

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата Јована Галића, дипл. инж. електротехнике - мастер електронике и телекомуникација (на основу решења о еквиваленцији раније стеченог звања са новим звањем), за израду докторске дисертације.

Одлуком Наставно-научног већа Електротехничког факултета бр. 5036/16-1 од 27.2.2018. године, именовани смо за чланове Комисије за оцену услова и прихватање теме докторске дисертације под насловом „Препознавање мултимодалног говора засновано на статистичком приступу“ (енгл. „The recognition of multimodal speech based on statistical approach“), кандидата Јована Галића.

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Јован (Неђо) Галић је рођен 1981. године у Травнику, Република Босна и Херцеговина. Основну школу је завршио у Бањој Луци са Вуковом дипломом. Средњу електротехничку школу је завршио у Бањој Луци одличним успехом. Петогодишњи студиј електротехнике је успешно окончао у марту 2007. године са средњим просеком оцена 8,80, на Одсеку за електронику и телекомуникације.

Јован Галић је крајем 2008. године уписао докторске студије на Електротехничком факултету у Београду, модул Телекомуникације, са руководиоцем научно-истраживачког рада проф. др Слободаном Јовичићем. Тренутно је на трећој години докторских студија са положеним свим испитима и просечном оценом 9,90.

Радну каријеру је започео непосредно по завршетку студија на Електротехничком факултету Универзитета у Бањој Луци. У звање асистента је изабран у фебруару 2008. године, а у звање вишег асистента у априлу 2012. године. У настави је ангажован на извођењу аудиторних и лабораторијских вјежби из предмета: Основи комуникација, Дигиталне телекомуникације, Електроакустика и обрада аудио сигнала, Теорија информација и Телекомуникациона електроника. Екипа из Телекомуникација, чији је био вођа, остваривала је запажен резултат на сусретима студената Електротехнике (Електријада). Научно интересовање је фокусирао на говорне технологије и комуникацију човек-рачунар, са аспектом на аутоматско препознавање мултимодалног говора (нормални говор и шапат). До сада има објављених 26 научних и стручних радова. Члан је удружења IEEE, секције за обраду сигнала (енгл. *Signal Processing*). Ожењен је и отац двоје деце.

1.2. Стечено научноистраживачко искуство

Активности на докторским студијама

Од децембра 2008. године, на докторским студијама Јован Галић је одradio све обавезе везане за студијски истраживачки рад и положио испите из следећих предмета:

1. Основи говорне комуникације (допунски испит)	оцена 9 (девет)
2. Обрада говорног сигнала (допунски испит)	оцена 10 (десет)
3. Примењена статистика (9 ЕСПБ)	оцена 9 (девет)
4. Дигитална обрада сигнала (9 ЕСПБ)	оцена 10 (десет)
5. Акустика просторија (9 ЕСПБ)	оцена 10 (десет)
6. Робусна обрада говорних сигнала (9 ЕСПБ)	оцена 10 (десет)
7. Говорне технологије (9 ЕСПБ)	оцена 10 (десет)
8. Енглески језик (6 ЕСПБ)	оцена 10 (десет)
9. Технике обраде и препознавања говорног сигнала (9 ЕСПБ)	оцена 10 (десет)
10. Одабрана поглавља из акустике просторија (9 ЕСПБ)	оцена 10 (десет)
11. Методе моделовања звучног поља (9 ЕСПБ)	оцена 10 (десет)
12. Акустички дизајн просторија (9 ЕСПБ)	оцена 10 (десет)

Поред положених испита, кандидат је обавио и 2 студијска истраживачка рада и један научно-стручни рад, што је предвиђено планом докторских студија.

Због истека законског рока за завршетак докторских академских студија, на молбу студента, кандидату је одобрен поновни упис докторског студија у октобру 2016. године, сагласно Статуту Универзитета у Београду и Статуту Електротехничког факултета.

