



Posudek diplomové práce

předložené ve Fonetickém ústavu Filozofické fakulty Univerzity Karlovy

Jméno a tituly posuzujícího vedoucího/opponenta: Ing. Tomáš Božil, Ph.D.

posudek vedoucího/vedoucí posudek oponenta/oponentky

Autor/autorka: Richard Mazúr

Název práce: Acoustic measures of glottalization in Czech

Rok odevzdání: 2023

Odborná úroveň práce:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Členění práce:

logické vykazuje mírné nedostatky vykazuje závažnější nedostatky nevhodné

Věcné chyby:

téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné četné závažné

Zvolená metoda:

původní a adekvátní vhodně zvolená nepřilíš vhodná nevhodně zvolená

Rozsah práce:

příliš velký přiměřený tématu dostatečný nedostatečný

Rozsah použité literatury:

nadprůměrný průměrný podprůměrný nevyhovující

Způsob používání odborné literatury:

adekvátní nevyhovující podezření na plagiát

Grafická a formální úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Jazyková úroveň:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Tiskové chyby:

téměř žádné vzhledem k rozsahu a tématu přiměřený počet četné

Celková úroveň práce:

vynikající velmi dobrá průměrná podprůměrná nevyhovující

Zadání práce:

mimořádně náročné náročnější průměrně náročné lehčí nedostatečně náročné
 splněno splněno s drobnými výhradami splněno s většími výhradami nesplněno

Samostatnost při zpracování práce (vyplňuje jen vedoucí práce)

nadprůměrná průměrná podprůměrná

Průběžné konzultace, dodržování termínů, připravenost (vyplňuje jen vedoucí práce)

výborná spolupráce standardní spolupráce spolupráce s výhradami nefungující spolupráce



Podrobné zhodnocení:

Richard Mazúr v úvodní teoretické kapitole rozebírá názvosloví spojené s hlasivkovým rázem, a to jak v české, tak i zahraniční literatuře. Dále shrnuje jeho funkce, rozšíření a možné akustické realizace popsané v odborné literatuře. V praktické části vybírá potenciálně vhodné akustické parametry pro jeho odlišení od hlásek, které se v jeho okolí vyskytují, s motivací, že by takové parametry mohly sloužit k detekci, zda byl ráz v potenciálních místech výskytu realizován či nikoliv. Vlastnosti parametrů pak proměřuje na nahrávkách hraných krátkých dialogů, které byly v minulosti mimo souvislost s touto diplomovou prací nahrány na Fonetickém ústavu. Na závěr autor natrénoval klasifikátor s algoritmem random forest, pomocí kterého vyhodnotil dopad jednotlivých vybraných parametrů na úspěšnost rozpoznání rázu od ostatních hlásek.

Ačkoliv text splňuje nároky kladené na diplomovou práci a autor je nepochybně erudovaný, bohužel se na výsledné podobě negativně projevila časová nerovnoměrnost její tvorby. V prvním a nejdelším období se student převážně seznamoval s řečovým materiálem a konzultace probíhaly zřídka, takže možná i z toho důvodu byly některé poznámky a náměty vedoucího práce zapomenuty. A naopak v druhé a výrazně kratší fázi tempo činnosti i množství konzultací enormně narůstalo a vlivem této překotné práce některé mé rady zapadly nebo je již zkrátka nebylo časově možné důsledně implementovat.

Důsledkem toho můžeme v práci pozorovat méně závažné záležitosti, jako je větší množství překlepů i gramatických chyb nebo například matoucí desetinná místa s nulami u celočíselných počtů položek v tabulce 2.4 či záhadné parametry X1-X13 s překlepem X2 v tabulce 2.6 (str. 33) bez textového vysvětlení, o kterých si čtenář musí domyslet, že se jedná o MFCC koeficienty atd.

