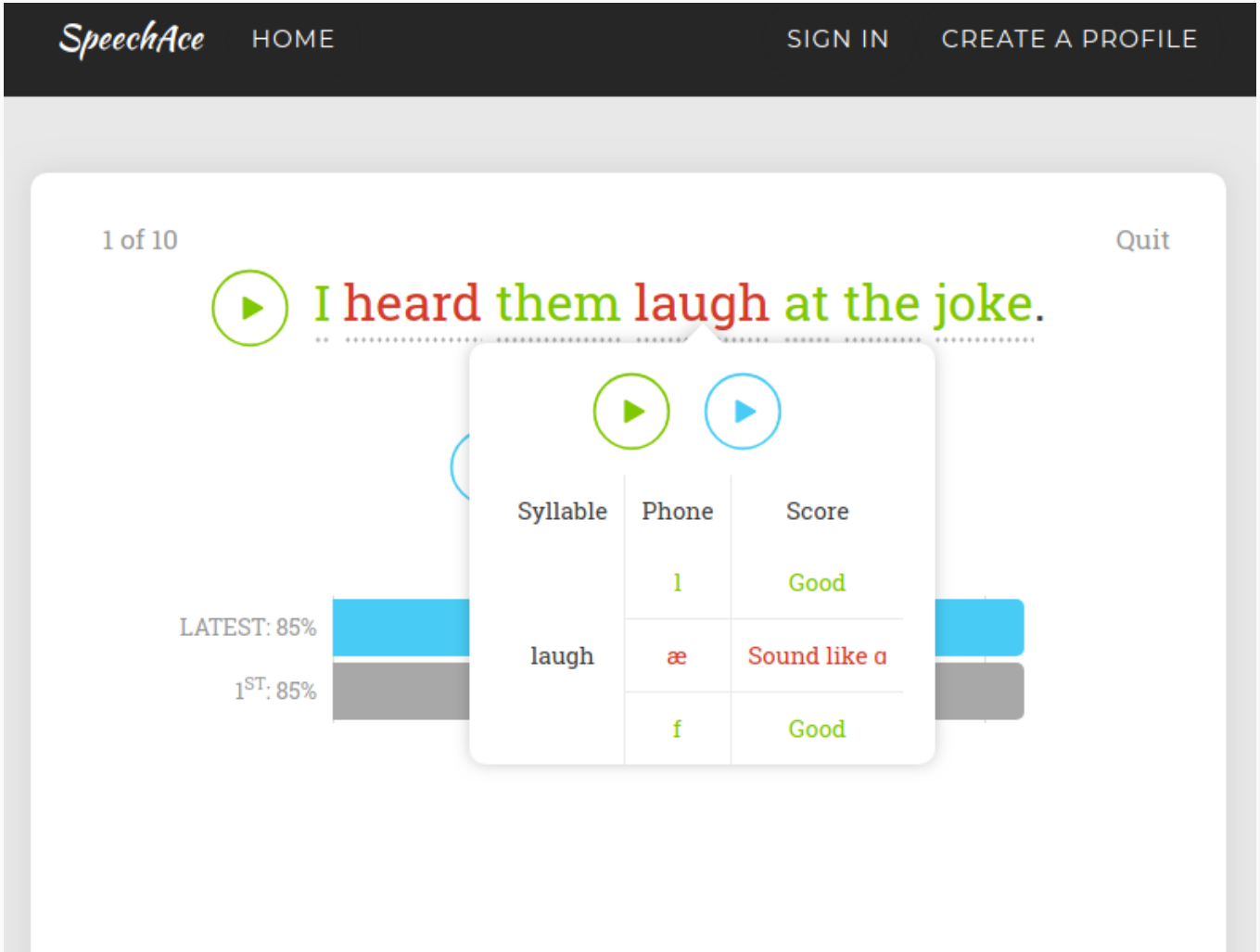
Hodnotenie segmentálnych chýb

Text	Fonémový prepis
hello world	heləʊ wɜːld

Nahrávka



Hodnotenie segmentálnych chýb: Existujúci nástroj



Obr. 1: Webové rozhranie nástroja SpeechAce.

