

AUTOMATINIO ASMENS ATPAŽINIMO IŠ BALSŲ PROBLEMOS IR PERSPEKTYVOS KRIMINALISTIKOJE

Dr. Bernardas Šalna, Juozas Kamarauskas

Lietuvos teismo ekspertizės centras
Lvovo g. 19a, LT-09313 Vilnius
Telefonas 263 85 50
Elektroninis paštas bernardas@centras.lt

Pagrindinės sąvokos: asmens atpažinimas iš balsų, kriminalistinis asmens atpažinimas iš balsų, automatinės atpažinimo sistemos.

S a n t r a u k a

Straipsnyje apžvelgti pagrindiniai metodai, šiuo metu taikomi automatiniam asmens atpažinimui iš balsų, ir atlikta teismo asmens identifikavimo iš balsų ir automatinio asmens identifikavimo sistemų analizė. Nagrinėjamos naujų automatinių asmens atpažinimo sistemų taikymo perspektyvos teismo ekspertizėje ir kriminalinėje paieškoje bei operatyviniame darbe. Pateikiami automatinio asmens identifikavimo rezultatai, naudojant balsų bazę. Gautų rezultatų pagrindu daroma išvada – automatinio asmens atpažinimo iš balsų sistema yra gana patogi vartotojui, nes ji iš vartotojo reikalauja minimalių žinių. Tokia sistema gali būti taikoma vykdant kriminalinę paiešką, tačiau atliekant teismo ekspertizę jos taikymas yra ribotas.

1. Įvadas

Asmens atpažinimo iš balsų technologijos kriminalistikoje pradėtos taikyti palyginti ne taip seniai. Pirmoji fonoskopinė ekspertizė, teisme priimta kaip įkaltis, buvo atlikta JAV 1960 metais, o Europoje – 1972 metais atlikta Lenkijoje ir šios rūšies ekspertizės imtos vadinti fonoskopinėmis. Nors kariškiai ir specialiosios tarnybos asmens atpažinimo iš balsų technologijas naudojo jau gerokai anksčiau, tačiau tai buvo tik pavieniai atvejai ir tik nuo 1970 metų, kai garso įrašymo technika ir kalbinės technologijos pasiekė atitinkamą technologinį lygį, asmens atpažinimas iš balsų buvo pradėtas taikyti kriminalistikoje. Šiuo metu kovoje su terorizmu asmens atpažinimo iš balsų technologijos tampa labai aktualios daugelyje Europos valstybių ir ypač JAV, kur šioms technologijoms skiriama labai daug dėmesio, materialiniai ir intelektualiniai išteklių. Platus šių technologijų taikymas susiduria su daugybe problemų, pavyzdžiui, dažniausiai labai skiriasi tiriamasis ir lyginamasis garso įrašai. Šis skirtumas atsiranda dėl skirtingų garso įrašų darymo sąlygų, skirtingų asmens emocinių būsenų, triukšmo įtakos, garso įrašymo kanalų nesutapimo ir t. t. [1, p. 1–26; 2].

2. Asmens atpažinimas iš balsų kriminalistikoje

Asmens atpažinimo iš balsų technologijas, kurios atitinkamai yra susijusios su kalbinėmis technologijomis, pagal taikymo arba panaudojimo pobūdį reikėtų skirstyti į dvi sritis:

1. Kriminalinė paieška arba operatyvinis darbas.
2. Teismo kriminalistika arba teismo ekspertizė (*speaker identification by experts*).

Balsas yra viena iš žmogaus biometrinių charakteristikų, todėl atpažinimas iš balsų taikomas ne tik kriminalistikoje, bet ir naudojant bankines technologijas, įėjimo kontrolės punktuose, duomenų bazėms apsaugoti ir t. t. [3].

