

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ			
1. Датум и орган који је именовao комисију:			
Декан Факултета техничких наука у Новом Саду, на предлог Наставно-научног већа Факултета техничких наука у Новом Саду, Решење број: 012-199/16-2021, дана 30.03.2023.			
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :			
1.	др Грбић Татјана	редовни професор	Теоријска и примењена математика, 19.02.2019.
	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 Лончар-Турукало Татјана	редовни професор	Телекомуникације и обрада сигнала, 26.04.2022.
	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 Јаковљевић Никша	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
4.	др Ђурић Никола	редовни професор	Теоријска електротехника, 22.04.2020.
	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 Николић Јелена	ванредни професор	Телекомуникације, 02.03.2020.
	презиме и име	звање	ужа научна област и датум избора
	Електронски факултет, Ниш		члан
	установа у којој је запослен-а		функција у комисији
6.	др Сечујски Милан	редовни професор	Телекомуникације и обрада сигнала, 11.03.2021.
	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
<p>1. Име, име једног родитеља, презиме: Тијана, Владо, Носек</p> <p>2. Датум рођења, општина, држава: 21.10.1992, Нови Сад, Србија</p> <p>3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив: Факултет техничких наука, Универзитет у Новом Саду; Мастер академске студије; Мастер инжењер електротехнике и рачунарства</p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија: 2016, Докторске академске студије Енергетика, електроника и телекомуникације</p>
III НАСЛОВ ДОКТОРСKE ДИСЕРТАЦИЈЕ:
Експресивни вишејезични синтетизатор говора
IV ПРЕГЛЕД ДОКТОРСKE ДИСЕРТАЦИЈЕ:
<p>Навести кратак садржај са знаком броја страница, поглавља, слика, схема, графикона и сл.</p> <p>Докторска дисертација кандидаткиње Тијане Носек, под насловом „Експресивни вишејезични синтетизатор говора”, написана је на 93 странице и структурирана у шест поглавља. Дисертација садржи 108 референци, 5 табела и 40 слика, а не садржи прилоге.</p> <p>На самом почетку дисертације, пре текстуалног дела, налазе се насловна страна и пратећи уводни материјал који садржи: кључну документацијску информацију на српском и енглеском језику, сажетак рада на српском и енглеском језику, захвалницу, садржај рада, те списак слика, списак табела и списак скраћеница.</p> <p>Докторска дисертација структурирана је кроз следећих шест поглавља:</p> <ol style="list-style-type: none"> 1. Увод; 2. Систем за синтезу говора на основу текста; 3. Примена неуралних мрежа у TTS; 4. Проширење основног TTS модела на бази DNN; 5. Експресивни вишејезични модел; 6. Закључак. <p>На крају дисертације наведен је списак коришћене научне литературе.</p>
V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСKE ДИСЕРТАЦИЈЕ:
<p>Наслов дисертације</p> <p>Комисија сматра да наслов јасно назначавља тематику која је предмет истраживања које је у докторској дисертацији приказано.</p> <p>Поглавље 1 – Увод</p> <p>У уводном поглављу објашњен је предмет докторске дисертације и изложени су циљеви истраживања које је у оквиру ње спроведено. Дат је сажет преглед досадашњег развоја у области синтезе говора на основу текста, као технологије која омогућује рачунару да се обрати човеку говором, и изложен је проблем недовољне флексибилности синтетизатора говора који су реализовани на основу класичних метода. Истакнут је значај машинског учења, посебно неуралних мрежа, у реализацији знатно флексибилнијих модела, који омогућују промену битних параметара говора као што су карактеристике гласа говорника или говорног стила. Дат је преглед примена синтезе говора на основу текста, са посебним освртом на примене синтетизатора високог квалитета и флексибилности, а истакнути су и главни научно-истраживачки изазови и практичне препреке у развоју таквих синтетизатора. Уводно поглавље закључено је кратким прегледом садржаја остатка дисертације.</p> <p>Мишљење Комисије је да је уводно поглавље пружио адекватну слику о мотивацији и потреби за истраживањем, као и о самом току и циљевима истраживања.</p>

Поглавље 2 – Систем за синтезу говора на основу текста

У оквиру овог поглавља дат је детаљан приказ типичне унутрашње структуре синтетизатора говора на основу текста и указано је на високу језичку зависност појединих његових елемената. Представљени су најважнији изазови евалуације квалитета синтетизованог говора и описане су методе које се за то најчешће користе, а које су коришћене и у оквиру самог истраживања. Истакнута је потреба за језичком обрадом текста који је потребно конвертовати у говор и представљени су конкретни задаци које она типично обухвата, као и поједине методе којима се они решавају. Детаљно су изложени приступи акустичком моделовању код различитих метода синтезе говора, као и методе анализе и синтезе говорног сигнала помоћу вокодера.

