

## НАСТАВНО-НАУЧНОМ ВЕЋУ

**Предмет:** Подобност теме и кандидата Бранка Марковића, магистра електротехничких наука за израду докторске дисертације

Одлуком Наставно-научног већа Електротехничког факултета бр. 955/1 од 30.06.2015. године, именовани смо за чланове Комисије за оцену подобности теме и кандидата Бранка Марковића за израду докторске дисертације и научне заснованости теме **„Анализа обележја у говорном сигналу за потребе аутоматског препознавања мултимодалног говора“**.

На основу исказа кандидата у току усмене одбране теме докторске дисертације Комисија је предложила корекцију теме у **„Анализа обележја у говорном сигналу за потребе препознавања мултимодалног говора“**.

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

## ИЗВЕШТАЈ

### 1. Подаци о кандидату

#### 1.1. Биографски подаци

Бранко Марковић је рођен 1966. године. Основну школу је завршио у Слатини, општина Чачак, а гимназију у Чачку. На Електротехничком факултету у Београду дипломирао је 1992. године, одсек Телекомуникације. Просечна оцена студија му је била 9.13, и на дипломском раду 10.

Исте године наставља постдипломске студије на смеру „Дигитални пренос информација“. Након 10 година проведених на раду у иностранству (Канада и САД), постдипломске студије завршава 2004. године. Наслов магистарске тезе био је „Функција ‘Позивање говором’ у мобилној телефонији“, а ментор је био Слободан Јовичић. Све испите на магистарским студијама положио је са оценом 10.

У периоду 1992 - 1993. године Бранко Марковић је био запослен у Институту Михајло Пупин у Београду где је радио на развоју мрежних уређаја: мултиплексера и модема. Од 1993. до 2002. године радио је у Канади и САД за телекомуникационе компаније „Bell Canada“, „Architel“ и „Nortel Networks“. У њима је радио на развоју комуникационог и управљачког софтвера и на пројектима као што су: Video Conferencing Manager, ASAP, NYNEX OLP Project, AIMS, Telezone, OMS и другима.

Почев од 2004. године па до данас Бранко Марковић ради као предавач на Високој школи техничких струковних студија Чачак, где је ангажован на групи предмета Информациони системи и технологије. Креирао је одређен број нових предмета за потребе ове школе: Рачунарске мреже, Интернет технологије (Основне студије) као и Софтверски алати, Мултимедијалне комуникације, Вишеслојна софтверска архитектура (Специјалистичке студије). Био је ментор већег броја дипломских и специјалистичких радова. Аутор је уџбеника/скрипти за одређени број предмета.

## 1.2. Стечено научноистраживачко искуство

Током израде магистарског рада под називом „Функција ‘Позивање говором’ у мобилној телефонији“ кандидат се бавио проблематиком дигиталног процесирања говора, његове предобраде и развојем система за аутоматско препознавање говора који су зависни и независни од говорника. Магистарски рад је одбранио 7. маја 2004. године на Електротехничком факултету у Београду, а ментор овог рада је био професор др Слободан Јовичић.

Под руководством истог ментора, кандидат је активно учествовао у кретању базе говорних узорака Whi-Spe која садржи 10.000 говорних узорака. База је намењена за обуку и препознавање мултимодалног говора, са акцентом на шапат и нормалан говор. Ову базу кандидат је користио и за своја истраживања

Кандидат Бранко Марковић је до данас објавио преко 40 радова, од којих су за предложену тему докторске дисертације релевантни следећи:

