

NASTAVNO-NAUČNOM VEĆU ELEKTROTEHNIČKOG FAKULTETA U BEOGRADU

Na 726. sednici Nastavno-naučnog veća Elektrotehničkog fakulteta u Beogradu održanoj 21.11.2010. godine, određeni smo za članove Komisije za ocenu uslova i prihvatanje teme doktorske disertacije mr Zorana Ćirovića pod naslovom "Sistem za verifikaciju govornika zasnovan na elektroglografskom signalu". Nakon pregleda dobijenog materijala, Nastavno-naučnom veću Elektrotehničkog fakulteta podnosimo sledeći

I Z V E Š T A J

PODACI O KANDIDATU

Zoran Ćirović je rođen 11.03.1970. godine u Novoj Varoši. Osnovnu i srednju školu je završio u Beogradu. Elektrotehnički fakultet u Beogradu upisao je 1988. godine. Diplomirao je 1993. godine, na smeru za Elektroniku i telekomunikacije, ostvarivši prosečnu ocenu tokom studija 9.62 i dobitnik je priznanja "Sreten Nedeljković" od Elektrotehničkog fakulteta. Od novembra 1993. godine radi u Institutu Mihajlo Pupin u Beogradu u laboratoriji za telekomunikacije, u grupi za digitalnu obradu signala. Završio je postdiplomske studije na Elektrotehničkom fakultetu u Beogradu u aprilu 1996. godine sa prosečnom ocenom 10, a magistarski rad pod naslovom "Obučavanje skrivenih Markovljevih modela u sistemima za prepoznavanje govora" je odbranio kod mentora profesora dr Milan Savića. Do sada je objavio jedan rad u međunarodnom časopisu, četiri rada na međunarodnim konferencijama, tri rada u domaćim časopisima i petnaest radova na domaćim konferencijama.

U okviru posla bavio se programiranjem na DSP procesorima TMS serije, radio je na naučno-istraživačkom projektu u okviru koga je radio na sistemu za prepoznavanje izolovano izgovorenih reči. Radio je i na projektovanje jednog TTS sistema za srpski jezik. Učestvovao je u izradi digitalnog snimača za glas do 30 analoginih kanala, (Registron-IMP-REG 8/16/30A). Radio je na izradi programa za skremblovanje govora i ugradnji istog na stanicama GP-300 koristeći DSP320F206 procesor. Radio je softver za IMBE vokoder kao na njegovoj primeni. Bavio se i prenosom govora u mrežnom okruženju u realnom vremenu. Izvesno vreme radio je i kao tehnički direktor firme „Partizan InfoPlus media“ gde je imao prilike da učestvuje i projektuje komercijalnih VAS aplikacija zasnovanih na dijalogu čovek-mašina.

Pojedini rezultati dosadašnjih istraživanja su publikovani u više radova od kojih je jedan sa IF. U toku dosadašnjeg rada, Zoran Ćirović je objavio sledeće naučne radove:

I Rad u časopisu sa IF

1. Zoran Ćirović, Milan Milosavljević, and Zoran Banjac, "Multimodal Speaker Verification Based on Electroglottograph Signal and Glottal Activity Detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 930376, 8 pages, 2010. doi:10.1155/2010/930376

II Radovi na međunarodnim konferencijama

2. Z. Ćirović, Lj. Stanimirović „Man-Machine Communication: an Isolated Word Recognition System Based On Hidden Markov Models“, *Professional Days on Signal Processing and Multimedia Technology Workshop*, Budapest, Hungary, October 1997.
3. Lj. Stanimirović, Z. Ćirović, M. Savić „Isolated Serbian word recognition system“, *Int. Conf. of Signal Proc. and Comm. – ICSPC'98*, Feb. 11-14, 1998, Las Palmas, Spain
4. Z. Ćirović, Lj. Stanimirović „A new model for text independent speaker identification“, *SSCC'98*, 22-24 September 1998, Durban, South Africa

5. Lj. Stanimirović, Z. Čirović, D. Đonin, M. Savić „Serbian Spoken Dialogue System For Bus Travel Information Retrieval“, *EUSIPCO*, 2000.

