

ZLATKO KOVAČ\*

## Prepoznavanje osobe u forenzici na temelju glasa

### *Sažetak*

*Članak opisuje analizu ljudskog glasa sa svrhom identifikacije osobe po njezinom glasu, odnosno govoru. Navedene su osnovne značajke govora, načini ispitivanja ljudskog glasa, osnovne komponente glasa kao i računalni alati za obradu glasa. Cilj je članka potaknuti interes za razvoj vještačenja zvuka u Ministarstvu unutarnjih poslova RH.*

***Ključne riječi:** glas, spektrogram, formanti, osnovna frekvencija glasa, Fourierova analiza, Bayesian model, audioforenzika.*

### UVOD

Glas određene osobe pokazuje mnoge elemente posebnosti odnosno jedinstvenosti pa se može, poput npr. otisaka prstiju, koristiti za identifikaciju ljudi. Svima je poznato da veoma lagano prepoznamo, s velikom točnošću, glas odnosno govor poznatih nam osoba pa i kad ga čujemo samo na trenutak. Ova činjenica bila je osnova za razvoj metoda identifikacije počinitelja kaznenog djela na temelju prepoznavanja njegovog glasa.

Počelo se s identifikacijom govornika – osumnjičenika uz pomoć svjedoka koji je čuo glas počinitelja. Prepoznavanje po glasu izvodilo bi se slično kao i vizualno prepoznavanje – svjedok bi slušao glas osumnjičenika u seriji s određenim brojem sličnih glasova drugih, neutralnih osoba, pokušavajući prepoznati glas počinitelja (engl. *voice line up*).

Ovdje valja odmah istaknuti kako postoje ozbiljna ograničenja kod dokaza baziranih na govoru. Naime, glavni je problem u tome kako dva ista govorna izričaja iste osobe, izgovorena npr. jedan za drugim, nisu jednaka. Glas je definiran kombinacijom varijabli koje, svaka za sebe imaju određeni stupanj slobode. Govor je kombinacija precizno kontroliranih pokreta organa za govor, od kojih neki, npr. jezik i usne, pokazuju veliku pokretljivost. Artikulacija neke riječi ili (dijela) rečenice neće uvijek biti ista. Isto tako glas može odstupati od svog "normalnog" i kao rezultat čimbenika i emocionalnih stanja koji

---

\* Zlatko Kovač, dipl. ing. elektrotehnike – voditelj programa specijalizacije u Odjelu za stručno usavršavanje i specijalizaciju Policijske akademije MUP-a RH.

su izvan kontrole govornika kao što je npr. prehlada, stres, panika, strah, ljutnja, umor itd. Tako će npr. prehladen čovjek govoriti promuklo, "kroz nos" ili će mu glas zvučati dublji nego obično. Nadalje, čovjek može svjesno utjecati na svoj govor. U slučaju da se želi npr. prikriti, govorit će šapatom ili povišenim glasom ("*u falsetu*"). Čovjek može oponašati glas druge osobe. Intonacija govora bit će različita kod npr. spontane konverzacije u odnosu na formalno čitanje nekog teksta. Ukratko govor je prilično varijabilna veličina te je za njegovo proučavanje potrebna posebna stručnost, pažnja i ozbiljnost.

## 1. VRSTE ISPITIVANJA GLASA

Identifikacijom govornika mogu se baviti stručnjaci koji imaju primarno akademsko obrazovanje u području fonetike, lingvistike, patologije glasa i govora (logopedije) i/ili (elektro)akustike te stručno iskustvo i znanje u forenzičnoj analizi glasa, govora i jezika. Audioforenzičari počinju kombinirati slušnu i spektrografijsku metodu za identifikaciju glasa. Pri tome se koriste i različita druga akustička mjerenja te se uspoređuju rezultati kako bi se povećala točnost identifikacije. Tako se raspoznavanje glasa (engl. *voiceprint identification*) definira kao kombinacija zvučne (slušne) i spektrografijske (mjerne) usporedbe jednog ili više poznatih glasova s nepoznatim glasom u svrhu identifikacije ili eliminacije<sup>1</sup>.