Kalbos signalų moksliniuose tyrimuose šio dalijimo nėra, nes moksliniai rezultatai sėkmingai taikomi tiek naudojant technologijas, kurios skirtos operatyviam darbui, tiek atliekant teismo eks-

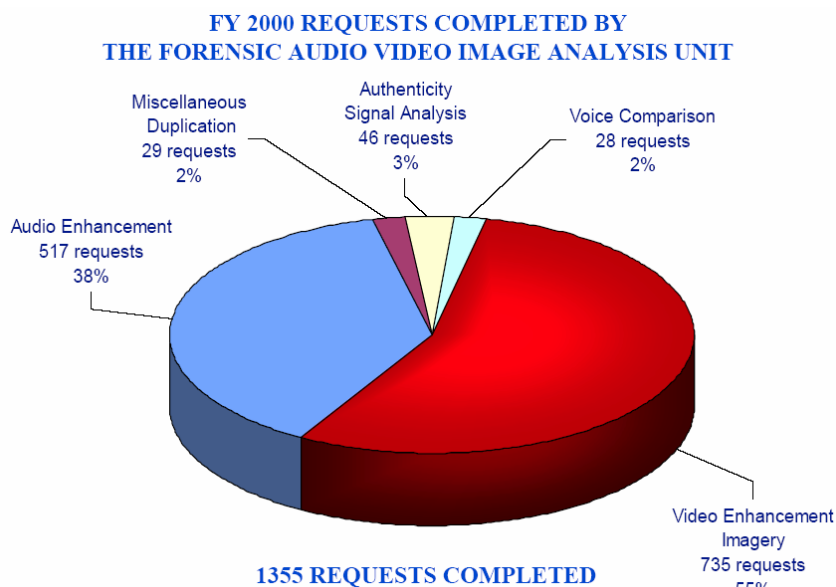
pertizę ar kitose su kriminalistika nesusijusiose srityse. Tačiau čia vis dėlto yra keletas esminių skirtumų, susijusių su pačiu asmens atpažinimo iš balso technologijų naudojimu.

Visų pirma kriminalinei paieškai reikalingi patogūs darbu technologiniai sprendimai, nereikalaujantys iš vartotojo ypatingų žinių ir gebėjimų. Operatyvinio darbuotojo arba kriminalisto techniko uždaviniai yra kokybiškai padaryti garso (vaizdo) įrašus, užtikrinti jų autentiškumą, naudojant specialiąsias technologijas (skaitmeninių vandens ženklų), greitai rasti reikiamus įrašus balsų bazėje arba automatiškai būdu sulygtinti su turimais asmens balso pavyzdžiais, esančiais balsų bazėje (jeigu tokia egzistuoja), pasinaudoti turimomis priemonėmis garso įrašų suprantamumui gerinti ir padaryti garso įrašų stenogramas. O teismo ekspertizės metu atliekamas ekspertinis garso įrašų tyrimas. Ekspertas privalo turėti nuodugnių atitinkamų mokslo sričių žinių, atlikti tyrimą, pasinaudodamas naujausiomis technologijomis bei metodais, ir teismui pateikti šio tyrimo moksliskai pagrįstą ataskaitą (jei reikia atlikti kompleksinį tyrimą), o prireikus apginti savo išvadas teisme. Todėl ekspertinis tyrimas arba ekspertinis asmens identifikavimas iš balso yra daug laiko užimantis procesas.

1 lentelėje pateiktas laikas, kurį sugaišta daugumos Europos šalių fonoskopinių ekspertizų laboratorijų ekspertai vienam asmeniui identifikuoti, kai turime geros kokybės, apytikriai 2 minučių trukmės tiriamąjį garso įrašą. 1 paveiksle pateiktas asmens identifikavimo klausimų skaičius, šį tyrimą atliko JAV FTB fonoskopijos laboratorija per 2001 metus [2; 4].

1 lentelė. Vieno asmens identifikavimo iš balso, užfiksuoto 2 minučių garso įrašė, trukmė

Valstybė	Tyrimo trukmė, val.
1. Belgija	40
2. Italija	40
3. Ispanija	60
4. Lenkija	40
5. Lietuva	28
6. Olandija	40
7. Prancūzija	100
8. Rusija	50
9. Vokietija	40



1 pav. JAV FTB fonoskopijos laboratorijos 2001 metais atliktos fonoskopinės ekspertizės klausimų skaičius

Kaip matome iš šios lentelės, ekspertinis asmens identifikavimas yra labai imlus darbu ir ilgai trunkantis procesas. Todėl šiuo metu pradėta intensyviai plėtoti automatinio asmens atpažinimo iš balso technologijas ir jas taikyti kriminalistikoje.