III Radovi u domaćim časopisima

6. Lj. Stanimirović, Z. Čirović „Sistemi za prepoznavanje pojedinačko izgovorenih reči bazirani na skrivenim Markovljevim modelima“, *ORIGINALNI NAUČNI ČLANAK, časopis NTP*, Beograd, septembar 1996.
7. Lj. Stanimirović, Z. Čirović, M. Savić „Sistem za prepoznavanje izolovano izgovorenih reči srpskog jezika“ *PREGLEDNI RAD, časopis INFO-Science*, 1997.
8. Lj. Stanimirović, Z. Čirović, M. Savić „Sistem za prepoznavanje izolovano izgovorenih reči srpskog jezika“, *časopis INFO*, br. 2, 1997.

IV Radovi na domaćim konferencijama

9. Z. Čirović, Lj. Stanimirović „Klasterovanje u sistemima za prepoznavanje izolovano izgovorenih reči baziranim na skrivenim Markovljevim modelima“, *Zbornik radova XXXIX konferencije ETRAN* (Zlatibor, 6-9.6.1995.), sveska II, str. 470-472, Zlatibor, 1995.
10. J. Drešević, Z. Čirović „Algoritam za utvrđivanje početka i kraja izolovano izgovorenih reči“, *Zbornik radova XXXIX konferencije ETRAN* (Zlatibor, 6-9.6.1995.), Zlatibor, 1995.
11. Lj. Stanimirović, Z. Čirović „Segmentacija reči na slogove korišćenjem funkcije glasnosti“, *Zbornik radova DoG* (Novi Sad, 10-11.1.1996.), Novi Sad, 1996.
12. „Upoređivanje funkcija raspodele verovatnoće kod sistema za prepoznavanje izolovano izgovorenih reči baziranim na kontinualnim SMM“, *Zbornik radova DoG* (Novi Sad, 10-11.1.1996.), Novi Sad, 1996.
13. Lj. Stanimirović, Z. Čirović „Segmentacija reči na slogove bazirana na mehanizmu čujenja“, *Zbornik radova naučno-stručnog skupa IT'96* (Žabljak, 11-15.3.1996.), str. 250-251, Žabljak, 1996.
14. Z. Čirović, Lj. Stanimirović „Realizacija sistema za prepoznavanje izolovano izgovorenih reči srpskog jezika“, *Zbornik radova naučno-stručnog skupa IT'96* (Žabljak, 11-15.3.1996.), str. 252-254, Žabljak, 1996.
15. Lj. Stanimirović, Z. Čirović „Prepoznavanje vezano izgovorenih reči“, *Zbornik radova XL konferencije ETRAN* (Budva, 4-7.6.1996.), sveska III, str. 357-359, Budva, 1996.
16. Lj. Stanimirović, Z. Čirović, M. Savić „Sistem za prepoznavanje izolovano izgovorenih reči srpskog jezika“, *Nagrađeni rad – Katalog INFOFEST-a* (Budva, 22-28.9.1996.), Budva, 1996.
17. Lj. Stanimirović, Z. Čirović, M. Savić „Prepoznavanje ključnih reči u spontanom govoru“, *DOGS'98*, Fruška Gora, 1998.
18. Z. Čirović „Jedan model za identifikaciju govornika nezavisno od teksta“, *DOGS'98*, Fruška Gora, 1998.
19. Z. Čirović „Obučavanje SMM-ova korišćenjem MCE kriterijuma“, *Zbornik radova XLII konferencije ETRAN* (V. Banja, 2-5.5.1998.), sveska II, str. 402-404, V. Banja, 1998.
20. Z. Čirović, M. Milosavljević, Z. Banjac „Analiza EGG signala pri verifikaciji govornika nezavisno od teksta u uslovima jakog šuma“, *TELFOR2009*
21. 1. Z. Čirović, M. Milosavljević, Z. Banjac, "Analiza uticaja detekcije govora i broja smeša na tačnost verifikacije govornika", p.39-42, *DOGS2008*, Kelebija
22. 2. Z. Čirović, M. Milosavljević, Z. Banjac, "Analiza EGG signala u multimodalnoj verifikaciji govornika na nepoznatom tekstu", p.42-45, *DOGS2008*, Kelebija
23. Z. Čirović, M. Milosavljević, Z. Banjac „Diskriminativna svojstava EGG parametara u sistemima za verifikaciju govornika baziranim na GMM modelima“, *TELFOR-2010*

Na osnovu dosadašnjeg stručnog rada i publikovanih naučnih rezultata može se zalljučiti da je kandidat Zoran Čirović kvalifikovan za rad na izradi algoritama za verifikaciju govornika uz dopunsko uključivanje elektroglogotografskog signala i izradi doktorske disertacije na predloženu temu.