Slušno ispitivanje mora procjenjivati one elemente govora koji su karakteristični za pojedinu osobu, a to su visina i boja glasa, naglasak, dijalekt, brzina i prekidi u govoru, način izgovaranja glasova (artikulacija), kvaliteta glasa, karakteristične riječi (poštapalice) i moguće nepravilnosti u govoru. Valja paziti koji su uzorci za slušno ispitivanje odabrani – govor spontane konverzacije ili govor čitanog teksta. Naime, o ovome ovisi visina glasa i ukupna melodioznost ispitivanog uzorka. Slušna se identifikacija govornika provodi po protokolu forenzične fonetike koji propisuje kvalifikacije ispitivača, elemente ispitivanja i način prikazivanja rezultata (Varošanec-Škarić, 2008).

Spektrografijsko ispitivanje je metoda prikaza ljudskog glasa na zaslonu računala ili specijalnih instrumenata – (akustičkih) analizatora zvuka. Ideja razvoja spektrometrije bila je da se dobije slika glasa neke osobe – spektrogram (engl. *voiceprint*) kao što se može dobiti sliku otisaka papilarnih linija te osobe (engl. *fingerprint*). Na taj način moglo bi se identificirati počinitelja kaznenog djela jednostavnim uspoređivanjem zvučne slike nepoznate osobe (počinitelja) sa zvučnim slikama iz baze podataka. Međutim, ovakva identifikacija danas još nije dovoljno pouzdana jer je ljudski glas promjenjiva veličina, pa se njegovo vještačenje, nakon složene akustičko fonetsko statističke analize, izražava samo određenim stupnjem vjerojatnosti i pouzdanosti identifikacije odnosno eliminacije.

Spektrometrija je bazirana na Fourierovoj transformaciji zvučnog zapisa, što je u biti prijelaz iz vremenske domene zvuka u frekvencijsku domenu. Naime, svaki se periodički akustički signal daje rastaviti na svoje frekvencijske komponente – harmonike koje su višekratnici osnovne frekvencije – osnovnog tona analiziranog zvuka. Na taj način moguće je analizirati zvuk na temelju raspodjele njegovih frekvencijskih komponenta s pripadajućim amplitudama (jačinama). Za spektrografijsku analizu govora, kao i ostalu

