

Diskriminativna svojstva EGG parametara u sistemima za verifikaciju govornika zasnovanim na GMM modelima

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

Sadržaj — U ovom radu je prikazana jedna metoda za utvrđivanje doprinosa određenih parametara vektora obeležja govornog signala u procesu verifikacije govornika u sistemima koji su zasnovani na smeši Gausovih raspodela. Metoda je primenjena za analizu EGG parametara i služi kao osnova za dalja istraživanja u oblasti primene u sistemima za verifikaciju govornika.

Ključne reči — divergencija, EGG, GMM, govor, Kullback-Leibler, verifikacija.

I. UVOD

GOVOR sadrži više različitih informacija koje se zajedno razmenjuju u procesu komunikacije. Automatsko izdvajanje tih informacija je još uvek u potpunosti nerešen problem i predmet je obimnih istraživanja.

Prepoznavanje govornika je automatizovani proces i u opštem slučaju mogu se razlikovati dve vrste prepoznavanja:

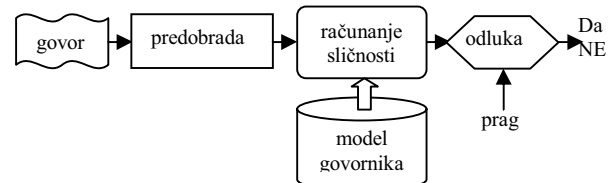
- identifikacija: postupak donošenja odluke da li analizirani govorni signal pripada jednom ili nijednom od govornika iz definisanog skupa,
- verifikacija: potvrda/odbacivanje pretpostavke da analizirani govorni signal pripada unapred određenom govorniku.

Verifikacija govornika se naziva još i biometrija govornika. Tačnost verifikacije nije još uvek dovoljno visoka u odnosu na druge biometrije (otisak prsta, DNK,...), [1]. Na tačnost verifikacije utiču: velika osetljivost na šum, promene u prenosnom kanalu kao i potreba za velikim brojem uzoraka tokom obučavanja sistema. I pored navedenih slabosti, sistemi za verifikaciju govornika imaju značajnu upotrebnu vrednost.

Tradicionalni sistemi za verifikaciju govornika zasnivaju se na obradi audio signala govora. U cilju

povećanja tačnosti, sistemi za verifikaciju mogu obuhvatiti parametre koji potiču iz više različitih izvora, čime se uvodi multimodalnost. U procesu verifikacije govornika postoje neaudio signali koji se dobijaju primenom posebnih senzora (neaudio senzora), kao što su EGG (*electroglotograph*), GEMS (*glotalelectromagnetic sensor*), P mikrofoni, [2]-[4]. Pretpostavka o povećanju tačnosti može biti zasnovana na ispitivanju njihove informativnosti i činjenici da su neaudio senzori robustniji na spoljna ometanja, kao što je šum koji je jedan od najvećih ometača u postupcima prepoznavanja govornika odnosno govora, [5], [6].

Na Sl. 1 je prikazana blok šema sistema za verifikaciju govornika. U bloku za predobradu se vrši filtriranje negovornih segmenata tj. izbacivanje pauza i šuma kao i izvesna kompenzacija različitosti prenosnog kanala. Zatim se vrši parametrizacija, tj. prebacivanje odmeraka iz vremenskog domena u neki skup obeležja pogodan za dalju obradu. Dobijena sekvenca vektora obeležja se koristi, u bloku za izračunavanje sličnosti, za poređenje sa unapred pripremljenim statističkim modelom (iz baze unapred pripremljenih modela). Izračunate verovatnoće se upoređuju i donosi se konačna odluka da li analizirani govor pripada određenom govorniku ili ne.



Sl. 1. Šematski prikaz verifikacije govornika

Uvođenje dodatnih parametara u proces verifikacije govornika radi povećanja tačnosti zahteva analizu njihovih diskriminativnih svojstava. Analiza ovih svojstava se u opštem slučaju može postići na dva načina: (i) *a priori*, bez kreiranja sistema za klasifikaciju i procene tačnosti estimacije sistema; (ii) *a posteriori*, poredeći tačnost verifikacije sistema sa i bez uključivanja dodatnih parametara. Nedostatak pristupa (i) je da unapred procenjena diskriminativna svojstva po uvođenju u realan sistem i usklađivanju sa drugim parametrima

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).

