

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидаткиње Милане Милошевић, дипл. инж. електротехнике – мастер за израду докторске дисертације „Идентификација говорника у условима емотивног говора“ (“Speaker recognition in conditions of emotional speech”)

Одлуком 5026/12-1 бр. од 19.07.2018. године, именовани смо за чланове Комисије за оцену подобности теме и кандидаткиње Милане Милошевић за израду докторске дисертације и научне заснованости теме „Идентификација говорника у условима емотивног говора“. На основу материјала приложеног уз Захтев кандидаткиње, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Милана Милошевић рођена је 23.05.1988. Била је ђак генерације и носилац Вукове дипломе у основној школи, као и у Математичкој гимназији. Освајала је награде на такмичењима из математике, физике и информатике и представљала Србију на Међународној Олимпијади из физике 2007. године у Ирану. Електротехнички факултет, Одсек за сигнале и системе, завршила је у року – основне студије са просечном оценом 9.91 и мастер студије са просечном оценом 10. Њен дипломски рад награђен је трећом наградом од стране удружења БАФА. Активно је учествовала у раду Факултета као представница студената у Наставно-научном већу, Савету факултета и као студенткиња продеканка.

Докторске студије је уписала 2012 године на Електротехничком факултету на Одсеку управљање системима и обрада сигнала. Положила је све испите са просечном оценом 10 и тренутно је посвећена изради докторске дисертације.

Течно говори енглески, шпански и немачки језик. Води блог о личном и професионалном развоју и активан је тркач дугопругаш. Иза себе има три маратонске и три полумаратонске трке.

1.2. Сечено научно-истраживачко искуство

Област истраживања Милане Милошевић укључује обраду сигнала, моделирање и идентификацију система, препознавање облика, карактеризацију сигнала говора и напредне технике за обраду сигнала говора. Међу важним истраживањима могу се сврстати истраживања у области препознавања емоција у говору, као и идентификација говорника на основу сегментних карактеристика гласа под руководством проф. Жељка Ђуровића. Током јесени и зиме 2015. године и пролећа и лета 2016. године била је на радној студијској пракси у Цириху, Швајцарска у Швајцарским федералним лабораторијама за материјале, науку и технологије – ЕМПА на одељењу за Акустику, под надзором инжењера електротехнике Улрике Главич. Милана Милошевић је докторске студије уписала 2012. године на Електротехничком факултету Универзитета у Београду, где је положила све испите са просечном оценом 10.00. Предмети које је положила су Нелинеарни мултиваријабилни системи, Говорне технологије, Вештачка интелигенција, Технике обраде и препознавања говорног сигнала, Комуникологија, Линеарни робусни системи, Интернет програмирање, Оптимални и адаптивни стохастички системи, Класификација и есимација сигнала и Софтверски системи за рад у реалном времену. Одређен јој је ментор дисертације, проф. др Жељко Ђуровић и период од 2012. до 2018. године провела је у истраживању и припреми докторске дисертације. Тема која је предложена за докторску дисертацију ове кандидаткиње има мултидисциплинарни карактер и домен интересовања се налази у областима обраде сигнала говора, статистичке анализе сигнала, психологије људског понашања и природе емоција. Досадашњи резултати кандидаткиње су приказани кроз неколико публикација. У даљем тексту је наведен списак публикација на којима је Милана Милошевић коаутор.

Саопштење са међународних скупова штампано у целини (Категорија М33)

- [1] **Milošević M.**, Glavitsch U. "Combining Gaussian mixture models and segmental feature models for speaker recognition", INTERSPEECH, 20 – 24 August 2017, Stockholm, Sweden
- [2] **Milošević M.**, Nedeljković Ž. "Emotional speech databases in Slavic languages – an overview", 26th IAFPA, 9 – 12 July 2017, Split, Croatia
- [3] Nedeljković Ž., **Milošević M.**, Đurović Ž., "Baseline system for speaker recognition - parameter analysis" – IcETRAN, June 5-8, 2017, Kladovo, Serbia
- [4] **Milošević M.**, Glavitsch U., He L. & Dellwo V. "Segmental features for automatic speaker recognition in a flexible software framework", 25th IAFPA, 24 – 27 July 2016, York, UK
- [5] **Milošević M.**, Nedeljković Ž., Đurović Ž., "SVM Classifier for Emotional Speech Recognition in Software Environment SEBAS" – IcETRAN, June 13 – 16, 2016, Zlatibor, Serbia
- [6] **Milošević M.**, Đurović Ž., "Challenges in Emotional Speech Recognition, " – IcETRAN, June 8-11, 2015, Silver Lake, Serbia