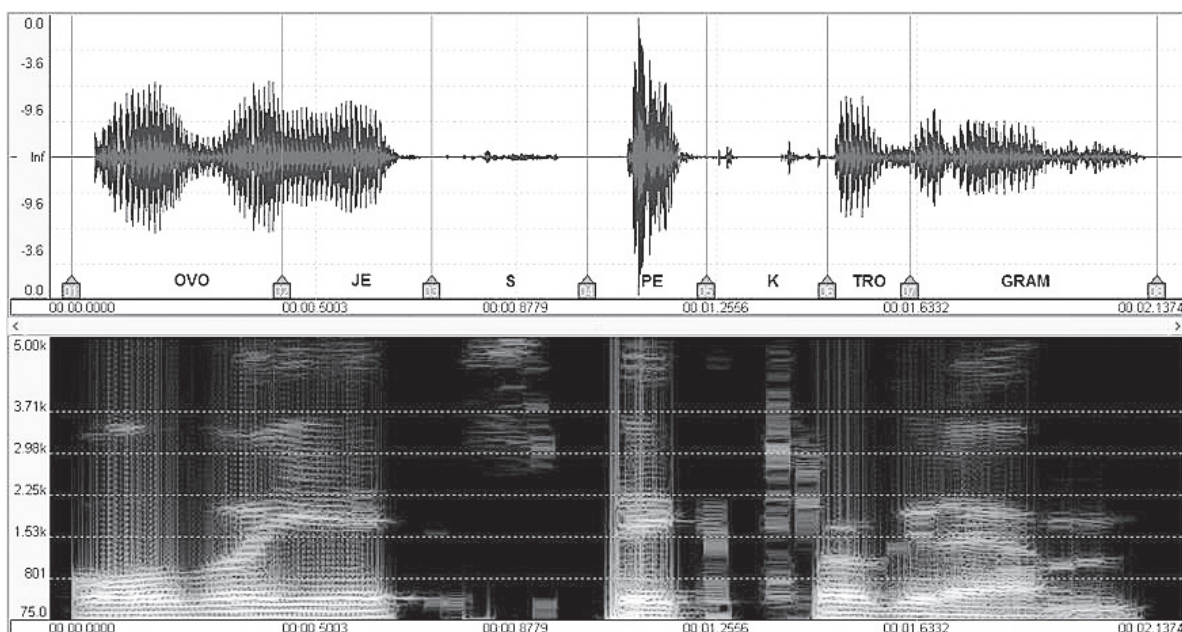
---

<sup>1</sup> [http://www.forensictapeexpert.com/published/voice\\_print.htm](http://www.forensictapeexpert.com/published/voice_print.htm) - 17. 10. 2010.

obradu audiosignala, u današnje se vrijeme najčešće koriste osobna računala (PC) s prikladnom programskom podrškom. Jedan od takvih programa je i DC Live Forensics 7.5 Audio Laboratory, tvrtke Diamond Cut Production Inc. U donjem dijelu slike 1 prikazan je spektrogram dobiven navedenim programom, za primjer snimke autorova izgovora rečenice: "Ovo je spektrogram."

## 2. SPEKTROGRAM, FORMANTI I OSNOVNA FREKVENCIJA LJUDSKOG GLASA

Spektrogram je dijagram koji prikazuje tri veličine govornog signala. Na x-osi nalazi se vrijeme trajanja govornog signala u sekundama, y-os sadrži frekvencijske komponente govora u hercima (Hz), a nijansa sive površine na dijagramu, od crne – preko tonova sive – do bijele boje, mjera je jačine frekvencijskih komponenata govornog signala u određenom trenutku. To znači da je dinamika jačine, od najmanje do najveće, prikazana prijelazom od crnog, preko nijansa sivog do bijelog. To znači da bijela područja u dijagramu predstavljaju one frekvencije u govoru koje su najizraženije – energetski najjače. Gornji dio slike 1 prikazuje audiosnimku izgovorene rečenice u standardnom prikazu kako ga daju audioprocorski programi gdje je, opet, na x-osi vrijeme u sekundama, a na y-osi amplituda (jačina, glasnoća) snimljenog glasa u decibelima.



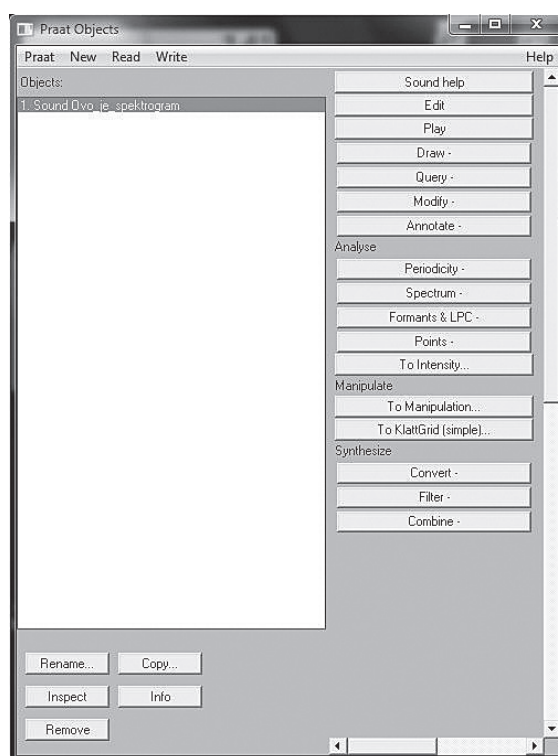
Slika 1: Vremenski dijagram i spektrogram izgovora rečenice: "Ovo je spektrogram."