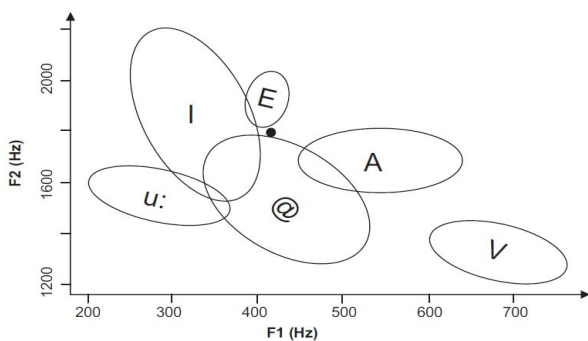
M. Milosavljević, Elektrotehnički fakultet u Beogradu, Bulevar kralja Aleksandra 73, Srbija; (e-mail: mmilan@etf.bg.ac.rs).

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

sistema mogu da budu narušena. Sa druge strane, nedostatak pristupa (ii) je da nastaju promene analiziranih svojstava usled prethodnog prilagođavanja sistema novim parametrima. U ovom radu će biti analizirana diskriminativna svojstva EGG parametara pristupom (i) u cilju kasnijeg uključivanja u sistem za verifikaciju govornika, [7].

II. GMM MODELI

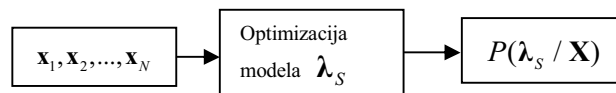
GMM modeli, koji se koriste za verifikaciju govornika, predstavljaju skup parametara koji statistički opisuju određenog govornika. U klasifikatoru, modeli se koriste kako bi se izračunala kvantitativna mera sličnosti tj. verovatnoća da govor, odnosno, sekvenca vektora obeležja govornog signala, pripada specifičnom govorniku. Izračunata verovatnoća koristi se kao rastojanje između sekvence vektora i statističkog modela koji se koristi za poređenje. Potreban uslov da neki statistički modeli mogu biti korišćeni u fazi odlučivanja je da svaki model prezentuje jedan deo prostora vektora obeležja i da ne postoji nedefinisanost pripadnosti nekog dela prostora. Modeli zasnovani na smeši Gausovih raspodela ispunjavaju gore pomenuti uslov i do danas su se pokazali veoma uspešnim u praksi, pa će biti korišćeni za kreiranje unapređenog sistema za verifikaciju, [5], [8], [10].



Sl. 2. Regioni karakterističnih frekvencija prvog formanta, [9]

Na Sl. 2 je prikazan primer klasifikacije jedne tačke iz prostora karakterističnih frekvencija (F1, F2), [9]. Ako se kao rastojanje koristi prosto euklidsko, tada će označena tačka da pripada regionu "E". Međutim, očigledno da je verovatnije da ta tačka pripada region "@" jer je na samoj ivici tog regiona. Označeni regionu u obliku elipsi, predstavljaju oblasti formantnih frekvencija određenih glasova. Na sličan način se određuje verovatnoća da li neki skup vektora obeležja pripada specifičnom govorniku ili ne. Za svakog govornika se kreiraju modeli sa najvećom sličnošću, odnosno verovatnoćom za tog govornika. Model govornika se kreira koristeći govor u fazi obučavanja i matematičke procedure za optimizaciju modela, tačnije parametara koji čine model. Kada se u iterativnom postupku postigne dovoljno dobar model, on se skladišti u bazi modela i kasnije koristi u postupku testiranja odnosno neposredne verifikacije. Iterativni postupak određivanja modela tj. njegovih parametara, nazivamo obučavanje. Šematski je obučavanje prikazano

na Sl. 3. Drugim rečima, ako je $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ skup vektora obeležja, a λ_s model govornika s , obučavanje je određivanje parametara modela tako da se maksimizira (optimizuje) verovatnoća $P(\lambda_s / \mathbf{X})$.



Sl. 3. Obučavanje kao postupak određivanja parametara modela λ_s na osnovu niza ulaznih vektora \mathbf{X} .