[7] **Milošević M.**, Đurović Ž., "Feature Extraction for Emotion Classification from Speech Signal," IcETRAN, June 2-5, 2014, Vrnjačka Banja, Serbia

Саопштења са националних скупова штампано у целини (Категорија M63)

[1] **Milošević M.**, Đurović Ž., "Social Signal Processing: Elements of Analysis and Applications," ETRAN, June 2013, Zlatibor, Serbia

Дана 24.09.2018. године Милана Милошевић полагала је испит о подобности теме и кандидата (докторски испит) пред комисијом у саставу: ванредни професор др Горан Квашчев, професор др Зоран Перић, ванредни професор др Милош Цветановић. Оцена Комисије је да је кандидаткиња Милана Милошевић на том испиту добила оцену: задовољила.

1.3. Оцена подобности кандидата за рад на предложеној теми

Током писања овог Извештаја, Комисија је узела у обзир све претходно наведене чињенице и резултате досадашњег рада кандидаткиње Милане Милошевић. Такође су са посебном пажњом размотрена искуства која је кандидаткиња стекла, како кроз образовни, тако и истраживачки процес. Узимајући у обзир све наведене компоненте које чине истраживачки профил кандидаткиње за докторску дисертацију, Комисија је стекла увид у способности кандидаткиње за научно-истраживачки рад, као и њене могућности да током израде предложене докторске дисертације, оствари резултате који могу бити приказани и валоризовани од стране међународне научне заједнице. Став Комисије је позитиван и након детаљног образложења саме теме предложиће да она буде усвојена, а кандидаткињи дата могућност да настави са израдом докторске дисертације.

2. Предмет и циљ истраживања

Препознавање говорника поново је добило значајну пажњу са појавом паметних, персонализованих система и уређаја. Системи за идентификацију говорника постали су саставни део сервиса аутентификације за на пример телефосно банкарство, претрага интернета и приступање обезбеђеним информацијама. Поред примена у свакодневном животу, идентификација говорника је од интереса у форензичким истраживањима, телефонским сервисним центрима, центрима за хитне случајеве, у трансферу поверљивих података итд. У тим случајевима, говорници углавном не користе уобичајени тон гласа. Када је говорник узнемирен, под стресом, плаче, смеје се, шапуће или виче, карактеристике његовог гласа су измењене. У системима за идентификацију говорника, то може довести до грешке у идентификацији говорника. Такве ситуације изазивају фрустрацију корисника, а у хитним случајевима, нарочито када су војне и безбедносне примене у питању, ово може довести до озбиљних последица. Циљ је креирање система за препознавање говорника који је што робуснији, а при томе користи што мање података.

Предмет истраживања кандидата је утицај емоција на варијације карактеристика сигнала говора и начина на који се ове промене одражавају на прецизност система за препознавање

Предмет истраживања кандидаткиње је утицај емоција на варијације карактеристика сигнала говора и начина на који се ове промене одражавају на прецизност система за препознавање говорника. Дистанце између говорника, у координатном систему вектора карактеристика, мењају се под утицајима емоција, што за последицу има измене перформанси система. Са друге стране, ако се овакве варијације могу моделирати кроз систем, макар одглумљеним говором, биће могуће урадити и адаптацију система. Последње, биће испитивано и колико на квалитет модела утиче количина искоришћених података за обуку модела.

Циљ истраживања кандидаткиње је употреба емотивног говора за повећану прецизност система за препознавање говорника. На основу анализе утицаја емотивног стања на измене карактеристика говора конструисани систем би за обуку осим уобичајеног неутралног говора користио и емотивни говор говорника. Осим тога, циљ је и да се систем за препознавање говора учини робуснијим на промене емотивног стања говорника.

У циљу илустрације актуелности предложене теме у наставку су наведене значајне референце из области на чијим резултатима ће се базирати предложено истраживање.