3. Kriminalistiniai asmens atpažinimo iš balso metodai

Moksliniuose kalbinių technologijų tyrinėjimuose kalbėtojo atpažinimas iš jo balso yra suprantamas kaip procesas, kurio metu iš asmens kalbos signalo išskiriami identifikaciniai požymiai, iš jų ir atpažįstamas konkretus asmuo. Kalbėtojo atpažinimas savo ruožtu dar skirstomas į identifikavimą ir verifikavimą. Identifikavimas gali būti priklausomas arba nepriklausomas nuo teksto. Atliekant teismo ekspertizę asmens atpažinimas dažniausiai yra nepriklausomas nuo teksto ir vadinamas asmens identifikavimu. Pagrindinis bruožas, skiriantis teisminį asmens identifikavimą nuo kitų identifikavimo sistemų, yra tas, kad paprastai turimas „nelinkusio bendrauti“ (*non cooperative speaker*) asmens balsas. Taip pat labai dažnai skiriasi tiriamasis ir lyginamasis garso įrašai. Šis skirtumas susidaro dėl skirtingų garso įrašų darymo sąlygų, skirtingų asmens emocinių būsenų, triukšmo įtakos, garso įrašymo kanalų nesutapimo ir t. t.

Šiuo metu kriminalistikoje naudojamus asmens atpažinimo iš balso metodus sąlyginai galima suskirstyti šitaip [2]:

1. sonografinis (*voiceprints*);
2. fonetinis-akustinis;
3. kombinuotasis (*combined*);
4. automatinis.

3.1. Sonografinis metodas

Sonograma – tai signalo momentinių spektrų seka, pavaizduota trimatėje erdvėje, kurią sudaro laikas, dažnis ir intensyvumas. Sonograma gana gerai atspindi kalbos signalo struktūrą, kadangi jau apie 1960 metus egzistavo specialus prietaisas – sonografas, kuriuo specialiaame popieriuje buvo atvaizduojamos sonogramos, tai ilgą laiką sonografinis metodas buvo vienas pagrindinių asmens atpažinimo iš balso metodų atliekant teisminę ekspertizę. 1979 metais JAV nacionalinės mokslų akademijos Balso identifikavimo ir audioanalizės pakomitetas (VIAAS) paskelbė asmens identifikavimo, naudojant sonogramas, standartą [5]. Šiame standarte nurodyta, kiek žodžių turi sutapti tiriamajame ir lyginamajame garso įrašuose, kad būtų gautos atitinkamos išvados, nustatyti garso atkūrimo aparatūros, eksperto kvalifikacijos reikalavimai ir t. t.

3.2. Fonetinis-akustinis metodas

Fonetinis-akustinis metodas yra tam tikra sonografinio metodo modifikacija. Šio metodo esmę sudaro tai, kad ekspertas iš klausos stengiasi nustatyti tam tikrus artikuliacijos bei tarties ypatumus ir atitinkamai juos atvaizduoti naudodamas sonogramas [6].

3.3. Kombinuotasis metodas

Kombinuotasis metodas, arba dar kartais vadinamas pusiau automatinis (*semiautomatic*) – tai metodas, kurio pagrindą sudaro auditoryvinė kalbos signalo analizė, atitinkanti psichoakustikos dėsnius, ir akustinė analizė. Akustinės analizės metu iš kalbos signalų išskiriami identifikaciniai asmens balso požymiai ir po to atliekami atitinkami statistiniai skaičiavimai. Galutiniame tyrimo etape ekspertas, žinodamas šių požymių fizikinę prasmę, jų statistinio pasiskirstymo ribas ir atsižvelgdamas į auditoryvinės analizės rezultatus, pateikia atitinkamą išvadą. Šiuo metu pasaulinėje ekspertinėje praktikoje naudojamos trys pagrindinės asmens identifikavimo iš balso sistemos, besiremiančios kombinuotuoju metodu: DIALECT (Rusija), IDEM (Italija) ir SIVE (Lietuva) [7]. Kombinuotasis metodas šiuo metu daugiausia taikomas Europoje: Rusijoje, Ispanijoje, Italijoje, Lenkijoje, Lietuvoje, Suomijoje ir kitur.

