

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ		
1. Датум и орган који је именовao комисију:		
Решењем бр. 012-199/33-2019 од 01.04.2021. године, на основу Одлуке Научно-наставног већа Факултета техничких наука, декан Факултета техничких наука проф. др Раде Дорословачки, именовao је Комисију за оцену и одбрану докторске дисертације.		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :		
1. др Милан Сечујски	редовни професор	Телекомуникације и обрада сигнала, 11.03.2021.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Нови Сад		Председник
установа у којој је запослен-а		функција у комисији
2. др Никша Јаковљевић	ванредни професор	Телекомуникације и обрада сигнала, 11.10.2019.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Нови Сад		Члан
установа у којој је запослен-а		функција у комисији
3. др Јернеја Жганец-Грос	редовни професор	Информациони системи, 16.01.2017.
презиме и име	звање	ужа научна област и датум избора
Факултет за индустријски инжењеринг, Ново Место, Словенија		Члан
установа у којој је запослен-а		функција у комисији
4. др Бранислав Поповић	виши научни сарадник	Телекомуникације и обрада сигнала, 24.02.2020.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Нови Сад		Члан
установа у којој је запослен-а		функција у комисији
5. др Марко Јанев	виши научни сарадник	Математика, 20.12.2017.
презиме и име	звање	ужа научна област и датум избора
Математички институт САНУ, Београд		Ментор
установа у којој је запослен-а		функција у комисији
6. др Владо Делић	редовни професор	Телекомуникације и обрада сигнала, 28.03.2013.
презиме и име	звање	ужа научна област и датум избора
Факултет техничких наука, Нови Сад		Ментор
установа у којој је запослен-а		функција у комисији

II ПОДАЦИ О КАНДИДАТУ
<ol style="list-style-type: none"> 1. Име, име једног родитеља, презиме: Дарко, Јован, Пекар 2. Датум рођења, општина, држава: 06.03.1972, Нови Сад, Србија 3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив: Факултет техничких наука, Универзитет у Новом Саду, Енергетика, електроника и телекомуникације, Мастер инжењер електротехнике и рачунарства 4. Година уписа на докторске студије и назив студијског програма докторских студија: 2012. година, Енергетика, електроника и телекомуникације
III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:
Novel method for speaker adaptation in parametric speech synthesis (Нова метода адаптације на говорника у параметарској синтези говора)
IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:
Навести кратак садржај са назнаком броја страница, поглавља, слика, схема, графикона и сл.
<p>У докторској дисертацији под називом „Нова метода адаптације на говорника у параметарској синтези говора” кандидата Дарка Пекара приказани су поступци за синтезу говора коришћењем параметарског приступа, а на бази веома малог узорка говора жељеног говорника. Дисертација је написана на 103 странице на енглеском (са пратећом документацијом и проширеним изводом на српском дисертација садржи укупно 145 страница). Садржај је подељен у 7 поглавља. Дисертација садржи 39 слика, 4 табеле и 76 научних референци. На почетку тезе су дати: наслов, кључна документацијска информација на српском и на енглеском језику, сажетак на српском и на енглеском језику, проширени извод на српском језику, садржај рада, списак скрећеница, као и захвалница.</p> <p>Дисертација обухвата следећих 7 поглавља:</p> <ol style="list-style-type: none"> 1. Увод 2. Увод у синтезу говора 3. Статистичка параметарска синтеза говора 4. ДНН синтеза говора 5. Адаптација са полазног на циљног говорника 6. Адаптација у два корака 7. Закључак
V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:
<p>У уводном поглављу дефинисан је предмет истраживања. Наведена је дефиниција синтезе говора на основу текста и дати су неки примери употребе ове технологије. Назначена је потреба за једноставним моделовањем нових гласова и стилова, уз употребу мале количине снимљеног материјала. У складу са овом потребом дефинисани су циљеви истраживања: унапређење алата за тренирање модела за синтезу говора, базираних на дубоким неуронским мрежама; развој техника за адаптацију већ истренираних модела на нове говорнике и стилове, коришћењем врло мале количине говорног материјала. У овом поглављу је описана и организација тезе.</p> <p><i>Комисија сматра да су предмет и циљеви истраживања дефинисани јасно и концизно.</i></p> <p>У другом поглављу укратко је представљена историја синтезе говора, као и преглед основних приступа у синтези говора, са посебним освртом на параметарски приступ. Набројани су и поступци који се користе у оцени квалитета синтетизованог говора. Направљен је и увод у адаптацију на говорника и објашњена главна проблематика којом се теза бави.</p> <p><i>Комисија сматра да садржај овог поглавља омогућава разумевање теоријских основа</i></p>

истраживања. Наведене су предности и мане свих приступа и назначено због чега су у истраживању коришћени параметарски приступи.

