

Tehnologije govornega jezika v pametnih nadzornih sistemih

Simon Dobrišek*, Boštjan Vesnicer†, France Mihelič*

*Univerza v Ljubljani, Fakulteta za elektrotehniko, LUKS
Tržaška 25, Ljubljana, Slovenija
{simon.dobrisek, france.mihelic}@fe.uni-lj.si

†Alpineon d.o.o.
Ulica Iga Grudna 15, Ljubljana, Slovenija
bostjan.vesnicer@alpineon.si

Povzetek

V članku je podan pregled obstoječe in možne uporabe tehnologij govornega jezika v t.i pametnih (inteligentnih) nadzornih sistemih. Tehnologije, ki omogočajo samodejno razpoznavanje govora, govorcev in njihovega psihofizičnega stanja s pomočjo računalniške analize govornega zvočnega signala, odpirajo povsem nove dimenzije razvoja pametnih nadzornih sistemov. Predstavljen je trenutno stanje razvoja teh tehnologij ter različni scenariji uporabe in primeri tovrstnih sistemov. Naslovljena so tudi širša pravna in etična vprašanja, ki jih odpira razvoj in uporaba tovrstnih tehnologij.

Spoken language technologies in smart surveillance systems

The paper provides an overview of existing and potential use of spoken language technologies in so-called smart (intelligent) surveillance systems. 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. The paper investigates the current state of development of these technologies and different application scenarios as well as existing examples of such systems. It also addresses the broader legal and ethical issues raised by the development and use of such technologies.

1. Uvod

Izjemen tehnološki napredek v zadnjih desetletjih je omogočil razvoj vedno bolj kompleksnih in vseprisotnih nadzornih tehnologij, katerih glavni namen je izboljšanje učinkovitosti pri zaznavanju in preprečevanju kriminala in terorizma. Pri sodobni varnostni paradigmi 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 nadzornih 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 varnostnih operaterjev na podlagi rezultatov samodejne analize zbranih podatkov. Ti sistemi predstavljajo 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 video-nadzorni sistemi.

Pri novejših CCTV sistemih se danes z metodami samodejne računalniške analize video vsebin že poskuša doseči zmožnost samodejnega zaznavanja in razpoznavanja varnostno sumljivih dogodkov, okoliščin ali obnašanja ljudi (Picciarelli et al., 2011). Mnogih varnostnih incidentov pa ni mogoče zaznati zgolj z analizo videa. Povsem novo dimenzijo inteligentnega nadzora ponuja integracija video-nadzornih sistemov z inteligentnimi avdio-

nadzornimi sistemi, ki omogočajo samodejno varnostno analizo zajetega zvočnega signala (Onut et al., 2011).

Avdio-nadzorni sistemi z zmožnostjo tristošestdeset stopinjskega pokrivanja prostora omogočajo razširitev nadzorovanega prostora preko vidnega polja navadnih nadzornih kamer. S samodejnim zaznavanjem in razpoznavanjem varnostno sumljivih zvokov, kot je kričanje, klicanje na pomoč, glasno izgovarjanje groženj ali pa zvoki razbijanja predmetov, korakov, odpiranje vrat, poka 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 umetno govorno opozarjanje in obveščanje prisotnih ljudi o zaznanem varnostnem incidentu.

Tehnologija	Varnostno-nadzorna uporaba
Razpoznavanje govora	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 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 varnostno-nadzornih uporab tehnologij govornega jezika.

Med razvijajočimi se tehnologijami govornega jezika je precej takšnih, ki jih je mogoče neposredno uporabiti v inteligentnih avdio-nadzornih sistemih. Osnovni pregled teh tehnologij in njenih možnih uporab v različnih varnostno-nadzornih scenarijih je podan v tabeli **Tabela 1**, v

naslednjem poglavju pa bolj podrobno obravnavamo nekaj različnih možnih varnostno-nadzornih scenarijev, pri katerih pridejo v poštev navedene tehnologije.

