

## Posudek oponenta diplomové práce

**Student:** Paliesek Jakub, Bc.  
**Téma:** Vliv akustiky prostředí na úspěšnost rozpoznávače řeči (id 23934)  
**Oponent:** Karafiát Martin, Ing., Ph.D., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno s podstatným rozšířením  
Student rozšířil zadání o trénování a implementaci nedávno publikovaných r-vektorů (embedings popisující charakteristiky místnosti), což je zajímavý přínos. Student sice ukazuje minimální zlepšení, ale výraznou komplementaritu s klasickými adaptačními technikami jako jsou i-vectory.
3. **Rozsah technické zprávy** je v obvyklém rozmezí  
Práce je v obvyklém rozmezí.
4. **Prezentační úroveň předložené práce** 85 b. (B)  
Práce je dobře strukturovaná, po úvodu, teoretické části, se student zaměřuje na popis experimentů, nejprve na malém množství dat, které pak dále rozšiřuje s tím že využívá závěrů z předchozích kapitol.  
Nalezl sem malé množství technických nepřesností:
  - str.13 - v případě monofonního modelu: zde jedna hláska neodpovídá jednomu stavu (viz text), ale jednomu modelu který se zpravidla skládá ze 3-5 stavů.
  - odhad adaptačních i-vectorů pro mluvčí a prostředí není přesný - zde se podrobně rozepisují na konci.
5. **Formální úprava technické zprávy** 80 b. (B)  
Práce je psána ve slovenštině, tedy nemohu plně posoudit jazykovou stránku práce.  
Nalezl sem jen par překlepů:
  - str.7 a 13: vrámci - v rámcí
  - str.13 : Nvidia - NVidia
  - str.24 : sa cielove prostredie - sa na cielove prostredie
6. **Práce s literaturou** 80 b. (B)  
Práce je doplněna velkým množstvím správně ocitované literatury.
7. **Realizační výstup** 90 b. (A)  
Jednalo se zejména o experimentální práci, využívající volně dostupný software. Všechny studentem vytvořené skripty jsou dodány a dokumentovány.
8. **Využitelnost výsledků**  
Diplomová práce navazuje na výsledky bakalářské práce. Jsou zde prezentovány analýzy které byly publikovány v prestižním časopise (IEEE) a budou nadále sloužit pro další výzkum.
9. **Otázky k obhajobě**
  1. V práci ukazujete i analyzujete výraznou degradaci systému za předpokladu že nebyl trénován ve stejných nebo podobných podmínkách. Zvažoval jste užití de-reverberačních technik jako například Weight Prediction Error (WPE)
  2. V případě použití alignmentů z čistých dat pro trénování sítě na augmentovaných datech, použil ste jen alignmenty nebo i lattice? Z pohledu LF-MMI objektivní funkce lze zvážit pouze čisté alignmenty (slouží pro cross-entropy regularizaci) a nechat lattice vygenerovat na cílových datech.
  3. r- a zejména x- vektory trénujete na malém množství dat (100h). Pro tuto architekturu je dobré mít aspoň 0.5-1k hodin (bez augmentace). Můžete zdůvodnit proč?
10. **Souhrnné hodnocení** 90 b. výborně (A)  
Student udělal velký kus práce a dostal se hluboko do problematiky přepisu mluvené řeči a Kaldi skriptů. Tyto skripty nejsou často úplně snadno čitelné, dobře dokumentované a často očekávají expertní znalost problematiky. Zejména, jako v této práci, pokud je třeba provést výraznou modifikaci.  
  
Student pěkně analyzuje destrukční efekt na přepis řeči způsobený reverberací. Ukazuje jako velice důležité, systém trénovat za stejných akustických podmínek (místnost, pozice mikrofonu, šum...). Což je bohužel nepraktické, proto se dále zaměřuje na augmentaci pomocí náhodně simulovaných prostředí a adaptace

systemu založená na využití informativního vektoru příznaků popisující mluvího či prostředí.

Zde sem narazil na pravděpodobnou chybu při odhadu i-vektorů pro různé prostředí či mluví. Autor správně definuje soubor "utt2spk" popisující identifikátor mluvího pro danou nahrávku. Nicméně výsledný odhad je stále online, tedy systém není schopen generovat adaptační vektory pro různé identifikátory a změna je pouze v normalizaci příznaků při trénování extraktoru, což nemá výrazný cílový efekt. Správně by bylo generovat i-vectory offline a rozkopírovat je podle online verze. Pak by byli rozdíly v grafech markantnější. Nicméně to vyžaduje větší expertní znalosti, proto bych to nebral jako zásadní nedostatek v diplomové práci.

Na závěr student implementuje a analyzuje další techniky pro adaptaci: takzvané "x-" a nedávno prezentované "r-" vektory. Oba přístupy jsou nad rámec zadání. Zde student ukazuje komplementaritu "r-" a "i-" vektorů, která bude pravděpodobně dále zkoumána a analyzována v právě běžících projektech zde na FITu.

Přes zmíněné nedostatky, student prezentuje kvalitní analýzy, rozšiřuje zadání diplomové práce a ukazuje nadstandardní znalosti v oblasti. Pro se příkládám k vyššímu bodovému ohodnocení.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 9. června 2021

Karafiát Martin, Ing., Ph.D.  
oponent