У складу са Законом о високом образовању, односно Правилником о докторским студијама Електротехничког факултета у Београду, Наставно-научно веће Електротехничког факултета, је на предлог Комисије за студије трећег степена, именovalo Комисију за оцену услова и прихватање теме докторске дисертације кандидата Јована Галића у саставу:

1. др Миомир Мијић, редовни професор, Универзитет у Београду - Електротехнички факултет,
2. др Милан Војновић, научни сарадник, Центар за унапређење животних активности, Београд,
3. др Горан Квашчев, ванредни професор, Универзитет у Београду – Електротехнички факултет.

За ментора је именована др Драгана Шумарац Павловић, ванредни професор Електротехничког факултета Универзитета у Београду.

Усмена одбрана теме је одржана 07.03.2018. на Електротехничком факултету Београду уз присуство свих чланова комисије. Одбрана је укључила презентацију предмета, циља и значаја предложене теме уз навођење полазних хипотеза, коришћених научних метода и очекиваних доприноса. Након презентације уследила су питања чланова комисије на које је кандидат успешно одговорио, што је комисија потврдила оценом "задовољно".

На Електротехничком факултету Универзитета у Бањој Луци, учествовао је у следећим активностима:

- учешће на пројектима:

1. „Примјена нових технологија у радио-комуникационим системима са фреквенцијским скакањем”, пројекат Министарства науке и технологије Републике Српске, 2008-2010.
2. „*RFID* технологије”, пројекат Министарства науке и технологије Републике Српске, 2008-2010.
3. „*NORBOTECH* - NORwegian-BOSnian TECHnology Transfer based on Sustainable Systems Engineering and Embedded Systems in the fields of Cloud Computing and Digital Signal Processing”, пројекат министарства вањских послова Краљевине Норвешке, 2011-2015.
4. Истраживање и развој рјешења за имплементацију *Smart City* концепта у граду Бања Лука, Министарство науке и технологије Републике Српске, 2017-2020. год.

- учешће на курсевима:

1. „*Real-Time Digital Signal Processing*”, Електротехнички факултет Бања Лука, од 11. до 14. јуна 2013.
2. „*CISCO* академија - *CCNA* курс”, ICT академија Електротехничког Факултета у Бањој Луци, од октобра 2007. до јуна 2008.

Списак објављених радова

Током докторских студија, Јован Галић је до данас објавио укупно 26 радова, од којих су за предложену тему докторске дисертације релевантни следећи:

- у међународним часописима:

1. Ђ. Grozdić, S. Jovičić, D. Šumarac Pavlović, J. Galić, B. Marković: „Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition“, *ADVANCES IN ELECTRICAL AND COMPUTER ENGINEERING*, Vol. 17, No. 1, ISSN 1582-7445, str. 21-26, 2017.
2. B. Marković, J. Galić, Ђ. Grozdić, S. Jovičić, M. Mijić: „Whispered speech recognition based on gammatone filterbank cepstral coefficients“, *JOURNAL OF COMMUNICATIONS TECHNOLOGY AND ELECTRONICS*, Vol. 62, No. 11, ISSN 1064-2269, str. 1255-1261, 2017.
3. B. Marković, J. Galić, M. Mijić: „Application of Teager Energy Operator on Linear and Mel Scales for Whispered Speech Recognition“, *ARCHIVES OF ACOUSTICS*, Vol. 43, No. 1, ISSN 0137-5075, str. 3-9, 2018.

- у домаћим часописима:

1. Ђ.Т. Grozdić, J. Galić, B. Marković, S.Т. Jovičić: *Application of Neural Networks in Whispered Speech Recognition*, Telfor Journal, Vol. 5, No. 2, ISSN 1821-3251, str. 103-106, 2013.

- у монографским студијама:

1. Ђ. Т. Grozdić, S. Т. Jovičić, J. Galić, B. Marković: *Experiments in whisper recognition using neural networks*. In: S.Т. Jovičić, M. Subotić, M. Sovilj (eds): *VERBAL COMMUNICATION QUALITY*, Interdisciplinary Research, II; CUŽA, Belgrade, ISBN 978-86-81879-46-7, 91-110, 2013.

