Analiza EGG signala pri verifikaciji govornika nezavisno od teksta u uslovima jakog šuma

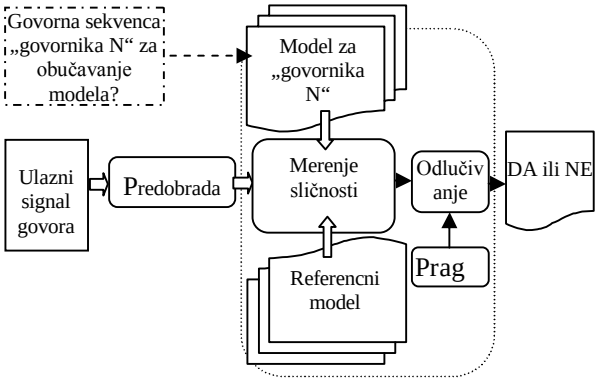
Zoran Ćirović, Milan Milosavljević, *Senior Member, IEEE,* Zoran Banjac *Member, IEEE*

***Sadržaj* — Većina tradicionalnih pristupa u verifikaciji govornika zasnovana je na karakteristikama audio signala, mada su takvi sistemi prilično osetljivi na prisustvo šuma. Da bi uvećali robustnost verifikacije, pedlažemo novi multimodalni metod koji uključuje neaudio karakteristike govora. Kao neaudio senzor koristili smo elektroglotograf (EGG). Algoritam za parametarizaciju EGG signala je zasnovan na aproksimaciji idealizovanog oblika i detektoru aktivnosti glasnica. Dobijeni rezultati pokazuju značajno unapređenje u verifikaciji govornika u uslovima jakog šuma i daju osnovu za dalje istraživanje u ovoj oblasti.**

***Ključne reči* — verifikacija, govor, egg, šum, multimodalni.**

I. UVOD

SAVREMENI sistemi za verifikaciju govornika (SV) nezavisno od teksta se sastoje iz dve osnovne faze: (1) faza formiranja modela tj. obučavanje, (2) faza testiranja tj.sama verifikacija. Neposredna verifikacija predstavlja drugu fazu ukupnog procesa. Sastoji se od stepena za parametarizaciju tj. izdvajanje govornih obeležja (specifičnih za govornika) i od stepena za klasifikaciju. Govorna obeležja treba da sadrže sve specifičnosti govornika, ali istovremeno i da odbacuju negovorničke specifičnosti (šum, prenosni kanal, emocije i sl.). Na Sl. 1. je prikazana struktura jednog sistema zasnovanog na verifikaciji nezavisno od teksta.



Sl. 1. Blok šema sistema za verifikaciju govornika

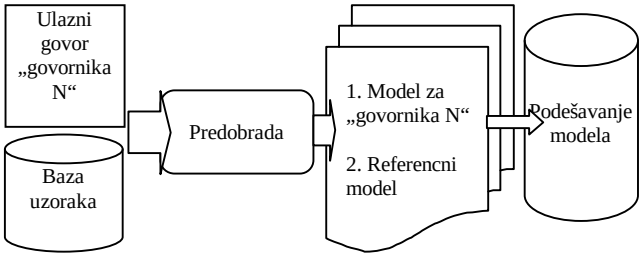
Z. Ćirović, Visoka škola za elektrotehniku i računarstvo u Beogradu, Srbija (telefon: 381-11-3950022; faks: 381-11-2471099; e-mail: zoran.cirovic@gmail.com).

M. Milosavljević, Elektrotehnički fakultet u Beogradu, Bulevar kralja Aleksandra 73, Srbija; (e-mail: mmilan@etf.bg.ac.rs).