- [1] Hasan, T., Sadjadi, S. O., Liu, G., Shokouhi, N., Bořil, H., Hansen, J. H. L. 'CRSS systems for 2012 NIST Speaker Recognition Evaluation.' IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2013, pp. 6783-6787.
- [2] Zhang, C., Bahmaninezhad, F., Ranjan, S., Yu, C., Shokouhi, N. and Hansen, J. H. L. 'UTD-CRSS Systems for 2016 NIST Speaker Recognition Evaluation,' Proceedings of Interspeech, 2017, pp. 1343-1347.
- [3] Lee, K. A. et al 'I4U Mega Fusion and Collaboration for NIST Speaker Recognition Evaluation 2016.' Proceedings of the Interspeech, 2017, pp. 1328-1332.
- [4] T. Wu, Y. Yang, and Z. Wu, 'Improving speaker recognition by training on emotion-added models' in Int. Conf. on Affective Computing and Intelligent Interaction, Memphis, TN, USA, 9-12 October 2005, pp. 382-389.
- [5] D. Li, Y. Yang, and W. Dai, 'Cost-sensitive learning for emotion robust speaker recognition' in The Scientific World Journal, vol. 2014(2014).
- [6] P. Song, Y. Jin, C. Zha, and L. Zhao, 'Speech emotion recognition method based on hidden factor analysis,' in Electronics Letters, 2014, vol 51(1), pp. 112-114.
- [7] Leemann, A., Kolly, M. J., Dellwo, V.: 'Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison', Forensic Science International, 2014, 238, pp. 59-67.

- [8] Markova V., Pertushin, V.A., 'RUSLANA: a database of russian emotional utterances', Proc. 2002 Int. Conf. Spoken Language Processing (ICSLP), Colorado, USA, September 2002, pp. 2041-2044
- [9] Alluri, K. R., Achanta, S., Prasath, R., Gangashetty, S. V., & Vuppala, A. K. (2016, November). A Study on Text-Independent Speaker Recognition Systems in Emotional Conditions Using Different Pattern Recognition Models. In International Conference on Mining Intelligence and Knowledge Exploration (pp. 66-73). Springer, Cham.
- [10] Nagrani, A., Chung, J. S., & Zisserman, A. (2017). VoxCeleb: a large-scale speaker identification dataset. arXiv preprint arXiv:1706.08612.
- [11] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 788–798, 2011
- [12] Reynolds, D. (2015). Gaussian mixture models. Encyclopedia of biometrics, 827-832.

3. Полазне хипотезе

Истраживање које се предлаже у оквиру ове докторске дисертације полази од хипотеза које су донекле поткрепљене резултатима доступним у литератури, али и неким прелиминарним резултатима саме кандидаткиње. Оне се могу формулисати на следећи начин:

- Систем за препознавање говорника обучен коришћењем искључиво неутралног говора има знатно ниже перформансе када је потребно урадити препознавање говорника који је у афективном стању.
- Стандардни систем за идентификацију говорника постаје робуснији у односу на говорнике који су у неком емотивном стању ако се емотивни говор укључи у процес обуке система.
- Квалитет модела говорника у великој мери зависи и од количине података који су искоришћени у обуци модела.
- Анализом карактеристика говора може се доћи до закључка и о емотивном стању и о говорнику.
- Коришћењем информације о емотивном стању говорника долази се до боље тачности у препознавању говорника.
- Мел-кепстрални коефицијенти који се уобичајено користе за идентификацију говорника су врло осетљиви на појаву афективног стања говорника.
- Укључивање неартикулисаних звукова говора, фреквенције појављивања појединих свесних или несвесних говорних сегмената, као и динамичких дескриптора говорног сигнала у вектор обележја који се класификује, може повећати ефикасност целе процедуре идентификације.

4. Научне методе истраживања

Током истраживања која су планирана предлогом ове докторске дисертације, доминанто ће бити коришћене методе анализе, апстракције, спецификације и синтезе. Конкретне појединости о појединим фазама истраживања и научним методама којима ће оне бити реализоване, дате су у следећим пасусима.