PREDMET I CILJ ISTRAŽIVANJA

Poslednjih godina savremene govorne tehnologije nalaze svoju primenu u mnogim oblastima kako svakodnevnog života tako i u tehnologijama posebne namene. Govor je jedan od osnovnih načina prirodne razmene informacija, a telefonske mreže su sveprisutne i lako dostupna prosečnom korisniku. Verifikacija govornika predstavlja jednu od govornih tehnologija koja se svrstava u oblast prepoznavanja govornika. To je automatizovan proces provere identiteta osobe koristeći govorni signal. Po izvoru signala i načinu donošenja odluke spada u biometrijske metode, pa se ova tehnologija naziva i biometrijom govornika.

Primena sistema za verifikaciju govornika zavisi i od odnosa korisnika prema sistemu. Ukoliko se od korisnika očekuje saradnja, onda se najčešće projektuju sistemi za verifikaciju koji proveravaju identitet na osnovu govora koji je unapred očekivanog sadržaja - sistemi za verifikaciju zavisno od teksta. Druga grupa su sistemi za verifikaciju kod kojih je tekst koji se izgovara unapred nepoznat, dakle od korisnika se ne očekuje saradnja. Ovakvi sistemi su obično manje tačnosti od prethodnih, a često je njihova primena praćena i većim prisustvom šuma. U praktičnoj primeni, sistemi za verifikaciju se koriste za zaštitu pristupa, najčešće u kombinaciji sa još nekom vrstom zaštite. Dodatna zaštita može biti zasnovana na govoru, nekoj drugoj biometriji ili kriptološkim mehanizmima autentifikacije. Ako se dodatna zaštita vrši putem govora, onda se od govornika očekuje da izgovara određene reči ili rečenice koje su deo sigurnosnog protokola. Osnovne metode klasifikovanja u sistemima za verifikaciju se mogu podeliti na generativne i diskriminativne. U generativne metode klasifikovanja spadaju skriveni Markovljevi modeli (*Hidden Markov Model* - HMM), modeli Gausovih smeša (*Gaussian Mixture Model* - GMM), dok u diskriminativne spada, na primer, metod nosećih vektora (*Support Vector Machine* - SVM). Ukoliko se koristi generativni klasifikator, vrsta verifikacije ujedno opredeljuje i vrstu modela koji se koriste. Ako je tekst unapred poznat, sistem se priprema tj. obučava za određene govorne sekvence pa se za takve sisteme po pravilu koriste skriveni Markovljevi modeli. U slučaju kada je tekst unapred nepoznat, sistem se prilagođava za proizvoljan tekst, pa se po pravilu koriste modeli Gausovih smeša.

Govor nosi informacije o sadržaju, jeziku, emocijama, govorniku itd. Osim toga, govor je i slučajan proces, pa se razlikuje i kada isti tekst izgovori ista osoba sa istim emocijama. Standardni sistemi za verifikaciju govornika koriste audio signal sniman mikrofonom, kao jedini izvor informacija. Odmerci takvog signala se transformišu u niz vektora obeležja koji su pogodniji za dalju obradu. U ovim sistemima su obeležja po pravilu kepstalni koeficijenti i njihove vremenske diferencije. Pokazalo se da ovi parametri daju najbolje rezultate u sistemima za automatizovano izdvajanje informacija iz govora. U procesu verifikacije nezavisno od teksta, modeli govornika treba da sadrže samo informacije koje su specifične za tu osobu, a sve ostale treba da budu potisnute.