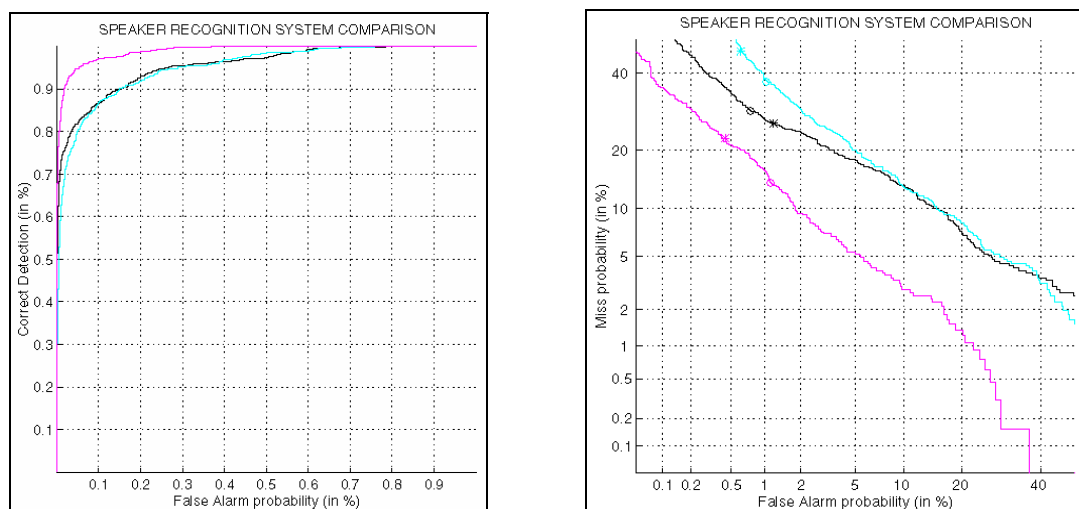
3.4. Automatinis metodas

Automatinis asmens atpažinimo metodas – kai į kompiuterį arba specialų įrenginį įrašomas nežinomo asmens balso garso įrašas ir žinomo asmens balso garso įrašas ir atliekamas tyrimas, kurio rezultatas – šių balsų sutapimo tikimybė arba tikėtinumų santykis (*likelihood ratio*) [8]. Tai yra iš balso garso signalų automatiškai išskiriami identifikaciniai požymiai ir naudojant atitinkamus sprendimų algoritmus gaunamas galutinis rezultatas – balsų sutapimo tikimybė arba atitinkama kreivė [8].

Kol kas šio metodo pagrindinė taikymo sritis – kriminalinė paieška ir operatyvinis darbas. Teisminėje praktikoje (teismo ekspertizėje) automatinis metodas nėra paplitęs, bet jis gana sėkmingai gali būti taikomas kaip sudėtinė kombinuotojo metodo dalis arba komponentė. Taip pat jau šiuo metu asmens verifikavimo iš balso metodai sėkmingai taikomi komerciniais tikslais [3], o juos kuriančios firmos tikisi nemažo pelno. Taigi netolimoje ateityje šie automatinio asmens atpažinimo iš balso algoritmai bei sistemos bus sėkmingai panaudoti ir kriminalistikoje.

Šiuo metu automatinio asmens atpažinimo sistemų pagrindą dažniausiai sudaro vadinamoji mel-kepstrinių požymių sistema (MFCC-39 akustinių požymių vektorius) [9] ir lyginimo algoritmai, veikiantys dinaminio programavimo (DP), paslėptų Markovo grandinių (HMM) ar gausinių mišinių modeliai (GMM) ir dirbtinių neuronų tinklai (ANN) arba įvairios šių algoritmų kombinacijos. Čia reikia pažymėti, kad tie patys požymiai naudojami ir kalbai atpažinti, o tai, mūsų nuomone, yra visai kitas uždavinys. Deja, šiuo metu labai mažai darbų atliekama specialiųjų požymių, priklausančių nuo asmens savybių, paieškai ir dauguma darbų skiriama GMM tobulinti.

