BODO PAMETNI NADZORNI SISTEMI PRISLUHNILI, RAZUMELI IN SPREGOVORILI SLOVENSKO?

Simon DOBRIŠEK, Vitomir ŠTRUC, France MIHELIČ
Univerza v Ljubljani, Fakulteta za elektrotehniko

Boštjan VESNICER
Alpineon d. o. o.

Dobrišek, S., Vesnicer, B., Štruc, V., Mihelič, F. (2013): Bodo pametni nadzorni sistemi prisluhnili in spregovorili slovensko? Slovenščina 2.0, 1 (2): 165–180.
URL: http://www.trojina.org/slovenscina2.0/arhiv/2013/2/Slo2.0_2013_2_08.pdf.

Članek obravnava tehnologije govorjenega jezika, ki bi lahko omogočile t. i. pametnim nadzornim sistemom, da bi nekoč prisluhnili, razumeli in spregovorili slovensko. Tovrstni sistemi se z uporabo senzorjev in naprednih računalniških metod umetnega zaznavanja in razpoznavanja vzorcev do neke mere zavedajo okolja ter prisotnosti ljudi in drugih pojavov, ki bi lahko bili predmet varnostnega nadzora. Med tovrstne pojave spada tudi govor, ki lahko predstavlja ključni vir informacije pri določenih varnostnonadzornih okoliščinah. Tehnologije, ki omogočajo samodejno razpoznavanje in tvorjenje govora ter samodejno razpoznavanje govorcev in njihovega psihofizičnega stanja s pomočjo napredne računalniške analize govornega zvočnega signala, odpirajo povsem nove dimenzije razvoja pametnih nadzornih sistemov. Samodejno razpoznavanje varnostno sumljivih govornih izjav, kričanja in klicev na pomoč ter samodejno zaznavanje varnostno sumljivega psihofizičnega stanja govorcev tovrstnim sistemom doda pridih umetne inteligence. Članek predstavlja trenutno stanje razvoja omenjenih tehnologij in možnosti njihove uporabe za slovenski govorjeni jezik ter različne varnostnonadzorne scenarije uporabe tovrstnih sistemov. Naslovljena so tudi širša pravna in etična vprašanja, ki jih odpira razvoj in uporaba tovrstnih tehnologij. Govorni nadzor je namreč eno najbolj občutljivih vprašanj varstva zasebnosti.

Ključne besede: tehnologije govorjenega jezika, pametni nadzorni sistemi, samodejno razpoznavanje govora, tvorjenje umetnega govora, samodejno razpoznavanje govorca

1 UVOD

Izjemen tehnološki napredek v zadnjih desetletjih je omogočil razvoj vedno bolj zapletenih in vseprisotnih nadzornih tehnologij, katerih glavni namen je izboljšanje učinkovitosti varnostno-obveščevalnih služb pri zaznavanju in preprečevanju kriminala in terorizma. Pri sodobnem prizadevanju za zagotavljanje varnosti se pojavlja potreba po prehodu iz retroaktivnega forenzičnega preiskovanja preteklih varnostnih incidentov v proaktivno sprotno odzivanje na samodejno zaznane varnostne incidente in grožnje s pomočjo t. i. inteligentnih oziroma pametnih nazornih tehnologij.

Pametne nadzorne tehnologije so integrirani računalniški sistemi, ki vključujejo tehnologije za zajem raznih senzorskih in drugih nadzornih podatkov ter računalniške postopke za njihovo samodejno obdelavo, ovrednotenje in analizo, kakor tudi postopke za samodejno odločanje oziroma podporo odločanju na osnovi rezultatov analize zbranih podatkov. Ti sistemi predstavljajo tehnološki razvojni napredek v primerjavi s tradicionalnimi nadzornimi sistemi, ki navadno vključujejo le osnovno infrastrukturo za zajemanje, shranjevanje in distribucijo nadzornih podatkov, nalogo zaznavanja oziroma preiskovanja varnostnih incidentov in groženj pa v glavnem še vedno prepuščajo razmeroma neučinkovitim človeškim operaterjem. Tipični tovrstni tradicionalni sistemi so t. i. CCTV videonadzorni sistemi.