У **трећем поглављу** описана је синтеза говора коришћењем скривених Марковљевих модела (енгл. *Hidden Markov Models – HMM*) и методе адаптације које се могу применити у овом приступу. Прво је описана *HMM* метода синтезе и дати мотиви за њену примену. Потом су описане две најчешће коришћене методе за адаптацију *HMM* модела на новог говорника, које се користе како у синтези, тако и у препознавању говора.

Комисија сматра да је примена скривених Марковљевих модела у синтези говора изложена на јасан и адекватан начин. Технике адаптације су такође детаљно описане, уз адекватне илустрације и референце. Садржај овог поглавља неопходан је за свеобухватно разумевање параметарских приступа.

У **четвртном поглављу** описана је синтеза говора коришћењем дубоких неуронских мрежа (енгл. *Deep Neural Networks – DNN*). Прво је дат опис функционисања дубоких неуронских мрежа, а потом је описана њихова примена у синтези говора. Описан је алат отвореног кода „Мерлин“, који је коришћен као основа за моделовање и обуку, као и унапређења која су спроведена у оквиру рада на овој дисертацији. Цитирана је литература и радови у којима је вршено поређење квалитета говора синтетизованог коришћењем скривених Марковљевих модела и дубоких неуронских мрежа на примеру говора за српски језик. У истим објективним и субјективним тестовима показано је да се употребом синтетизатора на бази дубоких неуронских мрежа добија говор бољег квалитета. Дат је и преглед тренутног стања технике кад је у питању адаптација на говорника, у случају да се за синтезу говора користе дубоке неуронске мреже.

Комисија сматра да су неуронске мреже веома детаљно описане, као и да је примена дубоких неуронских мрежа у синтези говора изложена на јасан и адекватан начин. Заједно са садржајем претходног поглавља даје свеобухватан преглед параметарских приступа. Стање технике везано за адаптацију на говорника је изложено на прегледан начин, уз одговарајуће референце.

У **петом поглављу** предложена је метода конверзије модела за синтезу говора, базираног на дубоким неуронским мрежама, са једног говорника на другог. Описан је начин обуке иницијалног модела, базираног на већој количини снимљеног материјала полазног говорника, као и накнадне адаптације тог модела коришћењем релативно мало материјала новог говорника. Прво је извршено поређење квалитета поравнања текста и снимка (на нивоу фонема) коришћењем стандардне технике преко монофона са предложеном техником и показано је да предложена техника изазива мање грешке у поравнању, нарочито када је на располагању мала количина говорног материјала новог говорника. Потом су извршени експерименти адаптације, објективна и субјективна евалуација резултата. Показано је да се, предложеном методом, сличан квалитет синтезе говора може добити са 2-6 пута мање материјала (зависно од методе евалуације), у поређењу са обуком модела од нуле. На крају овог одељка је извршена и процена утицаја карактеристика полазног говорника на резултат и добијен је закључак да карактеристике полазног говорника немају велик утицај (барем на скупу на којем су вршени експерименти).

Комисија сматра да је предложени приступ за конверзију модела за синтезу говора описан на јасан и адекватан начин. Предложена метода успешно је примењена у поравнању, као и у моделовању говора када је количина расположивог материјала мала. Описани експерименти спроведени су коректно, а резултати интерпретирани објективно.