Automatinio asmens atpažinimo iš balso sistemų darbingumui vertinti naudojamos vadinamosios priėmimo vykdymo kreivės (PVK) (*receiver operating curves* – ROC ir *detection error trade-off* – DET). Šių kreivių pavyzdžiai pateikti 2 paveiksle.

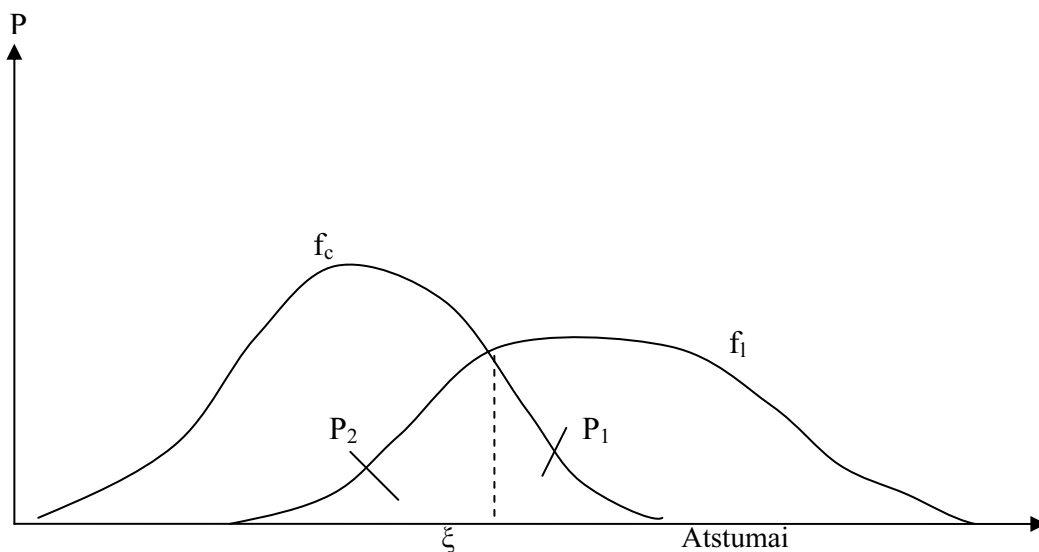


2 pav. Priėmimo ir vykdymo ROC bei DET kreivės

Braižant šias kreives, abscisių ašyje atidedamas klaidingų atmetimų („savojo“ palaikymas „svetimu“) lygis, procentais – FR („*false reject*“, arba kai kada „*false alarm*“), o ordinačių ašyje, braižant ROC kreivę, atidedamas tinkamų priėmimų lygis, procentais – CD (*correct detection*). Braižant DET kreivę, atidedamas klaidingų priėmimų („svetimo“ palaikymas „savu“) – FA („*false accept*“ arba kai kada „*miss probability*“). $FA = CA/M$ ir $FR = CR/N$, kur CA – klaidų skaičius, kai „svetimo“ žmogaus balsas priimamas kaip „savo“, ir M – matavimų skaičius. Atitinkamai CR – klaidų skaičius, kai „savas“ balsas priimamas kaip „svetimas“, ir N – matavimų skaičius. Bendru atveju $M = N$ ir atitinkamai dauginama iš šimto norint išreikšti procentais. Nustatę pasirinktą slenkstį – T (*threshold*), gauname vieną ROC arba DET kreivės tašką. Kreivės DET taškas, kur $FA = FR$, dar vadinamas lygių klaidų tašku – EER ir yra vienas iš pagrindinių sistemos darbingumo savybių. Kuo mažesnis EER, tuo sistema yra geresnė. Taip pat sistemos darbingumui labai didelę reikšmę turi slenkščio parinkimas. Paprastai slenkstis parenkamas atsižvelgiant į vadinamuosius intra- ir interindividualius atstumų pasiskirstymus.

