Západočeská univerzita v Plzni
Doručeno: 22.01.2013
ZCU 001850/2013
listy: 1 přílohy:
druh:
zcupesc66640

Univerzita Karlova v Praze
Filozofická fakulta
Fonetický ústav
Doc. Mgr. Radek Skarnitzl, Ph.D.

Posudek disertační práce

Concatenation cost in unit selection speech synthesis,

kterou předkládá

Ing. Milan Legát

Předložená práce Ing. Milana Legáta se zabývá řetězením jednotek v konkatenační syntéze založené na dynamickém výběru jednotek, konkrétně cenou řetězení při spojování vokalických segmentů. Z hlediska volby tématu se jedná o vhodný a i dnes aktuální cíl, protože konkatenační syntéza představuje stále v oblasti řečových technologií standard. Další zvyšování její kvality snahou o eliminaci náhodně se vyskytujících chyb při řetězení je proto nepochybně přínosné.

Disertační práce je rozčleněna na pět hlavních kapitol. V první z nich autor obsáhle představuje dosavadní výzkum v oblasti konkatenační syntézy se zaměřením především na cenu řetězení, na různé způsoby jejího vyjádření a na její hodnocení. Druhá kapitola zavádí nový způsob získávání data od posluchačů týkající se plynulosti či neplynulosti při řetězení – metodu *polovět* (*half-sentence method*). Tato metoda autorovi dovoluje zadávat poslechové testy, které umožňují jednoznačnou a spolehlivou kvantifikaci nespojitosti při řetězení u jednotlivých položek a zároveň není trvání jednotlivých položek omezeno jen na několik hlásek jako u jiných způsobů zkoumání plynulosti řetězení. V druhé kapitole autor dále popisuje metodické pozadí poslechových testů, včetně zkoumání spolehlivosti posluchačů a stanovení kritéria pro posuzování jednotlivých položek jako *fakt*.

V následujících dvou kapitolách jsou již výsledky percepčních testů – položky identifikované jako *fakta* – porovnávány s různými způsoby vyjádření diskontinuity. Za velmi chvályhodné považuji skutečnost, že Milan Legát do svých analýz zahrnuje foneticky informovaný pohled, že se snaží nacházet i lingvisticky interpretovatelné parametry a nespoléhá pouze na blíže neuchopitelné parametry jako např. koeficienty MFCC. Třetí kapitola se zabývá možnou nespojitostí důsledkem nesouladu v oblasti základní frekvence (F0). Na rozdíl od přechozích výzkumů, které pro určení ceny řetězení počítaly pouze se statickým rozdílem mezi dvěma

sousedními datovými body, navrhuje Ing. Legát využití trajektorií základní frekvence, tedy dynamického průběhu F0 v blízkosti bodu řetězení. Jeho výsledky naznačují, že právě dynamické vlastnosti jako sklon křivky jsou pro percepci nespojitosti důležitější než statické rozdíly. Autor tak potvrzuje významnost dynamických vlastností řeči, tedy změn akustických parametrů v čase pro vnímání řeči, na niž v posledních letech poukazují i foneticky orientované výzkumy. Za vyzdvihnutí stojí i to, že autor do svých analýz zahrnul i psychoakustické jednotky vyjadřující vnímání výšky. Dovolil bych si však polemizovat s užitečností melů pro účely analýzy F0: v oblasti, v níž se běžně hodnoty F0 pohybují, mely prakticky odpovídají objektivní veličině, hertzům. Pro psychoakustické vyjádření intonačních rozdílů jsou nejvhodnější půltóny (Nolan, 2003) – ty však Ing. Legát ve své práci rovněž používá. Pro fonetika je však výsledek porovnání psychoakustických a objektivních jednotek zklamáním – ke zlepšení reprezentace F0 v půltónech oproti hertzům bohužel nedošlo.

Ve čtvrté kapitole se Milan Legát zabývá rolí různých konsonantických kontextů na spojitost řetězení. V první části kapitoly se zabývá nesouladem vzniklým řetězením vokálů pocházejících z nazálních a nenazálních kontextů, v druhé části pak analyzuje vliv nesouladu v jednotlivých způsobech a místech artikulace. Zatímco v první části nachází významné výsledky alespoň u ženského hlasu a ukazuje zajímavý vliv fáze, v případě druhé části systematické a interpretovatelné vlivy nesouladu v konsonantickém kontextu nenachází. Vydělení nazálních kontextů považuji za přínosné; nazalitu v každém případě můžeme považovat za rys, který může výrazně ovlivňovat vokalickou kvalitu. Je proto trochu škoda, že se autor v druhé části kapitoly nepokusil nalézt podobné rysy a zůstal pouze u tradičního dělení konsonantických kontextů podle místa a způsobu artikulace. Domnívám se, že výraznější vliv především způsobu artikulace okolních konsonantů na řetězení vokálů ani nelze očekávat a že by hledání mohlo směřovat spíše k podobným obecnějším „rysům", jako je právě nazalita. Rád bych autora poprosil, aby se pokusil při obhajobě nějaký takový rys navrhnout.

Pátá kapitola disertační práce Ing. Legáta se týká aspektů spojených se samotným fungováním konkatenační syntézy založené na výběru jednotek. Po krátkém představení systému ARTIC autor zkoumá, nakolik extrémní hodnoty ceny řetězení a ceny cíle odpovídají neplynulým, slyšitelným spojením. Výsledky v této oblasti nenaplnily očekávání – přibližně 70 % slyšitelných spojení neodpovídalo extrémním hodnotám identifikovaným pomocí ceny řetězení – a výzkum v této oblasti ceny řetězení rozhodně není ukončen.