Osnovni problem u praktičnoj primeni verifikacije govornika je ambijentalni šum. Prisustvo šuma se ne može izbeći ni zanemariti. Sistemi koji se danas koriste obično imaju realizovane algoritme za smanjivanje efekata šuma. Veliki broj naučnih radova prikazuje algoritme za smanjivanje uticaja šuma. Neki od tih algoritama pretpostavljaju prethodno poznavanje prirode šuma kao što su: tehnike filtriranja šuma, kombinovanje paralelnih modela, adaptacija na okruženje, upotreba polja mikrofona ili tehnike izdvajanja govora koje modeluju raspodelu govora i šuma. Za slučajeve kada ne postoji dovoljno informacija o šumu, predstavljene su metode koje odbacuju jako zašumljene delove govora ili kombinuju obučavanje modela uz više ograničenja sa modelovanjem šuma nepoznatih spektralnih karakteristika. Novija istraživanja pokazuju da uključivanje dopunskih signala u proces automatizovanog prepoznavanja govornika smanjuje uticaj šuma. Takvi signali se dobijaju primenom neaudio senzora u toku procesa proizvodjenja govora.

Cilj istraživanja u preloženoj doktorskoj disertaciji je primena jednog neaudio senzora, elektroglotografa (EGG), u sistemima za verifikaciju govornika nezavisno od teksta zasnovanim na GMM modelima. Istraživanje treba da pokaže da li EGG signal sadrži diskriminativna svojstva u odnosu na različite govornike, kao i način na koji treba uključiti obeležja EGG signala u sistem za verifikaciju. Pošto elektroglotograf meri promenu električne provodljivosti usled promene dodirne površine glasnica, EGG signal sadrži informacije samo u zvučnim segmentima govora, pa treba dosatno ispitati i načine kombinovanja EGG obeležja sa kepstalnim parametrima audio signala u

formiranju sistema verifikacije govornika zasnovanog na GMM modelima. Realizacija ovakvog sistema bi omogućila veću tačnost verifikacije u prisustvu ambijentalnog šuma.

Osvrt na relevantne bibliografske izvore

Tradicionalni sistemi za verifikaciju govornika koriste snimljeni audio signal govora kao jedini izvor informacija. Audio signal se, nakon odmeravanja, u postupku predobrade, transformiše u vektore obeležja zasnovane na LPCC (*Linear Prediction Cepstral Coefficients*), MFCC (*Mel Frequency Cepstral Coefficients*) ili LAR (*Log Area Ratio*) koeficijentima, na vremenskim intervalima od oko 20ms, [1-3]. Koristeći vektore obeležja, vrši se statističko modelovanje govornika. Modeli govornika se koriste za izračunavanje verovatnoće da neka druga sekvenca vektora pripada određenom govorniku tj. osobi. U poslednjih nekoliko godina izdvojili su se modeli Gausovih smeša (GMM) kao jedan od najuspešnije i najčešće korišćenih modela govornika u procesu verifikacije nezavisno od teksta, [1], [4]-[7].

U slučaju da je govor zašumljen, vektor obeležja gubi informativnost i povećana je greška prepoznavanja. Prezentovano je više algoritama u rešavanju ovog problema koji pretpostavljaju poznavanje šuma, kao što su: tehnike filtriranja, [8],[9], kombinovanje više paralelnih modela, [10]-[12], Jacobian prilagođavanje, [13],[14], korišćenje niza mikrofona, [15],[16] ili tehnike isticanja govora zasnovane na modelovanju funkcije gustine verovatnoće govora i šuma, [17],[18]. Kada ne postoji dovoljno informacija o šumu, može se ignorisati jako zašumljeni deo govora [19],[20] ili se vršiti kombinacija obučavanja modela uz više uslova sa isključivanjem pojedinih karakteristika kako bi se modelovao šum nepoznatih karakteristika, [21].