Mezi závažnější prohřešky vůči stylu bohužel patří veliké množství plovoucích objektů (obrázky a tabulky), které nejsou žádným způsobem odkazovány v textu, resp. je skutečně jen velmi malé množství výjimek, které v textu ukotveny jsou, a to ještě větší část z nich pouze s uvedeným číslem bez specifikace, zda jde o obrázek, tabulku či snad odkaz na jinou sekci textu. Jedná se o ukázkový příklad problému, na který jsem ještě před odevzdáním práce stihl upozornit, ale v závěrečném spěchu byl zkrátka opomenut a je jisté, že příslušný týden času navíc by zde jistě k výsledné úrovni výrazně pomohl.

Co se týče náročnosti rozsahu provedené práce, musím zdůraznit, že diplomant do začátku získal z velké části zpracovaný řečový materiál a oproti původnímu předpokladu rozsah manuálních segmentací (tzv. labelování) nerozšířil, ale pouze existující anotace rychlými automatickými skripty sjednotil z hlediska typických problémů (velikost písmen, nadbytečné mezery, překlepy transkripčních znaků mimo zvolenou sadu symbolů, srovnání hranic slov s nejbližšími hranicemi hlásek) bez větších manuálních zásahů. Pochopitelně uznávám, že povinnost manuálního zpracování materiálu nebyla v zadání práce uvedena, tudíž toto není myšleno jako výhrada, pouze upozornění, že v této konkrétní oblasti nakonec diplomová práce nebyla náročná a nic výrazného nepřinesla. Možnost vyjádřit se a případně upřesnit tuto záležitost dávám prostřednictvím první otázky k obhajobě, viz dále.

Nyní se zaměřím na další výhrady k textu. V úvodní kapitole (str. 3 a 4) je z hlediska rázu opomenuto důležité dílo Bohuslava Hály (1967) – Výslovnost spisovné češtiny, které užívání rázu věnuje celý jeden oddíl. Oproti tomu je zde uvedena informace, že ráz byl začleněn do české ortoepické normy s citací Vývoje kodifikace české výslovnosti od Veroniky Štěpánové z roku 2019. Způsob, jakým je to v textu uvedeno, v neznalém čtenáři téměř určitě vytvoří pocit, že jakási norma stanovila pravidla povinného používání rázu v češtině až v roce 2019. Je jisté, že autor tuto referenci myslel spíše tak, že se v tomto



díle o historii rázu dočteme detailní informace, ale přímá citace Hály (1967) by zde zkrátka měla být. Ten je v textu práce zmíněn až mnohem později (str. 12), a to jen velmi okrajově a ve zcela jiném kontextu.

V rešeršní sekci 1.3.3 (str. 19) se hovoří o používaných akustických parametrech vhodných pro vyhodnocování přítomnosti rázu a je zmíněna studie, ve které se jako nejslibnější ukazatel jeví autokorelační funkce na rozdíl od průběhu f_0 a energie. Protože se jedná o nosnou kapitolu vzhledem k tématu diplomové práce, očekával bych zde detailnější vysvětlení oproti tomuto přílišnému zjednodušení, jelikož autokorelační funkce je jasně definovaná matematická funkce (která se mimochodem nejčastěji používá právě pro určování f_0), zatímco v odkazovaném článku se dozvíme, že se ve skutečnosti jedná o „short-time peak autocorrelation“. A autoři následně vysvětlují, že se jedná o speciální algoritmus fungující jako nadstavba nad autokorelační funkcí; konkrétně je v krátkodobém časovém segmentu vypočtena autokorelační funkce, nalezeno maximum nehledě na zpoždění a amplitudy těchto maxim jsou v čase vynášeny do následně analyzovaného grafu. Kromě takového vysvětlení mi chybí i vyjádření, zda diplomant takový postup ve své metodě použije či ne a proč.

V tabulce 1.4 (str. 19) se objevuje zkratka CPP, jež bude definována v textu až mnohem později, a její popis „noise measure“ je evidentně nepřesný či zcela mylný, nejčastěji je CPP uváděno jako „voice quality measure“ v souvislosti s dysfonií.

