

## POSUDEK NA DISERTAČNÍ PRÁCI

**Téma práce:** Identifikace mluvčího v temporální doméně řeči

**Doktorand:** Mgr. Lenka WEINGARTOVÁ

**Posudek vypracoval:** Doc. Ing. Petr POLLÁK, CSc.  
ČVUT FEL K13131, Technická 2, 166 27 Praha 6

Předložená práce Mgr. Lenky Weingartové se věnuje problematice popisu temporálních charakteristik řeči s užším zaměřením na využití v systémech identifikace mluvčího. Problematika identifikace mluvčího zejména pro forenzní účely je v současné době velmi intenzivně studována a rozvíjena, a to jak na úrovni automatické identifikace tak na úrovni expertní identifikace s podporou moderních analýz akustického řečového signálu. Předkládaná detailní studie temporálních vlastností českých hlásek s globálním obecným cílem vytvoření obecnějšího popisu temporální struktury řeči resp. modelu popisujícího referenční temporální charakteristiky pro účely analýzy individuálních odlišností jednotlivých mluvčích je nepochybně vysoce aktuální na úrovni obecné fonetiky i pro návrhy uvedených automatických systémů zpracování řeči.

Práce je členěna do 11 kapitol, kde prvních 5 stručnějších kapitol uvádí čtenáře do problematiky temporálního popisu řeči. V 6. kapitole si doktorandka stanovuje výzkumné otázky a cíle práce, které lze stručně shrnout následovně:

- prvním významným cílem práce je popis temporálních charakteristik jednotlivých českých hlásek spolu s vlivem různých faktorů na popisované charakteristiky,
- druhým cílem je návrh temporálního modelu češtiny na základě provedených a popsáných analýz,
- posledním cílem pak je analýza temporálních charakteristik s ohledem na individuální rysy typické pro jednotlivé mluvčí a diskuse nad jejich možným využitím pro identifikaci mluvčího.

Lze jistě konstatovat, že výše stanovené cíle jsou disertabilní a doktorandka tyto cíle v předložené práci naplnila. Předložená práce představuje přínos pro obecný popis temporální struktury řeči a dosažené výsledky mohou posloužit pro další rozvoj popisu obecné temporální struktury řeči i pro další jazyky. Nejvýznamnější přínosy předložené práce pak vidím především v následujících bodech.

- V první řadě autorka naplňuje první cíl a předkládá přehledný a systematický popis temporálních charakteristik jednotlivých českých hlásek. Toto tvoří stěžejní část práce a autorka tomuto popisu věnuje nejrozsáhlejší 8. kapitolu. Na pro tyto účely vytvořeném korpusu Minidialogy-H autorka podrobně popsala doby trvání jednotlivých hlásek včetně podrobných statistických analýz. Autorka popisuje detailně výskyty hlásek ve všech možných pozicích, tj. z hlediska pozice ve slabice, pozice slova v prozodické frázi, pozice slabiky ve finálním slově, hláskového okolí, délky slova, apod. Dosažené výsledky srovnává s dříve prezentovanými různými studiemi, některými i velmi historickými (např. Josef Chlumský z roku 1911). Autorka zmiňuje rozdíly ve studiích dané především odlišnostmi analyzovaných materiálů či stylem mluvy (formální zřetelně vyslovovaná řeč vs. čtená až semispontánní řeč v korpusu minidialogy). V této souvislosti mne napadá otázka, do jaké míry hraje roli různý historický kontext (tj. nejenom prezentované rozdílné možnosti měření doby trvání, ale i obecný styl mluvy před 50-100 lety a v současnosti)? Jak moc jsou prezentované výsledky srovnatelné z tohoto hlediska?

Doktorandka dále zde i ve finálním závěru zmiňuje nemožnost systematického zobecnění dosažených výsledků v důsledku omezeného zastoupení všech možných výskytů hlásek

v dostupném korpusu. Toto omezení je pochopitelné a pramení z nemožnosti vytvoření řádově rozsáhlejšího korpusu s velmi precizní ruční anotací. Otázkou je, zda by nebylo vhodné provést i ekvivalentní analýzy na automaticky segmentovaných datech (pro srovnání třeba i na korpusu Minidialogy-H, a v případě korelovanosti výsledků pak i na reprezentativnějším rozsáhlejším korpusu, kde již není možné provést ruční fonetickou segmentaci). Nejedná se o principiální výhradu snižující význam této práce, chápu nutnost volby určitého kompromisu v každé realizované studii, která má být dokončena v reálném čase. Analýzu automaticky segmentovaných dat bych však považoval za velmi zajímavou a účelovou, zejména z hlediska následného použití semiautomatických a automatických metod identifikace mluvčího, kde ruční segmentace dat pro analýzu je vždy velmi omezující.

K této části bych měl ještě další otázku do diskuse. V obrázcích v kapitole 8 autorka označuje 95% intervaly spolehlivosti odhadu střední hodnoty trvání jednotlivých hlásek, což je jistě významná informace s ohledem na odhad středních hodnot z relativně malého souboru dat. Jedná se zřejmě o standardní popis podobných charakteristik ve fonetické literatuře, nicméně mohl být ovšem v textu zmíněn přesněji i s podrobnějším vysvětlením (zejména uvedeným vztahem pro výpočet). Možná pod vlivem mé techničtější odbornosti a následujících histogramů prezentovaných v práci se domnívám, že může dojít k záměně za intervaly spolehlivosti variability vlastní doby trvání hlásek, nikoliv variability odhadu průměru z omezeného množství realizací. Mimochodem, není uváděná chyba odhadu střední hodnoty méně podstatná vzhledem k výrazně vyšší variabilitě trvání jednotlivých hlásek v různých realizacích?

