

# SPEAKER'S GENDER RECOGNITION: THE INFLUENCE OF AGE AND EMOTIONAL STATE

**Tomáš Rendek**

Master Degree Programme (2), FEEC BUT  
E-mail: xrende00@stud.feec.vutbr.cz

Supervised by: Hicham Atassi

E-mail: atassi@feec.vutbr.cz

## ABSTRACT

The presented paper describes two experiments made on speaker's gender recognition. The first experiment deals with the influence of speaker's age on the classification process. The second experiment is on finding the impact of the emotional state on the accuracy of gender recognition. Two different feature sets namely: segmental and suprasegmental are exploited and the gender classification is performed by using Gaussian Mixture Model (GMM) with diagonal covariance matrix.

## 1. ÚVOD

Reč ako základný dorozumievací prostriedok ľudí v sebe skrýva množstvo informácií, charakterizujúcich rečníka. Jedná sa hlavne o informácie charakterizujúce vek, pohlavie alebo emočné rozpoloženie. Algoritmy umožňujúce získať uvedené informácie nachádzajú široké uplatnenie v každodennej praxi. Príkladom môžu byť call centrá, v ktorých sa každodenne odohrá niekoľko tisíc hovorov ktoré je nutné indexovať. Vlastné primárne experimenty ukázali, že na presnosť odhadu pohlavia výrazne vplýva ako vyjadrenie emócie tak aj vek rečníka. Oba experimenty sú zhrnuté na nasledujúcich riadkoch.

## 2. ROZBOR EXPERIMENTOV

Odhad pohlavia rečníkov z hlasových nahrávok spočíval vo výbere príznakov, natrénovaníu klasifikátora a samotnej klasifikácii. Pre výpočet kvality príznakov sme použili metódu  $F$ -pomer, ktorá vychádza z výpočtu štatistických údajov o triedach. Príznak je považovaný za kvalitný, ak sa prvky jednej triedy vyskytujú v okolí strednej hodnoty a súčasne stredné hodnoty uvažovaných tried sa čo najviac líšia. Postupne sme tak pridávali príznaky a sledovali dosiahnutú úspešnosť. Experimenty boli vykonané v prostredí Matlab.

### 2.1. DATABÁZE NAHRÁVOK

Ako podklad pre náš výskum nám poslúžili nasledujúce databázy nahrávok:

- VAGD – česká databáza 408 nahrávok, nahovorených rodenými hovorcami bez vyjadrenia emócie, rozdelená do nasledovných piatich vekových skupín [1]:

A1: 7 až 12 rokov, A2: 13 až 15 rokov, A3: 16 až 20 rokov, A4: 21 až 30 rokov, A5: 30 až 50 rokov.

- Emo-DB – nemecká databáza siedmich emočných stavov človeka nahovorená hercami. Pozostáva z 535 vzoriek predstavujúcich súvislé vety [1].

## 2.2. PRÍZNAKY A KLASIFIKÁTOR

V experimente sme sa zamerali na nasledujúce dve skupiny príznakov:

- Suprasegmentálne príznaky vychádzajúce z priebehov základného tónu reči, prvých troch formantov a ich šíriek pásma.
- Segmentálne príznaky – mélovské kepstrálne koeficienty (MFCC), percepčné lineárne predikčné koeficienty (PLP).

V roli klasifikátora vystupuje Gaussov model zmesí (GMM). Jeho výhodou je, že je všeobecne známy a jeho chovanie bolo dôkladne preskúmané [2]. Princíp GMM klasifikátora je založený na modelovaní príznakov z tréningovej množiny jednou alebo viacerými Gaussovými funkciami rozloženia pravdepodobnosti [3]. Parametre klasifikátora boli odvodené z predchádzajúcich experimentov [4], kde bolo dosiahnutých najlepších výsledkov pri použití diagonálnej kovariančnej matice a jednej Gaussovej funkcie.

## 3. POPIS EXPERIMENTOV

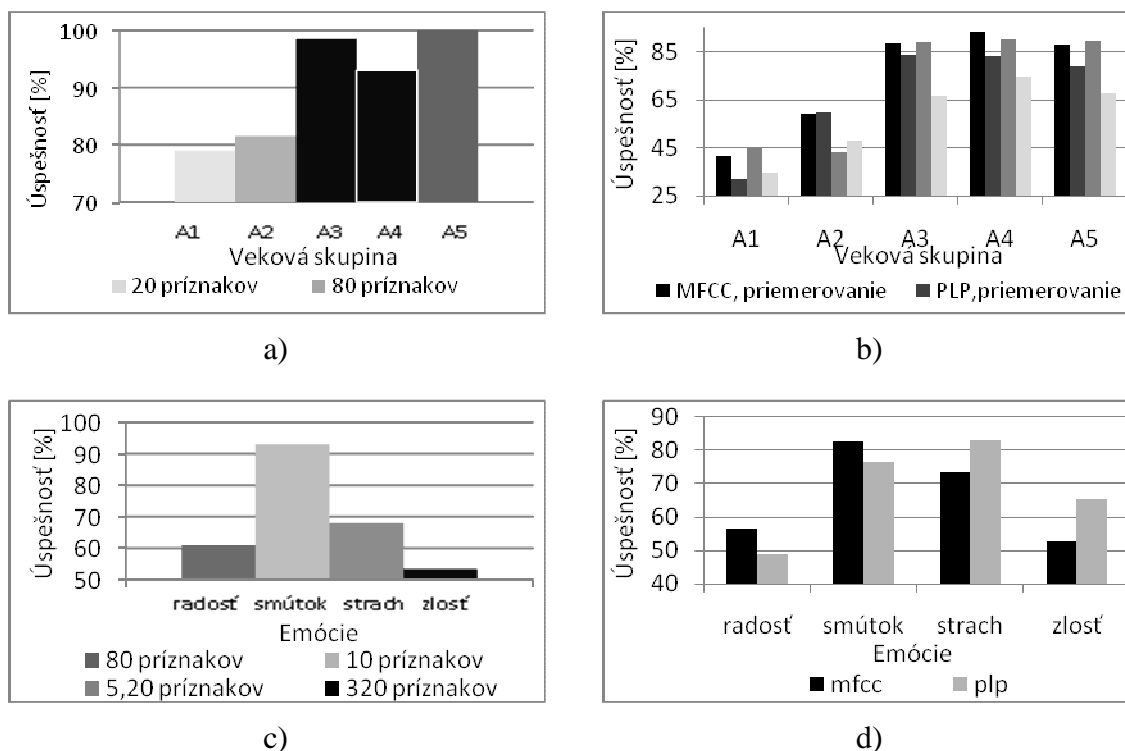
### 3.1. VPLYV VEKU NA ÚSPEŠNOSŤ ODHADU POHLAVIA