Navedene svijetle površine na dijagramu predstavljaju formante ljudskog glasa. Formanti odnosno frekvencije formanata pokazatelji su rezonancije (energetskog pojačanja) vokalnog trakta (grlo, usna i nosna šupljina) za vrijeme govora. To su frekvencije kod kojih zračna struja, koja prolazi vokalnim traktom, ima najveću jačinu. Formanti se pojavljuju kod izgovaranja vokala (a, e, i, o, u) i nazala (m, n), i dobar su element za identifikaciju govornika jer odražavaju nepromjenjivu dimenziju govornikova vokalnog trakta kao i način kako artikulira svaki glas. Za fonetsku analizu najvažnija su prva tri

formanta – F1, F2 i F3, kao i relativan razmak među njima. Frekvencijsko područje formanta približno je od 250 Hz do 2,5 kHz<sup>2</sup>. Žene i djeca imaju više iznose frekvencije formanta u odnosu na muškarce. Ljudski glas ima i svoju osnovnu frekvenciju – F0, koja nam određuje visinu glasa (engl. *pitch*). Za odrasle muške osobe ona je približno 125 Hz, dok je za žene približno 250 Hz<sup>3</sup>.

### 3. PRAAT – ALAT ZA ANALIZU LJUDSKOG GLASA

U programu DC Live Forensics 7.5 moguće je vidjeti samo grubu sliku formanta na spektrogramu jer ovaj softver i ne služi za detaljne fonetske analize. Program je primarno namijenjen za čišćenje audiosnimki od ometajućih šumova te poboljšanje razumljivosti govora kada se radi o lošim tonskim snimkama. Za tu svrhu daleko moćniji je poznati program PRAAT<sup>4</sup> – alat široko primjenjiv na sva područja koja se bave zvukom, glasom i jezikom. Slika 2 prikazuje početni (glavni) ekran PRAAT-a.



Slika 2: Glavni ekran programa PRAAT

Već iz samog izbornika na desnoj strani polaznog prozora, bez dubljeg ulaženja u opis programa, vidljive su široke mogućnosti ovog alata. Program, između ostalog, obrađuje učitani audiosignal u smislu filtriranja (potiskivanja) ili isticanja pojedinog frekvencijskog područja, označuje segmente audiozapisa koji su dijelovi rečenica, riječi, slogova ili samih glasova. Nadalje, program analizira valni oblik zvuka po frekvenciji,

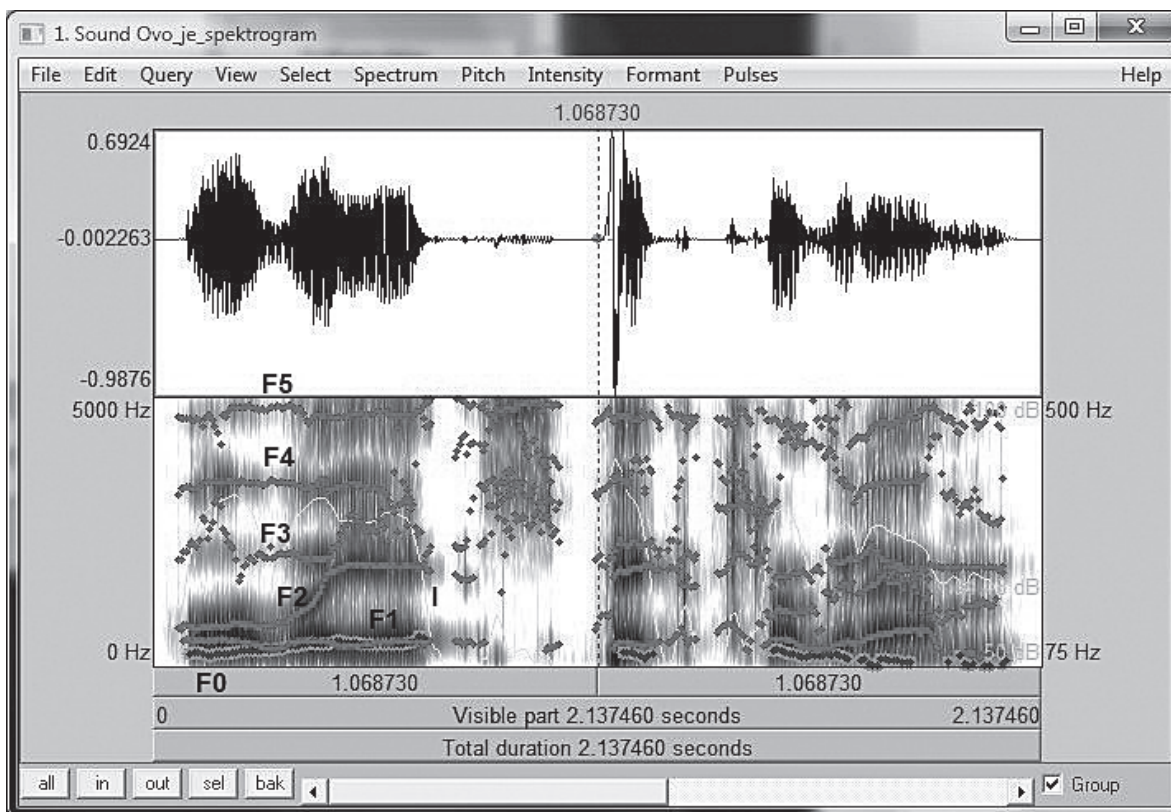
<sup>2</sup> <http://www.scribd.com/doc/53614464/Forensic-Voice-Identification> - 17. 10. 2010.