Pri novejših CCTV sistemih se danes z metodami računalniške analize video vsebin že poskuša doseči zmožnost samodejnega zaznavanja in razpoznavanja varnostno sumljivih dogodkov, okoliščin ali obnašanja ljudi (Piciarelli in Foresti 2011). Mnogih varnostnih incidentov pa ni mogoče zaznati zgolj z analizo videa. Povsem novo dimenzijo inteligentnega nadzora ponuja integracija videonadzornih sistemov z inteligentnimi avdionadzornimi sistemi, ki omogočajo samodejno varnostno analizo zajetega zvočnega signala (Onut in dr. 2011)

Avdionadzorni sistemi z zmožnostjo tristošestdesetstopinjskega pokrivanja prostora omogočajo razširitev nadzorovanega prostora preko vidnega polja navadnih nadzornih kamer. S samodejnim zaznavanjem in razpoznavanjem varnostno sumljivih zvokov, kot so kričanje v stiski, klicanje na pomoč, glasno

izgovarjanje groženj, hrup razbijanja stekla in drugih predmetov, odmevanje korakov, zvok odpiranja vrat, pok pištole ipd., lahko nadzornemu sistemu dodamo zmožnost samodejnega osredotočanja pozornosti v smeri izvorov teh sumljivih zvokov. Samodejno razpoznani varnostni incidenti bi lahko sprožili ustrezen odziv sistema, kot je samodejni klic policije in reševalnih služb ali opozarjanje in obveščanje prisotnih ljudi o zaznanem varnostnem incidentu s tvorjenjem umetnega govora.

Tehnologija	**Varnostnonadzorna uporaba**
Razpoznavanje govora	Samodejno razpoznavanje izgovorjenih groženj in drugih varnostno sumljivih izjav ter neposrednih in prikritih klicev na pomoč
Razpoznavanje govorcev	Razpoznavanje znanih kriminalcev in varnostno sumljivih posameznikov
Razpoznavanje govorjenega jezika	Razpoznavanje govorjenega jezika govorca za prilagajanje nadzornega sistema njegovim govornim značilnostim
Razpoznavanje psihofizičnega stanja govorca	Razpoznavanje agresivnega in drugače varnostno sumljivega obnašanja ali prestrašenosti ljudi
Umetno tvorjenje govora	Govorno obveščanje prisotnih ljudi o zaznanem varnostnem incidentu

Tabela 1: Pregled možnih varnostnonadzornih uporab tehnologij govorjenega jezika.

Med razvijajočimi se tehnologijami govorjenega jezika je precej takšnih, ki jih je mogoče neposredno uporabiti v inteligentnih avdionadzornih sistemih. Osnovni pregled različnih tovrstnih tehnologij in njenih možnih uporab v različnih varnostnonadzornih scenarijih je podan v Tabeli 1. Večina navedenih tehnologij je odvisna od govorjenega jezika, to pa pomeni, da je za njihovo prilagoditev značilnostim govorjene slovenščine potrebno izvesti dodatno raziskovalno in razvojno delo. Zaradi relativne majhnosti slovenskega

govornega področja in z njim povezanega trga ter svojskih značilnosti našega govorjenega jezika ni pričakovati, da bodo tuji razvijalci in tuje korporacije v doglednem času našli tržni interes za razvoj oziroma prilagoditev teh tehnologij za naš jezik. Inteligentni avdionadzorni sistemi bodo tako prisluhnili, razumeli in spregovorili v slovenščini kvečjemu po zaslugi ustrezno usposobljenih slovenskih raziskovalcev in razvijalcev, katerih delo bo financirano predvsem netržno, torej iz javnih sredstev, ki se namenjajo razvojnim projektom v nacionalnem interesu, kamor spada tudi ohranjanje slovenskega govorjenega jezika in s tem tudi slovenske kulturne dediščine.

V nadaljevanju bolj podrobno obravnavamo nekaj različnih možnih varnostnonadzornih scenarijev, pri katerih pridejo v poštev obravnavane tehnologije, in izpostavljamo morebitne posebnosti pri njihovi prilagoditvi značilnostim govorjene slovenščine.

Ne glede na govorjeni jezik uporaba pametnih avdionadzornih tehnologij odpira precej izjemno občutljivih pravnih in etičnih vprašanj, ki jih obravnavamo v zadnjem delu članka. Vse oblike prisluha in samodejne analize govora namreč lahko predstavljajo grožnjo osnovni človekovi pravici do varstva zasebnosti. V Evropi in po svetu se kljub temu zaradi različnih smiselnih in upravičenih varnostnonadzornih in drugih razlogov te tehnologije intenzivno razvijajo in tudi že uporabljajo v vedno bolj integriranih pametnih nadzornih sistemih. Razvoja teh tehnologij zato ne smemo ignorirati, saj ni pričakovati, da bo dolgoročno zaobšel slovenski govorjeni jezik.