Z. Banjac, Visoka škola za elektrotehniku i računarstvo u Beogradu, Srbija (e-mail: zbanjac@viser.edu.rs).

Stepen za parametarizaciju vrši transformaciju ulaznog govornog signala u niz višedimenzionalnih vektora (vektori obeležja). Svaki vektor odgovara kratkom akustičkom segmentu (par desetina milisekundi). Govor se smatra kvazistacionarnim u tom intervalu. Stepen za klasifikaciju (klasifikator) daje automatizovanu odluku da li niz vektora govornih obeležja, dobijen na izlazu stepena za predobradu, odgovara određenom govorniku. Za takvu odluku se koriste pripremljeni modeli: testiranog gvornika i referencni (pozadinski) model. Na osnovu procenjene sličnosti da sekvenca vektora pripada jednom od ta dva modela i praga odlučivanja sistem daje konačnu odluku da li je govorna sekvenca od tog govornika ili ne. Modeli koji se koriste tokom verifikacije su kreirani tokom faze obučavanja.

Na Sl. 2. je prikazana generalizovana blok šema procesa obučavanja za verifikaciju govornika.



Sl. 2. Blok šema faze obučavanja modela

Za što bolju obuku potrebna je što veća baza govornih uzoraka od različitih govornika. Baza uzoraka se koristi za formiranje referencnog (pozadinskog) modela. Nakon kreiranja referencnog modela, kreiraju se modeli govornika koji će biti testirani u procesu verifikacije.

II. MULTIMODALNI SISTEMI

Multimodalni sistemi zasnivaju se na korišćenju više različitih vrsta signala. U sistemima za prepoznavanje govora i govornika prvi multimodalni sistemi su bili zasnovani na kombinaciji audio signala i signala slike lica govornika. Na taj način se pokušalo da prenese na mašinu prepoznavanje koje sam čovek koristi. Kasnija istaživanja su rađena i sa drugim neaudio signalima. Neaudio signali govora dobijaju se primenom odgovarajućih neaudio senzora. Doprinos ovakih signala može značajno da utiče na tačnost prepoznavanja govornika odnosno govora, a pokazalo se da su neki neaudio senzori značajno robustniji na spoljna ometanja, recimo na šum koji je

jedan od najvećih ometača u procesu analize govora.

Od neaudio signala danas se najviše koriste 3 vrste senzora: EGG , GEMS, P-mic[2],[3].

EGG – *Electroglottograph*- uređaj koji meri promene provodnosti, odnosno rad glasnica, praćenjem promena impedanse tkiva na vratu u predelu glasnica. Za merenje se koriste dve sonde između kojih se propušta visokofrekventna struja male jačine.

GEMS – *Glottal Electromagentic Micropower Sensors* – uređaj koji radi na principu radara u opsegu mikro talasa. Meri se refleksija od vibracija delova vokalnog trakta.

P-mic – *physiological microhone* – uređaj koji se sastoji od gela koji menja impedansu i piezoelektričnog elementa. Tipično se smešta ispod glasnica i meri vibracije.

U ovom radu je prikazano istraživanje EGG signala u sistemu za verifikaciju govornika nezavisno od teksta.

III. Elektroglotografija

Elektroglotograf je uređaj za merenje kretanja (varijacija) i stepena kontakta između vibrirajućih glasnica tokom govora. Tačnije, uređaj meri promene impendanse između dve elektrode koje se postavljaju na vrat u predelu grkljana, a promene su proporcionalne promenama stepena kontakta između glasnica. Tokom merenja, između pomenutih elektroda se uspostavlja stujni tok visokofrekventne struje (frekvencije u području MHz), malog napona i jačine potpuno bezopasne za čoveka.

