

## Oponentský posudek na disertační práci

**Mgr. Davida Klusáčka**

### **„New Methods in Statistical Speech Recognition“**

---

Disertační práce Mgr. Davida Klusáčka s názvem „New Methods in Statistical Speech Recognition“ se zabývá problematikou automatického rozpoznávání mluvené řeči, je psána v angličtině a její rozsah je 226 stran. Práce má na rozdíl od jiných disertačních prací nestandardní kompozici – je psána převážně jako monografická publikace pojednávající o problematice automatického rozpoznávání řeči, a to v některých svých částech z nezvykle rigorózního matematického pohledu. V disertaci není uvedena obvyklá kapitola „Cíle disertace“, ale tyto cíle jsou velmi stručně (cca 2. řádky textu) shrnuty v krátkém abstraktu v úvodním Summary disertace. Lze tak dovodit, že vlastní problematika, kterou disertant ve své práci zkoumá, je věnována návrhu zcela nového front-end s názvem NUFIBA, jenž by měl řešit některé nedostatky současných vstupních modulů zpracovávajících řečový signál v systémech rozpoznávání řeči. Téma disertační práce považuji jednoznačně za disertabilní. Nevím, jestli je možné tvrdit, že nejslabším článkem současných rozpoznávačů je právě front-end, nicméně je jasné, že vědecká komunita zabývající se rozpoznáváním řeči se dá říci „ustrnula“ na řešení vstupního parametrizátoru řeči, které pochází ještě z minulého století. Dva dnes stále prioritně využívané parametrizátory – MFCC a PLP jsou již více než 20 let staré a v posledních létech nikdo nepřišel s takovým řešením, které by jednoznačně vylepšilo dříve zavedené přístupy. Nedá se přitom říci, že se o taková řešení výzkumníci nepokoušejí. Pro některé speciální případy/prostředí byla navržena řešení, která vykazují lepší funkcionalitu, nicméně nově pojatý front-end, který by vykazoval vysokou robustnost k řečníkovi, přenosovému kanálu a prostředí a který by jednoznačně zvyšoval klíčový parametr při rozpoznávání řeči, jímž je přesnost, tak ten se v posledních létech nepodařilo navrhnout a prosadit do aplikací. Snahu Davida Klusáčka, vylepšit vlastnosti současných front-endů považuji proto za hodnou podpory a zájmu. Disertační práce nemusí mít v tomto případě ambice masivně pohnout světovým věděním, ale například přispět k vylepšení funkcionality parametrizátoru pro vybraný typ řešené úlohy.

Disertační práce je rozdělena do 10 kapitol. První přehledové kapitoly, které se zabývají problematikou pravděpodobnostního přístupu k rozpoznávání řeči (2. kapitola), teorie pravděpodobnosti a informace (3. kapitola) a teorie zpracování signálu (4. kapitola), jasně naznačují, že disertant je teoreticky velmi dobře vybaven a schopen jasně a srozumitelně popsat vybranou problémovou oblast. Zvláště poměrně rozsáhlá kapitola zabývající se zpracováním signálu (74 stran) přesahuje svým rozsahem a velmi rigorózním zpracováním běžnou kapitolu disertační práce.

Jádro disertace lze nalézt v kapitole 6 nazvané „NUFIBA Front-end“. Disertant velmi zevrubně vysvětluje důvody návrhu nového parametrizátoru, který by měl přispět zejména ke

zlepšení některých negativních jevů, které disertant identifikoval u standardního rozpoznávače. Návrh vychází z optimalizace banky filtrů, které jsou podrobeny Hilbertově transformaci. Optimalizace je nicméně prováděna s ohledem na odezvy filtrů a není sledován klíčový parametr, jímž je vliv na přesnost rozpoznávání.

Kapitola 8 je věnována shlukování a na třídách založeným jazykovým modelům. Tato kapitola je opět dobře zpracována, nicméně z hlediska cíle disertace příliš nezapadá do kontextu celé práce. Kapitoly 7 a 9 „Akustický model“ a „Experimenty“ jsou extrémně krátké a je zřejmé, že disertant neměl pro jejich zpracování buď dostatek času anebo podkladových údajů. Zejména chybějící výsledky experimentů jsou závažné a významně degradují celou práci.

Mgr. David Klusáček je nepochybně velmi kvalitní teoreticky fundovaný výzkumný pracovník. Jeho disertační práce má rozsáhlé výborně zpracované pasáže střídané nikoli špatně zpracovanými, ale spíše vůbec nezpracovanými či nedopracovanými částmi.

Protože disertant nestanovil v disertaci explicitně cíle své práce, bylo by dobré, kdyby tak učinil v průběhu obhajoby. Měl by též přesněji uvést, jak tyto cíle konkrétně naplnil. Disertabilita úlohy pak bude jednoznačěji posouditelná. V případě, že disertant v průběhu obhajoby upřesní cíle a přínosy disertace pro naplnění zákona č. 111/98 Sb., přikloním se k udělení titulu Ph.D.

Otázky:

1. V disertační práci popisujete v kap. 2 strukturu Markovského modelu fonému (obr.3 a 4) z „dílny“ IBM z 80. a 90. let. Tato struktura je velmi logická a funkční pro fonetické modelování, nicméně méně vhodná pro následné dekódování. V kapitole 7 uvádíte další typ struktury, a to model difónu s nulovými přechody. Používáte ve svých experimentech stejné struktury modelů anebo pracujete se strukturou jinou, která možná lépe koresponduje s Vámi navrhovaným front-end?
2. Uveďte alespoň dílčí výsledky experimentů (z hlediska přesnosti rozpoznávání) s Vámi navrženým NUFIBA front-end.
3. Porovnejte výpočetní nároky běžně používaných front-end v korespondenci s Vámi navrženým typem NUFIBA.

Prof.Ing. Josef Psutka, CSc.  
ZČU Plzeň

V Plzni, 2.9.2012