2 VARNOSTNONADZORNI SCENARIJI

Pri razvoju novih tehnologij navadno najprej izvedemo študijo in ovrednotenje možnih scenarijev njihove smiselne in upravičene uporabe. Pri pametnih nadzornih sistemih, ki vključujejo tehnologije govorjenega jezika, pridejo v poštev varnostnonadzorni scenariji, ki se kakorkoli nanašajo na samodejno računalniško analizo zajetih zvočnih govornih signalov. V nadaljevanju obravnavamo nekaj izbranih primerov takšnih scenarijev.

2.1 Nadzor avdiokomunikacijskih kanalov

Uporaba tehnologij govorjenega jezika za nadzor avdiokomunikacijskih kanalov je med vsemi obravnavanimi varnostnonadzornimi scenariji še najbolj znana in tudi razvita. Zaradi nacionalnih varnostnih interesov razvoj teh tehnologij v največji meri neposredno podpirajo kar vlade različnih razvitih držav. Namen te podpore je predvsem večanje učinkovitosti njihovih nacionalnih varnostno-obveščevalnih služb pri preprečevanju, zatiranju in preganjanju kriminala in terorizma.

Na tem področju se za proaktivne nadzorne sisteme štejejo predvsem sistemi za samodejno zaznavanje in razpoznavanje (identifikacijo) govorcev, ki jih varnostno-obveščevalne službe obravnavajo in spremljajo zaradi utemeljenih sumov storitve oziroma precej verjetne možnosti storitve kaznivih dejanj in katerih govor bi se lahko pojavil v avdiokomunikacijskih kanalih oziroma omrežjih. S tehnologijo samodejnega razpoznavanja govora pa se poskuša samodejno zaznati in razpoznati izgovorjena sporočila, ki so varnostno sumljiva (denimo, napeljevanje in napovedovanje kriminalnih ali terorističnih dejanj ipd.) in so potrebna proaktivne in preventivne obravnave varnostno-obveščevalnih služb.

Poleg navedenih tehnologij je za tovrstne varnostnonadzorne scenarije uporabna tudi tehnologija razpoznavanja govorjenega jezika, s katero je mogoče doseči, da nadzorni sistem samodejno zazna in razpozna jezik govorca ali celo njegov materni jezik, ko govorec govori tuji jezik.

Ni razloga, zakaj te tehnologije ne bi bile zanimive za uporabo za govorjeno slovenščino. Primerno operativno učinkovito (in primerno nadzorovano) delovanje naših varnostno-obveščevalnih služb in policije pri preprečevanju, zatiranju in preganjanju kriminala in terorizma v naši državi je prav gotovo v slovenskem nacionalen interesu. Razvoj tovrstnih tehnologij za govorjeno slovenščino bi zagotovo povečal njihovo učinkovitost in s tem tudi varnost slovenskih državljanov. Tako kot v mnogih podobnih primerih pa moramo tudi pri odločitvi za podporo razvoju teh tehnologij za govorjeno slovenščino seveda pretehtati ekonomske in socialne stroške uvajanja teh tehnologij na eni strani ter dejansko stopnjo varnostne ogroženosti in z njo povezane

ekonomske in socialne stroške na drugi strani.

2.2 Integrirani avdiovizualni nadzor prostorov

Varnostni nadzor odprtih in zaprtih prostorov se danes izvaja predvsem z videonadzornimi sistemi. Večino tovrstnih varnostno-nadzornih scenarijev je mogoče razširiti z dodajanjem funkcije pametnega avdionadzora. Ta razširitev predvideva obstoj možnosti dodatne namestitve mikrofonov v prostor za zajemanje zvočnih signalov. Najsodobnejše motorizirane mrežne nadzorne kamere imajo pogosto že vgrajen mikrofon ali vsaj mikrofonski vhod in z njihovo primerno namestitvijo lahko vzpostavimo nadzorno polje mikrofonov. S sodobnimi postopki obdelave zvočnih signalov lahko z računalniško analizo zvočnih signalov, zajetih iz polja mikrofonov, izvedemo časovno in prostorsko lokalizacijo zvočnih virov, ki se pojavljajo v nadzorovanem prostoru (Keyrouz in dr. 2007). Po lokalizaciji zvočnih virov lahko izvedemo še postopke samodejnega razpoznavanja varnostno sumljivih zvokov, med katerimi so lahko tudi govor in drugi človeški glasovi, kot so kričanje, izgovarjanje groženj, klici na pomoč ipd., ki odražajo govorčevo psihofizično stanje.

Tipični varnostnonadzorni scenariji, ki bi vključevali takšne razširjene sisteme, so danes že skoraj običajni varnostni nadzori javnih prostorov, kot so mestne ulice, potniške postaje, podhodi in javna dvigala, parkirišča, garaže, igrišča in tudi javna prevozna sredstva.