Druhý odstavec sekce 2.1.1 (str. 20) obsahuje nedůležitou, ale zavádějící informaci. Digitalizace (vzorkování a kvantizace) určitě není to, co zvuk převádí do časové domény. Digitalizace pouze převádí spojitý signál na diskrétní, ale zmiňovaný oscilogram by bylo možné zobrazit i pro analogový záznam (např. gramofon).

V této sekci se také hovoří o 4 kategoriích, do kterých by bylo možné následně vybrané akustické míry rozdělit. Není ale jasné, zda to mají být nejdříve zmíněné 4 typy domén zobrazení signálu (oscilogram, spektrum, spektrogram, keprstrum) nebo dále uvedené 4 základní akustické domény (frekvence, trvání, intenzita, spektrum). Je jisté, že tato dvě dělení spolu nelze jednoznačně spárovat a postrádají logiku i vzhledem k členění nadpisů následujících odstavců: Duration a Fundamental Frequency by smysl dávaly dle druhého způsobu dělení, na stejnou úroveň je ovšem dále položen Jitter, který jistě spadá pod Fundamental Frequency, a pak stejnou úroveň pokračuje HNR (které může spadat fakticky kamkoliv), pak se hovoří o spectral tilt týkající se spektra, a hned následují Cepstral peak prominence a MFCC, které by odpovídaly keprstrální doméně dle prvního způsobu dělení.

Tato část textu vzhledem k tématu práce by měla být zpracována nejpodrobněji a nejpečlivěji, ale bohužel je velmi stručná a svým způsobem ne úplně logicky členěná. Také je škoda, že dříve zmíněný výstup z autokorelační funkce zde nebyl zařazen.

Na str. 24 je mylně uvedeno, že program VoiceSauce je implementací Matlabu pro fonetický výzkum. Ve skutečnosti je VoiceSauce program pro fonetický výzkum, který je implementován v Matlabu.

Na stejné straně je i nepřesný popis balíčku rPraat. Není to knihovna, která by ovládala Praat z R (na rozdíl od nepoužitého balíčku PraatR); rPraat je pouze samostatný interface umožňující importovat a exportovat některé typy souborů mezi Praatem a R. A k tomu obsahuje řadu užitečných funkcí pro zpracování těchto objektů v R nezávisle na Praatu.

Co se týče exploratorní analýzy, je škoda, že nebylo zvoleno podrobnější dělení skupin (např. alespoň na konsonanty a vokály), jak je zmíněno v diskusi na str. 45. Minimálně podrobnější zdůvodnění tohoto rozhodnutí by bylo vhodné.



FILOZOFICKÁ FAKULTA
Univerzita Karlova

Fonetický ústav

Otázky k obhajobě a náměty do diskuze:

- Na str. 28 hovoříte o tom, že segmenty mohou být špatně zarovnány a provedl jste manuální kontrolu dle pravidel v Machač a Skarnitzl (2009). V jakém rozsahu a s jakým výsledkem? Chybně zpracované položky byly vyřazeny nebo jste je zarovnal lépe? Byly namátkově objeveny potenciální problematičtí anotátoři a jejich TextGridy jste z hlediska segmentů manuálně opravil?
- Na str. 31 je zmíněno, že nedefinované hodnoty f_0 nebo jitteru na výstupu Praatu a VoiceSauce jste nahradil nulami. Proč jste je nenechal jako NA? Pro jitter jsou hodnoty blízké nule typické. Jakým způsobem jste s těmito hodnotami dále nakládal – např. průměrné hodnoty v tabulce 2.6 (str. 33) jsou počítány i z těchto nul?
- Na str. 37 uvádíte: „Zdá se, že hodnoty f_0 větší než 300 Hz svědčí o glotalizaci“. Čím si to vysvětlujete?

Práci

doporučuji nedoporučuji uznat jako diplomovou.

Vzhledem k výše uvedenému zhodnocení navrhuji hodnocení stupněm: Velmi dobře až Dobře.

V Českém Brodě, 20. ledna 2023

Ing. Tomáš Bořil, Ph.D.