<sup>3</sup> <http://www.cslu.ogi.edu/tutordemos/SpectrogramReading/ipa/formants.html> - 17. 10. 2010.

<sup>4</sup> <http://www.fon.hum.uva.nl/praat/> - 17. 10. 2010.

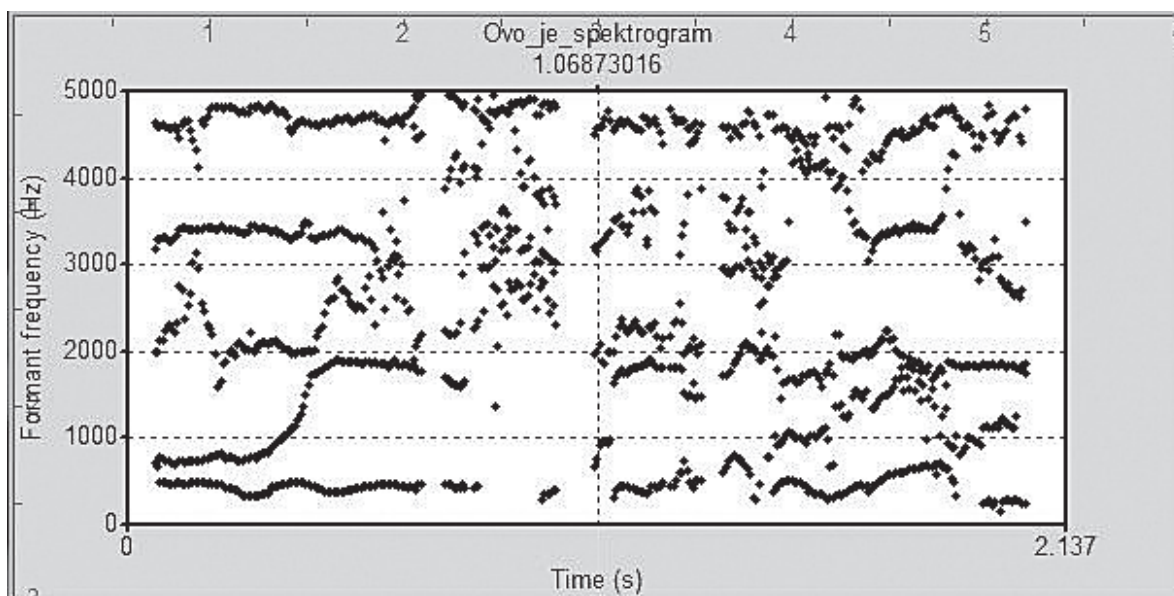


amplitudi, trajanju, snazi, prekidima u signalu, udjelu šuma u signalu itd. On izračunava i prikazuje, u tabelarnom i grafičkom obliku, intenzitet zvuka, spektrogram, visinu glasa-zvuka, trajanja i prekide različitih dijelova zvuka, glasa ili govora, podrhtavanja u frekvenciji i amplitudi glasa (engl. *jitter* i *shimmer*), a isto tako može obrađivati i audiozapise dugačkog trajanja. PRAAT se inicijalno otvara u dva prozora – prozor objekata i prozor slika. Svaki ulazni element analize, kao i rezultat proračuna, program tretira kao objekt i smješta ga u prozor objekata, dok se rezultati proračuna grafički prikazuju u prozoru slika ili tabelarno u posebnim prozorima.