V primeru zaprtih varovanih javnih prostorov, kjer je večja možnost poskusov nasilnega ropa (to so na primer zlatarne, pošte, banke ipd.), bi samodejnemu razpoznavanju izgovorjenih groženj, kričanja in klicev na pomoč lahko dodali tudi funkcijo za samodejno govorno proženje tihega alarma ter klic policije in reševalnih služb z izgovarjanjem vnaprej predvidenih prikritih prožilnih govornih izjav. S tehnologijami samodejnega razpoznavanja psihofizičnega stanja govorca pa bi bilo mogoče sistem usposobiti, da bi zaznal izrazito agresivno obnašanje ali prestrašenost prisotnih govorečih ljudi. Tovrstni sistemi bi tako lahko celo reševali življenja, saj so znani primeri (ropi), ko ranjeni ljudje, niso uspeli sprožiti klasičnega alarma ali pravočasno priklicati pomoči.

2.3 Samostojni avdionadzor prostorov

V primerih, ko video nadzor še ni ali ne more/sme biti vzpostavljen (slaba vidljivost ali varovanje zasebnosti) oziroma predstavlja prevelik poseg v zasebnost ljudi (denimo javna stranišča), je mogoče razmišljati o uporabi samostojnih pametnih avdionadzornih sistemov. Takšni sistemi bi prišli denimo v poštev v javnih prostorih, kjer je večja možnost kriminalnih dejanj v nočnem času in ob slabi vidljivosti (spolno nadlegovanje, poskusi ropa ipd.). Primeri takšnih prostorov so odprta ali pokrita slabo osvetljena parkirišča, garažni koridorji, cestni podhodi in prehodi, javna dvigala, javna stranišča ipd. V vseh teh primerih bi prišel v poštev samostojni avdionadzorni sistem, ki bi imel poleg samodejnega razpoznavanja varnostno sumljivih zvokov, kot je razbijanje in neobičajen hrup, tudi zmožnost razpoznavanja kričanja, izgovorjenih groženj, klicev na pomoč ipd. V primeru, ko so v prostor nameščeni tudi zvočniki (tipično je to v dvigalih), bi lahko sistem prisotne tudi govorno opozoril na zaznano neobičajno obnašanje, kar bi jih lahko odvrnilo od izvedbe kriminalnih dejanj.

Avdionadzorne sisteme je mogoče uporabiti tudi za druge namene, ne le za preprečevanje kriminala in terorizma. Mogoče jih je namreč uporabiti tudi v zasebnih varovanih stanovanjih, v katerih bivajo ostareli, bolni ali onemogli ljudje, ki jih zaradi varovanja zasebnosti moti videonadzor. V teh primerih bi avdionadzorni sistem lahko uporabili za samodejno razpoznavanje klica na pomoč ali zaznavanje poslabšanega psihofizičnega stanja oziroma stiske varovancev. Tak sistem bi lahko samodejno razpoznal tudi druge neobičajne zvoke v prostoru, kot je hrup padajočih predmetov ipd. Tovrstni sistemi bi lahko preprečili pogoste neprijetne dogodke, ko osamljeni onemogli starejši varovanci po padcu na svojem domu ležijo tudi po nekaj dni na tleh in ne uspejo priklicati pomoči.

Vsi navedeni varnostnonadzorni scenariji ponujajo precej možnosti bolj intenzivnega razvoja tehnologij govorjenega jezika, ki presega uporabo v klasičnih uporabniških komunikacijskih vmesnikih za mobilne in druge informacijsko-komunikacijske platforme.

3 VARNOSTNONADZORNE GOVORNE TEHNOLOGIJE

Pri razvoju in uporabi obstoječih tehnologij govorjenega jezika v pametnih

nadzornih sistemih se izkaže, da jedro teh tehnologij navadno ni potrebno posebej prilagajati varnostnonadzornemu področju uporabe. Še največ težav se pojavlja pri pridobivanju primernih zbirk zvočnih govornih posnetkov, ki ustrezajo izbranim varnostnonadzornim scenarijem in so nujno potrebne za izvedbo raznih učnih postopkov in ovrednotenje zanesljivosti delovanja sistemov.

Posebno težavo pa predstavlja prilagoditev teh tehnologij manj razširjenim govorjenim jezikom, kot je govorjena slovenščina. V prejšnjem poglavju opisani avdionadzorni scenariji namreč predvidevajo zajem in analizo izrazito spontanega govora, zato se pri razvoju teh tehnologij ne moremo zanašati na obstoječe slovenske govorne podatkovne zbirke, ki večinoma vključujejo bran ali medijski govor, denimo (Mihelič in dr. 2003) ali (Žgank in dr. 2004). V nadaljevanju obravnavamo posebnosti govornih zbirk, ki so primerne za razvoj pametnih avdionadzornih tehnologij.