- na međunarodnim konferencijama:

1. Đ. Grozdić, B. Marković, **J. Galić**, S.T. Jovičić: „Primena neuralnih mreža u prepoznavanju govora u šapatu“, Zbornik radova konferencije TELFOR 2012, str. 728-731, ISBN 978-1-4673-2984-2, Beograd, Srbija, 2012.
2. **J. Galić**, T. Pešić-Brđanin, „The Voice Fundamental Frequency Statistical Parameters under Noisy Conditions with the Cepstrum Method“, 10th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services TELSIS 2011, str. 769-772, ISBN 978-1-4577-2018-5, sveska 2, Niš, Srbija, 2011.
3. B. Marković, S.T. Jovičić, **J. Galić**, Đ. Grozdić: „Whispered Speech Database: Design, Processing and Application“, Zbornik radova 16th International Conference, TSD 2013, str. 591-598, ISBN 978-3-642-40584-6, Pilsen, Czech Republic, 2013.
4. B. Marković, **J. Galić**, Đ. Grozdić, S.T. Jovičić: „Application of DTW Method for Whispered Speech Recognition“, Zbornik radova konferencije SPEECH AND LANGUAGE 2013, str. 308-315, ISBN 978-86-81879-45-0, Beograd, Srbija, 2013.
5. **J. Galić**, B. Marković, Đ. Grozdić, S.T. Jovičić: „The Influence of Feature Vector Selection on Performance of Automatic Recognition of Whispered Speech“, Zbornik radova konferencije SPEECH AND LANGUAGE 2013, str. 258-264, ISBN 978-86-81879-45-0, Beograd, Srbija, 2013.
6. Đ. Grozdić, B. Marković, **J. Galić**, S.T. Jovičić, D. Furundžić: „Neural network-based recognition of whispered speech“, Zbornik radova konferencije SPEECH AND LANGUAGE 2013, str. 223-229, ISBN 978-86-81879-45-0, Beograd, Srbija, 2013.
7. **J. Galić**, S.T. Jovičić, Đ. Grozdić, B. Marković: „HTK-Based Recognition of Whispered Speech“, Zbornik radova 16th International Conference on Speech and Computer SPECOM 2014, str. 251-258, ISBN 978-3-319-11580-1, Novi Sad, Srbija, 2014.
8. **J. Galić**, S.T. Jovičić, Đ. Grozdić, B. Marković: „Constrained Lexicon Speaker Dependent Recognition of Whispered Speech“, Zbornik radova konferencije INDEL 2014, str. 180-184, ISBN 978-99955-46-22-9, Banja Luka, Republika Srpska, 2014.
9. Đ. Grozdić, S.T. Jovičić, **J. Galić**, B. Marković: „Application of inverse filtering in enhancement of whisper recognition“, Zbornik radova konferencije NEUREL 2014, str. 157-161, ISBN 978-1-4799-5887-0, Beograd, Srbija, 2014.
10. B. R. Marković, S. T. Jovičić, **J. Galić**, Đ.T. Grozdić: „Recognition of the Multimodal Speech Based on the GFCC features“, Zbornik radova konferencije IcETRAN 2015, str. AK1 1.3 1-5, ISBN 978-86-80509-71-6, Srebrno Jezero, Srbija, 2015.
11. B. Marković, S. Jovičić, M. Mijić, **J. Galić**, Đ. Grozdić: „Recognition of Whispered Speech Based on PLP Features and DTW Algorithm“, Zbornik radova konferencije IcETRAN 2016, str. AK1.3 1-4, ISBN 978-86-7466-618-0, Zlatibor, Srbija, 2016.
12. **J. Galić**, D. Šumarac Pavlović, S. Jovičić, B. Marković, Đ. Grozdić: „Prepoznavanje bimodalnog govora bazirano na metodi potpornih vektora“, Zbornik radova konferencije TELFOR, ISBN: 978-1-5386-3072-3, Beograd, Srbija, 2017.
13. **J. Galić**, S. Jovičić, B. Marković, D. Šumarac Pavlović, Đ. Grozdić: „Speaker dependent recognition of whispered speech based on MLLR adaptation“, Zbornik radova konferencije DOGS, str. 29-32, ISBN: 978-86-7892-993-9, Novi Sad, Srbija, 2017.
14. S. Kovačević, **J. Galić**, T. Pešić-Brđanin: „Logatome intelligibility analysis of the FM demodulated signal“, Zbornik radova konferencije IcETRAN, str. TE11.6, 1-4, ISBN 978-86-7466-692-0, Kladovo, Srbija, 2017.
15. B. Marković, G. Stevanović, S. Jovičić, M. Mijić, **J. Galić**, Đ. Grozdić: „Recognition of Normal and Whispered Speech Based on RASTA Filtering and DTW Algorithm“, Zbornik