### **- у домаћим часописима:**

1. **Branko Marković**, „*Upotreba skrivenih Markovljevih modela u prepoznavanju govora*“, Interdisciplinarni časopis „Nauka Tehnika Bezbednost“, broj 2, Institut bezbednosti, Beograd, 2002, ISSN 0353-5517, str. 3-23.
2. Đorđe T. Grozdić, **Branko Marković**, Jovan Galić, Slobodan T. Jovičić, „*Application of Neural Networks in Whispered Speech Recognition*“, TELFOR Journal, Vol. 5, No. 2, 2013, ISSN 1821-3251, pp. 103-106.
3. **Branko Marković**, Danijel Lacmanović, Marko Tatović, Gordana Marković, „*Analiza uticaja vrste kepralnih koeficijenata na prepoznavanje izolovano izgovorenih reči srpskog jezika*“, „Tehnika i praksa“, Broj 9, 2013, VŠTSS Čačak, ISSN 2217-2130, str. 97-104.
4. **Branko Marković**, Mirjana Mitić, Milan Pajkić, Nenad Milenković, Gordana Marković, „*Analiza uticaja globalnog ograničenja CE2-1 na prepoznavanje izolovano izgovorenih reči srpskog jezika*“, „Tehnika i praksa“, Broj 10, 2013, VŠTSS Čačak, ISSN 2217-2130, str. 147-153.
5. Goran Ljubić, Marko Surudžić, **Branko Marković**, „*Realizacija DTW algoritma korišćenjem MATLAB-a*“, „Tehnika i praksa“, Broj 12, 2014, VŠTSS Čačak, ISSN 2217-2130, str. 125-132.

### **- у монографским студијама:**

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

### **- на међународним конференцијама:**

1. **Branko Marković**, Gordana Marković, Nebojša Stanković, „*Osnovne karakteristike mikrofonskih redova simulirane u MATLAB-u*“, Tehnologija informatika obrazovanje za društvo učenja i znanja, 5. međunarodni simpozijum, Zbornik, II deo, str. 80-89, Fakultet Tehničkih Nauka, Novi Sad, 19-20. juna 2009.

2. **Branko Marković**, "*The beamforming techniques through the visual tools*", The Sixteenth International Congress on Sound and Vibration, Proceedings of ICSV16, Paper 989, Krakow, Poland, 5-9 July 2009.
3. Rajko Mitrović, Ivan Živanović, Ivan Radeljić, **Branko Marković**, "*Korišćenje asinhronog dinamičkog programiranja u rešavanju problema optimalne staze*", Tehnika i informatika u obrazovanju, TIO 2012, 4. internacionalna konferencija, Zbornik, str. 193-199, Tehnički fakultet, Čačak, 1-3. juna 2012.
4. Željko Borić, **Branko Marković**, "*Simulator govornog termometra baziran na DS1820 senzoru i PIC18F45K22 mikrokontroleru*", Zbornik radova 20tog Telekomunikacionog foruma TELFOR 2012, str. 544-547, Beograd, 20-22. novembra 2012.
5. Đorđe T. Grozdić, **Branko Marković**, Jovan Galić, Slobodan T. Jovičić, "*Primena neuralnih mreža u prepoznavanju govora u šapatu*", Zbornik radova 20tog Telekomunikacionog foruma TELFOR 2012, str. 728-731, Beograd, 20-22. novembra 2012.
6. **Branko Marković**, Slobodan T. Jovičić, Jovan Galić and Đorđe Grozdić, "*Whispered Speech Database: Design, Processing and Application*", I. Habernal and V. Matousek (Eds.): TSD 2013, LNAI 8082, pp. 591-598, 2013, Springer-Verlag Berlin Heidelberg 2013.
7. **Branko Marković**, Jovan Galić, Đorđe Grozdić, Slobodan T. Jovičić, "*Application of DTW method for whispered speech recognition*", Proceedings Speech and Language 2013, 4<sup>th</sup> International Conference on Fundamental and Applied Aspects of Speech and Language, pp. 308-315, Belgrade, October 25-26, 2013.
8. Jovan Galić, Slobodan Jovičić, Đorđe Grozdić, **Branko Marković**, "*The influence of feature vector selection on performance of automatic recognition of whispered speech*", Proceedings Speech and Language 2013, 4<sup>th</sup> International Conference on Fundamental and Applied Aspects of Speech and Language, pp. 258-264, Belgrade, October 25-26, 2013.
9. Đorđe Grozdić, **Branko Marković**, Jovan Galić, Slobodan Jovičić, D. Furundžić, "*Neural Network-Based recognition of whispered speech*", Proceedings Speech and Language 2013, 4th International Conference on Fundamental and Applied Aspects of Speech and Language, pp. 223-229, Belgrade, October 25-26, 2013.
10. **Branko R. Marković** and Đorđe T. Grozdić, "*The LPCC-DTW Analysis for Whispered Speech Recognition*", Proceedings of 1<sup>st</sup> International Conference of Electrical, Electronic and Computer Engineering, IcETRAN 2014, pp. AKI1.1.1-4, Vrnjačka Banja, Serbia, June 2-5, 2014.
11. Jovan Galić, Slobodan T. Jovičić, Đorđe Grozdić and **Branko Marković**, "*HTK-Based Recognition of Whispered Speech*", A. Ronzhin et al. (Eds.): SPECOM 2014, LNAI 8773, pp. 251-258, 2014, Springer International Publishing Switzerland 2014.
12. Jovan Galić, Slobodan T. Jovičić, Đorđe Grozdić, **Branko Marković**: "*Constrained Lexicon Speaker Dependent Recognition of Whispered Speech*", Zbornik radova konferencije INDEL 2014, pp. 180-184, Banja Luka, Republika Srpska, 2014.
13. Đorđe Grozdić, Slobodan Jovičić, Jovan Galić, **Branko Marković**: "*Application of inverse filtering in enhancement of whisper recognition*", Zbornik radova konferencije NEUREL 2014, str. 157-161, Beograd, Srbija, 2014.
14. **Branko R. Marković**, Slobodan T. Jovičić, Jovan Galić and Đorđe T. Grozdić, "*Recognition of Multimodal Speech Based on the GFCC Feature*", Proceedings of 2<sup>nd</sup>