Veću otpornost na šum je moguće postići primenom dodatnih senzora, tj. odgovarajućih signala govora. Pažljivo projektovan multimodalni sistem može biti uspešniji od bilo kog sistema zasnovanog na korišćenju samo jednog od signala, [1],[22]. Jedna grupa multimodalnih sistema je zasnovana na senzorima koji nisu u kontaktu sa govornikom, kao na primer: *Glottal Electromagnetic Micropower Sensors* (GEMS), ultrasonični ili video signal, [23]-[25]. Druga grupa je zasnovana na senzorima koji su fizički u kontaktu sa govornikom, kao na primer: elektroglograf (EGG), P-mikrofon, mikrofon koji naleže na kosti (obično na kosti lobanje ili lica), [22],[24],[26]. Upotreba senzora koji su fizički povezani sa govornikom je u specifičnim primenama (na primer u vojsci).

EGG senzor meri promene električne provodnosti na vratu, usled vibracija glasnica, [27]-[29]. Pošto EGG signal sadrži korisne informacije samo na zvučnim segmentima govora, primena ovog signala zahteva i primenu posebnog detektora aktivnosti. Zato, osim standardnog (VAD) detektora, [30], koji se koristi za izdvajanje govornih segmenata audio signala, ovaj sistem treba da poseduje i detektor aktivnosti glasnica (GAD),[31], za izdvajanje zvučnih segmenata govora u kojima EGG signal sadrži korisne informacije. Pre projektovanja multimodalnog sistema za verifikaciju, važno je utvrditi diskriminativna svojstva samo EGG signala. Pošto se koriste GMM modeli koji opisuju raspodele verovatnoća, merenje diskriminativnih svojstava se može izvesti primenom Kullback Leibler divergencije, [32].

Za projektovanje sistema za verifikaciju nezavisno od teksta, važno je koristiti bazu govornih uzoraka koja je glasovno slična tipičnom govoru za koji se sistem projektuje, [33].

- [1] A. R. Douglas and C. R. Richard, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions On Speech And Audio Processing*, January 1995, vol. 3, no. 1, pp. 72-83.
- [2] D. Chow and W. H. Abdulla, "Robust speaker identification based on perceptual log area ratio and gaussian mixture models," in *Proc. INTERSPEECH 2004 - ICSLP*, Jeju Island, Korea, October 2004, pp. 1761-1764.
- [3] W.B. Mikhael, and Pravinkumar Premakanthan, "Speaker verification/recognition and the importance of selective feature extraction: review," in *44th IEEE Proceedings on Midwest Symposium on Circuits and Systems*, Ohio, 2001, vol. 1, pp. 57 -61.