radova konferencije IcETRA, str. AK1.8.2-4, ISBN 978-86-7466-692-0, Kladovo, Srbija, 2017.

- на домаћим конференцијама:

1. **J. Galić**, M. Popović, S.T. Jovičić, B. Marković, Đ. Grozdić: „Primjena skrivenih Markovljevih modela u prepoznavanju govora u šapatu“, Zbornik radova konferencije INFOTEH 2013, str. 387-390, ISBN 978-99955-763-1-8, Jahorina, Republika Srpska, 2013.
2. **J. Galić**, T. Pešić-Brdanin, “Uporedna analiza metoda za estimaciju osnovne frekvencije govornog signala u prisustvu bijelog šuma“, Zbornik radova 55. Konferencije za ETRA, AK2.1-1-4, ISBN 978-86-80509-66-2, Banja Vrućica, Republika Srpska, 2011.
3. **J. Galić**, T. Pešić-Brdanin, I. Janković, “Statistička analiza osnovne frekvencije kod vokala srpskog jezika“, Zbornik radova konferencije INDEL 2010, str. 236-239, ISBN 978-99955-46-03-8, Banjaluka, Republika Srpska, 2010.
4. Đ. T. Grozdić, D. Šumarac Pavlović, S. T. Jovičić, **J. Galić**, B. Marković: „Komparacija tehnika normalizacije kepsralnih koeficijenata u automatskom prepoznavanju šapata“, Zbornik radova konferencije ETRA 2015, str. AK1 1.8 1-5, ISBN 978-86-80509-71-6, Srebrno Jezero, Srbija, 2015.

1.3. Оцена подобности кандидата за рад на предложеној теми

Комисија је размотрила све претходно наведене резултате досадашњег научно-истраживачког рада и стечена искуства кандидата Јована Галића. На основу тога Комисија је стекла увид у способности кандидата за научно-истраживачки рад и његове могућности да током израде докторске дисертације оствари резултате који ће бити приказани и оцењени од стране међународне научне заједнице. Став Комисије је да Јован Галић испуњава потребне услове да приступи изради докторске дисертације под наведеном темом.

2. Предмет и циљ истраживања

Говорна комуникација представља најприроднији и најпогоднији начин комуникације између људи. Као таква, она заузима значајно место у интерфејсима између човека и машине. Према начину продуковања, говор се може поделити у 5 модова: шапат, тихи говор, говор уобичајеног интензитета (нормални говор), гласни говор и вика. Побројани модови се међусобно разликују по нивоу звука, трајању изговора и паузе, спектралној расподели енергије и спектралном нагибу [1].

Шапат представља специфичан вид говорне комуникације који је у све чешћој употреби. Користи се у ситуацијама када је потребно остварити дискретну атмосферу у комуникацији (нпр. у позориштима, читаоницама, на пословним састанцима, итд.) или када се жели сакрити информација од других присутних особа. Такође, шапат може да буде артикулисан и као последица измењеног здравственог стања (ларингитис и ранитис), а није ретка употреба шапата у криминалним активностима. Извор звука код генерисања шапатавог говора је турбулентан, апериодични проток ваздуха између гласница које не вибрирају. У основи, шапат представља мод говора у којем не постоји ларингеална, већ шумна побуда вокалног тракта. У основне карактеристике говора артикулисаним шапатам поред одсуства ларингеалних вибрација спадају значајно нижа енергија, равнији спектар и померене формантне учестаности [2-3]. Ипак, и поред повећаног напора у перцепцији, разумљивост шапата у комуникацији између људи је врло висока [4].