**- на домаћим конференцијама:**

1. **Branko Marković**, “*Neke mogućnosti za sinhronizaciju rama korišćenjem informacione suvišnosti*”, V sveska, ETAN, XXXVI Konferencija, str. 345-352, Kopaonik, 1992.
2. **Branko R. Marković**, “*Grafički interfejs za poređenje govornih signala korišćenjem DTW metoda*”, Zbornik radova DOGS2002, Bečej, 16-17. maj 2002.
3. **Бранко Р. Марковић**, “*WiseHMM – Графички интерфејс за обучавање скривених Марковљевих модела (HMM) континуалног типа*”, Zbornik radova XLVI za ETRAN, Vol. II, str. 325-328, Banja Vrućica-Teslić, 2002.
4. **Branko R. Marković**, “*WiseDigit – Softver za prepoznavanje vezano izgovorenih cifara u uslovima šuma*”, Zbornik radova DOGS2004, ASR.3, Sombor, 9-10. septembra 2004.
5. **Бранко Р. Марковић**, “*WiseEdit 1.0 – софтверски пакет за аквизицију говорних узорака у облику wave фајлова*”, Zbornik radova XLVI za ETRAN, tom II, str. 388-391, Budva, 5-10. juna 2005.
6. Gordana Marković, **Branko Marković**, “*Vizuelni DTW kao nastavno sredstvo za poređenje govornih uzoraka*”, Tehnika i informatika u obrazovanju, TIO '08, str. 409-415, Tehnički fakultet, Čačak, 9-11. maja.
7. **Branko Marković**, Gordana Marković, Nebojša Stanković, “*Realizacija grafičkog interfejsa za Delay-And-Sum Beamformer*”, Zbornik radova 53. Konferencije za ETRAN, str. AK3.5-1-3, Vrnjačka Banja, 15-18. juna 2009.
8. Jovan Galić, Slobodan Jovičić, **Branko Marković**, “*Uporedna analiza tri modela za automatsko prepoznavanje govora pri šapatu*”, Zbornik apstrakta konferencije TAKTONS 2011, str. 12, Novi Sad, Srbija, 2011.
9. **Branko Marković**, Goran Luković, “*WiseWave 1.4 – softver za analizu uticaja prozorovanja na prepoznavanje izolovano izgovorenih reči*”, Zbornik radova 56. Konferencije za ETRAN, str. AK1.6-1-4, Zlatibor, 11-14. juna 2012.
10. Jovan Galić, Milan Popović, **Branko Marković**, Đorđe T. Grozdić, Slobodan T. Jovičić, “*Primjena skrivenih Markovljevih modela u prepoznavanju govora u šapatu*”, Zbornik radova konferencije INFOTEN 2013, str. 387-390, Jahorina, Republika Srpska, 2013.