Na osnovu Bayes-ovog pravila važi

$$P(\lambda_s / \mathbf{X}) = \frac{P(\mathbf{X} / \lambda_s) P(\lambda_s)}{P(\mathbf{X})} \quad (1)$$

gde je $P(\lambda_s)$ verovatnoća modela λ_s govornika s . U praksi je često ispunjen uslov da su svi govornici iz referentnog skupa (baze) jednako verovatni tokom verifikacije, pa ova vrednost tada nema uticaja na konačnu odluku. $P(\mathbf{X})$ je verovatnoća sekvence vektora obeležja \mathbf{X} . Ova vrednost je ista za različite modele i ne utiče na ishod verifikacije. Ako su ispunjeni uslovi, tokom obučavanja i same verifikacije, može se koristiti verovatnoća $P(\mathbf{X} / \lambda_s)$.

Uvodeći pretpostavku da je sekvenca vektora obeležja \mathbf{X} nezavisna važi:

$$P(\mathbf{X} / \lambda_s) = \prod_{t=1}^N p(\mathbf{x}_t / \lambda_s) \quad (2)$$

Uvedene pretpostavke su neophodne za iterativni postupak obučavanja. U slučaju vektora obeležja koga čine kepralni koeficijenti i njihove promene u vremenu, što je i najčešći izbor parametara, eksperimentalno je pokazano da su raspodele verovatnoće Gausovog tipa ili se mogu veoma dobro aproksimirati smešom Gausovih raspodela:

$$p(\mathbf{x}_t / \lambda_s) = \sum_{i=1}^M w_i p_i(\mathbf{x}_t / \lambda_s) \quad (3)$$

Svaka smeša (komponenta) ima svoj težinski koeficijent w_i sa kojim učestvuje u ukupnoj raspodeli. Pri tome, svaka komponenta smeše je jedna multivarijantna Gausova funkcija raspodele verovatnoće:

$$p_i(\mathbf{x} / \lambda_s) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (4)$$

gde je Σ_i kovarijansna matrica, Σ_i^{-1} inverzna kovarijansna matrica, $|\Sigma_i|$ determinanta, $\boldsymbol{\mu}$ vektor srednjih vrednosti, a d dimenzija vektora obeležja \mathbf{x} . Parametri smeše multivarijantnih raspodela predstavljaju parametre modela λ_s jednog govornika. Ukupan broj parametara jednog GMM modela sa M smeša je M vektora srednjih vrednosti, isto toliko kovarijansnih matrica odnosno težinskih koeficijenata:

$$\lambda_s = \{w_i, \boldsymbol{\mu}_i, \Sigma_i\}, i=1, \dots, M \quad (5)$$

Pošto se pri verifikaciji govornika, zasnovanoj na smeši Gausovih raspodela, porede verovatnoće da GMM model

generiše određenu sekvencu vektora, za merenje specifičnosti parametara može se koristiti Kullback-Leibler divergencija (*KLD*) za odgovarajuće verovatnoće raspodele.

III. KULLBACK-LEIBLER DIVERGENCIJA

Kullback-Leibler divergencija (*KLD*), poznata i kao relativna entropija, predstavlja osnovnu meru divergencije između dve raspodele $p(x)$, $q(x)$, definiše se jednačinom:

$$KLD(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad (6)$$

Divergencija ima nekoliko važnih osobina:

(i) $KLD(p \parallel p) = 0$, (ii) $KLD(p \parallel q) = 0$ samo ako je $p = q$ (iii) $KLD(p \parallel q) \geq 0$, za svako p, q . Mada se *KLD* intuitivno doživljava kao mera rastojanja, jednačina (6) ne zadovoljava uslov simetričnosti, $KLD(p \parallel q) \neq KLD(q \parallel p)$, pa se *KLD* ne može koristiti kao mera rastojanja. U tu svrhu obično se koristi $(KLD(p \parallel q) + KLD(q \parallel p)) * 1/2$, što zadovoljava uslov simetričnosti, [10].

IV. OČEKIVANA INFORMATIVNA RAZLIČITOST

$KLD(p(\mathbf{x}_n/H) \parallel p(\mathbf{x}_n/\bar{H}))$ se može interpretirati kao očekivano razdvajanje informacija između statističkih hipoteza, tj. pretpostavke H , u odnosu na alternativnu pretpostavku \bar{H} , pri čemu je pretpostavka H tačna. Ako H predstavlja model označen sa λ_i , koji je definisan za pretpostavljenog govornika u prostoru vektora obeležja \mathbf{x}_{EGG}^n , a \bar{H} predstavlja neki drugi model λ_j , $j \neq i$, očekivano razdvajanje informacija postaje:

$$KLD(i \parallel j) = KLD\left(p(\mathbf{x}_{EGG}^n/\lambda_i) \parallel p(\mathbf{x}_{EGG}^n/\lambda_j)\right), j \neq i \quad (7)$$

Modeli λ_i , λ_j predstavljaju GMM modele kreirane samo za testirane parametre u procesu verifikacije.