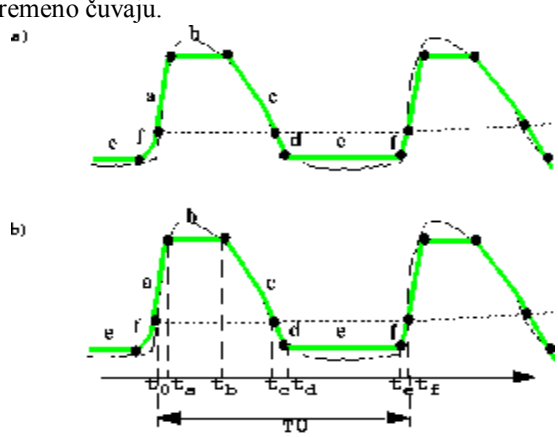
Sl. 3. Izgled elektroda glotografa

Ukupni EGG signal je superpozicija komponente brzog kretanja (vibracija) glasnica i komponente sporih kretanja koja su po svojoj prirodi dugačija. Sporo promenljiva komponenta signala potiče, na primer, od kretanja grkljana tokom gutanja, od vertikalnog pomeranja grkljana koje je povezano sa generisanjem glasova tokom govora, ali i od promena u kontaktu između elektroda i kože. Brzo promenljiva komponenta nosi informacije o kretanju glasnica pa talasni oblik ove komponente koristimo za parametarizaciju EGG signala.

Primena EGG signala vezana je za mogućnost korišćenja EGG uređaja u verifikaciji, što može da predstavlja problem pri praktičnoj realizaciji.

IV. Predobrada EGG signala

Da bi se dobio kvantitativni opis EGG signala, uvodi se model zasnovan na idealizovanoj formi talasnog oblika, Rothenberg (1983) i Baken (1992), Marasek, 1995a, 1996, Sl. 4. Oštar rastući nagib se smatra početkom jedne periode EGG signala. Idealizovan talasni oblik ima ravan maksimum, segment (b), mada postoje varijacije u provodnosti kod originalnog signala koji ima tipično parabolički oblik, a može takođe imati male slučajne oscilacije usled promene kapacitivnosti koje nisu rezultat kretanja glasnica (Esling, 1984). Da bi se eliminisao taj efekat, postavlja se prag na 90% maksimuma i minimuma amplitude. Ova vrednost praga je postavljena na osnovu Esling-ove preporuke, ali su kasnija istraživanja pokazala potvrdu vrednosti ovog emirijskog praga. Maksimum kontaktne faze (faze zatvorenosti na Sl. 3.) je između preseka EGG signala sa 90% pragom. Trenutak maksimalne amplitude, isto kao i minimumi se privremeno čuvaju.



Sl. 4. a) Model EGG signala sa ravnim segmentima b) Vremenski intervali - faze kretanja glasnica

Opadajuća ivica EGG signala je podeljena na dve faze tokom otvaranja. Definicija tačke koja razdvaja ove dve faze u EGG domenu je teška usled metodoloških ograničenja i ometajućih efekata. Određivanje karakterističnih tačaka je od zanačaja za parametarizaciju EGG signala.

V. Eksperimentalni rezultati

A. Govorna baza

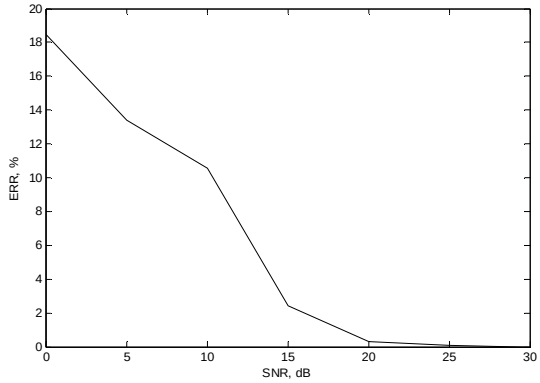
Baza koja je korišćena u eksperimentima sastoji se od 50 snimaka (fajlova) napravljenih sa 16 govornika. Za svakog govornika je napravljeno do 4 snimka. Rečenice korišćene u govoru su pažljivo birane tako da dobro predstavljaju tipičan srpski jezik [9]. Audio i EGG signali su snimani istovremeno i sinhrono pomoću mikrofona i EGG uređaja (model EG-PC3 - Tiger DRS, Inc., USA). Oba signala se odmeravaju sa 44kHz. U toku faze testiranja, svaki od 50 snimaka se koristi kao ulazni u sistem tj. sinimak čija se verifikacija testira. Na ovaj način se vrši 49*50 = 2450 merenja, a to istovremeno i definiše tačnost eksperimenta. Od ovog broja testova, manji broj je sa uzorcima istog govornika, svega 49*3=147.

