Oponentní posudek habilitační práce

Téma: **Adaptation of speech recognition systems to selected real-word deployment conditions**

Autorka: **Ing. Petr Červa, Ph.D.**

Oponent: ***Doc. Ing. Petr Pollák, CSc.***

Předložená habilitační práce se zabývá velmi aktuální problematikou v oblasti zpracování řečového signálu, a to adaptací systémů rozpoznávání řeči (ASR). Uvedená problematika je velmi aktuální a výzkum v dané oblasti probíhá celosvětově velmi intenzivně, neboť adaptace ASR systémů je nezbytná pro zvýšení přesnosti rozpoznávání v reálných podmínkách.

Práce je psána formou rozšířeného shrnutí souboru 12 publikací autora. Toto pojetí považuji za velmi vhodné, zejména pak s ohledem na vyšší kvalitu zmíněného souboru publikací, z nichž je nutné vyzdvihnout v první řadě 2 články v prestižních časopisech Computer Speech and Language (2021) a Speech Communication (2013), článek v impaktovaném časopise Radioengineering (2013), celkem 8 publikací na prestižních konferencích Interspeech (2011, 2012 - 2x, 2018, 2021) a ICASSP (2017, 2018 - 2x). Konečně i publikace z konference TSD 2021 je kvalitní, publikovaná ve sborníku vydávaném nakladetelstvím Springer a zaindexovaná ve WoS.

Uvedené publikace pokrývají 3 užší oblasti zájmu obecně širších výzkumných aktivit autora v oblasti rozpoznávání řeči (ASR), konkrétně problematiky neřízené adaptace na mluvčího, práci s neřečovými událostmi v analyzovaném řečovém signálu a nakonec i práci s multilingválními systémy rozpoznávání řeči. Všechny 3 oblasti představují výzkum aktuálně řešený v širším mezinárodním měřítku a uvedené publikace potvrzují autorovu erudici v dané oblasti, nepochybnou originalitu jeho vědecké práce a dávají detailní přehled o vysoce kvalitních výzkumných aktivitách autora v oblasti rozpoznávání řeči v průběhu uplynulých 10 let.

Každá z výše zmíněných užších oblastí je reprezentována 4 publikacemi a k jejich reprintům autor přidává vždy velmi stručné shrnutí řešeného problému i dosažených výsledků prezentovaných v každé jednotlivé publikaci. V úvodní kapitole 2 pak autor stručně shrnuje problematiku rozpoznávání řeči obecně a definuje základní pojmy a problémy, jejichž řešení jsou diskutována v prezentovaných publikacích. Tento způsob považuji za šťastný a speciálně autorovy dodatečné komentáře jsou velmi zdařilé, výstižné a dobře srozumitelné, a pomáhají tak k rychlé orientaci čtenáře v každém konkrétním řešeném problému. Nenalezl jsem zásadní typografické ani gramatické chyby, což jsem však u habilitační práce již ani neočekával.

V konečném důsledku práce srozumitelným způsobem uvádí čtenáře do diskutované problematiky a může být velmi dobrým studijním materiálem pro všechny, kdo se zadanou problematikou začínají seznamovat. Metodika výkladu a popisu potvrzuje didaktické schopnosti autora.

Co se týče vlastních řešení popisovaných v jednotlivých publikacích, nemám zásadních připomínek a předpokládám, že jejich kvalita je potvrzena přijetím na významných mezinárodních fórech a nebudu tedy duplikovat již proběhlá recenzní řízení.

Na základě výše uvedených skutečností předloženou práci jednoznačně doporučuji k přijetí v rámci habilitačního řízení.

Do obecné diskuse při vlastním habilitačním řízení bych položil následující otázky k vybraným úlohám:

1. Při rozpoznávání řeči s hudbou na pozadí uvádíte v článku zajímavé výsledky, kdy pro přizpůsobené podmínky (matched train-test conditions) je dosaženo významného zvýšení přesnosti rozpoznávání. V závěru zmiňujete mimo jiné další možné zvýšení robustnosti při zahrnutí více hudby různých žánrů s různým SNR do trénovací množiny. Zajímalo by mě, jaké rozšíření by podle Vás dávalo praktický smysl resp. zda by při větší různosti hudby na pozadí bylo postačující pouze zmíněné rozšíření trénovací množiny nebo zda by bylo nutné zasáhnout i do struktury sítě, tj. zvýšit počet vrstev či počet neuronů v jednotlivých skrytých vrstvách?

2. V publikacích věnujících se rozpoznávání jazyka je dosaženo odlišné přesnosti v případě identifikace slovanských vs. skandinávských jazyků (byť použité klasifikátory nejsou zcela totožné). Je to dáno zejména uváděným rozdílným množstvím trénovacích dat či je u těchto jazyků větší míra podobnosti (zde nemyslím větší podobnost diskutovaných variant norštiny)? Také by mě zajímalo, jak rychle stoupá přesnost LID při analýze delších záznamů a zda je případná identifikace z delších záznamů prakticky použitelná ve vámi používaných sytémech automatického přepisu zpravodajství?

V Praze dne 7. října 2021