Návrh systému: Metódy

Metódy

1. Goodness of Pronunciation (GOP) skóre

$$\text{LR GOP}(p) = \log\left(\frac{p(\boldsymbol{O}^{(p)}|p)}{\max_{q\in Q, q\neq p} p(\boldsymbol{O}^{(p)}|q)}\right) \Big/ d.$$

2. Klasifikačné metódy:
   - **Klasifikátor**: neurónové siete,
   - **Príznaky**
     - vierohodnosti HMM stavov,
     - pravdepodobnosti fonologických rysov.

Experimenty

Dataset

- ISLE – dataset nenatívnej angličtiny (talianský a nemecký rečníci).

Vyhodnotenie

- **ROC krivka** – závislosť medzi mierou chybného prijatia (FAR) a mierou chybného odmietnutia (FRR)
  - **FAR** – pomer medzi počtom zle vyslovených foném klasifikovaných ako správne vyslovené k celkovému počtu zle vyslovených,
  - **FRR** – pomer medzi počtom správne vyslovených foném klasifikovaných ako nesprávne vyslovené k celkovému počtu správne vyslovených.
- **Rovnaká miera chyby (EER)**
  - $EER := FAR = FRR$

Experimenty

1. Porovnanie základných metód s ref. prácou.
2. Porovnanie monofónového a trifónového AM.
3. Multilingválne trénovanie.
4. Porovnanie rôznych GOP skóre.
5. Porovnanie dopredných a LSTM neurónových sietí.