B. Konvencionalni sistem za verifikaciju (Sistem 1)

Konvencionalni sistem za verifikaciju govornika karakteriše predobrada govornog signala zasnovana na

mel-keprstalnim vekorima. Kepstralni vector se sastoji od 14 MFCC koeficijenata uključujući nulti kepstralni koeficijent. Ukupno 13 koeficijenata, plus promene vrednosti pojedinih komponenata (delta i delta-delta koeficijenti) daju vektor diumenzija D=42 koeficijenta po svakom okviru. Svaki okvir se sastoji od 1024 odmerka na vremenskom intervalu od $t_W \approx 23.2ms$. Okviri su preklopljeni kako bi se izbegao rizik gubitka korisnih informacija na prelazima. U ovom sistemu, okviri su preklopljeni na polovinu dužine, tj. 512 odmeraka. Nakon računanja MFCC kepstralnih koeficijenata, radi se normalizacija oduzimanjem srednje vrednosti za svaki koeficijent posebno, (*cepstral mean subtraction)* [10]. I na kraju, konvencionalni sistem za verifikaciju koristi konvencionalni detektor glasovne aktivnosti (VAD) za odvajanje okvira govornog signala od šuma. Ovaj VAD detektor se zasniva na praćenju promena energije i broja preseka sa nulom.

Snimanje uzoraka je obavljeno u kancelarijskom radnom okruženju. Slični uzorci su se koristili i u fazi obučavanja modela. U toku faze testiranja signalu je dodat beli aditivni Gausov šum. Greška verifikacije ERR, je računata za različite vrednosti odnosa signal šum SNR, od 0 do 30dB sa korakom od 5dB kao na Sl. 5.

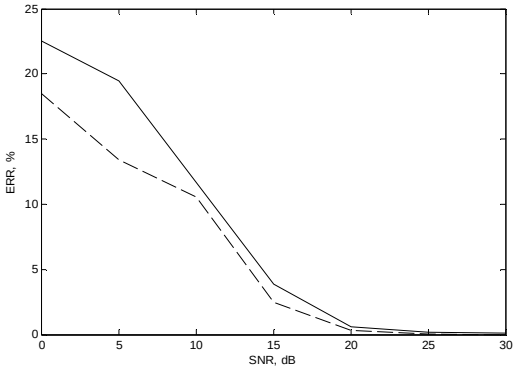


Sl. 5. Greška verifikacije za različite vrednosti SNR u konvencionalnom sistemu za verifikaciju (Sistem 1)

Na osnovu dobijenih rezultata može se zaključiti da je konvencionalni SV sistem prilično osteljiv na beli šum. U uslovima jakog šuma kada je SNR<15dB, uticaj šuma postaje značajan.

C. GAD vs. VAD (Sistem 2)

Drugi eksperiment pokazuje grešku verifikacije za različite odnose signal šum kada se VAD detektor zameni sa detektorom glotalne aktivnosti (GAD) (Sl. 6.). Dobijeni rezultati su prikazani punom linijom. Mada šum nema uticaja na GAD, greška verifikacije govornika je uvećana, čak i u uslovima jakog šuma. Očigledno, rezultai su posledica izbora detektora govora. VAD detektor odvaja govor koristeći adaptivne pragove u zavisnosti od nivoa pozadinskog šuma. Sa druge strane, GAD detektor govora izdvaja govor nezavisno od tog nivoa. Na taj način, segmenti dobijeni VAD detektorom su sa većim efektivnim SNR nego u slučaju GAD detektora.