Комисија се слаже да је у оквиру овог поглавља дат јасан увид у стање истраживања у области синтезе говора, као и да је коришћена литература актуелна и релевантна.

Поглавље 3 – Примена неуралних мрежа у TTS

Ово поглавље даје преглед примена неуралних мрежа за решавање појединих задатака у оквиру синтезе говора, укључујући и приступе код којих је читав систем реализован у виду неуралне мреже. Следи преглед теоријских основа машинског учења на основу неуралних мрежа, укључујући структуру основних градивних јединица – неурона, начине њиховог повезивања, као и различите методе обуке неуралних мрежа на обимним скуповима података у циљу њиховог прилагођавања решавању конкретних проблема као што је синтеза говора. Поглавље се завршава детаљним прегледом класичне структуре синтетизатора говора заснованог на неуралним мрежама, односно модула намењеног синтези говорног сигнала на основу језички обрађеног текста.

Комисија сматра да су у поглављу 3 јасно приказане теоријске основе дисертације, у виду модела који је у оквиру истраживања проширен, односно модификован у циљу повећања његове флексибилности и увођења других напредних могућности.

Поглавље 4 – Проширење основног TTS модела на бази DNN

На почетку поглавља 4 дат је преглед начина на које је могуће остварити синтезу говора различитим говорним стилем или гласом неког другог говорника, представљен је концепт адаптације неуралне мреже на други говорни стил или глас неког другог говорника, као и приступи на основу којих се синтетизовани говор једног говорника трансформише у синтетизовани говор неког другог говорника на нивоу акустичких обележја или на нивоу таласног облика сигнала. Затим је дат детаљан приказ метода којима је могуће постићи синтезу говора на више језика уз ефикасно комбиновање знања стечених о различитим језицима, а дат је и детаљан приказ истраживања на тему повећања експресивности синтезе.

Оцена комисије је да је у оквиру поглавља 4 јасно и детаљно представљено стање истраживања у области синтезе говора напредних могућности као што су променљив говорни стил, глас говорника или језик синтезе.

Поглавље 5 – Експресивни вишејезични модел

У поглављу 5 представљен је модел који је главна тема докторске дисертације. Овај модел подржава синтезу говора у више говорних стилова и гласовима више говорника, као и на више језика, а омогућава синтезу како комбинације говорник-језик која постоји у скупу за обуку, тако и оних комбинација које у њему не постоје. Стандардни модел изложен у претходном поглављу прво је проширен са *embedding* слојем за јединствене комбинације говорник-стил-кластер, затим и са фонетским *embedding* слојевима, а у случајевима када је било потребно обухватити и језике за које се користе различите конвенције прозодијске анотације додат је и прозодијски *embedding*.

У поглављу су систематично представљени резултати низа експеримената са циљем евалуације квалитета говора синтетизованог под различитим условима, са јасном назнаком о циљу сваког експеримента. Истакнута је разлика између сценарија у ком је потребно синтетизовати говор гласом одређеног говорника на језику за који су на располагању снимци тог говорника и сценарија у ком ти снимци нису на располагању (тзв. *cross-lingual* сценарио). Експерименти, који укључују евалуацију на основу поређења објективних акустичких параметара као и евалуацију на основу тестова слушања, потврђују могућност

синтезе разумљивог говора који звучи природно на основу предложених архитектура и метода обуке.

Комисија сматра да су описани експерименти спроведени коректно, а резултати интерпретирани објективно. Описани експерименти имају јасан допринос целокупном истраживању које представља значајан корак у развоју говорних технологија.

Поглавље 6 – Закључак

У закључку су сумирани резултати истраживања и назначени су даљи правци рада.

Комисија сматра да су коректно и логично представљени наредни кораци у развоју вишејезичне експресивне синтезе говора на основу неуралних мрежа.

Литература

Комисија сматра да коришћена литература осликава систематичан приступ истраживању. Коришћена литература је актуелна и обухвата истраживања на тему експресивне синтезе говора са посебним освртом на методе засноване на неуралним мрежама, што је у складу са циљевима ове 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 истакнутом међународном часопису (M22)