Experiment 1: Porovnanie základných metód s ref. prácou

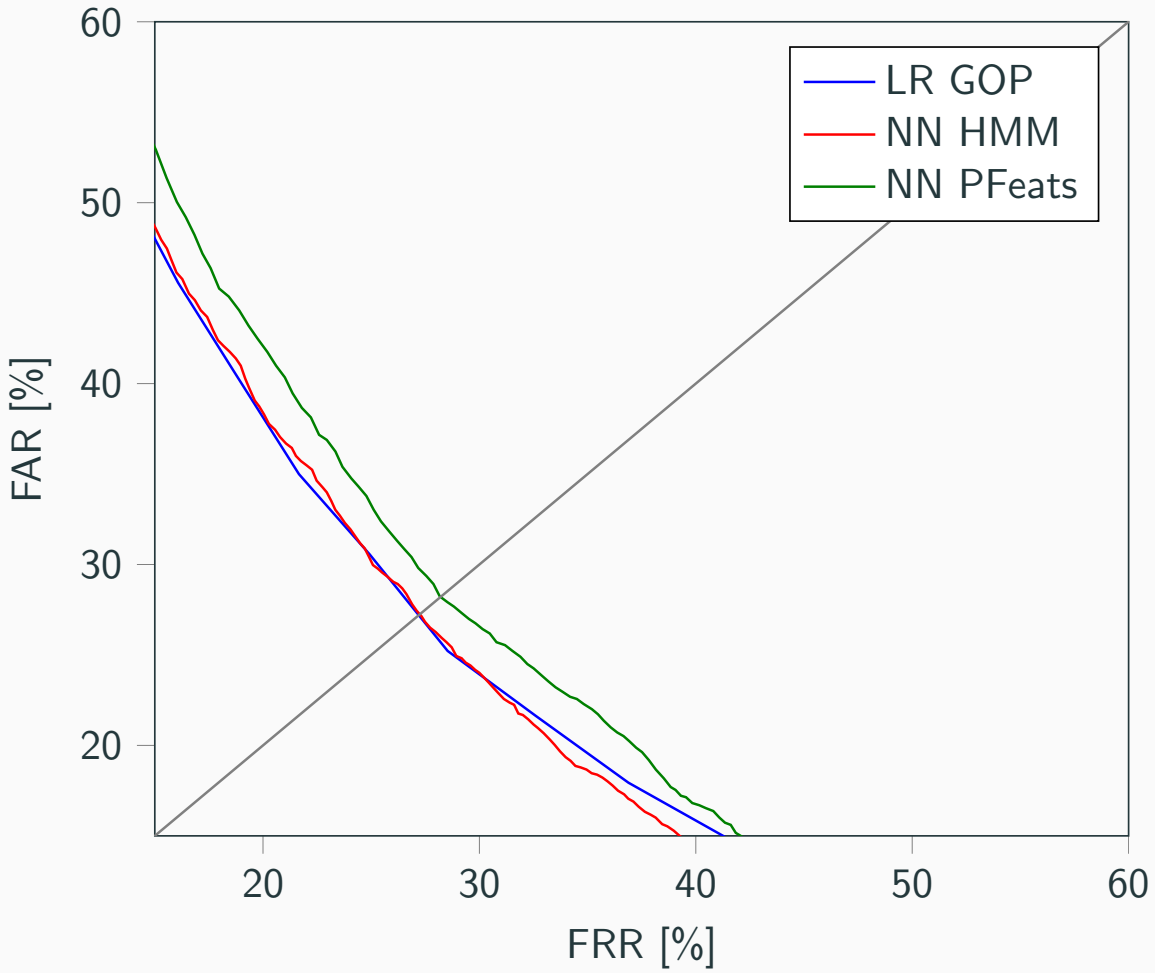
Metódy

- LR GOP skóre,
- neurónová sieť + vierohodnosti HMM stavov (NN HMM),
- neurónová sieť + pravdepodobnosti fonologických rysov (NN Pfeats).

Rozpoznávač reči

- monofónový DNN-HMM model.

Experiment 1: Porovnanie základných metód s ref. prácou



Obr. 2: Graf závislosti FAR a FRR pre základné metódy.

Experiment 1: Porovnanie základných metód s ref. prácou

EER [%]	LR GOP	NN HMM	NN PFeats
Navrhnutý systém	**27,23**	**27,57** $\pm$ 0,16	28,49 $\pm$ 0,14
Referenčný systém[1]	39,00	31,80	**28,30**

Tabuľka 1: Dosiahnuté výsledky pomocou základných metód.

[1] Arora, V.; Lahiri, A.; Reetz, H.: Phonological Feature Based Mispronunciation Detection and Diagnosis using Multi-Task DNNs and Active Learning. 2017.