Сви резултати кандидаткиње биће верификовани кроз одговарајуће експерименте. Као материјал за експерименте, биће коришћене стандардизоване базе неутралног говора, конструисане са наменом за експерименте у области препознавања говора и говорника, као и базе емотивног говора српског, руског и/или других језика. Чињеница је да се на интернету могу наћи доступне говорне базе одређеног броја језика, међутим нису све оне једнако употребљиве. Многе од тих база су некавалитетне, са интензивним позадинским шумом, многе од њих нису на одговарајући начин лабелиране, те се не могу са потпуном сигурношћу разврстати присуство и типови емоција. Такође су неке од база снимане од стране волонтера који нису на одговарајући начин припремљени за ваљано испољавање одговарајућих емотивних експресија, те поново резултати добијени на таквим базама морају бити узети са резервом. Због тога ће се у раду кандидаткиња ослонити само на базе које су снимане у контролисаним условима који обезбеђују одговарајући квалитет експеримената и резултата.

Експерименти ће бити изведени програмским кодом имплементираним од стране кандидаткиње, као и прилагођеним коришћењем програмских пакета OpenSmile;) и PRAAT. Основне карактеристике гласа које ће се користити за даљу обраду биће мел-фреквенцијски кепстрални коефицијенти. Међутим, с обзиром на доказану осетљивост ових коефицијената у условима афективног говора, значајан део истраживања биће посвећен укључивању неартикулисаних елемената говора као и динамичких дескриптора акустичког сигнала у вектор обележја. Као основни алгоритам за идентификацију говорника биће употребљене Гаусове мешавине. Избор класификатора, односно технике за идентификацију, ће такође бити једна од озбиљних тачака истраживања. У литератури је уобичајен и најчешће коришћен модел Гаусове мешавине који једноставно имплицира употребу Бајесовог класификатора за идентификацију, међутим овакав приступ показује и извесне недостатке. Првенствено у смислу квалитета процене кроскорелационих коефицијената између појединих обележја. Зато ће кандидаткиња посветити одређену пажњу и примењивости других техника за класификацију.

Коначно, посебна ће пажња бити посвећена евалуацији добијеног система за идентификацију. Значајан је моменат поделе доступних снимака на обучавајуће и тестирајуће, јер се врло често ова подела јасно одражава на веродостојност процене тачности система. Идеја је да се након пројектовања целокупног алгоритма за идентификацију, изгради и интерактивни модул намењен корисницима који нису у бази, чиме би се евалуацији развијеног система сагледала и са другог, практичног аспекта.

5. Очекивани научни доприноси

Међу очекиваним научним доприносима који ће се појавити као резултат истраживања које је предложено овом докторском дисертацијом, издвајају се следећи.

- Процена успешности коришћења гласа као биометријског податка за идентификацију говорника у условима емотивног говора. За ову намену биће искоришћени системи за идентификацију који су тренутно доступни у литератури.
- Детаљно испитивање степена одступања модела говорника у зависности од емотивног стања говорника са становишта језика, пола, квалитета снимака, типа емоција, величине базе.
- Анализа и избор адекватног модела емотивног говора у обуке модела говорника, полазећи од стандардних модела као што је Гаусова мешавина у простору мел-кепстралних коефицијената, али и анализа повећања ефикасности класификатора у условима додавања других лингвистичких атрибута и неартикулисаних говорних сегмената.
- Анализа и избор процеса идентификације/класификације говорника у условима емотивног говора. Анализа ће кренути од једноставнијих класификатора као што су "наивни Бајес", према сложенијим техникама класификације и кластеризације.
- Искрпна евалуација квалитета развијеног система за идентификацију говорника при емотивном говору а са различитих аспеката, као што су величина базе, квалитет снимака у бази, пола говорника, интензитета емоција и типа емоција.

6. План истраживања и структура рада

Свакако је први и врло важан сегмент плана истраживања да се кандидаткиња упозна са литературом о препознавању говорника и емоција у говору, као и постојећим базама података емотивног говора.

У првој фази истраживања кандидаткиња ће вршити имплементацију и тестирање номиналног система за препознавање говорника коришћењем Гаусових мешавина и мел-фреквенцијских кепстралних коефицијената. Обука ће се вршити коришћењем базе података неутралног говора ТЕВОИД, која је намењена за истраживања идентификације говорника, зарад оптималног подешавања параметра система. Након тога упознаће се са интерпретацијом добијених резултата и писањем научног чланка.

У другој фази истраживања, кандидаткиња ће у експерименте укључити остале базе података емотивног говора. Поменути ситем биће најпре обучаван и тестиран коришћењем неутралног говора из поменутих база. Затим ће се тестирање извршити и емотивним говором. Након тога, кандидаткиња ће на различите начине укључити и емотивни говор у обуку модела. Тако добијени модели биће тестирани и неутралним и

емотивним говором како би се одредила успешност у односу на почетни модел говорника обучаван коришћењем искључиво неутралног говора.