- 1) **Nosek T.**, Suzić S., Pekar D., Obradović R., Sečujski M., Delić V.: Cross-Lingual Neural Network Speech Synthesis Based on Multiple Embeddings, *International Journal of Interactive Multimedia and Artificial Intelligence*, ISSN 1989-1660, Vol. 7, No.2, 2021, DOI: 10.9781/ijimai.2021.11.005. (IF: 4.936, 2021)
- 2) Simić N., Suzić S., **Nosek T.**, Vujović M., Perić Z., Savić M., Delić V.: Speaker recognition using constrained convolutional neural networks in emotional speech, *Entropy: international and interdisciplinary journal of entropy and information studies – an Open Access Journal by MDPI*, ISSN: 1099-4300, *Special Issue “Methods in Artificial Intelligence and Information Processing,”*, Vol. 24, No.3, 2022, DOI: 10.3390/e24030414. (IF: 2.738, 2021)

Рад у међународном часопису (M23)

- 3) Sečujski M., Pekar D., Suzić S., Smirnov A., **Nosek T.**: Speaker/Style-Dependent Neural Network Speech Synthesis Based on Speaker/Style Embedding, *Journal of Universal Computer Science*, ISSN: 0948-695X, Vol. 26, No. 4, DOI: 10.3897/jucs.2020.023, 2020. (IF: 1.139, 2020)
- 4) Suzić S., **Delić T.**, Pekar D., Delić V., Sečujski M.: Style Transplantation in Neural Network-based Speech Synthesis, *Acta Polytechnica Hungarica*, ISSN: 1785-8860, Vol. 16, No. 6, DOI: 10.12700/APH.16.6.2019.6.11, 2019. (IF: 1.219, 2019)
- 5) Delić V., Perić Z., Sečujski M., Jakovljević N., Nikolić J., Mišković D., Simić N., Suzić S., **Delić T.**: Speech Technology Progress Based on New Machine Learning Paradigm, *Computational Intelligence and Neuroscience*, ISSN: 1687-5265, Vol. 2019, DOI: 10.1155/2019/4368036, 2019. (IF: 2.284, 2019)

Рад у часопису националног значаја (M52)

- 6) **Delić T.**, Sečujski M., Suzić S.: A Review of Serbian Parametric Speech Synthesis Based on Deep Neural Networks, *TELFOR Journal*, ISSN: 1821-3251, Vol. 9, No. 1, 2017.

Саопштење са међународног скупа штампано у целини (M33)

- 7) Simić N., Suzić S., **Nosek T.**, Vujović M., Sečujski M.: Impact of different voting strategies in CNN based speech emotion recognition, 30th European Signal Processing Conference – EUSIPCO 2022, Београд, Србија, 2022.
- 8) **Nosek T.**, Suzić S., Vujović M., Pekar D., Sečujski M., Delić V.: Explicit control of the level of expressiveness in DNN-based speech synthesis by embedding interpolation, The 23th International Conference Speech and Computer – SPECOM 2021, Санкт Петербург, Русија, 2021.
- 9) Delić V., Mišković D., Popović B., Sečujski M., Suzić S., **Delić T.**, Jakovljević N.: Central Audio-Library of the University of Novi Sad, International Symposium on Intelligent Distributed Computing (IDC), Санкт Петербург, Русија, 2019.
- 10) Suzić S., **Delić T.**, Jovanović V., Sečujski M., Pekar D., Delić V.: A comparison of multi-style DNN-based TTS approaches using small datasets, International Conference on Electromechanics and Robotics – ER(ZR), Санкт Петербург, Русија, 2018.

Саопштење са скупа националног значаја штампано у целини (M63)

- 11) Mandarić I., Vujović M., Suzić S., **Nosek T.**, Simić N., Delić V.: Initial Analysis of the Impact of Emotional Speech on the Performance of Speaker Recognition on New Serbian Emotional Database, 29th Telecommunications forum TELFOR 2021, Београд, Србија, 2021.
- 12) Đurkić T., Lojaničić A., Suzić S., Popović B., Sečujski M., **Nosek T.**: Emotion recognition from speech based on ML algorithms applied on two Serbian datasets, 29th Telecommunications forum TELFOR 2021, Београд, Србија, 2021.
- 13) Sečujski M., **Delić T.**, Suzić S., Pekar D.: Improvement of the Quality of Neural Network Based Speech Synthesis through Intra-Speaker Clustering, International Acoustics and Audio Engineering Conference (TAKTONS), Нови Сад, Србија, 2019.
- 14) **Delić T.**, Suzić S., Sečujski M., Ostojić V.: Deep neural network speech synthesis based on adaptation to amateur speech data, International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), Палић, Србија, 2018.
- 15) **Delić T.**, Suzić S., Sečujski M., Pekar D.: Rapid Development of New TTS Voices by Neural Network Adaptation, Међународни симпозијум ИНФОТЕХ, Јахорина, БиХ, 2018.
- 16) Suzić S., **Delić T.**, Pekar D., Ostojić V.: Novel alignment method for DNN TTS training using HMM synthesis models, IEEE International Symposium on Intelligent Systems and Informatics (SISY), Суботица, Србија, 2017.
- 17) **Delić T.**, Suzić S., Ostrogonac S., Etinski S., Pekar D.: Multi-style Statistical Parametric TTS, Дигитална обрада говора и слике (ДОГС), Нови Сад, Србија, 2017.