Experiment 2: Porovnanie monofónového a trifónového AM

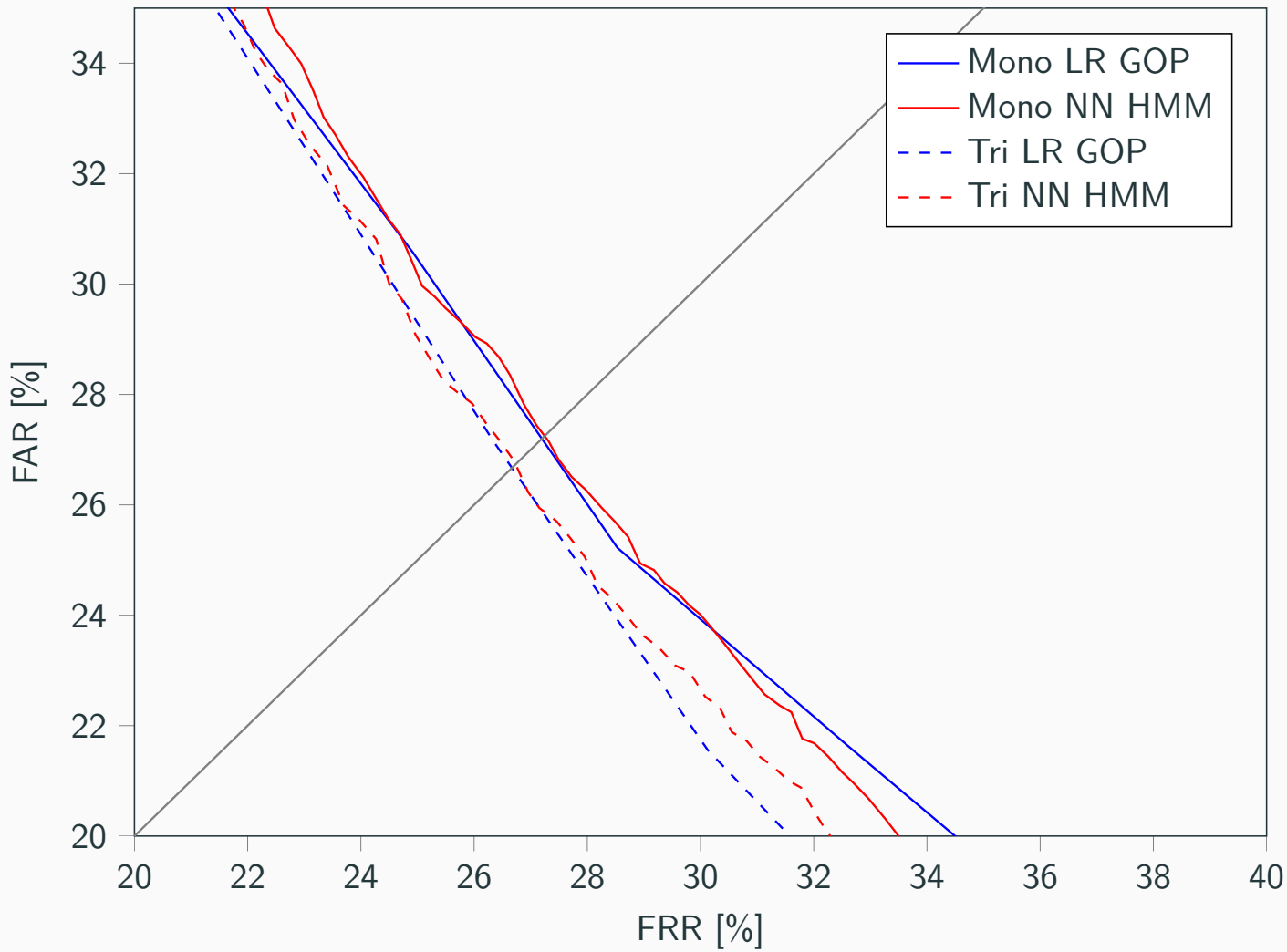
Metódy

- LR GOP skóre,
- neurónová sieť + vierohodnosti HMM stavov (NN HMM).

Rozpoznávač reči

- DNN-HMM monofónový akustický model,
- DNN-HMM trifónový akustický model.

Experiment 2: Porovnanie monofónového a trifónového AM



Obr. 2: Graf závislosti FAR a FRR pri použití monofónového a trifónového AM.

Experiment 2: Porovnanie monofónového a trifónového AM

EER [%]	LR GOP	NN HMM
Mono AM	27,23	$27{,}57 \pm 0{,}16$
Tri AM	**26,91**	**27,06** $\pm 0{,}18$

Tabuľka 2: Dosiahnuté výsledky pri použití monofónového a trifónového AM.

Experiment 3: Multilingválne trénovanie

Jazyky

- natívne jazyky: angličtina, nemčina, taliančina,
- nenatívne jazyky: angličitina.

Metódy

- Mono LR GOP,
- Tri LR GOP.

Experiment 3: Multilingválne trénovanie

EER [%]	Mono LR GOP	Tri LR GOP
NN-EN	27,75	26,91
EN $\rightarrow$ NN-EN	27,27	26,35
DE $\rightarrow$ NN-EN	**26,55**	26,63
IT $\rightarrow$ NN-EN	26,71	26,35
EN $\rightarrow$ DE $\rightarrow$ NN-EN	**26,55**	26,43
EN $\rightarrow$ IT $\rightarrow$ NN-EN	26,83	26,47
EN $\rightarrow$ DE $\rightarrow$ IT $\rightarrow$NN-EN	**26,55**	**25,78**
EN $\rightarrow$ IT $\rightarrow$ DE $\rightarrow$NN-EN	26,67	26,31

Tabuľka 3: Dosiahnuté výsledky pri použití multilingválnych akustických modeloch.