$KLD(i \parallel j)$, definisan jednačinom (7), se meri kada modeli λ_i i λ_j : (a) pripadaju istom govorniku, označava se sa $KLD_{intra}(i \parallel j)$ (intra-speaker variability), odnosno (b) pripadaju različitim govornicima, označava se sa $KLD_{inter}(i \parallel j)$ (inter-speaker variability).

V. EKSPERIMENT

A. Baza

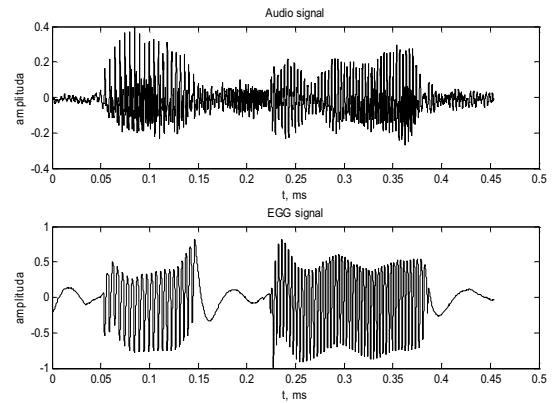
Baza se sastoji od snimaka govornog signala koji su snimani u kancelarijskim uslovima sa 44100 odmeraka u sekundi. Osim audio signala istovremeno je sniman i EGG signal. Baza se sastoji od 50 uzoraka sa 16

govornika koji su do 4 puta čitali tekst u trajanju od 40sek do 1minut. Tekst koji su govornici čitali, predstavlja dobro odabran skup rečenica sa izbalansiranim prisustvom svih glasova srpskog jezika, [11].

Od 50 govornih uzoraka formira se 50 modela, po jedan za svaki snimljeni uzorak. Svaki govorni uzorak se testira u odnosu na 49 preostalih. Na ovaj način se dobija ukupan broj testova $49 * 50 = 2450$ testova. Za slučaj testiranja greške pogrešnog prihvatanja imamo samo testiranje na uzorcima istog govornika. Takvih testova je mnogo manje jer je i broj uzoraka za istog govornika znatno manji od broja uzoraka za ostale govornike. U našem slučaju je to svega oko 150 testova. Drugim rečima ovaj rezultat ima manju statističku tačnost od prethodnog.

B. Parametrizacija EGG signala

EGG signal predstavlja intenzitet visokofrekvencijske struje između dve sonde postavljene na vrat govornika između kojih se nalaze glasnice. Promene intenziteta struje su posledica promena provodnosti između sonde, a promena provodnosti posledica je vibracija glasnica odnosno njihovog spajanja ili odvajanja u zvučnim delovima govora. Na Sl. 4. prikazan je uporedno audio i EGG signal jednog segmenta govora.



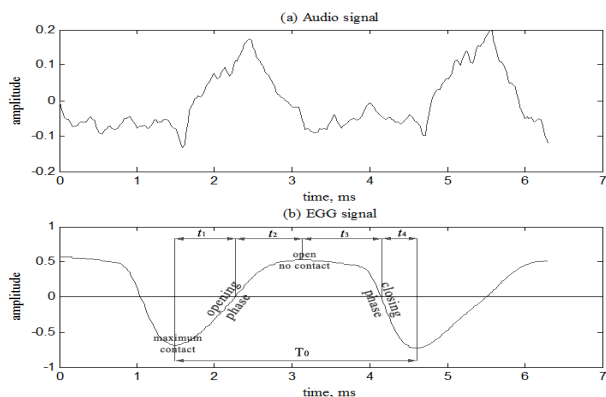
Sl. 4. Vremenski oblik EGG signala u odnosu na audio signal u zvučnim odnosno bezzvučnim segmentima

Perioda EGG signala zvučnog segmenta jednog govornika prikazana je na Sl. 5. Na slici su označene značajne tačke EGG signala: minimum, prolasci kroz nulu i maksimum. Ove tačke definišu specifična rastojanja kojima se može opisati EGG signal i ta rastojanja će se koristiti kao parametri za verifikaciju govornika, [7].