Кандидат је дана 28.06.2015. године пред комисијом у саставу др Драгана Шумарац Павловић, ванредни професор, др Милан Војновић, научни сарадник и др Срђан Станковић, професор емеритур, успешно одбранио предложену тему за докторску дисертацију.

1.3. Оцена подобности кандидата за рад на предложеној теми

Комисија је размотрила све претходно наведене резултате досадашњег рада и стечена искуства Бранка Марковића. 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 Бранко Марковић испуњава потребне услове да приступи изради докторске дисертације под наведеном темом.

## 2. Предмет и циљ истраживања

Говорник током продукције говора може генерисати различите модове и они се могу сврстати у пет основних категорија: шапат, тих говор, нормалан говор, наглашен говор и вика. Ови модови се међусобно разликују по енергији, а такође и према спектралним карактеристикама, односу звучних и беззвучних делова, паузама, итд [1]. У фокусу истраживања система за аутоматско препознавање говора (АПГ) већ деценијама се налази нормалан говор. Међутим, у новије време све више интереса добијају и други модови, а међу њима посебно је интересантан шапат.

При креирању робусног система за препознавање говора посебна пажња се обраћа одабиру његових акустичких обележја. У литератури је приказано коришћење обележја која биолошки описују функције ува, али и одговарајућих процеса који могу бити повезани са неуронима који преносе сигнале од ува до одговарајућих центара мозга. Најчешће коришћена обележја у препознавању нормалног говора су LPCC (*Linear Predictive Cepstral Coefficients*), LFCC (*Linear-Frequency Cepstral Coefficients*), MFCC (*Mel-Frequency Cepstral Coefficients*) [2] и њихове модификације као што су M-MFCC (*Modified Mel-Frequency Cepstral Coefficients*) и ECC (*Exponential-Frequency Cepstral Coefficients*) [3]. У употреби је и GFCC (*Gammation Frequency Cepstral Coefficients*) [4] који моделира слушни механизам низом Гаматон филтера пропусника опсега учестаности који се међусобно преклапају. У интересантна обележја могу се убројати и PLP (*Perceptive Linear Predictive coefficients*) [5], и RASTA (*Relative Spectral Transform coefficients*) које је предложио Hermansky [6], која омогућавају додатне карактеристике као што су коришћење функције једнаке гласности и снаге, а такође и релативну спектралну трансформацију. Применом нових оператора у стандардној процедури предобrade говора, а пре свега нелинеарног оператора Тигер енергије (ТЕО – *Teager Energy Operator*) [7], постоји могућност да се учини одређен помак у препознавању различитих говорних модова.

Предмет истраживања у овом раду је проналажење скупа векторских обележја која могу знатније допринети побољшању резултата препознавања шапата, али и потенцијално коришћење таквих обележја за остале модове говора.