- [4] L. Burget, P. Matejka, P. Schwarz, O. Glembek, and J. Cernocky, "Analysis of feature extraction and channel compensation in a GMM speaker recognition system," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 1979-1986, 2007.
- [5] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19-41, 2000.
- [6] W.M. Campbell, D.E. Sturim, and D.A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308-311, 2006.
- [7] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of inter-speaker variability in speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 980-988, July 2008.
- [8] J. Ortega-Garcia and L. Gonzalez-Rodriguez, "Overview of speaker enhancement techniques for automatic speaker recognition," in *Proc. ICSLP'96*, Philadelphia, PA, 1996, pp. 929-932.
- [9] S. Suhadi, S. Stan, T. Fingscheidt, and C. Beaugeant, "An evaluation of VTS and IMM for speaker verification in noise," in *Proc. Eurospeech'03*, Geneva, Switzerland, 2003, pp. 1669-1672.
- [10] M. J. F. Gales and S. Young, "HMM recognition in noise using parallel model combination," in *Proc. Eurospeech'93*, Berlin, Germany, 1993, pp. 837-840.
- [11] T. Matsui, T. Kanno, and S. Furui, "Speaker recognition using HMM composition in noisy environments," *Comput. Speech Lang.*, vol. 10, pp. 107-116, 1996.
- [12] L. P. Wong and M. Russell, "Text-dependent speaker verification under noisy conditions using parallel model combination," in *Proc. ICASSP'01*, Salt Lake City, UT, 2003, pp. 457-460.
- [13] S. Sagayama, Y. Yamaguchi, S. Takahashi, and J. Takahashi, "Jacobian approach to fast acoustic model adaptation," in *Proc. ICASSP'97*, Munich, Germany, 1997, pp. 835-838.
- [14] C. Cerisara, L. Rigaziob, and J.-C. Junqua, "a-Jacobian environmental adaptation," *Speech Commun.*, vol. 42, pp. 25-41, 2004.
- [15] L. Gonzalez-Rodriguez and J. Ortega-Garcia, "Robust speaker recognition through acoustic array processing and spectral normalization," in *Proc. ICASSP'97*, Munich, Germany, 1997, pp. 1103-1106.
- [16] I. McCowan, J. Pelecanos, and S. Scridha, "Robust speaker recognition using microphone arrays," in *Proc. A Speaker Odyssey—The Speaker Recognition Workshop*, Crete, Greece, 2001, pp. 101-106.
- [17] Y. Hu and P.C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 4, pp. 334 - 341, July 2003.
- [18] A. Kundu, S. Chatterjee, A.S. Murthy and T.V. Sreenivas, "GMM based Bayesian approach to speech enhancement in signal/transform domain," *IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 4893-4896, March 2008.
- [19] A. Drygajlo and M. El-Maliki, "Speaker verification in noisy environment with combined spectral subtraction and missing data theory," in *Proc. ICASSP'98*, Seattle, WA, 1998, pp. 121-124.
- [20] L. Besacier, J. F. Bonastre, and C. Fredouille, "Localization and selection of speaker-specific information with statistical modelling," *Speech Commun.*, vol. 31, pp. 89-106, 2000.
- [21] J. Ming, T. J. Hazen, J. R. Glass and D. A. Reynolds, "Robust Speaker Recognition in Noisy Conditions," *IEEE Trans. On Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1711-1722, July. 2007.
- [22] W. M. Campbell, T. F. Quatieri, J. P. Campbell, and C. J. Weinstein, "Multimodal Speaker Authentication using Nonacoustic Sensors," *Proc. Int. Workshop on Multimodal User Authentication*, Santa Barbara, California, December 2003, pp. 215-222.
- [23] H.E. Cetingul, E. Erzin, Y. Yemez and A. M. Tekalp, "Multimodal Speaker/Speech Recognition using Lip Motion, Lip Texture and Audio," *Signal Processing (Elsevier)*, Special Issue on Multimodal Human-Computer Interfaces, vol. 86, No. 12, pp. 3549-3558, 2006.
- [24] T. F. Quatieri, D. P. Messing, K. Brady, W. M. Campbell, J. P. Campbell, M. S. Brandstein, C. J. Weinstein, J. D. Tardelli, P. D. Gatewood, "Exploiting Nonacoustic Sensors for Speech

- Enhancement”, *In Proc. Workshop on Multimodal User Authentication*, Santa Barbara, California, pp. 66–73, 11–12 December 2003.
- [25] B. Zhu, T. J. Hazen, J. R. Glass, „Multimodal Speech Recognition with Ultrasonic Sensors“, *in Proc. Interspeech 2007*, Antwerp, Belgium, August 2007, pp.662-665.
- [26] A. Subramanya, Z. Zhang, Z. Liu, J. Droppo, and A. Acero, “A Graphical Model for Multi-Sensory Speech Processing in Air-and-Bone Conductive Microphones,” *in Proc. Eurospeech2005*, Lisbon, Portugal, Sept. 2005, pp. 2361–2364.
- [27] D.G. Chlders,: *Speech Processing and Synthesis Toolboxes*, John Wilez & Sons, 2000.
- [28] M. Rothenberg, and J. J. Mahshie, “Monitoring Vocal Fold Abduction through Vocal Fold Contact Area,” *Journal of Speech and Hearing Research*, vol. 31, pp. 338-351, 1988.
- [29] R.J. Baken, “Electroglottography,” *Journal of Voice*, no 6, pp. 98-110, 1992.
- [30] M. Hahn, C.K. Park, “An improved speech detection algorithm for isolated Korean utterances,” *in Proc. ICASSP '92*, San Francisco, Calif, USA, March 1992, Vol. 1, pp. 525–528.
- [31] M. R. P. Thomas and P. A. Naylor, “The SIGMA Algorithm: A Glottal Activity Detector for Electroglottographic Signals,” *IEEE Audio, Speech, and Language Processing*, vol. 17, no. 8, November 2009
- [32] J. R. Hershey and P.A. Olsen, “Approximating the Kullback Leibler divergence between gaussian mixture models,” *in: IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2007, vol 4, pp. 317-320.
- [33] S.T. Jovičić, Z. Kašić, M. Đorđević, M. Rajković; "Serbian emotional speech database: design, processing and evaluation"; *SPECOM-04*, 2004, St.Petersburg, Russia.