Системи за аутоматско препознавање говора (енг. *Automatic Speech Recognition* - ASR) имају значајан низ предности у односу на остале интерфејсе човек-машина: руке и очи остају слободне за обављање других активности, микрофон има значајно мање габарите од тастатуре, кориснике система није потребно обучавати, итд. Ипак, и поред огромног напретка још увек имају низ недостатака које се највише огледају у значајно сниженим перформансама приликом различитих услова при обуци и тестирању (бучно окружење, различити дијалекти и начини изговора говорника, различита емотивна стања говорника итд.).

ASR системи су намењени препознавању типичног мода говора, а то је нормални говор. Из тог разлога перформансе система значајно опадају у препознавању било ког нетипичног мода говора. Од побројаних 5 модова, највећа деградација перформанси система је управо при шапату, као говорном моду који се по својим акустичким карактеристикама највише дистанцира од преостала 4 мода. У ASR системима се ради заштите приватности неретко говори шапатом, и то изолованим или повезаним речима из ограниченог речника (диктирање цифара, адресе становања, датума рођења, итд.). Мотивација за овим истраживањем је проистекла из нарастајуће потребе да се дијалог човек-машина подигне на виши ниво, који укључује комуникацију са машином говором другачије артикулисаним од говора уобичајеног интензитета.

У препознавању говора је заступљено неколико приступа и техника, међу којима су најпознатије методе базиране на скривеним Марковљевим моделима (*Hidden Markov Models* - HMM), вештачким неуралним мрежама (*Artificial Neural Networks* - ANN) и потпорним векторима (*Support Vector Machines* - SVM). Због динамичке природе говорног сигнала, у ASR системима доминира статистички приступ базиран на HMM, при чему је честа комбинација са моделима мешавина Гаусових расподела (*Gaussian Mixture Models* - GMM) или у савременим хибридни системима са дубоким неуралним мрежама [5].

Најзначајнија досадашња истраживања у вези са препознавањем мултимодалног говора се огледају у прецизној детекцији граница између нормалног говора и шапата у континуалном говору који поред нормалног говора садржи и сегменте шапата [6], препознавању мултимодалног говора из малог речника без адаптације и обуке на говору шапата [7-8], као и повећању робустности система за препознавање мултимодалног говора коришћењем малог дела базе за адаптацију [9].

Предмет овог истраживања је аутоматско препознавање мултимодалног говора базирано на статистичком приступу, са акцентом на препознавање изолованих речи. Истраживање обухвата два говорна мода: нормалан (неутралан) говор и шапат, као говорни мод који се највише дистанцира од остала четири мода. Циљ истраживања је повећање робустности ASR система на српском језику, обученог на нормални говор, уколико се тестира упоредо са оба мода. Значај истраживања може да буде вишеструк, првенствено у побољшању комуникације човек-машина, у оптимизацији препознавача на бази Марковљевих модела за бимодалну верзију говорне експресије до утврђивања обележја у говорном сигналу која могу омогућити већу робустност HMM алгорита у препознавању мултимодалног говора. Такође, говорна база која је снимљена на српском језику (чија је почетна форма приказана у [10]), може да се користи и у другим врстама истраживања мултимодалне експресије говора. Прелиминарни резултати препознавања мултимодалног говора применом HMM алгорита су представљени у [11-12], за поменућу говорну базу.

Релевантни иницијални библиографски извори за предложено истраживање:

- [1] C. Zhang, J.H.L. Hansen, "Analysis and Classification of Speech Mode: Whisper through Shouted," pp. 2289-2292, Proceedings of INTERSPEECH, 2007.