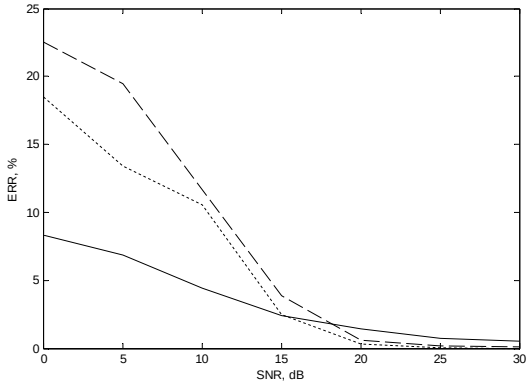


Sl. 6. Greška SV sistema zasnovanog na kepstralnim vekorima i GAD (Sistem 2 – puna linija), odnosno VAD detekotrom (Sistem 1 – isprekidana linija).

Ukoliko je govorna sekvenca dovoljno velika tako da za testiranje postoji dovoljan broj vektora u segmentima govora sa većom energijom, očekivano je da i rezultati dobijeni VAD detekotrom budu bolji, odnosno da će efekat šuma će biti manji u tom slučaju.

D. Spajanje EGG parametara sa kepstralnim koeficijentima (Sistem 3)

Skup normalizovanih vremenskih intervala, dat na Sl. 4. i osnovna perioda T_0, se koriste kao parametri EGG signala. U ovom eksperimentu, konvencionalni vektor obeležja se proširuje sa 5 parametara (ukupno 47 koeficijenata). Za detektor aktivnosti koristi se GAD. Obučavanje i testiranje je urađeno kao kod konvencionalnog SV sistema. Rezultati su dati na Sl. 7.



Sl. 7. Greška SV sistema koji koristi: GAD i EGG parametre sa kepstralnim vektorima (Sistem 3 – puna linija), samo GAD i kepstralne vektore (Sistem 2 – isprekidana linija), VAD detektor i kepstralni vektori (Sistem 1 – tačkasta linija).

Greška prepoznavanja za različite vrednosti SNR je prikazana u Tabeli 1.

Rezultat prikazan u tabeli pokazuje ostvareni dobitak δ, kao razliku između tačnosti verifikacije tradicionalnog SV sistema (1) i unapređenog (2).

Kroz analizu rezultata koji su prezentovani ovde, jasno se vidi da EGG parametri mogu dati doprinos SV sistemima u šumovitim uslovima.

TABELA 1: SV GREŠKA ZA (1) TRADICIONALNI VEKTOR, (2) SPOJENI VEKTOR SA GAD I POSTIGNUTI DOBITAK δ .

	0dB	*5dB*	*10dB*	*15dB*
(1) Sistem 1	18.46	13.40	10.54	2.45
(2) Sistem 3	8.33	6.85	4.4	2.37
$\delta = (1) - (2)$	**10.13**	**6.55**	**6.14**	**0.08**

Kako je prikazano na Sl. 7 odnosno tabeli 1, dobijen je značajan doprinos od čaki 10% tačnosti verifikaciji govornika pri jakom šumu.

VI. ZAKLJUČAK

U ovom radu prikazali smo istraživanje u pogledu unapređenja verifikacije govornika u uslovima jakog šuma. Eksperimenti su rađeni za srpski jezik, nezavisno od teksta, a korišćeni su statistički modeli Gausovih smeša.

Rezultati nedvosmisleno pokazuju da postoji značajan doprinos EGG parametara u SV sistemima, na način kako je to pokazano, u uslovima visokog šuma i potvrđuju opravdanost daljih istraživanja u ovoj oblasti.

LITERATURA

[1] Campbell W. M "Multimodal Speaker Authentication using Nonacoustic Sensors"**.** *Proc. Int. Workshop on Multimodal User Authentication*, Santa Barbara, Calif, Dec. 2003, pp. 215-222.

[2] Douglas A. R „Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", *IEEE Transactions On Speech And Audio Processing*, January 1995, Vol. 3, NO. 1.