POLAZNE HIPOTEZE

U radu se polazi od pretpostavke da EGG signal sadrži dovoljno informacija koje su specifične za govornika i da se ovaj signal može iskoristiti u kreiranju novog sistema za verifikaciju govornika nezavisno od teksta, koji će biti otporniji na spoljni šum.

Druga pretpostavka je da se vektori obeležja audio signala mogu proširiti sa obeležjima EGG signala i da tako dobijeni, zajednički vektor obeležja može biti iskorišćen za kreiranje zdeuženog sistema za verifikaciju. Značaj ova pretpostavke je što bi omogućila (i) primenu standardnih algoritama obučavanja GMM modela, koji su optimizovani za jednu vrstu signala, na zajednički sistem, (ii) u delu sistema za odlučivanje koristile bi se samo aposteriorne verovatnoće zajedničkog sistema obeležja. Pri tome neophodno je uvažiti različite prirode audio i EGG signala govora, imajući u vidu da EGG signal sadrži korisne informacije samo u zvučnim segmentima govora.

NAUČNE METODE ISTRAŽIVANJA

Metodologija istraživanja u okviru predložene doktorske disertacije sastojaće se u sledećem:

- analiza nedostataka standardnih sistema za verifikaciju govornika,
- analiza elektroglografskog signala, koja uključuje i prethodno snimanje baze uzoraka sa istovremenim zapisom audio i EGG signala,
- razvoj algoritama za predobradu elektroglografskog signala,
- razvoj algoritama za detekciju aktivnosti glasnica,
- razvoj algoritama za merenje diskriminativnih osobina EGG signala,
- razvoj referencnog sistema za verifikaciju zasnovanog na audio signalu i standardnom detektoru aktivnosti govora koristeći prethodno snimljenu bazu uzoraka,
- eksperimentalna provera predloženog rešenja u odnosu na referencni sistem.

Osnovne karakteristike predloženog rešenja:

- verifikacija se zasniva na primeni GMM modela,
- obučavanje GMM modela se vrši standardno, tj. primenom EM algoritma,
- verifikacija koristi dopunski signal govora dobijen primenom elektroglografa – EGG signal,
- izdvajanje informativnih segmenata govora se obavlja primenom dva detektora aktivnosti: (i) standardnog detektora koji se primenjuje na audio signal, (ii) detektora aktivnosti glasnica koji se primenjuje na EGG signal.
- Sama verifikacija koristi zajednički GMM model za audio i/ili EGG parametre. Pri odlučivanju se računaju dve vrednosti: (i) verovatnoća da sekvenca vektora obeležja govora pripada pretpostavljenom modelu, (ii) verovatnoća da ta sekvenca pripada univerzalnom tj. pozadinskom modelu.

OČEKIVANI NAUČNI DOPRINOS

Cilj istraživanja ove disertacije je razvoj sistema za verifikaciju govornika nezavisno od teksta, zasnovanog na GMM modelima, a uz upotrebu dopunskog elektroglografskog signala.

Očekuje se da istraživanja pokažu: (i) koji je skup EGG obeležja najdiskriminativniji za verifikaciju govornika, (ii) način uvođenja dodatnih EGG obeležja pri kreiranju sistema sa proširenim obeležjima zasnovanog na GMM modelima, (iii) procenu stepena dobiti u tačnosti i otpornosti na ambijentalni šum novog predloženog sistema verifikacije.

Naučni doprinos obuhvata istraživanja koja će omogućiti realizaciju robustnog sistema za verifikaciju govornika nezavisno od teksta primenom dopunskih obeležja EGG signala otpornog na ambijentalni šum. Napomenimo da je šum okoline osnovna prepreka za šire uvođenje sistema za verifikaciju na osnovu govorne biometrije na aerodromima u cilju podizanja opšteg nivoa bezbednosti prometa putnika u uslovima globalnog terorizma. Osim toga, ambijentalni šum je osnovna prepreka za kontinualnu biometrijsku verifikaciju na osnovu govorne komunikacije izvešioa specijalnih borbenih dejstava. Uvođenje sistema verifikacije govornika očekivane otpornosti na ambijentalni šum bi u velikoj meri smanjilo bezbednosne rizike u svim ovim i sličnim situacijama.