Jak již naznačují i mé dosavadní komentáře, disertační práce Milana Legáta dle mého názoru neobsahuje žádné výraznější problematické aspekty. Protože je práce psaná v angličtině, považuji za vhodné zmínit se i o jazykovém aspektu. Ačkoli se v práci objevují občasné

chyby – nacházíme je převážně v oblasti členů a dále pak u jednotlivých výrazů, např. *extend* namísto *extent* (ve spojeních jako *to some extent*, *to what extent*; např. str. 39, 95, 142) nebo *thrill* namísto *trill* (str. 127,128) – rád bych vyzdvihnul, že autorova angličtina je obecně na velmi dobré úrovni.

Na závěr tedy shrnuji, že Milan Legát v předložené disertační práci přesvědčivým a srozumitelným způsobem popsal výsledky svého výzkumu v oblasti ceny řetězení v konkatenační syntéze řeči. Za přínosnou považuji snahu autora o využití i foneticky interpretovatelných parametrů a o nacházení přijatelných vysvětlení. Z podání práce je zřejmá výrazná znalostní úroveň autora, jeho vynalézavost při řešení problémů a metodický přístup k výzkumu. Části práce byly již autorem publikovány na předních odborných konferencích a týkají se různých dílčích témat v oblasti konkatenační syntézy. Disertační práce Milana Legáta splnila svůj cíl a zároveň nechává otevřený prostor pro pokračování výzkumu v budoucnosti.

Na základě výše uvedených skutečností jednoznačně doporučuji, aby předložená práce byla přijata jako práce disertační.

V Praze dne 21. ledna 2013

Doc. Mgr. Radek Skarnitzl, Ph.D.

Použitý odkaz:

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. Proc. of 15th ICPhS, pp. 771-774. Barcelona: ICPhS.

Ing. Petr HORÁK, Ph.D.
ÚFE AV ČR, v.v.i.
Chaberská 57
182 51 PRAHA 8 - Kobylisy

Oponentský posudek na disertační práci **Ing. Milana Legáta**

Concatenation Cost in Unit Selection Speech Synthesis

Disertační práce práce se zabývá tématem optimalizace ceny řetězení řečových jednotek při syntéze řeči metodou dynamického výběru řečových jednotek. Autor si vzal za hlavní úkol navrhnout novou metodiku určení ceny řetězení s ohledem na její maximální korelaci s percepcí rušivých artefaktů v bodech řetězení posluchačem. S ohledem na rozsah práce se autor zabýval cenou řetězení pěti českých samohlásek v provedení od dvou mluvčích – jednoho mužského a jednoho ženského.

V samotné práci se autor v úvodní kapitole po krátkém úvodu široce zabývá současným stavem problematiky určení ceny řetězení řečových jednotek při syntéze řeči metodou dynamického výběru řečových jednotek. Ve druhé kapitole se pak autor zabývá metodikou sestavování testovacího korpusu a poslechových testů. Třetí kapitola je věnována vlivu nespojitostí základního tónu v bodech řetězení na jejich percepci v syntetizovaném signálu. Autor zde dochází k závěru, že pro vnímání kvality řetězení v samohláskách je nejdůležitější průběh základního tónu v oblastech řetězení a ne jako tradičně používaný rozdíl hodnot základního tónu v bodě řetězení. Ve čtvrté kapitole se autor zabývá vlivem okolního souhláskového kontextu řetězené samohlásky. V páté kapitole autor navrhl metodiku měření percepční důležitosti různých metod určování ceny řetězení. V šesté kapitole pak autor shrnuje přínos vlastní práce a zabývá se výhledem do budoucna.

V závěru práce je uveden seznam použité literatury čítající 86 odkazů z toho v jednom případě je disertant autorem a ve třech případech spoluautorem. Na konci první kapitoly je uveden souhrn deseti publikací, které vznikly v rámci řešení disertační práce, z toho jedna publikace v impaktovaném odborném časopise a devět publikací na mezinárodních konferencích.

Vlastní disertační práce je psána v anglickém jazyce a má 178 stran s 56 obrázky a 32 tabulkami. Disertační práce působí uceleným dojmem, a je obsahově, jazykově i po grafické stránce na velmi vysoké úrovni. K disertační práci nebylo přiloženo datové médium.

Podle mého názoru disertace beze zbytku splnila sledovaný cíl a autorem publikované výsledky představují jednoznačný přínos pro obor syntézy řeči. Autorem zvolený postup řešení i použité metody považuji za vhodné pro danou problematiku. Za nejdůležitější výsledky v recenzované práci považuji především zjištění, že současné metody určování ceny řetězení nekorespondují s vnímáním rušivých artefaktů při řetězení řečových jednotek a následně navržené metody, které lépe korespondují s percepcí rušivých artefaktů v bodech řetězení. Tato disertační práce je zároveň příslibem pro další autorovu práci v oboru zpracování řeči.

Závěrem bych chtěl konstatovat, že autor splnil beze zbytku požadavky zadání. Disertace splňuje podmínky samostatné tvůrčí vědecké práce a obsahuje původní výsledky, které byly autorem publikovány na domácích i zahraničních konferencích. Práci doporučuji k obhajobě a na základě celkového dojmu ji hodnotím jako výbornou.

V Praze, dne 20. 2. 2013

.....

podpis oponenta