Experimenty: Zhrnutie výsledkov

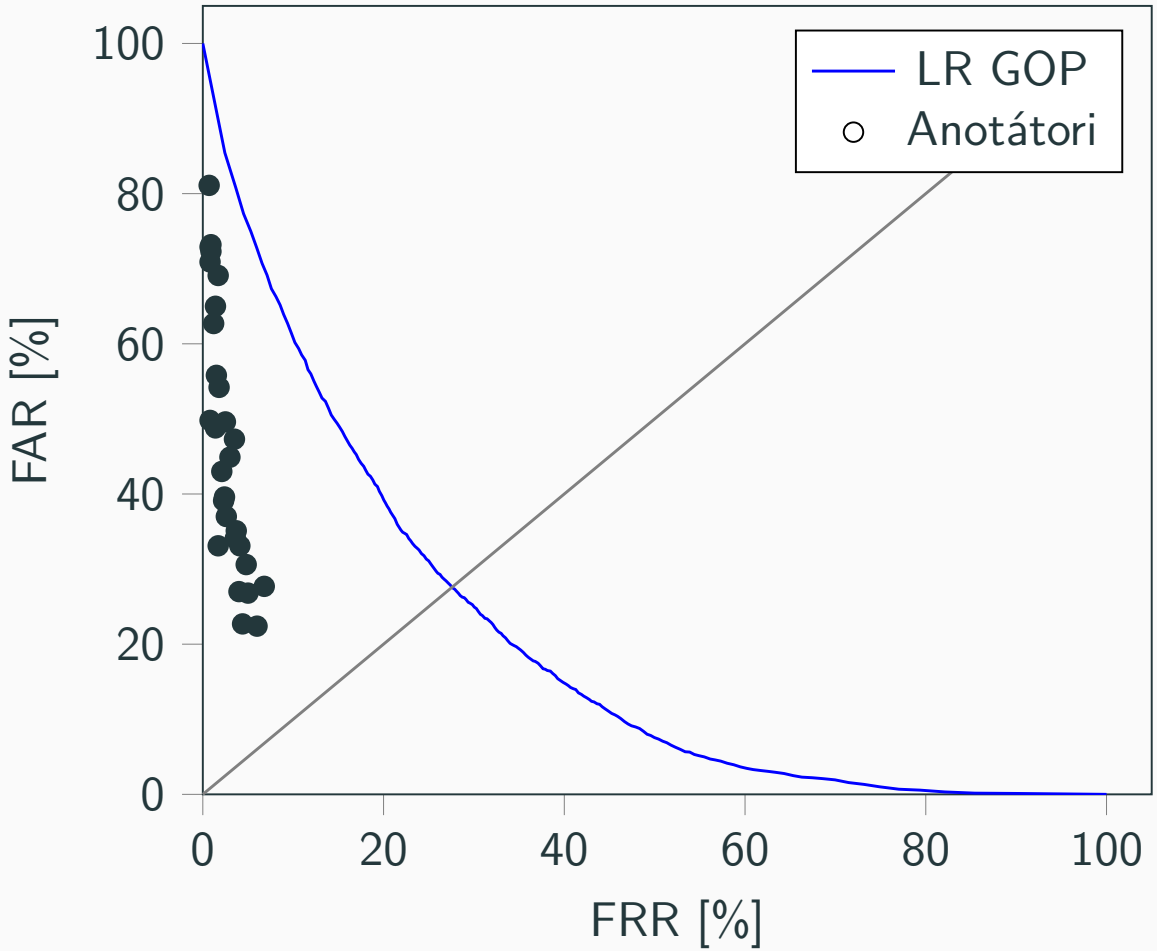
- Najlepšia metóda: LR GOP skóre.
- Ďalšie zlepšenie vďaka použitiu trifónového AM a multilingválneho trénovania.
- Najepší výsledok: EER = 25,78 % (ref. systém 28,30 %).