Korišćeno je 5 vremenskih parametara: perioda EGG signala, rastojanje do prvog preseka sa nulom, do maksimuma, do drugog preseka sa nulom i do kraja periode. Poslednje 4 vrednosti se koriste kao relativne u odnosu na periodu tj. u odnosu na prvi parametar. Ovi parametri su korišćeni za formiranje modela a zatim merenje inter odnosno intra Kullback-Leibler divergencije.

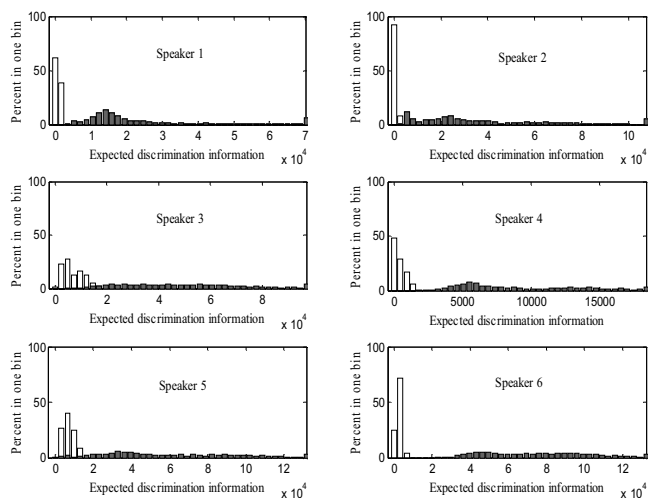
Rezultati su prikazani u obliku histograma za svaki govorni uzorak i za svakog govornika posebno. Pri tome je korišćena bela boja za rezultate na istom govorniku

odnosno siva za različite govornike.



Sl. 5. Perioda EGG signala i karakteristični vremenski parametri

Na Sl. 6 prikazani su rezultati za šest govornika pojedinačno.



Sl. 6. Očekivana diskriminacija za 6 govornika:

- (a) kada modeli pripadaju istom govorniku (beli binovi)
- (b) kada modeli pripadaju različitim govornicima (sivi binovi)

Sa Sl. 6 se jasno vidi da skoro ne postoji preklapanje divergencija između različitih grupa: $KLD_{intra}(i||j)$ odnosno $KLD_{inter}(i||j)$. Za sve govornike iz baze, $\max(KLD_{intra}(i||j)) < KLD_{inter}(i||j)$ je tačno za 94.2%. Na osnovu ovog rezultata, može se zaključiti da EGG parametri imaju dovoljnu informativnost za različite govornike koja opravdava dalja istraživanja u pogledu kreiranja optimalnog sistema za verifikaciju koji bi koristio ove parametre.

VI. ZAKLJUČAK

U ovom radu su analizirana diskriminativna svojstva EGG parametara radi verifikovanja njihove informativnosti i potencijalnog doprinosa u sistemima za

verifikaciju govornika. Postupak je prilagođen multimodalnim sistemima za verifikaciju nezavisno od teksta koji su zasnovani na GMM modelima. Opisani metod daje meru različitosti gustina smeše Gausovih raspodela EGG parametara za istog govornika u odnosu na različite govornike, primenom Kullback-Leibler divergencije. Dobijeni rezultati pokazuju da korišćeni EGG parametri imaju izražena diskriminativna svojstva i nagoveštavaju mogućnost njihove primene u sistemima za verifikaciju govornika radi povećanja tačnosti.

LITERATURA

- [1] Available: <http://en.wikipedia.org/wiki/Biometric>
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Z. Ćirović, M. Milosavljević, and Z. 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
- [8] 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.
- [9] Cassidy, Steve (2001). Slp806: speech recognition. Available: www.ling.mq.edu.au/units/slp806/unit_notes/index.html.
- [10] 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.
- [11] S.T. Jovičić, "Serbian emotional speech database: design, processing and evaluation"; *SPECOM-04*, 2004, St.Peter, Russia.

ABSTRACT

This paper presents a method for measure of discriminatory properties for a set of parameters in speaker verification proces based on mixtures of gaussian probabilities. The method is applied for analysis of EGG parameters and presents the proof for future research.

DISCRIMINATORY PROPERTIES OF EGG PARAMETERS IN SPEAKER VERIFICATION SYSTEMS BASED ON GMM MODELS

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