- [2] T. Ito, K. Takeda and F. Itakura, "Analysis and Recognition of Whispered Speech," *Speech Communication*, 45, pp.129-152, 2005.
- [3] S.T. Jovičić, "Formant feature differences between whispered and voiced sustained vowels," *Acta Acustica*, vol. 84, no. 4, pp. 739-743, 1998.
- [4] S. T. Jovičić, Z.M. Šarić, "Acoustic analysis of consonants in whispered speech," *Journal of Voice*, vol. 22, no. 3, pp. 263-274, 2008.
- [5] Đ. Grozdić, S.T. Jovičić, M. Subotić, "Whispered speech recognition using deep denoising autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 15-22, 2017.
- [6] C. Zhang, J. H.L. Hansen, "Whisper-Island Detection Based on Unsupervised Segmentation With Entropy-Based Speech Feature Processing," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 4, pp. 883-894, 2011.
- [7] S. Ghaffarzadegan, H. Boril, J. H. L. Hansen, "Model and Feature Based Compensation for Whispered Speech Recognition," *Proceedings of INTERSPEECH 2014*, Singapore, pp. 2420-2424, 2014.
- [8] S. Ghaffarzadegan, H. Boril, J. H.L. Hansen, "UT-Vocal Effort II: Analysis and constrained-lexicon recognition of whispered speech," *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, pp. 2544-2548, 2014.
- [9] S. Ghaffarzadegan, H. Boril, J. H. L. Hansen, "Generative Modeling of Pseudo-Whisper for Robust Whispered Speech Recognition," *IEEE/ACM Transactions on Speech and Language Processing*, vol. 24, no. 10, pp. 1705-1720, 2016.
- [10] B. Marković, S. Jovičić, J. Galić, Đ. Grozdić, "Whispered speech database: design, processing and application", in: *The 16th International Conference TSD 2013*, I. Habernal and V. Matousek (Eds.): TSD 2013, LNAI 8082, pp. 591-598, 2013.
- [11] J. Galić, S.T. Jovičić, Đ. Grozdić, B. Marković, "HTK-Based Recognition of Whispered Speech", in: *The 16th International Conference on Speech and Computer SPECOM 2014*, A. Rozhnin, R. Potapova and V. Delic (Eds.): SPECOM 2014, LNAI 8773, pp. 251-258, 2014.
- [12] Đ. Grozdić, S. Jovičić, D. Šumarac Pavlović, J. Galić, B. Marković, "Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition," *Advances in Electrical and Computer Engineering (AECE) Journal*, vol. 17, no. 1, pp. 21-26, 2017.
- [13] P.X. Lee, D. Wee, H. Si ZinToh, B.P. Lim, N. Chen, B.A. Ma, "Whispered Mandarin Corpus for Speech Technology Applications," in *Proc. of INTERSPEECH 2014*, Singapore, pp. 1598-1602, 2014.
- [14] P. Kozierski, T. Sadalla, S. Drgas, A. Dąbrowski, and D. Horla, "Kaldi Toolkit in Polish Whispery Speech Recognition," *Przegląd Elektrotechniczny*, pp. 301-304, 2016.

3. Полазне хипотезе

Основне хипотезе које се постављају пред предложену дисертацију су:

1. Постоји могућност побољшања успешности препознавања у неусаглашеним обука/тест условима, погодним избором јединице за моделовање и броја стања по фонему.
У аутоматском препознавању изолованих речи као јединице за моделовање се најчешће користе фонеме независни од контекста (монофони), фонеме зависни од контекста (трифони) и целе речи. И поред тога што се фонеме зависни од контекста више користе, њихов недостатак је мања робустност и потреба за што већом количином података за обуку. Претпоставка је да фонеме независни од контекста могу да буду најробустнији у препознавању шапата у систему обученом на нормални говор. Такође, уколико се уместо фиксног броја стања по моделу користи број стања који је сразмеран трајању фонема очекује се повећање успешности препознавања.
2. Перформансе система се могу побољшати лабелирањем дела базе за обуку.
Поред величине базе, перформансе ASR система базираног на НММ алгоритму су зависне и од тачности иницијализације параметара модела. Ради тачније иницијализације, део базе који се користи у обуци ће бити лабелиран, тј. у одређеном делу снимака фонетска транскрипција ће бити праћена временским тренуцима почетка и краја сваког фонема.