Циљ предложеног истраживања је да се на бази стандардних векторских обележја која подразумевају примену LFCC, MFCC, GFCC, PLP, RASTA и сличних изврши детаљна анализа. Увођењем нормализације (као што је CMS-*Cepstral Mean Subtraction*) [8], а затим и оператора ТЕО, дође до погодног решења којим се може побољшати систем за препознавање говора. При томе ће фокус бити на изоловано изговореним речима које зависе од говорника. Као препознавач у раду би се користио неки од класичних метода (као што је DTW (*Dynamic Time Warping*) и/или HMM (*Hidden Markov Models*) [9]-[10]). Тестирање препознавања би се вршило на говорној бази Whi-Spe [11]. Ова база садржи 10.000 говорних узорака. У раду би фокус би био на усаглашеним и неусаглашеним сценаријим за модове шапата и нормалног говора.

Значај овог истраживања огледао би се у унапређењу система за АПГ у околностима када се комбиновано користе модови нормалног говор и шапата, али се претпоставља да би таква анализа могла бити од значаја и за друге модалитете говора.

### Релевантни библиографски извори за предложено истраживање:

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

- [2] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 4, pp. 357–366, Aug. 1980.
- [3] S. E. Bou-Ghazle and J. H. L. Hansen, "A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress", vol. 8, No. 4, July, *IEEE Trans Speech and Audio*, 2000.
- [4] O. Cheng, W. Abdulla, and Z. Salcic, "Performance evaluation of front end algorithms for robust speech recognition," in *Proc. ISSPA*, 2005, pp.711-714 *Speech Communication*, 45, pp.129-152, 2005.
- [5] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech", *J. Acoust. Soc. Am.*, vol. 87, no. 4, pp. 1738-1752, Apr. 1990.
- [6] H. Hermansky and N. Morgan, "RASTA processing of speech", *IEEE Trans. on Speech and Audio Proc.*, vol. 2, no. 4, pp. 578-589, Oct. 1994.
- [7] A. Georgogiannis, V. Digalakis, "Speech Emotion Recognition Using Non-Linear Teager Energy Based Features in Noisy Environments," pp. 2045–2049, *EURASIP*, 2012.
- [8] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.
- [9] L. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition", pp. 200-226, Prentice Hall, 1993.
- [10] B. Marković, J. Galić, Đ. Grozdić, S. T. Jovičić, "Application of DTW method for whispered speech recognition", *Proceedings Speech and Language 2013, 4<sup>th</sup> International Conference on Fundamental and Applied Aspects of Speech and Language*, Зборник радова, pp. 308-315, Belgrade, October 25-26, 2013.
- [11] B. Marković, S. T. Jovičić, J. Galić, Đ. Grozdić, "Whispered speech database: design, processing and application", in: *The 16th International Conference TSD 2013, Proceedings*, LNAI 8082, pp. 591–598, 2013.

### 3. Полазне хипотезе

Основне хипотезе предложене докторске дисертације су:

- 1) Анализом скупа класичних векторских обележја могуће је пронаћи најповољнија с којима се постижу најбољи резултати при препознавању мултимодалног говора.

Анализа скупа класичних обележја подразумева коришћење:

- LFCC – који користе линеарну фреквенцијску скалу и одговарајуће кепстралне коефицијенте,
- MFCC – који користе Mel фреквенцијску скалу и одговарајуће кепстралне коефицијенте,
- GFCC – који користе Bark фреквенцијску скалу са Гаматон филтрима и одговарајуће кепстралне коефицијенте,
- PLP – који користи перцептивне линеарне предиктивне коефицијенте,
- RASTA – који користи коефицијенте релативне спектралне трансформације,

Несумњиво је да ова обележја нису подједнако ефикасна у препознавању мултимодалног говора, али тај аспект није до сада анализиран у литератури. У том смислу, могуће је у скупу ових обележја пронаћи она која дају најбољи резултати при препознавању мултимодалног говора.

- 2) Применом CMS-а или других метода нормализације на скупу класичних векторских обележја могуће је додатно побољшати препознавање мултимодалног говора.