У **шестом поглављу** предложена је још једна метода конверзије, која се ослања на неколико корака. Пре свега се обучи почетни модел за више говорника и креира одговарајући простор за уградњу говорника (*embedding space*). Овај модел је полазна тачка за све конверзије. У првом кораку конверзије за новог говорника се тражи одговарајућа тачка у простору говорника. Експериментално је утврђено да се већ након тог корака добија глас који је релативно сличан циљном говорнику. У другом кораку се врши адаптација остатка мреже на циљног говорника, слично као што је описано у петом поглављу. Спроведени су бројни експерименти, објективна и субјективна евалуација. Предложена метода је поређена са две основне методе: оне из петог

поглавља и још једне методе из литературе. Доказана је предност предложене методе, како објективним, тако и субјективним тестовима. Такође се испитивала важност два корака у адаптацији и утврђено је да је други важнији, али да и први корак доприноси квалитету говора који се на основу њих синтетизује. Такође се испитивао утицај количине материјала за адаптацију, и утврђено је да се разумљив говор сличан циљном говорнику може добити већ са 30 секунди говорног материјала, али да је за квалитетну синтезу ипак потребно 10 или више минута снимљеног говора.

Комисија сматра је нови предложени приступ за конверзију модела описан на јасан и адекватан начин. Предложени приступ успешно је примењен у конверзији и синтези говора, базираној на изузетно мало говорног материјала. Описани експерименти спроведени су коректно, а резултати интерпретирани објективно. Показана је предност предложеног приступа у поређењу са два основна и могућност добијања висококвалитетног модела за синтезу говора, базираног на само 10 минута говорног материјала.

У седмом поглављу сумирани су резултати истраживања и назначени су даљи правци истраживања.

Комисија сматра да су закључци донети на бази изложених резултата и да потврђују значај предложених приступа у генерисању модела за синтезу говора базираних на дубоким неуронским мрежама, а на основу веома мало расположивог говора циљног говорника. Предложени наредни кораци у истраживању су логични и мотивисани закључцима до којих се дошло током истраживања.

Литература садржи 76 прегледно систематизованих библиографских навода.

Комисија сматра да је литература адекватно одабрана, правилно коришћена и да одговара теми дисертације.

VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ:

Таксативно навести називе радова, где и када су објављени. Прво навести најмање један рад објављен или прихваћен за објављивање у складу са *Правилима докторских студија Универзитета у Новом Саду* који је повезан са садржајем докторске дисертације. У случају радова прихваћених за објављивање, таксативно навести називе радова, где и када ће бити објављени и приложити потврду уредника часописа о томе.

Рад у међународном часопису (M21-23)

1. Branislav Popović, Marko Janev, **Darko Pekar**, Nikša Jakovljević, Milan Gnjatović, Milan Sečujski, Vlado Delić, "A Novel Split-and-Merge Algorithm for Hierarchical Clustering of Gaussian Mixture Models", *Applied Intelligence*, ISSN: 0924-669X, Springer-Verlag New York, Inc., Vol. 37, No. 3, pp. 377-389 (2012) DOI: 10.1007/s10489-011-0333-9 (IF(2012)=1.853, Computer Science, Artificial Intelligence 32/115, M21)
2. Siniša Suzić, Tijana Delić, **Darko Pekar**, Vlado Delić, Milan Sečujski, "Style Transplantation in Neural Network Based Speech Synthesis", *Acta Polytechnica Hungarica*, ISSN: 1785-8860 Issue No. 16(6), pp. 171-189 (2019) (IF(2019)=1.219, Engineering, Multidisciplinary 63/91, M23)
3. Milan Sečujski, **Darko Pekar**, Siniša Suzić, Anton Smirnov, Tijana Nosek, "Speaker/Style-Dependent Neural Network Speech Synthesis Based on Speaker/Style Embedding", *Journal of Universal Computer Science*, ISSN: 0948-695X, Vol. 26, No. 4, pp. 434-453 (2020) (IF(2019)=0.701, Computer Science, Theory & Methods, 90/108, M23)
4. Tijana Nosek, Siniša Suzić, **Darko Pekar**, Radovan Obradović, Milan Sečujski, Vlado Delić, "Cross-Lingual Neural Network Speech Synthesis Based on Multiple Embeddings", *International Journal of Interactive Multimedia and Artificial Intelligence*, ISSN: 1989-1660, (accepted for publication in 2021), (IF(2019)=2.561, Computer Science, Artificial Intelligence 63/137, M22)

Рад у часопису националног значаја (M52)