3.1 Varnostnonadzorne zbirke govornih posnetkov

Pri pridobivanju varnostnonadzornih zbirk govornih posnetkov je možnih več pristopov in vsak ima svoje slabosti in prednosti. Pridobivanje govornih posnetkov, ki verodostojno odražajo obravnavan varnostnonadzorni scenarij, je razmeroma zahtevno in pri tem navadno uporabljamo eno od treh metodologij.

Pri prvem pristopu se zanašamo na snemanje govora v igranih razmerah, ki jih uredimo posebej za te potrebe, pogosto pa uporabimo kar posnetke igranih ali dokumentarnih filmov, ki vsebujejo primerne filmske sekvence.

Druga možnost je, da pri pridobivanju govorne zbirke sodelujejo prostovoljci, ki jih z različnimi psihološkimi tehnikami spodbudimo k želenemu obnašanju. Primer takšne zbirke za govorjeno slovenščino je govorna zbirka AvID (Gajšek in dr. 2008), ki je primerna za razvoj samodejnega razpoznavalnika emocionalnega stanja slovensko govorečega govorca.

Najtežje pa je pridobiti posnetke resničnih varnostnonadzornih razmer, kot so posnetki že nameščenih avdiovizualnih nadzornih sistemov, posnetki klicev ljudi v komunikacijski center policije ali reševalnih služb ipd. Te zbirke najbolj verno odražajo resnične varnostnonadzorne razmere, vendar jih je težko

pridobiti zaradi pravnih in drugih ovir. Takšno zbirko govorjene slovenščine bi načeloma lahko pridobili v sodelovanju s klicnimi centri slovenskih urgentnih in reševalnih služb ter policije, vendar bi gotovo trčili na precej ovir, ki se nanašajo na varstvo pravic in predvsem zasebnosti klicateljev v klicne centre. V takšni zbirki se ne bi smeli nahajati nobeni osebni podatki, ki bi omogočali identifikacijo klicateljev, predvsem pa ne bi smela biti javna, ampak na razpolago samo razvijalcem teh tehnologij.

S to razmeroma zahtevno problematiko se ubadajo predvsem raziskovalci na področju čustvenega računalništva (angl. affective computing), ki v okviru različnih projektov in mrež odličnosti (denimo, http://emotion-research.net) pridobivajo tovrstne zbirke avdiovizualnih posnetkov, denimo korpus SAFE (Clavel in dr. 2006) ipd.

3.2 Razpoznavanje govora

Obstoječo razmeroma razvito tehnologijo samodejnega razpoznavanja govora je mogoče neposredno uporabiti za različne varnostnonadzorne scenarije (razpoznavanje izgovorjenih groženj, klicev na pomoč in varnostno sumljivih izjav ter prikrito govorno proženje alarma ipd.). Samodejno razpoznavanje govora izvedemo z računalniškim dekodiranjem oz. pretvorbo govornega sporočila iz digitalno vzorčenega akustičnega signala v niz simbolov, ki predstavlja niz razpoznanih besed.

Dekodirnik navadno temelji na računalniški simulaciji matematičnega modela hierarhično strukturiranega končnega pretvornika (angl. Finite State Transducer), ki poenostavljeno povedano »preskakuje« med svojimi notranjimi stanji kot posledica prisotnosti določenih frekvenc v zajetem vhodnem zvočnem signalu, ki so značilne za posamezne glasove danega govorjenega jezika. Pri določenih preskokih pretvornik na izhod odda simbol ali nize simbolov in s tem se vhodni akustični signal pretvarja v nize simbolov.

Najvišji nivo hierarhične strukture končnega pretvornika modelira dani govorjeni jezik (»preskakovanje« med besedami), vmesni nivo besednjak in slovar izgovarjav besed (»preskakovanje« med fonemi oz. alofoni besed) in najnižji nivo akustične uresničitve posameznih fonemov, ki sestavljajo besede (»preskakovanje« med komponentami fonov oz. glasov, ki jih določa

prisotnost značilnih frekvenc v zvočnem signalu).