Напомена: У периоду до 2019. године кандидаткиња је научне радове објављивала под девојачким презименом **Делић**.

VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА:

Докторска дисертација представља истраживање које обухвата комплетан процес развоја, односно формирања вишејезичног експресивног модела који подржава употребу гласова више говорника. Овакав модел омогућава не само продукцију експресивног говора гласом различитих говорника на њиховом оригиналном језику, већ и на другим језицима из говорне базе. Најважнији закључци докторске дисертације су следећи:

- 1) Висок квалитет синтетизованог говора, а уз значајну флексибилност која се огледа у могућности промене говорног стила, идентитета говорника, па чак и језика синтезе може се постићи коришћењем различитих *embedding* простора, како фонетских (једна тачка представља један фонем) тако и говорничких (једна тачка представља једну комбинацију говорник-стил-кластер).
- 2) Формирањем прозодијских *embedding* простора може се постићи да се вишејезични модели могу креирати не само за језике за које су коришћене исте конвенције

<p>прозодијске аотације, већ и за језике код којих су конвенције прозодијске аотације различите.</p> <p>3) Могућност примене знања стечених о једном језику на други језик, захваљујући <i>embedding</i> просторима, значајно смањује захтеве у погледу величине говорних база које су неопходне за синтезу разумљивог говора природне интонације. Модел не захтева ни велике количине говорног материјала циљног говорника, а омогућава синтезу гласом циљног говорника на ма ком језику подржаном моделом.</p> <p>4) Евалуација предложених архитектура и метода кроз објективна мерења као и субјективне тестове слушања потврдила је висок квалитет синтетизованог говора, односно одржавање квалитета у погледу природности и разумљивости у кључном сценарију синтезе – у комбинацији говорник-језик која не постоји у скупу за обуку.</p>
<p>VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА: Експлицитно навести позитивну или негативну оцену начина приказа и тумачења резултата истраживања.</p> <p>На основу детаљног увида у садржај докторске дисертације од стране чланова Комисије, закључено је да је истраживање пажљиво испланирано и систематично извршено, да су експерименти адекватно спроведени и да су резултати истраживања интерпретирани коректно и објективно. Оцена комисије је, дакле, позитивна.</p> <p>Рад је проверен у софтверу за детекцију плагијаризма <i>iThenticate</i>, у Библиотеци Факултета техничких наука у Новом Саду. Анализом извештаја о подударности текста дисертације са другим изворима, Комисија је утврдила да је проценат подударности занемарљив и донела закључак да је докторска дисертација оригинално ауторско дело.</p>
<p>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:</p>
<p>1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме? Увидом у документацију приложену приликом пријаве теме, Комисија је утврдила да је докторска дисертација написана у складу са образложењем наведеним у пријави теме.</p>
<p>2. Да ли дисертација садржи све битне елементе? Комисија констатује да докторска дисертација садржи све битне елементе који се захтевају по Статуту Факултета техничких наука и Универзитета у Новом Саду, као и Закона о високом образовању.</p>
<p>3. По чему је дисертација оригиналан допринос науци? У оквиру дисертације предложене су нове архитектуре и методе обучавања неуронских мрежа које омогућују аутоматску синтезу експресивног говора на основу текста, чак и гласом говорника за кога није на располагању говорни материјал на траженом језику, а предложен је и начин адаптације таквог модела на новог говорника. У оквиру истраживања, које је обухватило и српски језик, креирани су вишејезични модели од којих неки по својим перформансама чак превазилазе стандардне моделе обучене на говорном материјалу на једном језику.</p> <p>Таква адаптација омогућује брзу и једноставну продукцију нових гласова задржавајући могућност синтезе на свим језицима подржаним моделом, без обзира на оригинални језик новог говорника.</p>
<p>4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања? У докторској дисертацији нису уочени значајни недостаци који би могли утицати на резултате истраживања.</p>

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, 17.04.2023.

1. **др Грбић Татјана**, редовни професор

_____, председник

2. **др Лончар-Турукало Татјана**, редовни професор

_____, члан

3. **др Јаковљевић Никша**, ванредни професор

_____, члан

4. **др Ђурић Никола**, редовни професор

_____, члан

5. **др Николић Јелена**, ванредни професор

_____, члан

6. **др Сечујски Милан**, 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.