Претпоставка је да се на овај начин могу добити тачнији иницијални модели и добити већа успешност препознавања.

3. Постоји могућност побољшања перформанси система модификацијом фреквенцијске скале за добијање вектора обележја.

Мел-фреквенцијски кепстрални коефицијенти (енгл. *Mel-Frequency Cepstral Coefficients - MFCC*) и њихове временске диференције (прва и друга) представљају стандардна обележја у ASR системима. Међутим, у значајно неусаглашеним условима обука/тест показују слабу робустност. Због своје акустичке природе, значајан део релевантних информација у шапату је садржан у горњем делу фреквенцијског опсега говорног сигнала у којем мел фреквенцијска скала има слабу резолуцију банке филтара. Због тога је претпоставка да се модификацијом фреквенцијске скале могу добити вектори обележја који ће повећати успешност препознавања шапата, а при том неће нарушити успешност препознавања нормалног говора.

4. Успешност препознавања се може побољшати у моду зависно и независно од говорника.

Препознавање у ASR системима се може обављати зависно и независно од говорника. Уколико се параметри препознавача подешавају према говорнику који ће бити тестиран препознавање је зависно од говорника, док код препознавања независно од говорника тестирани говорник не учествује у обуци система. Практичност употребе ASR система намеће природан услов препознавања независно од говорника. Претпоставка је да се побољшање успешности препознавања може остварити на оба начина препознавања: зависно и независно од говорника.

4. Научне методе истраживања

На основу прегледа литературе установљено је да је препознавање шапата актуелна истраживачка тема пре свега због недостатка велике, систематски креиране и анотиране базе за обуку. До сад је креирано свега неколико говорних база са изговорима нормалног говора и шапата и то за јапански [2], енглески [8], мандарински [13] и пољски језик [14].

С обзиром да су говорне технологије јако зависне од језика, те се као такве не могу увести из других земаља и применити на било ком језику, за потребе истраживања је снимљена прва верзија говорне базе *Whi-Spe (Whispered Speech)*, која је прва база такве врсте на српском језику [10]. Ова база садржи снимке 50 изолованих речи изговорених нормалним говором и шапатам. Десет говорника (пет женских и пет мушких) је сваку реч изговорило 10 пута, у одвојеним сесијама снимања. Након мануелне сегментације, која је урађена у софтверском алату *Adobe Audition*, и анализе квалитета снимака, говорна база садржи 10.000 снимака у укупном трајању од 2 часа.

Обим говорне базе је довољан за истраживачке активности на развоју препознавача базираног на НММ како зависног тако и независног од говорника. Софтверска реализација препознавача базираног на НММ алгоритму ће бити урађена у софтверском пакету *Matlab*. Експерименти ће бити урађени коришћењем НТК алата (енгл. *Hidden Markov Model Toolkit*) који су често заступљени у истраживањима везаним за препознавање говора. Део базе за обуку ће бити мануелно лабелиран увидом у таласни облик и спектрограм коришћењем софтверског пакета *Praat*. Ради добијања веродостојније оцене перформанси препознавача биће урађена кросвалидација. На крају, резултати препознавања ће бити статистички анализирани и приказани у форми процента тачно препознатих речи (енгл. *Word Recognition Rate - WRR*).