- Dalším významným přínosem předložené práce je detailní analýza individuálních specifícností temporálních charakteristik pro jednotlivé mluvčí v analyzovaném materiálu, které jsou popsány v následující kapitole, a to jak samotného trvání segmentů tak obecnějších temporálních ukazatelů. Z provedených analýz na korpusu Minidialogy-H předkládá autorka shrnutí nejvýznamnějších temporálních charakteristik, které hrají významnou roli při identifikaci mluvčího, neboť vykazují variabilitu mezi mluvčími a stabilitu pro jednoho mluvčího.

Určité omezení dané studie je dané ovšem omezeným zastoupením mluvčích v analyzovaném korpusu (mladí dospělí mluvčí, převážně ženy). Autorka toto ve své práci explicitně zmiňuje předpokládá budoucí studie na nově budovaném korpusu pokrývajícím lépe pohlaví, věk, vzdělání či region původu jednotlivých mluvčích, stejně tak jako různé mluvní styly.

- V 10. kapitole pak autorka definuje model temporálních charakteristik, což lze považovat za druhý hlavní přínos předložené práce. Model vychází z principu Klattova modelu, neaplikuje minimální trvání a modeluje 8 různých faktorů jako pozice ve slabice, slovo v prozodické frázi, hláskové okolí apod. Oceňuji korekci parametrů modelů pro hlásky málo zastoupené v analyzovaném korpusu na základě hodnot z jiných studií.

Modelová trvání byla použita k výpočtu modelových kontur lokálního artikulačního tempa (LAR) pro jednotlivé promluvy. Pro jednotlivé modely pak byly analyzovány odchylky hodnot skutečných od modelových (rezidua modelu), které pak byly podrobně analyzovány až na úroveň jednotlivých promluv jednotlivých mluvčích a bylo prokázáno, že nejsou náhodné. Závěrem této kapitoly je tedy potvrzení kvality navrženého temporálního modelu, který předpovídal trvání hlásek s průměrnou odchylkou 16,5 ms.

V této souvislosti bych měl další malou připomínku, a to k titulu práce. Nepoužil bych název "Identifikace mluvčího v temporální doméně řeči", ale spíše "Individuální charakteristiky mluvčího v temporální doméně řeči" či "Popis řeči v temporální doméně s ohledem na aplikace pro identifikaci mluvčího", neboť konkrétní identifikace resp. verifikace mluvčího zde nebyla realizována. Vlastní identifikace je už ovšem jistě téma na další samostatnou disertační práci.

- Velmi významným přínosem práce je nakonec i samotné vytvoření korpusu Minidialogy-H a jeho velice detailní anotace. To bylo jistě zásadní pro realizovanou studii temporálních charakteristik pro češtinu i pro analýzu individuálních odlišností dílčích charakteristik pro jednotlivé mluvčí. Uvedená databáze nalezne nepochybně významné uplatnění i v dalším fonetickém výzkumu. V této souvislosti by mě zajímalo, zda je korpus dostupný pro odbornou komunitu.

Po formální stránce je předložená práce na velmi dobré úrovni. Je členěna přehledně (v úvodní části možná do příliš mnoha krátkých kapitol), text je psaný velmi pěknou češtinou, grafická úprava je na velmi vysoké úrovni. Drobnou výhradu bych měl k formulacím v 1. osobě množného čísla. Je zřejmé, že výzkum tohoto druhu je vždy výsledkem kolektivní práce (přes jednoznačně nezpochybnitelný osobní přínos doktorandky), přesto v disertační práci považuji za vhodnější preferovat formulace v 1. osobě jednotného čísla či v trpném rodě. Drobný nesoulad je v číslování obrázků a tabulek (od 9. kapitoly nesedí číselné řady k číslům kapitol). Jako technik bych také uvítal přesnější popis vztahů použitých pro analýzy. Uvedené připomínky však nesnižují nijak zásadně celkově dobrý dojem z práce.

Přínos předložené práce potvrzuje i související publikační činnost, která zahrnuje 11 publikací za poslední 3 roky. Nejvýznamnější publikace představují 2 články v mezinárodním časopise *Research in Language* a 3 příspěvky na mezinárodní konferenci "International Conference on Speech Prosody" zaindexované v databázi Scopus.

Na základě výše uvedených skutečností lze tedy jistě konstatovat, že předložená práce popisuje originální výsledky vědecké práce. Z předložené práce je zřejmý jak teoretický, tak i praktický nadhled doktorandky nad danou problematikou. Výše zmíněné otázky a připomínky neprezentují jako zásadní výhrady, ale jako další otázky do diskuse započaté samotnou doktorandkou v předložené práci. Práci hodnotím celkově jako velmi kvalitní a **doporučuji** ji k obhajobě za účelem získání vědecké hodnosti doktora na Karlově univerzitě v Praze. Navrhuji klasifikaci stupněm **prospěla**.

V Praze dne 5. května 2015