Analýza datasetu

Postup

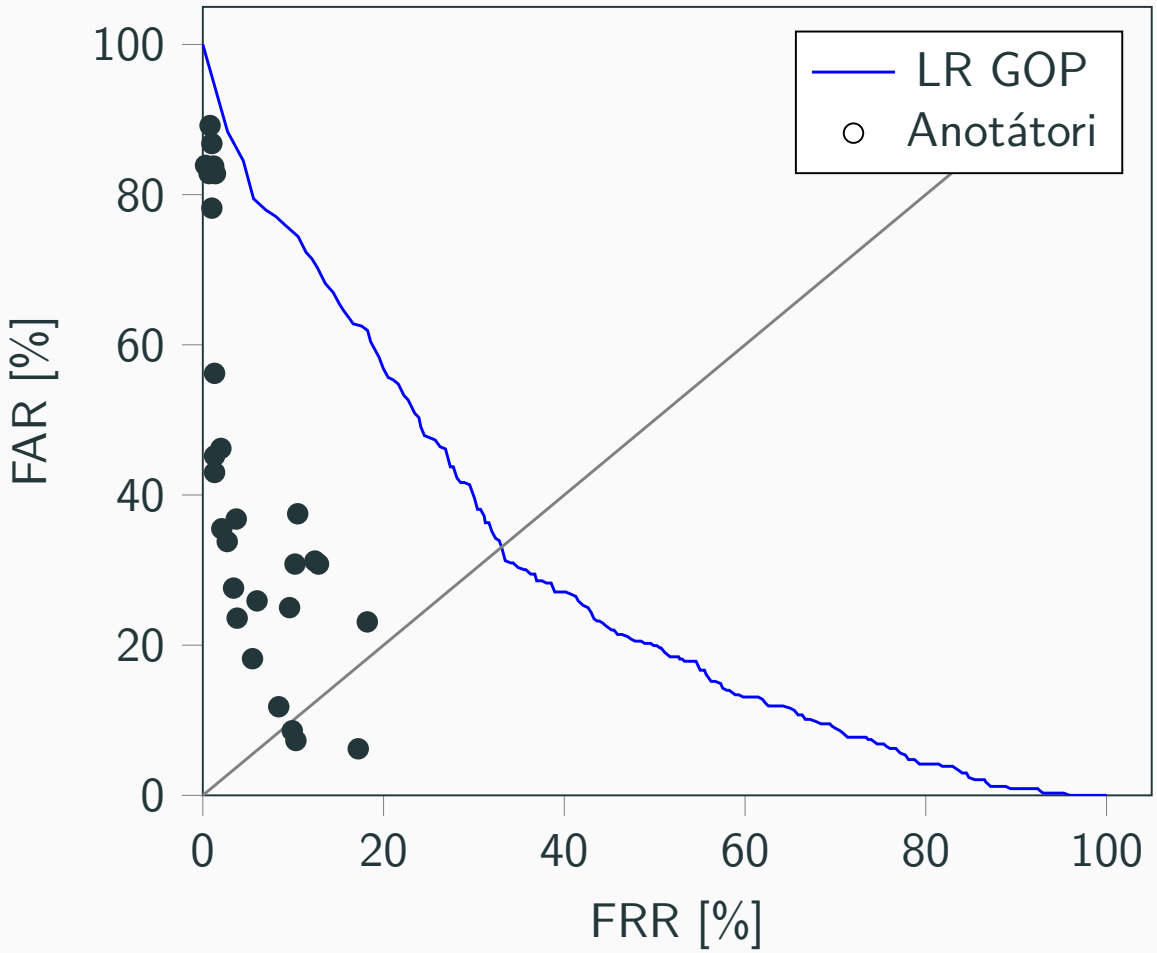
- Časť datasetu anotovaná viacerými anotátormi súčasne,
- FAR a FRR po dvojiciach anotátorov, hodnotenie jedného anotátora je uvažované ako referenčné.

Analýza datasetu



Obr. 2: Hodnoty FAR a FRR určené po dvojiciach anotátorov.

Analýza datasetu



Obr. 2: Hodnoty FAR a FRR určené po dvojiciach anotátorov – fonéma ə.

Otázky

Otázka oponenta 1

Otázka

V předposledním odstavci na straně 27 píšete, že pro multi-class výstup používáte na výstupní vrstvě softmax. Jasně prosím zdůvodněte motivaci této architektury a tříd které klasifikujete (vstupní a výstupní data).