Na najnižjem nivoju hierarhično strukturiranega končnega pretvornika so tako navadno stanja in prehodi med stanji t. i. prikritih Markovih modelov (Jelinek 1998) (Mohri in dr. 2008). Tako predstavljen govorni model se je izkazal kot zelo prilagodljiv različnim govorjenim jezikom in različnim področjem uporabe. Pri njegovi prilagoditvi govorjeni slovenščini moramo samo upoštevati osnovne jezikoslovne in glasoslovne značilnosti našega jezika, kot so posebnosti besedišča, izgovarjave posameznih besede, skladnje itd.

Za uporabo razpoznavalnika govora pri avdionadzoru odprtih in zaprtih prostorov je potrebnega še nekaj dodatnega razvojnega in raziskovalnega dela, ki se nanaša na zagotavljanje čim večje robustnosti delovanja v zahtevnih akustičnih okoljih (prisotnost uličnega hrupa, več virov zvoka v prostoru ipd.). To lahko dosežemo z uporabo naprednih metod časovne in prostorske lokalizacije govornih zvočnih virov v prostoru z uporabo polja mikrofonov in postopki slepega ločevanja med zvočnimi viri (angl. Blind Source Separation – BSS) ipd.

3.3 Razpoznavanje govorcev

Postopke razpoznavanja govorcev bi lahko v grobem razdelili v dve skupini. V prvo uvrščamo postopke, ki se uporabljajo za besedilno odvisno razpoznavanje, v drugo pa postopke, ki se uporabljajo za besedilno neodvisno razpoznavanje. Za avdionadzorne sisteme (razpoznavanje poljubnega govora znanih kriminalcev in osumljencev) pridejo v poštev predvsem besedilno neodvisni sistemi.

Tehnologija besedilno neodvisnega razpoznavanja govorcev je v zadnjem poldrugem desetletju doživela precejšen napredek. V veliki meri gre zasluge za to pripisati ameriški organizaciji NIST, ki je z rednim organiziranjem dogodkov, na katerih se med seboj »pomerijo« najboljši raziskovalci s tega področja, uspela privabiti ugledne raziskovalne ustanove s celega sveta.

Eden izmed ključnih prebojev na tem področju je bil dosežen z uvedbo t. i. splošnega modela govorcev (angl. Universal Background Model – UBM), ki iz več sto ur posnetkov govora velikega števila različnih govorcev strne večino pomembnih akustičnih lastnosti govorcev v relativno majhno množico

parametrov matematičnega statističnega modela, predstavljenega s t. i. mešanico Gaussovih porazdelitev (angl. Gaussian Mixture Models – GMM) (Reynolds in dr. 2000). Uvedba modela UBM pa je poleg številnih drugih prednosti preko postopka največje izkustvene verjetnosti (angl. Maximum A posteriori Probability – MAP) uspela zagotoviti visoko zanesljivost razpoznavanja.

Pristop z uporabo modela UBM je zelo učinkovit v primeru, ko so akustične razmere (šum, lastnosti mikrofona in prenosnih poti itd.) v govornih posnetkih enake, a se uspešnost razpoznavanja precej poslabša, kadar temu ni tako. Raziskovalci so zato predlagali številne rešitve, s katerimi so poskušali izničiti ali vsaj zmanjšati vpliv sejne spremenljivosti (razlike med akustičnimi in govornimi okoliščinami pri zajemanju govornih posnetkov). Med vsemi predlaganimi rešitvami je bila največje pozornosti deležna analiza vezanih faktorjev (angl. Joint Factor Analysis – JFA) (Kenny in dr. 2007; Dehak in dr. 2011), s katero je govorne posnetke danega govorca mogoče pretvoriti v manj razsežen vektor značilk (poimenovan i-vektor), ki ohrani večino diskriminatorne informacije, ki jo potrebujemo za ločevanje med različnimi govorci.

3.4 Razpoznavanje psihofizičnega stanja govorcev

Tehnologija razpoznavanja psiho-fizičnega stanja govorcev je v osnovi precej podobna tehnologiji razpoznavanja govorcev (uporaba UBM-MAP modela itd.), pri čemer se govorni posnetki razvrščajo namesto v razrede govorcev v nekaj razredov obravnavanih psihofizičnih stanj (Gajšek in dr. 2012). Za razločljiva psihofizična stanja govorcev, ki bi jih naj bilo mogoče razpoznati na podlagi akustične analize govornega signala, se navadno obravnava nekaj izbranih emocionalnih stanj (strah, jeza, presenečenje, ipd.) ter psihofizična stanja, ki so posledica alkoholiziranosti ali vpliva mamil.