Slika 3: Parametri glasa (vidi dalje tekst)

Samo radi ilustracije ozbiljnosti ovog programa, na slici 3 prikazana je obrada istog audiozapisa koji je korišten za ilustraciju spektrograma prikazanog u DC Live Forensics 7.5 (slika 1). Ovdje su u sivom spektru vidljiva područja formanata (crno – područja najveće energetske zastupljenosti, bijelo – područja najmanje energetske zastupljenosti), krivuljama F1-F5 su iscrtani njihovi tokovi (formanti od F1 do F5), F0 je krivulja osnovne frekvencije glasa (visina glasa), krivulja I prikazuje kretanje intenziteta glasa, a desno su dane i brojčane vrijednosti dinamičkog područja glasa u decibelima. Navedene krivulje ovdje su označene slovima zbog crno-bijelog ispisa, što je donekle nepregledno. Međutim, sam program iscrtava sve ovo u različitim bojama i preglednost zadovoljava. Krivulje formanata mogu se posebno iscrtati, što je prikazano na slici 4. Nadalje, u novom prozoru mogu se iscrtati formanti u drugom obliku – kao frekvencijsko-amplitudni dijagram. Isto tako se u drugom prozoru može izlučiti vremensko-frekvencijski dijagram osnovne frekvencije glasa F0 ili dijagram intenziteta glasa.



Slika 4: Krivulje formanata

PRAAT ima poseban modul za sintezu zvuka kojim se mogu generirati (simulirati) ljudski glasovi – vokali npr., prikladne matematičko-statističke alate, alate za crtanje osnovnih geometrijskih oblika, mogućnost programiranja i automatiziranja određenih postupaka uz pomoć svog programskog jezika (*PRAAT script*), pisanje različitih tekstualnih oznaka vezanih uz audiozapis u trakama – slojevima (engl. *tiers*) i tekst gridu i još mnogo opcija prilagođenih za specifične znanstvene discipline (npr. *Neural nets*).

Osim PRAAT-a, ovdje valja spomenuti još jedan programski paket – programi tvrtke KAYPENTAX. Radi se o dva programa (s dvije inačice: *standard* i *advanced*):

1. CSL (Computer Speech Lab – hardverska verzija) i Multi-Speech (verzija kao računalni softver)
2. MDVP (Multi-Dimensional Voice Program).

U svijetu se ovi programi često rabe za forenzične analize glasa i govora, a izračunavaju znatno više akustičkih parametara glasa i govora od PRAAT-a.

#### 4. NOVIJA ISTRAŽIVANJA

Novija istraživanja na području audioforenzike idu u smjeru potpune automatizacije analize govora. Novije su tendencije da se forenzična identifikacija govornika temelji na kombinaciji rezultata iz slušne analize i akustičke analize koja koristi Bayesian model za procjenu pouzdanosti dokaza<sup>5</sup>. Naime, umjesto da se daje izvješće o stupnju vjerojatnosti da sporni uzorak govora potječe od osumnjičenika, od vještaka se sada traži da procijeni vjerojatnost dokaza u svjetlu dviju suprotstavljenih pretpostavki. Jedna je hipoteza tužitelja: pretpostavka da sporni materijal potječe od osumnjičenika. Druga će hipoteza općenito biti da dokazni materijal potječe od nekog drugog iz moguće suspektne populacije, npr.

<sup>5</sup> <http://pdffinder.net/Forensic-audio-and-visual-evidence-2004-2007:-A-Review.html> - 17. 10. 2010.

populacije odraslih muških osoba iz grada ili regije. Omjer između tih dviju mogućnosti naziva se omjer vjerojatnosti (engl. *Likelihood Ratio*) i izražava se numeričkom vrijednošću. Ono što omjer vjerojatnosti zaista izražava jest relativna snaga dokaza, tj. dodatni element za tužilaštvo koji njihovu poziciju čini jačom ili slabijom<sup>6</sup>.