У трећој фази истраживања, кандидаткиња ће имплементирати кластеризацију емотивног стања говорника и тестирати успешност овог система. У овој фази истраживања се очекује врло исцрпна компаративна анализа између различитих приступа у параметарској и непараметарској кластеризацији.

У завршној фази истраживања кандидаткиња ће испитивати начине спајања система за детекцију емотивног стања и система за препознавање говорника у комбиновани систем препознавања говорника са повећаном тачношћу препознавања говорника.

Структура саме тезе ће у великој мери пратити предложене фазе истраживања. Осим уводног поглавља предвиђена су три поглавља прегледног типа од којих ће прво бити посвећено задацима области препознавања говорника и изазова у овој области, са посебним освртом на изворе варијабилности у гласу.

Друго поглавље биће посвећено моделима говора: физиолошком и психолингвистичком моделу генерисања сигнала говора. Треће поглавље садржаће преглед карактеризације сигнала говора и применама у идентификацији говорника. Четврто поглавље ће бити посвећено прегледу алгоритама примењиваним у идентификацији говорника са посебним освртом на досадашња истраживања у идентификацији говорника који су у одређеном емотивном стању. Пето поглавље ће се односити на структуру и опис предложеног алгорита за препознавање говорника у условима емотивног говора, опис експеримената и добијене резултате ових експеримената. Последња два поглавља садржаће закључке и коришћену литературу, респективно.

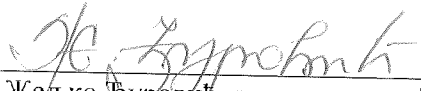
7. Закључак и предлог

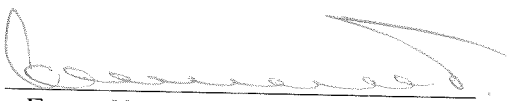
Током писања овог Извештаја Комисија је узела у обзир све релевантне чињенице које формирају слику о подобности теме и кандидата за реализацију предвиђених истраживања. Имајући у виду резултате досадашњег рада кандидаткиње Милане Милошевић који је показан у фази истраживања у оквиру докторске дисертације и њеног истраживачког искуства, Комисија је стекла неспорни утисак да кандидаткиња јесте способна за научно-истраживачки рад као и за реализацију предвиђених истраживања. Са друге стране, анализирајући предложену тему под називом "Идентификација говорника у условима емотивног говора", стављајући је у контекст савремених, актуелних научних истраживања, сагледавајући наведене непосредне циљеве, очекиване резултате и доприносе, Комисија је у потпуности сагласна да оваква тема задовољава све услове подобности за докторску дисертацију. При томе Комисија констатује да ова тема припада ужој научној области Управљање системима и обрада сигнала, да Електротехнички факултет Универзитета у Београду јесте матичан за ову научну област, и при томе за

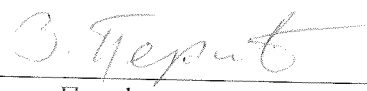
ментора рада предлаже проф. др Жељка Ђуровића запосленог на Електротехничком факултету Универзитета у Београду. Уз Извештај је приложен списак радова проф. Жељка Ђуровића који га квалификују за ментора. У закључку овог Извештаја, а имајући у виду све напред наведено, Комисија има част да Наставно-научном већу Електротехничког факултета Универзитета у Београду предложи да прихвати тему докторске дисертације "Идентификација говорника у условима емотивног говора" и да кандидаткињи Милани Милошевић омогући да приступи њеној изradi.

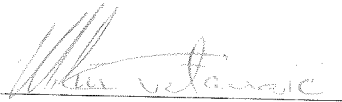
У Београду,
дана 30.08.2018.

ЧЛАНОВИ КОМИСИЈЕ


др Жељко Ђуровић, редовни професор
Универзитет у Београду – Електротехнички факултет


др Горан Квашчев, ванредни професор
Универзитет у Београду – Електротехнички факултет


др Зоран Перић, редовни професор
Универзитет у Нишу – Електронски факултет


др Милош Цветановић, ванредни професор
Универзитет у Београду – Електротехнички факултет