[3] Furui S. „1.7 Speaker recognition", *In 'Survey of the State of the Art in Human Language Technology*, 1996.

[1] Chlders D.G.: „Speech Processing and Synthesis Toolboxes", *John Wilez & Sons*, 2000.

[2] M. Rothenber, "Monitoring Vocal Fold Abduction through Vocal Fold Contact Area", *Journal of Speech and Hearing Research*, 1988, Vol. 31, pp. 338-351.

[3] Baken R.J., „Electroglottography", *Journal of Voice*, 1992, 6, pp. 98-110.

[4] Hahn M.: "An improved speech detection algorithm for isolated Korean utterances", *ICASSP '92, San Francisco, Calif, USA, March 1992*, Vol. 1, pp. 525–528.

[5] Dempster A., "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, 1977, pp. 1–38.

[6] S.T. Jovičić, „Serbian emotional speech database: design, processing and evaluation"; *SPECOM-04*, 2004, St.Peter, Russia.

[7] Reynolds D. "Speaker Verification Using Adapted Gaussian Mixture Models", *Digital Signal Processing*, 2000, pp. 19-41.

[8] A. Rosenberg, "The use of cohort normalized scores for speaker verification", *ICSLP '92*, 1992, Alberta, Canada, pp.599-602.

[9] F. Bimbot, "A Tutorial on Text-Independent Speaker Verification", *EURASIP Journal on Applied Signal Processing*, '04, pp. 430-451

[10] W. M. Campbell, **"**Multimodal Speaker Authentication Using Nonacoustic Sensors", *In Proc. Workshop on Multimodal User Authentication in Santa Barbara*, Calif., pp. 215–222, 2003.

[11] T. F. Quatieri, "Exploiting Nonacoustic Sensors for Speech Enhancement", *In Proc. Workshop on Multimodal User Authentication in Santa Barbara*, Calif., pp. 66–73, 2003

[12] J. Ming, "Robust Speaker Recognition in Unknown Noisy Conditions," *Proc. IEEE Trans. Audio, Speech, and Language*, 15(5), 1711-1723, 2007.

[13] Mark T., "The SIGMA Algorithm: A Glottal Activity Detektor for Electroglottographic Signals", *Audio, Speech, And Language Processing*, Vol. 17, no. 8, November 2009 1557.

[14] Patrick A., "Estimation of Glottal Closure Instants in Voiced Speech Using the DYPSA Algorithm", *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 15, no. 1, January 2007.

[15] K. Sri Rama Murty, "Epoch Extraction From Speech Signals ", IEEE Trans. on Audio, Speech, and Language Processing, Vol. 16, no. 8, November 2008

[16] A. Bouzid, N. Ellouze, "Electroglottographic Measures Based on GCI and GOI Detection Using Multiscale Product" *International Journal of Computers, Communications & Control* Vol. III (2008), No. 1, pp. 21-32.

[17] Holzrichter, F.; "Speech articulator measurements using low power EM-wave sensor", *J. Acoust Soc. Am.* 103 (1) 622, 1998.

[18] Burnett, G. C., "The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract", *Thesis UC Davis, Jan. 15th, 1999*, document #9925723 available from University Microfilms, Inc. Ann Arbor, Michigan;.

ABSTRACT

The majority of existing approaches in the speaker verification area is based on audio signal features, but these features could be pretty sensitive in high noise environment. To achieve robust verification, we propose new multimodal method which includes additional non-audio features. As non-audio sensor electroglottograph (EGG) is applied. Algorithm for EGG parameterization is based on the shape of the idealized waveform and glottal activity detektor. Obtained results show significant improvement of the text independent speaker verification and opportunity for future improvements in this area.

ANALYSIS OF EGG SINGAL FOR TEXT INDEPENDENT MULTIMODAL SPEAKER VERIFICATION IN HIGH NOISE ENVIROMENT

Zoran Ćirović, Milan Milosavljević, *Senior Member, IEEE*, Zoran Banjac, *Member, IEEE*