Bayesian model još je samo teoretski i matematički algoritam, koji tek treba testirati u stvarnoj praksi. Parametri koje bi ovaj model trebao uzeti u obzir nisu još u potpunosti definirani, kao niti način kako te parametre izmjeriti. Za sada, bez ljudskog nadzora (odabira parametara), nije pouzdano prepustiti odluku o identifikaciji samo matematičkom algoritmu. Lako bi moglo doći do pogrešnog prepoznavanja ili odbijanja u identifikaciji.

Također se dosta koriste rezultati iz druge, paralelne grane istraživanja glasa (engl. *Speaker Recognition* ili *Speaker Verification*) koja se razvijala za komercijalne svrhe – identifikaciju i autorizaciju osoba na temelju njihova glasa kao *passworda* za ulaze u objekte (npr. banke, trezore, vojne objekte itd.) ili pristup različitim računima u elektroničkim transakcijama. Budući da se u ovim automatiziranim sustavima radi o potpuno objektivnom procesu identifikacije, rezultat se izračunava uz pomoć statističkih algoritama i izražava kroz vjerojatnost poklapanja određenih parametara ispitivanog glasa s parametrima referentnih glasova iz glasovne baze. Zato takvi sustavi moraju imati prag prihvaćanja (ili odbijanja) osobe koju prepoznaju, a taj prag je optimum pouzdanosti sustava tj. kompromis između broja pogrešnog prepoznavanja ("provuče" im se nepoznata osoba kao prepoznata) i lažnog alarma (ne prepoznaju poznatu osobu). Algoritmi iz tog područja prilagođuju se i ugrađuju u programe za audioforenziku.

## 5. ZAKLJUČAK

Ovim tekstom dan je kratak pregled osnovnih postupaka kojima se audioforenzika služi u prepoznavanju ljudskog glasa u svrhu još jednog načina identifikacije osoba za kriminalističko procesnu i/ili sudsku praksu. Prepoznavanje glasa samo je jedna od mnogobrojnih drugih zadataka ove forenzične discipline. Tu se još mogu nabrojati npr. prepoznavanje različitih drugih zvukova osim ljudskog glasa, analiza čujnosti nekog akustičkog događaja, poboljšanja razumljivosti zvučnih zapisa na različitim medijima, analiza autentičnosti audiosnimaka ili audiouređaja, dekodiranje dijaloga (za transkribiranje), vještačenja mobilnih komunikacija, nadzornih kamera kao i vještačenja raznih drugih zvukova kao npr. pucnja, eksplozija, škripe kotača, zvukova određenih strojeva, šumova na prometnicama ili šumova određenih otvorenih ili zatvorenih prostora, vještačenja odjeka i drugih akustičkih karakteristika dvorana, crkava i drugih građevina (arhitektonska akustika) itd.

Cilj je ovog članka dati uvid u područje audioforenzike koja se bavi identifikacijom ljudskog glasa odnosno govora. Ova je disciplina važna i radi toga što je već mnogo puta u kriminalističkoj i sudskoj praksi doživljeno da je zapis glasa ili govora jedne ili više osoba, nažalost, ostao kao jedini trag i dokaz počinjenog kaznenog djela. Identifikaciju osobe po glasu neki sudovi u SAD-u i drugim državama u svijetu prihvaćaju kao dokaz, neki uopće ne prihvaćaju, a neki prihvaćaju djelomično i uvjetovano, s određenim stup-

---

<sup>6</sup> [http://www24.us.archive.org/stream/ForensicSpeechAndAudioAnalysis/ForensicSpeechAndAudioAnalysisForensicLinguistics-AReview\\_djvu.txt](http://www24.us.archive.org/stream/ForensicSpeechAndAudioAnalysis/ForensicSpeechAndAudioAnalysisForensicLinguistics-AReview_djvu.txt) - 17. 10. 2010.

njem dokazne snage. U Hrvatskoj, ovu materiju, u najširem smislu, spominje novi Zakon o kaznenom postupku u članku 301. stavku 1., koji definira razna prepoznavanja, pa tako i prepoznavanje zvuka općenito (što znači i ljudskog glasa). Dalje od ovoga, koliko je autoru poznato, u MUP-u RH se ne ide.