Uporaba teh tehnologij odpira precej pravnih in etičnih vprašanj, ki jih naslavljamo v zadnjem delu članka. Dejstvo pa je, da se v Evropi in po svetu te tehnologije razvija in uporablja tudi v vedno bolj integriranih nadzornih sistemih.

2. Varnostno-nadzorni scenariji

Pri razvoju novih tehnologij navadno izvedemo študijo in ovrednotenje scenarijev njihove smiselne uporabe. Pri pametnih nadzornih sistemih, ki vključujejo tehnologije govornega jezika, pridejo v poštev varnostno-nadzorni scenariji, ki se kakor koli nanašajo na samodejno računalniško analizo zajetih zvočnih govornih signalov.

2.1. Avdio nadzor komunikacijskih kanalov

Uporaba tehnologij govornega jezika za nadzor avdio-komunikacijskih in informacijskih kanalov je med vsemi obravnavanimi varnostno-nadzornimi scenariji še najbolj znana in tudi razvita. Zaradi nacionalnih varnostnih interesov razvoj teh tehnologij v največji meri neposredno podpirajo kar vlade različnih najbolj razvitih držav.

Na tem področju se za proaktivne nadzorne sistem štejejo predvsem sistemi za samodejno zaznavanje in razpoznavanje (identifikacijo) govorcev, ki jih varnostno-obveščevalne službe obravnavajo in spremljajo zaradi utemeljenih sumov storitve kaznivih dejanj in katerih govor se pojavi v avdio-komunikacijskih kanalih. S tehnologijo 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 varnostno-nadzorne scenarije uporabna tudi tehnologija razpoznavanja govornega jezika, s katero se lahko doseže, da nadzorni sistem samodejno zazna in razpozna jezik govorca ali celo njegov materni jezik, ko ta govori tuj jezik.

2.2. Integrirani avdio-vizualni nadzor prostorov

Varnostni nadzor prostorov se danes izvaja predvsem z video-nadzornimi sistemi. Večino tovrstnih varnostnih scenarijev je mogoče razširiti z dodajanjem funkcije pametnega avdio nadzora. Ta razširitev predvideva obstoj možnosti namestitve mikrofонов za zajemanje zvočnih signalov. Najsodobnejše motorizirane mrežne nadzorne kamere imajo pogosto že vgrajen mikrofonski vhod (slika 1) in z njihovo primerno namestitvijo lahko postavimo nadzorno polje mikrofонов. S sodobnimi postopki obdelave zvočnih signalov iz polja mikrofонов lahko izvedemo časovno in prostorsko lokalizacijo zvočnih virov, ki se pojavljajo v nadzorovanem prostoru (Keyrouz, 2007). Po lokalizaciji zvočnih virov lahko izvedemo postopke samodejnega razpoznavanja varnostno sumljivih zvokov, med katerimi je lahko tudi govor in drugi človeški glasovi, kot so kričanje, izgovarjanje groženj, klici na pomoč ipd.

Po lokalizaciji in razpoznavanju varnostno sumljivih zvokov ali glasov bi takšen sistem lahko preusmeril pozornost in vidno polje motoriziranih video-nadzornih kamer v smeri teh zvočnih virov in pritegnil pozornost morebitnih človeških varnostnih operatorjev in služb.



Slika 1: Primer mrežne motorizirane video-nadzorne kamere firme Axis z mikrofonskim vhodom.

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

V primeru zaprtih varovanih javnih prostorov, kjer je večja možnost poskusov 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 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 sistem možno usposobiti tudi tako, da bi samodejno zaznal izrazito agresivnost ali prestrašenost govorečih prisotnih ljudi. Tovrstni sistemi bi tako lahko reševali življenja, saj so znani primeri (ropi na poštah in v zlatarnah), ko v ropu ranjeni ljudje niso uspeli sprožiti klasičnega alarma ali pravočasno priklicati pomoči.