5. Очекивани научни допринос

Током израде ове докторске дисертације, очекују се следећи научни доприноси:

1. одређивање најпогодније јединице за моделовање (фонем независан од контекста, фонем зависан од контекста или цела реч) за постизање максималне тачности препознавања у усаглашеним и неусаглашеним обука/тест условима;
2. анализа утицаја фонетске транскрипције речи у моделу језика и повећање робустности препознавача лабелирањем дела базе за обуку;
3. анализа избора параметра за класификацију говорног мода (нормалан говор/шапат) и упоредна анализа перформанси препознавача базираног на класификацији и препознавача са измешаном базом за обуку (енг. *Multi-style Training*), која садржи изговоре оба мода говора; и
4. предлог новог алгоритма за екстракцију вектора обележја модификацијом фреквенцијске скале и упоредна анализа перформанси препознавача са предложеним алгоритмом и постојећим векторима обележја пре свега са традиционалним MFCC и PLP векторима обележја.

6. План истраживања и структура рада

Истраживање ће бити реализовано кроз неколико фаза:

- Истраживање почиње са полазном идејом да је могуће имплементирати препознавач шапата са задовољавајућом тачношћу (зависно и независно од говорника) за ограничену величину речника. Обзиром да се шапат као говорни мод комуникације јавља значајно ређе од нормалног говора, тежња је да перформансе препознавача у препознавању нормалног говора не буду деградиране.
- Прва фаза истраживања укључује преглед актуелних истраживања и досадашњих резултата из области аутоматског препознавања мултимодалног говора за друге говорне базе. Следи упознавање са коришћеном говорном базом (Whi-Spe) и методологијом препознавања говора која је у језгру свих савремених препознавача, скривеним Марковљевим моделима.
- Друга фаза истраживања је посвећена анализи могућности НТК скупа алата у препознавању говора и развоју софтвера за обуку и тестирање НММ модела. Следи анализа избора јединице за моделовање у препознавању нормалног говора и шапата у усаглашеним и неусаглашеним обука/тест сценаријима.
- Тежиште треће фазе је побољшање успешности препознавача додавањем динамичких обележја и погодним избором фонетске транскрипције речи. Такође, биће извршена анализа доприноса аутоматског и мануелног лабелирања у препознавању нормалног говора и шапата, али са обуком само на нормалном говору.
- Четврта фаза садржи предлог модификације фреквенцијске скале за екстракцију вектора обележја и анализу ефикасности предложеног алгоритма са постојећим векторима обележја који су најчешће заступљени у препознавању шапата.
- Последња фаза укључује анализу препознавања мултимодалног говора уколико су у обуци на располагању изговори оба говорна мода: нормалног говора и шапата. Биће извршен развој класификатора говорног мода (нормални говор/шапат) као и упоредна анализа успешности препознавача базираног на класификацији говорног мода и препознавача са измешаном базом за обуку.

7. Закључак и предлог

На основу приложене документације, као и непосредног увида у досадашњи научни рад кандидата Јована Галића, Комисија је закључила да:

1. кандидат испуњава све формалне и суштинске услове прописане Законом о високом образовању и Статутом Електротехничког факултета Универзитета у Београду за одобрење теме за израду докторске дисертације;
2. предложена тема Препознавање мултимодалног говора засновано на статистичком приступу“ је научно заснована, а резултати до којих ће се доћи израдом ове дисертације биће значајан и оригиналан допринос области.

Комисија констатује да предложена тема припада области Техничке акустике за коју је Електротехнички факултет у Београду матичан и за ментора рада се предлаже др Драгана Шумарац Павловић, ванредни професор, која је запослена на Електротехничком факултету у Београду. Због тога је уз овај Извештај приложен и списак радова проф. др Драгане Шумарац Павловић, који је квалификују за ментора.

Узимајући у обзир чињенице наведене у овом Извештају, Комисија предлаже Наставно-научном већу Електротехничког факултета Универзитета у Београду да прихвати тему докторске дисертације „Препознавање мултимодалног говора засновано на статистичком приступу“ и да кандидату Јовану Галићу омогући да приступи њеној изради.

Београд, 14.03.2018.

ЧЛАНОВИ КОМИСИЈЕ



др Драгана Шумарац Павловић, ванредни професор
Универзитет у Београду – Електротехнички факултет



др Миомир Мијић, редовни професор
Универзитет у Београду – Електротехнички факултет



др Милан Војновић, научни сарадник
Центар за унапређење животних активности, Београд



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