PLAN ISTRAŽIVANJA I STRUKTURA RADA

Plan istraživanja obuhvatiće analizu problema verifikacije govornika. Sadržaće pregled standardnih metoda za verifikaciju govornika i ukazaće na njihove nedostatke u primeni. Posebno će biti objašnjeni sistemi zasnovani na modelima Gausovih smeša, obzirom da je oblast istraživanja verifikacija nezavisno od teksta. Potrebno je analizirati celokupan postupak verifikacije koji uključuje: parametrizaciju, obučavanje GMM modela, formiranje pozadinskog modela kao i postupak odlučivanja na osnovu modela i ulazne sekvence govora.

U drugom delu biće prezentovan elektroglograf, kao i priroda signala koja se dobija primenom ovog senzora. Ovaj signal treba uporediti sa audio signalom na istim sekvencama govora i istaći razlike. Zatim, prezentovati karakteristične parametre koji se koriste za opis EGG signala kao i najpoznatije algoritme za dobijanje istih. Nakon toga, potrebno je uvesti algoritam za merenje diskriminativnih osobina ovog signala i uraditi odgovarajuće eksperimente. Imajući u vidu prirodu GMM modela, ovo merenje se može izvesti računanjem Kullback Leibler divergencije, (KLD). Merenje treba da uporedi KLD vrednosti za skup EGG parametara koji se koriste pri kreiranju GMM modela u dva slučaja: (i) kada su vektori obeležja od istog govornika (*intra-speaker variability*) (ii) kada su vektori obeležja od različitih govornika (*inter-speaker variability*).

U trećem delu biće prikazani rezultati eksperimenata koji treba da pokažu karakteristike novog sistema za verifikaciju i uporede standardni sistem sa novim. Pri tome je potrebno kreirati 4 sistema i uporediti dobijene rezultate. Prvi treba da bude standardni, sa upotrebom kepralnih vektora audio signala i standardnim detektorom aktivnosti govora zasnovanim na energiji i broju preseka sa

nulom. Drugi treba da bude isti kao prethodni, osim što će se detektor aktivnosti govora zameniti detektorom aktivnosti glasnica. Treći treba da bude zasnovan na zajedničkom vektoru obeležja audio signala i EGG signala, kao i na detektoru aktivnosti vibracija glasnica. Četvrti treba da bude isti kao i treći, osim što vektor obeležja čine samo EGG parametri. Svi sistemi treba da budu analizirani u pogledu osetljivosti na šum. Na osnovu dobijenih rezultata treba sagledati način korišćenja EGG parametara kao i dobit koja se postiže njihovom primenom.

Algoritmi za parametrizaciju audio odnosno EGG signala, detekciju aktivnosti govora odnosno glasnica, obučavanje modela, odlučivanje i na kraju upoređivanje rezultata, treba da bude urađeni u programskom jeziku MATLAB i testirani na bazi uzoraka koji su dobijeni istovremenim snimanjem audio i EGG signala.

ZAKLJUČAK I PREDLOG

Na osnovu izloženog, Komisija konstatuje da kandidat mr Zoran Ćirović ispunjava sve formalne i suštinske uslove za pristupanje izradi doktorske disertacije, kao i da je predložena tema “Sistem za verifikaciju govornika zasnovan na elektroglografskom signalu” naučno aktuelna u domenu govornih tehnologija i digitalne obrade signala. Zato Komisija predlaže Naučno-nastavnom veću Elektrotehničkog fakulteta u Beogradu da kandidatu mr Zoranu Ćiroviću odobri izradu doktorske disertacije pod gore navedenim naslovom.

U Beogradu, 19.01.2011. godine

Komisija:

dr Milan Milosavljević, redovni profesor

dr Branko Kovačević, redovni profesor

dr Vlado Delić, vanredni profesor