1. Siniša Suzić, Tijana Delić, Stevan Ostrogonac, Simona Đurić, **Darko Pekar**, "Style-Code Method

for Multi-Style Parametric Text-To-Speech Synthesis”, *SPIIRAS Proceedings*, Issue No. 5(60), 2018, pp. 216-240, ISSN 2078-9181 (print), ISSN 2078-9599 (online), DOI: 10.15622/sp.60.8

Саопштење са међународног скупа штампано у целини (M33)

1. Siniša Suzić, Tijana Delić, Vladimir Jovanović, Milan Sečujski, **Darko Pekar**, Vlado Delić, “A comparison of multi-style DNN-based TTS approaches using small datasets”, *13th International Conference on Electromechanics and Robotics "Zavalishin's Readings (ER(ZR)-2018)*, Saint Petersburg, Russia, April 18-21, 2018, pages 6, MATEC Web of Conferences 161, DOI: 10.1051/mateconf/201816103005
2. Tijana Delić, Siniša Suzić, Milan Sečujski, **Darko Pekar**, “Rapid Development of New TTS Voices by Neural Network Adaptation”, *17th International Symposium INFOTEH-JAHORINA*, Jahorina, Bosnia and Hercegovina, March 21-23, 2018, pp. 1-6, DOI: 10.1109/INFOTEH.2018.8345518
3. Siniša Suzić, Tijana Delić, **Darko Pekar**, Vladimir Ostojić, „Novel alignment method for DNN TTS training using HMM synthesis models“, *IEEE 15th International Symposium on Intelligent Systems and Informatics, SISO 2017*, Subotica, Serbia, September 14-16, 2017, pp. 271-276
4. Milan Sečujski, Tijana Delić, Siniša Suzić, **Darko Pekar**, “Improvement of the Quality of Neural Network Based Speech Synthesis through Intra-Speaker Clustering”, *Internat. Acoustics and Audio Engineering Conference (TAKTONS)*, RTV Vojvodine, Fakultet tehničkih nauka, Novi Sad, Srbija, Srpska sekcija Audio Engineering Society (AES), 2019, pp. 9-10, ISBN: 978-86-6022-223-9

Саопштење са скупа националног значаја штампано у целини (M63)

1. Tijana Delić, Siniša Suzić, Stevan Ostrogonac, Simona Đurić, **Darko Pekar**, “Multi-style Statistical Parametric TTS”, *Zbornik radova konferencije Digitalna obrada govora i slike (DOGS 2017)*, Novi Sad, Srbija, November 22-25, 2017, pp. 5-8, ISBN: 978-86-7892-993-9
2. Branislav Popović, Edvin Pakoci, Nikša Jakovljević, **Darko Pekar**, “Experiments With Regression Trees for Speaker Adaptation in Continuous Speech Recognition”, *10th Conference on Digital Speech and Image Processing DOGS 2014*, Novi Sad, Serbia, 2014, pp. 47-50

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА:

У дисертацији се истражује проблем стварања висококвалитетних синтетичких гласова за системе синтезе говора на основу текста, када је доступна само мала количина података за обуку на циљни глас. Након увода и поређења уобичајених приступа синтези говора, илустрован је низ претходних покушаја решавања овог проблема. Неки од њих су се заснивали на старијим приступима (нпр. *HMM*), док су неки новији покушали да понуде решење коришћењем *DNN* архитектуре.

Предложена су два различита приступа адаптацији гласа говорника. Оба су модели синтезе говора засновани на *DNN*, способни за прилагођавање одређеном говорнику и стилу говора. Прва метода иницијално обучава модел на релативно великој количини материјала за обуку (3 или више сати) и користи тај модел као полазну тачку за адаптацију на циљни глас. Самим тим, нови модел није обучен на случајно иницијализованој, већ на претходно обученој мрежи, што је резултирало већим квалитетом синтетизованог говора. Није примећена значајна разлика када су коришћени различити почетни модели, што захтева додатна испитивања. Такође су коришћене различите величине адаптационог материјала и упоређен је квалитет добијеног говора. Показало се да је више материјала дало боље резултате, али и да прилагођавање чак и са релативно малом количином података може пружити упоредиве резултате са моделима обученим од нуле са много више говорног материјала. Евалуација се заснивала на објективним мерама, али и на тестовима слушања.