Примена CMS-а омогућава да се у мултимодалном говору изврши спектрална нормализација модова. На бази ове примене могу се добити знатно бољи резултати у препознавању мултимодалног говора него коришћењем класичног приступа. Може се претпоставити да се посебно побољшање може постићи код неусаглашених сценарија и на тај начин формирати робусан систем за њихово препознавање.

3) Перформансе система за препознавање мултимодалног говора могу се побољшати коришћењем оператора Тигер енергије (ТЕО) којим се посебно добро може описати турбулентно кретање ваздуха током изговора у моду шапата. Додатно побољшање система препознавања може се очекивати истовременом применом статичких и динамичких кепстралних коефицијената.

Оператор Тигер енергије може се применити на већину класичних обележја. На основу тога би се добила обележја типа: TELFCC, TEMFCC, TEGFCC, TE-PLP, TE-RASTA, итд. Анализа би показала колико и како ова обележја утиче на препознавање говора, а посебно код усаглашеног сценарија шапат/шапат и неусаглашених сценарија (шапат/нормалан и нормалан/шапат). Рално је претпоставити да би се на овај начин добио скуп векторских обележја који омогућава боље резултате у препознавању.

#### **4. Научне методе истраживања**

Провера наведених полазних хипотеза и остварења задатих истраживачких циљева биће реализована коришћењем неколико научних метода и поступака:

- развој алгоритама за добијање статичких и динамичких кепстралних коефицијената за скуп класичних обележја типа LFCC, MFCC, GFCC, PLP, RASTA и сл. са и без CMS-а;
- развој алгоритама за добијање статичких и динамичких кепстралних коефицијената за скуп обележја типа TELFCC, TEMFCC, TEGFCC, TE-PLP, TE-RASTA и сл;
- развој софтверских модула на бази горе поменутих алгоритама који омогућавају генерисање одговарајућих векторских обележја;
- примена софтверских модула на говорну базу Whi-Spe за усаглашене и неусаглашене сценарије;
- анализа добијених векторских обележја коришћењем стандардних препознавача (као што су DTW и слични);
- статистичка обрада добијених резултата и њихово поређење;
- одређивање најповољнијег обележја за мултимодални говор, и то за сваки од четири сценарија: нормалан/нормалан, шапат/шапат, нормалан/шапат и шапат/нормалан.

Развој софтверских модула биће реализован коришћењем софтверских пакета MATLAB и *Visual Studio*, при чему ће се тежити максималној аутоматизацији обраде узорака.

#### **5. Очекивани научни допринос**

Развој система за аутоматско препознавање говора је стални процес који има за циљ све поузданије препознавање у околностима када оно што је изговорено одступа од нормалног говора. У том смислу је у новије време актуелизовано истраживање мултимодалних облика говора, међу којима значајно место заузима шапат. Од значаја је проналажење најповољнијих векторских обележја која током предобраде говорног сигнала могу утицати на побољшање резултата препознавања. У складу са тим, очекивани научни допринос докторске дисертације огледају се у следећем:

- Примена оператора Тигер енергије у развоју нових алгоритама за детекцију акустичких обележја у говорном сигналу и доказивање њихове предности у односу на стандардна обележја у околностима препознавања мултимодалног говора (нормалан/шапат).

на стандардна обележја у околностима препознавања мултимодалног говора (нормалан/шапат).

- Проналажење најповољнијег векторског обележја или више векторских обележја који омогућавају најбољи WRR у препознавању мултимодалног говора код усаглашених (нормалан/нормалан, шапат/шапат) и неусаглашених (нормалан/шапат и шапат/нормалан) сценарија.
- Повећање робусности препознавања говора делимичном унификацијом говорних модалитета применом CMS-а и других метода нормализације.