Tudi na tem področju so dala spodbudo razvoju predvsem tekmovanja, ki so organizirana v okviru serije največjih mednarodnih konferenc na področju tehnologij govorjenega jezika Interspeech (Schuller in dr. 2009). Tovrstne sisteme bi se dalo neposredno uporabiti za razpoznavanje agresivnega obnašanja ali prestrašenosti posameznikov, ki so prisotni v varovanem prostoru. Podobno kot sistemi za razpoznavanje govorcev, so tudi te

tehnologije sicer v glavnem neodvisne od govorjenega jezika. Se pa zanesljivost delovanja tovrstnih sistemov zagotovo izboljša, če so naučeni iz posnetkov emocionalnega govora v govorjenem jeziku, za katerega se uporabljajo. Odražanje čustev v akustičnih značilnostih govora je nedvomno drugačno v govorjeni slovenščini kot pa, denimo, v govorjeni japonščini ipd.

Na podoben način kot se izvaja razpoznavanje govorcev in njihovih psihofizičnih stanj, se lahko izvede tudi od besedila neodvisno razpoznavanje govorjenega jezika danega govorca. Izvedba takšnega razpoznavalnika temelji na predpostavki, da se že na nivoju osnovnih spektralnih značilnosti akustične uresničitve govora odražajo razlike med jeziki. Da je temu res tako, kažejo primeri, ko so ljudje sposobni prepoznati jezik govorca, ne da bi razumeli eno samo izgovorjeno besedo.

4 PRAVNA IN ETIČNA VPRAŠANJA

Predstavljene tehnologije in njihova uporaba v različnih varnostnonadzornih scenarijih odpirajo precej pravnih in etičnih vprašanj, še posebej v Evropi, kjer se daje precej poudarka človekovim pravicam do varovanja osebnih podatkov in pravici do zasebnosti. Veljavna zakonodaja v evropskih državah in v celotni EU je precej nedorečena glede vprašanja varovanja teh pravic pri samodejni obdelavi in izmenjavi varnostnonadzornih podatkov (Cannataci 2010). Veljavna zakonodaja tako praktično onemogoča uporabo pametnih avdionadzornih sistemov za varnostni nadzor javnih prostorov in to kljub temu, da bi ti sistemi lahko v precej primerih reševali življenja. Obstaja namreč upravičen strah, da bi tovrstni sistemi omogočili varnostno-obveščevalnim službam neupravičeno prisluškovanje pogovorom ljudi v nadziranih prostorih. To skrb je potrebno upoštevati, zato bi bilo potrebno v obravnavane tehnologije vgraditi sistemske varovalke, ki bi onemogočile zlorabo v druge namene, kot je bilo prvotno zamišljeno (denimo, onemogočanje nepotrebnega shranjevanja zajetih govornih posnetkov ipd.).

Avtorji prispevka sodelujemo pri evropskih projektih, ki se ukvarjajo s temi vprašanji in katerih cilj je podpora modernizaciji in izboljšanju učinkovitosti sredstev in delovanja organov kazenskega pregona ter izmenjavi informacij na tem področju za odkrivanje dobrih praks ter pripravo smernic in modelnih zakonov, ki bi vsebovali primerne zaščitne ukrepe za državljane pri razvoju in

uporabi pametnih nadzornih tehnologij. Pridobljeno znanje v okviru teh projektov upoštevamo tudi pri svojem razvojno-raziskovalnem delu, ki ga izvajamo na tem področju.

5 SKLEP

V članku je obravnavana problematika uporabe tehnologij govorjenega jezika v pametnih nadzornih sistemih. To področje ponuja precej priložnosti za intenzivno razvojno in raziskovalno delo tudi pri nas, saj so te tehnologije v veliki meri odvisne od govorjenega jezika in ni pričakovati, da bodo tuji razvijalci v kratkem razvili tovrstne sisteme, ki bodo uspešno delovali tudi za slovensko govorno področje. Avtorji prispevka imamo dolgoletne izkušnje z razvojem vseh omenjenih tehnologij za slovenski govorjeni jezik, vključno z razvojem sintetizatorja emocionalnega govora, ki se lahko uporabi pri pametnih nadzornih sistemih, ki vključujejo govorno obveščanje nadzorovanih ljudi v prostoru, z namenom njihovega odvračanja od izvedbe kaznivih dejanj. V prihodnosti se zato nameravamo bolj posvetiti razvoju teh tehnologij tudi za uporabo v izbranih primernih varnostno-nadzornih scenarijev.

ZAHVALA

Delo, predstavljeno v tem prispevku, je bilo deloma podprto s financiranjem iz Sedmega okvirnega programa Evropske unije (FP7-SEC-2011.10.6) na podlagi sporazuma o financiranju številka 285582 RESPECT. Raziskovalno delo drugega avtorja je delno financirala Evropska unija iz Evropskega sklada za regionalni razvoj v okviru Operativnega programa krepitve regionalnih razvojnih potencialov za obdobje 2007-2013 po pogodbi št. 3211-10-000468 KC OpComm.