Друга предложена метода је поступак адаптације у два корака у којем се прво проналази оптимално уграђивање (енгл. *embedding*) за циљни глас на итеративни начин. Пре тога се формира модел трениран на много говорника и током тог процеса се формира простор за уградњу (енгл. *speaker embedding*). Други корак састоји се од прилагођавања остатка неуронске мреже, оптимизацијом свих тежина и помераја (енгл. *weights and biases*), тако да резултујућа мрежа може произвести говор циљног говорника. Будући да је излаз након прве фазе већ близу циља, очекивана количина промена примењених на мрежи је релативно мала. Ово спречава преобучавање мреже и омогућава много бољу генерализацију претходно невиђених догађаја.

Друга метода је показала да надмашује два друга недавно предложена параметарска модела синтезе

говора зависна од говорника и стила, посебно у случају кад је количина доступних података за адаптацију изузетно мала. То се постигло захваљујући заједничкој репрезентацији говорника и стила и њиховим уграђивањем у простор ниже димензије, при чему је модел у стању да утврди сличности међу говорницима и стиловима.

Приступ са уграђивањем отвара низ занимљивих могућих примена предложеног модела у било ком домену где је потребно брзо и ефикасно прилагођавање синтезе говора новом говорнику или стилу.

VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА:

Експлицитно навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.

На основу детаљног увида у садржај докторске дисертације комисија сматра да је кандидат јасно и прегледно приказао резултате истраживања. Сви добијени резултати представљени су помоћу одговарајућих табела и графикана. Интерпретација резултата је објективна и коректна.

Дисертација је проверена у софтверу за детекцију плагијаризма *iThenticate*. Извештај о подударности је показао да је дисертација оригинално ауторско дело кандидата (*Similarity Index* 17%). Утврђен нешто већи степен поклапања са постојећом литературом у складу је са очекивањима имајући у виду да је теза писана на енглеском језику, као и да укључује и резултате два недавно објављена рада (на енглеском) у којима је кандидат коаутор.

У складу са наведеним, Комисија **ПОЗИТИВНО** оцењује начин на који су резултати истраживања приказани и тумачени.

IX КОНАЧНА ОЦЕНА ДОКТОРСKE ДИСЕРТАЦИЈЕ:

Експлицитно навести да ли дисертација јесте или није написана у складу са наведеним образложењем, као и да ли она садржи или не садржи све битне елементе. Дати јасне, прецизне и концизне одговоре на 3. и 4. питање:

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме?

***ДА**, дисертација је у потпуности написана у складу са образложењем наведеним у пријави теме.*

2. Да ли дисертација садржи све битне елементе?

***ДА**, дисертација садржи све битне елементе који се захтевају по Статуту Факултета техничких наука и Универзитета у Новом Саду, као и Закона о високом образовању.*

3. По чему је дисертација оригиналан допринос науци?

У дисертацији су предложене две методе за прављење модела за синтезу говора, базиране на веома малој количини говорног материјала за обуку (у неким случајевима и значајно испод 10 минута). Ово представља фундаменталну промену кад је у питању креирање нових говорника и стилова, што је донедавно био веома дуготрајан и скуп процес, и захтевао је више сати снимљеног материјала чија обрада изискује значајно време и не може се у потпуности аутоматизовати, већ захтева ангажовање стручњака.

*Извршено је и поређење са другим методама, које су до недавно представљале *state of the art* у овом домену. Закључено је да предложене методе (а нарочито друга), значајно надмашују тренутно стање технике.*

4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?

*У дисертацији **нису** уочени недостаци који би утицали на резултате истраживања.*

X ПРЕДЛОГ:
На основу наведеног, комисија предлаже:
а) да се докторска дисертација прихвати, а кандидату одобри одбрана;
б) да се докторска дисертација врати кандидату на дораду (да се допуни односно измени);
в) да се докторска дисертација одбије.

Нови Сад, 12.04.2021.:

1. др Милан Сечујски, редовни професор
_____, председник
2. др Никша Јаковљевић, ванредни професор
_____, члан
3. др Јернеја Жганец-Грос, редовни професор
_____, члан
4. др Бранислав Поповић, виши научни сарадник
_____, члан
5. др Марко Јанев, виши научни сарадник
_____, ментор
6. др Владо Делић, редовни професор
_____, ментор

НАПОМЕНА: Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај и да исти потпише.