- Klasifikácia binárnych fonologických rysov (multi-label N z M),
- **Vstupy:** fbank príznaky jedného rámca reči,
- **Výstupy:** 19 fonologických rysov.

Softmax

$$f_j(a_1, a_2, \ldots, a_m) = \frac{e^{a_j}}{\sum_{k=1}^{M} e^{a_k}}$$

Logistická sigmoida

$$f(a_j) = \frac{1}{1 + e^{-a_j}}$$

Otázka oponenta 2

Otázka

Popište stručně včem přesně se liší referenční a váš systém, a z čeho podle vás vzniklo zlepšení, kterého jste dosáhl.

GOP skóre

Navrhnutý systém

$$\mathrm{GOP}(p) = \log\left(\frac{p(\boldsymbol{O}^{(p)}|p)}{\max_{q\in Q, q\neq p} p(\boldsymbol{O}^{(p)}|q)}\right) \Big/ d.$$

Referenčný systém

$$\mathrm{GOP}(p) = \log\left(\frac{p(p|\boldsymbol{O}^{(p)})}{\max_{q\in Q, q\neq p} p(q|\boldsymbol{O}^{(p)})}\right) \Big/ d.$$

Otázka oponenta 2

NN HMM

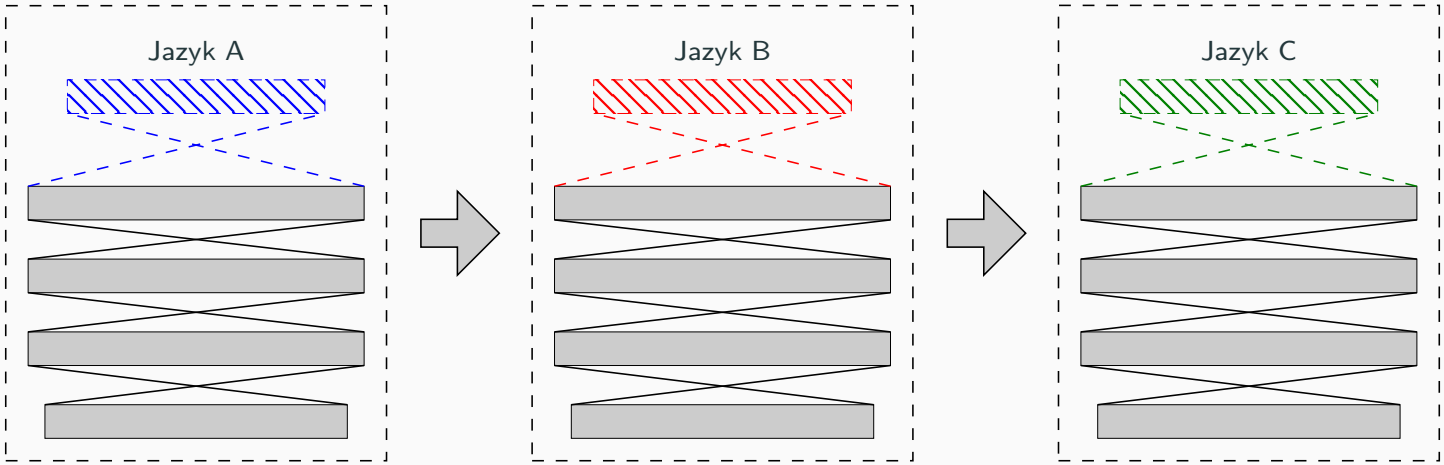
- odlišná objektívna funkcia – crossentropia vs. kvadratická stredná odchýlka,
- normalizácia príznakov (nulová stredná hodnota a jednotkový rozptyl).

NN PFeats

- porovnateľné výsledky,
- odlišná objektívna funkcia – crossentropia vs. kvadratická stredná odchýlka,
- normalizácia príznakov (nulová stredná hodnota a jednotkový rozptyl).

Princíp multilingválneho trénovania

Princíp



Obr. 3: Postupné, sekvenčné trénovania DNN akustického modelu na viacerých jazykoch.