LITERATURA

Cannataci, J. A. (2010): Squaring the circle of smart surveillance and privacy. *Fourth Inter. Conf. on Digital Society*: 323–328. St. Maarten.

Clavel, C., Vasilescu, I., Devillers, L., Ehrette, T., Richard, G. (2006): The SAFE Corpus: fear-type emotions detection for surveillance applications. *LREC'06*: 1099–1104. Genoa.

Dehak, N., Kenny, P., Dehak, R., Dumouchel, P. (2011): Front-End Factor Analysis for Speaker Verification. *IEEE Trans. Audio, Speech, Lang. Process, 19*(4): 788–798.

Gajšek, R., Mihelič, F., Dobrišek, S. (2013): Speaker state recognition using an HMM-based feature extraction method. *Computer Speech & Language, 27*(1): 135–150.

Gajšek, R., Podlesek, A., Komidar, L., Sočan, G., Bajec, B., Štruc, V., et al. (2008): AvID: Audio–Video Emotional Database. V *Proceedings of the 11th International Multiconference Information Society. C*: 70–74. Ljubljana, Slovenia: Jožef Stefan.

Jelinek, F. (1998): *Statistical Methods for Speech Recognition*. Cambridge, MA, USA: MIT Press.

Kenny, P., Boulianne, G., Ouellet, P., & Dumouchel, P. (2007): Speaker and Session Variability in GMM-Based Speaker Verification. *IEEE Trans. on Audio, Speech, and Language Processing, 15*(4): 1448–1460.

Keyrouz, F., Diepold, K., & Keyrouz, S. (2007): High performance 3D sound localization for surveillance applications. *Proc. IEEE AVSS '07*: 563–566). London.

Mihelič, F., Gros, J., Dobrišek, S., Žibert, J., & Pavešić, N. (2003, July): Spoken Language Resources at LUKS of the University of Ljubljana. *International Journal of Speech Technology, 6*(3): 221–232.

Mohri, M., Pereira, F. C., & Riley, M. (2008): Speech recognition with weighted finite-state transducers. V *Speech recognition*. Germany: Springer-Verlag.

Onut, I., Aldridge, D., Mondel, M., & Perelgut, S. (2011): The 2nd Workshop on Smart Surveillance System Applications. V *Proc. CASCON '11*: 382–384. Riverton: IBM Corp.

Piciarelli, C., & Foresti, G. L. (2011): Surveillance-oriented event detection in video streams. *IEEE Intelligent Systems, 26*(3): 32–41.

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000): Speaker Verification

using Adapted Gaussian Mixture Models. *Digital Signal Processing, 10*: 19–41.

Schuller, B., Steidl, S., & Batliner, A. (2009): The Interspeech 2009 emotion challenge. V *Proc. Interspeech 2009. 627*: pp. 312–315. Brighton: ISCA.

Žgank, A., Rotovnik, T., Sepesy Maučec, M., Verdonik, D., Kitak, J., Vlaj, D., et al. (2004): Acquisition and Annotation of Slovenian Broadcast News Database. V *Proceedings of the Fourth International Conference on Language:* 2103–2106. Lisbon, Portugal: European Language Resources Association.

WILL SMART SURVEILLANCE SYSTEMS LISTEN, UNDERSTAND AND SPEAK SLOVENE

The paper deals with the spoken language technologies that could enable the so-called smart (intelligent) surveillance systems to listen, understand and speak Slovenian in the near future. Advanced computational methods of artificial perception and pattern recognition enable such systems to be at least to some extent aware of the environment, the presence of people and other phenomena that could be subject to surveillance. Speech is one such phenomenon that has the potential to be a key source of information in certain security situations. Technologies that enable automatic speech and speaker recognition as well as their psychophysical state by computer analysis of acoustic speech signals provide an entirely new dimension to the development of smart surveillance systems. Automatic recognition of spoken threats, screaming and crying for help, as well as a suspicious psycho-physical state of a speaker provide such systems to some extent with intelligent behaviour. The paper investigates the current state of development of these technologies and the requirements and possibilities of these systems to be used for the Slovenian spoken language, as well as different possible security application scenarios. It also addresses the broader legal and ethical issues raised by the development and use of such technologies, especially as audio surveillance is one of the most sensitive issues of privacy protection.

Keywords: spoken language technology, smart surveillance systems, automatic speech recognition, automatic speaker recognition