Осим ових доприноса могуће је да ће истраживање резултирати формирањем одговарајућег хибридног решења које би дало повећање вероватноће препознавања мултимодалног говора на бази комбинације више векторских обележја у различитим сценаријима.

## 6. План истраживања и структура рада

Планирано истраживање обухватиће следеће активности:

- анализа и преглед актуелних истраживања у области детекције акустичких обележја у говорном сигналу и њихове примене у препознавању говора;
- анализа и преглед актуелних истраживања и резултата из области аутоматског препознавања мултимодалног говора;
- развој и реализација софтвера за скуп класичних обележја која укључују LFCC, MFCC, GFCC, PLP, RASTA и слична обележја без и са CMS-ом;
- развој и реализација софтвера за скуп обележја која садрже TE оператор примењен на LFCC, MFCC, GFCC, PLP, RASTA и слична са CMS-ом и на тај начин добијање нових векторских обележја: TELFCC, TEMFCC, TEGFCC, TE-PLP, TE\_RASTA;
- креирање одговарајућих векторских обележја за све елементе говорне базе Whi-Spe;
- верификација резултата на основу класичних препознавача као што је DTW и слични коришћењем усаглашених и неусаглашених сценарија;
- упоредна анализа резултата добијених на основу побројаних векторских обележја;
- анализа побољшања перформанси препознавања на бази коришћења динамичких параметара примењених заједно са TEO оператором;
- анализа неусаглашених сценарија који су од суштинске важности у препознавању мултимодалног говора;
- анализа и предлог могућег побољшања система предобраде говорног сигнала варијацијом одговарајућих параметара предобраде (величина рама, преклапање, отежавање, и слично) и/или могућом унификацијом мултимодалности говорног сигнала.

## 7. Закључак и предлог

Након увида у наведене резултате досадашњег рада кандидата Бранка Марковића, Комисија је донела одлуку да је он способан за научно-истраживачки рад и за реализацију предвиђених истраживања. Одлука је донета имајући у виду његове досадашње стручне резултате остварене током основних и магистарских студија, као и у фази истраживања везаним за докторат.

Такође, Комисија је мишљења да је предложена тема докторске дисертације „Анализа обележја у говорном сигналу за потребе аутоматског препознавања мултимодалног говора“, са коригованим насловом „Анализа обележја у говорном сигналу за потребе препознавања мултимодалног говора“ подобра у смислу наведених непосредних циљева

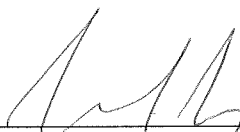
рада, очекиваних резултата и научних доприноса, као и у контексту савремених научних истраживања у предложеној области.

Комисија констатује да тема рада припада области Техничке акустике за коју је Електротехнички факултет у Београду матичан и за ментора рада се предлаже др Миомир Мијић, редовни професор, која је запослен на Електротехничком факултету у Београду. Из тог разлога уз овај Извештај приложен је и списак радова проф. др Миомира Мијића који га квалификују за ментора.

Имајући у виду чињенице наведене у овом извештају, Комисија предлаже Наставно-научном већу Електротехничког факултета у Београду да прихвати тему докторске дисертације „Анализа обележја у говорном сигналу за потребе препознавања мултимодалног говора“ и да кандидату Бранку Марковићу омогући да приступи њеној изради.

У Београду, 2.07.2015.

#### ЧЛАНОВИ КОМИСИЈЕ



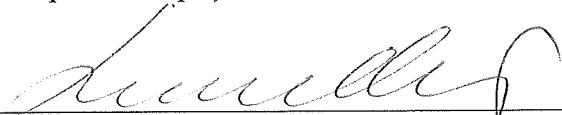
др Миомир Мијић, редовни професор  
Универзитет у Београду – Електротехнички факултет



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



др Милан Војновић, научни сарадник,  
ЦУЖА – Центар за Унапређење Животних Активности



др Срђан Станковић, професор емеритус  
Универзитет у Београду – Електротехнички факултет