Bez obzira na sudstvo, policija ovom forenzičnom disciplinom dobiva moćno oruđe u svojim istražnim radnjama. Stoga je želja autora da ovim člankom, kao i člankom Zvuk u kriminalističkom istraživanju (Radmilović, Kovač, 2010), potakne interes za razvoj audioforenzike u MUP-u RH. Opseg članaka prilagođen je također tom cilju. Na kraju je priložena i opširnija literatura (najviše s interneta) za sve one koje zanima ova materija te se žele s njom detaljnije upoznati.

## LITERATURA

1. Olsson, J. (2010). *Forenzička lingvistika*. Zagreb: Nakladni zavod Globus.
2. Heđever, M. (2009). *Uvod u forenzičnu fonetiku i akustiku*. Skripta iz kolegija Forenzična fonetika i akustika na Edukacijsko-rehabilitacijskom fakultetu u Zagrebu.
3. Jelaković, T. (1978). *Zvuk, sluh, arhitektonska akustika*. Zagreb: Školska knjiga.
4. Radmilović, Ž., Kovač, Z. (2010). *O nekim mogućnostima korištenja zvuka u kriminalističkom istraživanju*. *Policija i sigurnost*, 19(4), 403.-416.
5. Varošaneć-Škarić, G. (2008). *Verifikacija govornika u forenzičnoj fonetici*. *Govor – časopis za fonetiku*, 25(1), 31.-44.
6. [http://www.taracentar.hr/attachments/article/24/forenzika\\_skripta.pdf](http://www.taracentar.hr/attachments/article/24/forenzika_skripta.pdf) - 17. 10. 2010.
7. [http://expertpages.com/news/voiceprint\\_identification.htm](http://expertpages.com/news/voiceprint_identification.htm) - 17. 10. 2010.
8. <http://www.scribd.com/doc/53614464/Forensic-Voice-Identification> - 17. 10. 2010.
9. <http://www.enhancedaudio.com/livedvd.htm> - 17. 10. 2010.
10. <http://www.intechopen.com/books/authors/recent-advances-in-signal-processing/information-mining-from-speech-signal> - 17. 10. 2010.
11. <http://www.research-horizons.cam.ac.uk/features/can-a-voice-identify-a-criminal.aspx> - 17. 10. 2010.
12. <http://www.icphs2007.de/conference/Papers/1047/1047.pdf> - 17. 10. 2010.
13. [http://www.isca-speech.org/archive\\_open/odyssey\\_2010/od10\\_012.html](http://www.isca-speech.org/archive_open/odyssey_2010/od10_012.html) - 17. 10. 2010.
14. <http://www.tandfonline.com/doi/abs/10.1080/00450618.2011.630412> - 17. 10. 2010.
15. [http://www24.us.archive.org/stream/ForensicSpeechAndAudioAnalysis/ForensicSpeechAndAudioAnalysisForensicLinguistics-AReview\\_djvu.txt](http://www24.us.archive.org/stream/ForensicSpeechAndAudioAnalysis/ForensicSpeechAndAudioAnalysisForensicLinguistics-AReview_djvu.txt) - 17. 10. 2010.
16. <http://pdffinder.net/Forensic-audio-and-visual-evidence-2004-2007:-A-Review.html> - 17. 10. 2010.



Summary

---

**Zlatko Kovač**

**Forensic Speech Recognition**

The paper deals with human voice analysis for the purpose of person's identification on the basis of his/her voice or speech. The basic characteristics of speech are introduced, as well as the methods of speech analysis, basic components of speech and the relevant software. The aim of the paper is to raise interest for the development of audio forensics at the Ministry of the interior of the Republic of Croatia.

**Key words:** voice, spectrogram, formants, pitch, Fourier analysis, Bayesian method, audio forensics.