3 paveiksle pateikti atstumų pasiskirstymai: intraindividualūs (kreivė f_c) – tai to paties asmens balso kitimo ribos, o interindividualūs – tai atstumai su kitų asmenų balsais (kreivė f_i). Idealiu atveju šie pasiskirstymai neturėtų susikirsti, t. y. didžiausias atstumas tarp to paties asmens ištartų skirtingų frazių turėtų būti mažesnis už mažiausią atstumą tarp to ir kito asmens ištartų frazių. Deja, visų realiųjų biometrinių sistemų šie pasiskirstymai susikerta, išskyrus akies rainelės požymius. 2004 metais Toledėje (Ispanija) (ODYSSEY 04), tarptautinėje konferencijoje, skirtoje asmeniui identifikuoti iš balso, buvo pristatyti 14 sistemų testavimo rezultatai naudojant kriminalistinę balsų bazę. Tai yra testavimui buvo naudojama 40 asmenų balso įrašai, padaryti naudojant GSM telefoną. Šie įrašai buvo paimti iš Olandijos policijos duomenų bazės. Iš šios balsų bazės buvo padaryta 30, 60, 120 sekundžių trukmės įrašai sistemai parengti ir 7, 15, 30 sekundžių įrašai testavimui. Tai yra 521 įrašas parengimui ir 9676

testavimui [10]. Buvo pastebėta, kad labai mažai pavyko rasti to paties kalbėtojo įrašų, kurių trukmė būtų apie 120 sekundžių. Geriausi rezultatai – EER= 12,1 proc. klaidų, esant 60 s parengimui ir 15 s trukmės testui. Panašūs rezultatai gauti ir naudojant JAV FTB kuriamą sistemą – FASR (*forensic automatic speaker recognition*) [4]. Reikia pabrėžti, kad FTB pavyko sukurti metodą, kuris leidžia pašalinti telefono kanalo įtaką, tai yra per telefono kanalą padarytus garso įrašus lyginti su įrašytais naudojant diktofono/magnetofono mikrofoną. Įvertinus šiuos rezultatus, galima daryti išvadą, kad kol kas tokių sistemų taikymas kriminalistikoje yra ribotas. Geresni rezultatai gaunami, kai balsų lyginimas ir paieška balsų bazėje yra automatizuota, o paties asmens balso pavyzdys paruošiamas naudojant pusiau automatines procedūras. Kaip tokios sistemos pavyzdys galėtų būti TRAL, kuri veikia šitokių būdu: visų pirma operatorius iš turimo garso įrašo suformuoja balso etaloną (*voice cards*) – tam tikrą požymių matricą, kuri saugoma duomenų bazėje. Toliau sistema automatiškai kiekvienam telefono garso įrašui apskaičiuoja identifikacinių požymių matricą ir palygina su visų balsų etalonais, esančiais duomenų bazėje. Po to pagal nustatytą sutapimo slenkstį randami balsai, jei tokių yra balsų bazėje. Vėliau ekspertas jau gali palyginti tiriamąjį ir lyginamąjį garso įrašus, kadangi jie yra kaupiami šioje duomenų bazėje. Šiuo atveju galimas tiek operatyvinis, tiek ekspertinis tyrimas.