Základ tvorila databáza VAGD rozdelená do piatich vekových skupín (A1 až A5). Ďalším cieľom experimentov bolo porovnať presnosť odhadu pri oddelenom použití suprasegmentálnych a segmentálnych príznakov. V prípade suprasegmentálnych príznakov sme menili počet uvažovaných príznakov v celom ich rozsahu na základe metódy  $F$ -pomer. Najnižšia úspešnosť rozpoznania bola dosiahnutá u detí a mladistvých (A1, A2), pričom najlepších výsledkov bolo dosiahnutých u dospelých ľudí vid' obr.1a, ktorý znázorňuje najlepšie dosiahnuté úspešnosti a počet použitých príznakov, nutných k dosiahnutiu danej presnosti odhadu. Zvyšovanie počtu uvažovaných príznakov viedlo od určitého bodu k poklesu celkovej úspešnosti. Daný trend potvrdilo použitie MFCC a PLP príznakov pri užití redukčných metód priemerovanie a  $k$ -means, vid' obr. 1b.

### 3.2. VPLYV EMÓCIE NA ÚSPEŠNOSŤ ODHADU POHLAVIA

Experiment spočíval v natrénovaní klasifikátora za pomoci neutrálneho rečového prejavu, a následné testovanie nahrávkami vyjadrujúcimi nasledujúce 4 emočné stavy človeka: radosť smútok, strach a zlosť. Z výsledkov vyplýva, že vyjadrenie emócie znemožňuje presný odhad pohlavia rečníka. Najviac tento odhad negatívne ovplyvnili emócie zlosť a radosť pri ktorých dochádza k dynamickým zmenám v hlasovom prejave človeka. Presnosť odhadu sa pohybovala okolo 50 %, pri použití veľkého množstva príznakov. Najmenej tento odhad ovplyvnil emočný stav smútok, pri ktorom sa už pri použití desiatich najlepších príznakov dosiahla najlepšia úspešnosť až 93%, vid' obr.1c. V prípade MFCC a PLP príznakov sme takisto dosiahli najlepšie výsledky opäť pre emóciu smútok, vid' obr.1d. Jeden z možných riešení zlepšenia úspešnosti je postup, kedy sa najprv určí emócia a na jej základe sa vyberie vhodný klasifikátor.

V grafických závislostiach sú zaznačené počty príznakov pri ktorých bola dosiahnutá najlepšia úspešnosť. Tento počet sa líši pre jednotlivé skúmané triedy.



**Obrázok 1:** Najlepšie dosiahnuté úspešnosti v rámci odhadu pohlavia pre oba experimenty, s vyznačením minimálneho počtu príznakov, pri ktorom bola dosiahnutá daná úspešnosť.

#### 4. ZÁVER

V prvom experimente sme zistili, že nízky vek rečníkov má negatívny vplyv na presnosť odhadu pohlavia. Dôvodom môže byť rôzny stupeň vývoja hlasového traktu, keďže u dospelých rečníkoch sme dosiahli lepších úspešností. V druhom experimente sme zistili, že správny odhad znemožňuje i prítomnosť emócie v hlase, hlavne zlosť. Presnejší odhad sme dosiahli v prípade suprasegmentálnych príznakov oproti segmentálnym.

#### LITERATURA

- [1] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B.: A Database of German Emotional Speech, Proceedings of Interspeech, pp. 1517-1520, 2005
- [2] Zeng Y.-M., Wu Z.-Y., Falk T. & Chan W.-Y.: Robust GMM based gender classification using pitch and RASTA-PLP parameters of speech. Proceedings of 5<sup>th</sup> Int. Conference on Machine learning and Cybernetics, pp. 3376-3379, Dalian, 2006.
- [3] Hubeika, V. Estimation of Gender and Age from Recorded Speech. In Proc. ACM Student Research competition 2006. Prague: Czech Technical University, 2006. s. 25-32. ISBN: 80-01-03595-6
- [4] Rendek, T. Určení věku a pohlaví mluvčích. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 31 s. Vedoucí semestrálního projektu Ing. Hicham Atassi.