3 pav. Intra- ir interindividualūs atstumų pasiskirstymai

4. Išvados ir perspektyvos

Automatinio asmens atpažinimo iš balso sistemos yra gana patogios vartotojui, nes iš vartotojo reikalauja minimalių žinių. Tokios sistemos jau taikomos atliekant kriminalinę paiešką ir operatyviame darbe, tačiau atliekant teismo ekspertizę jų taikymas ribotas. Vienas iš pagrindinių automatinių sistemų darbo trūkumų – sąlyginai mažas atpažinimo tikslumas ir tai, kad sistemai parengti ir testavimui arba pačiam identifikavimui reikalingi sąlyginai ilgi (daugiau kaip 2 min.) garso įrašai. Kriminalistikoje šiuo metu aktualiausia yra trumpų garso įrašų (10–15 s) tyrimas. Nors automatinė sistema ir turi trūkumų, jos metodai gali būti sėkmingai taikomi kombinuotoje sistemoje [7] ir, kaip rodo mūsų bei kitų šalių patirtis, tai leidžia pagreitinti ekspertizės atlikimą bei objektyvizuoti tyrimą. Viena iš perspektyviausių sričių būtų kriminalistikos fonotekos arba balsų bazės, turinčios integruoto automatinio asmens atpažinimo galimybę, kūrimas. Toks centralizuotas garso įrašų tvarkymas leistų ne tik automatiškai atpažinti asmenį iš balso, kas irgi labai svarbu, bet ir greitai surasti reikiamą lyginamą bei tiriamą įrašus, operatyviai apsiukeisti saugomais garso įrašais, užtikrinti tokių garso įrašų autentiškumą ir t. t.



LITERATŪRA

1. **Broeders A. P. A.** Forensic Speech and Audio Analysis. 1998 to 2001 A Review. Proceedings of the 13-th INTERPOL Forensic Science Symposium. – Lyon, France, 16–19 October 2001.
2. **Delgado-Romero C.** La Identificacion de Locutres en el ambito forense, PhD thesis, 2001. Universidad Complutense de Madrid.
3. **Bimbot F. et. al.** An Overview of the PICASSO Project Research Activities in Speaker Verification for Telephone Applications. Proceedings of COST250 Workshop on Speaker Recognition in Telephony. – Rome, Italy, 2001. November 10–12.
4. **Nakasone H., Beck S. D.** Forensic Automatic Speaker Recognition. ESCA Workshop on Automatic Speaker Recognition, Verification. A Speaker Odyssey, Crete. – Greece, 2001.
5. **Voice Comparison Standards** of the VIAAS of the IAI. Vol. 41 // Journal Forensic Identification (September / October 1991).
6. **Kunzel H. J.** Sprechererkennung. Kriminalistik Verlag. – Heidelberg, 1987.
7. **Šalna B.** Criminalistic Person Identification by Voice System // Problems of Forensic Sciences. 2001.
8. **Rose P.** Forensic Speaker Identification. Taylor & Francis. – London and New York, 2002.
9. **Siafarikas M., Ganchev T., Fakotakis N.** Wavelet Packed Based Speaker Verification. A Speaker Odyssey. Proceedings ODYSSEY04. May 31–June 3, 2004. Toledo (Spain).
10. **Van Leeuwen D. A., Bouten J. S.** Results of the 2003 NFI-TNO Forensic Speaker Recognition Evaluation. A Speaker Odyssey. Proceedings ODYSSEY04. May 31–June 3, 2004. Toledo (Spain).



Problems and Prospects of Forensic Automatic Speaker Recognition by Voice

*Dr. Bernardas Šalna, Juozas Kamarauskas
Forensic Science Centre of Lithuania*

Keywords: speaker recognition by voice, forensic speaker recognition by voice, automatic recognition systems.

SUMMARY

Technologies of speaker recognition by voice in criminology were applied recently. Now they become topical in many countries.

Main methods, used in automatic speaker recognition by voice were considered in this article. Analysis of forensic speaker identification by voice and automatic identification system was performed. Outlook of novel automatic speaker recognition systems in forensic inspection, criminal search and strategical work were discussed. Results of automatic speaker identification evaluation, using database were described. Conclusion can be made, that automatic speaker recognition system is convenient, because it requires minimal knowledge from the user. This system can be applied in criminal search, but it has limitations in forensic inspection.

One of the most important lack of automatic speaker recognition systems is less accuracy of recognition and relatively long record length is required for training and testing of the system. In spite of lack of automatic recognition system, some methods that are used in this system can be applied in combined system. And how our and other countries experience shows, it allows to perform research rather and more objective. One of the most promising way could be development of the database, consisting of voice records and integrated automatic recognition system, what would allow to perform not only automatic speaker recognition, but finding test and template records too, and so on.