2.3. Samostojni avdio-nadzor prostorov

V primerih, ko video nadzor še ni ali ne more/sme biti vzpostavljen (slaba vidljivost ali varovanje zasebnosti), je mogoče razmišljati o uporabi samostojnih pametnih avdio-nadzornih sistemov. Takšni sistemi bi prišli denimo v poštev v javnih prostorih, kjer je večja možnost kriminalnih dejanj v nočnem času (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 avdio-nadzorni sistem, ki bi imel poleg samodejnega razpoznavanja varnostno sumljivih zvokov, kot je razbijanje in drug neobičajen hrup, tudi zmožnost samodejnega 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 tudi govorno opozoril prisotne na zaznano neobičajno obnašanje, kar bi lahko prisotne obvarovalo ali odvrnilo od kriminalnih dejanj.

Avdio-nadzorne sisteme je mogoče uporabiti tudi za druge namene, kot je le preprečevanje kriminala in terori-

zma. Možno jih je namreč uporabiti tudi v zasebnih varovanih stanovanjih, v katerih bivajo ostareli, bolni ali onemogli ljudje, ki jih zaradi zasebnosti moti video nadzor. V teh primerih bi avdio-nadzorni sistem lahko uporabili za samodejno razpoznavanje klica na pomoč ali 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 varnostno-nadzorni scenariji ponujajo precej možnosti bolj intenzivnega razvoja tehnologij govornega jezika, ki presega uporabo v klasičnih uporabniških vmesnikih za mobilne in druge informacijsko-komunikacijske platforme.

3. Varnostno-nadzorne govorne tehnologije

Pri razvoju in uporabi obstoječih tehnologij govornega jezika v pametnih nadzornih sistemih se izkaže, da jedro teh tehnologij navadno ni potrebno posebej prilagajati varnostno-nadzornemu področju uporabe. Še največ težav se pojavlja pri pridobivanju primernih zbirk zvočnih govornih posnetkov, ki ustrezajo izbranim varnostno-nadzornim scenarijem in so nujno potrebne za izvedbo raznih učnih postopkov in ovrednotenje zanesljivosti delovanja sistemov.

3.1. Varnostno-nadzorne zbirke govornih posnetkov

Pri pridobivanju varnostno-nadzornih zbirk govornih posnetkov je možnih več pristopov in vsak ima svoje slabosti in prednosti. Pridobivanje govornih posnetkov, ki verodostojno odražajo obravnavan varnostno-nadzorni scenarij, je razmeroma zahtevno in pri tem se navadno poslužujemo ene od treh metodologij.

Pri prvi se zanašamo na snemanje govora v igranih razmerah, ki jih izvedemo 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 pričakovanemu obnašanju. Najtežje pa je pridobiti posnetke resničnih varnostno-nadzornih razmer, kot so posnetki že nameščenih avdio-vizualnih nadzornih sistemov, posnetki klicev ljudi v komunikacijski center policije ali reševalnih služb ipd. Te zbirke najbolj verno odražajo resnične varnostno-nadzorne razmere, vendar jih je zelo težko pridobiti zaradi pravnih in drugih ovir.

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 avdio-vizualnih posnetkov (denimo, korpus SAFE – Clavel, 2006).

3.2. Razpoznavanje govora

Obstoječo razmeroma razvito tehnologijo samodejnega razpoznavanja govora je mogoče neposredno uporabiti za različne varnostno-nadzorne scenarije (razpoznavanje

izgovorjenih groženj, klicev na pomoč in varnostno sumljivih izjav ter prikrito govorno proženje alarma ipd). Ta tehnologija danes v glavnem temelji na računalniškem izvajanju postopka dekodiranja govornega signala z uporabo govornega modela, ki je predstavljen kot hierarhična struktura verjetnostnih končnih avtomatov. Najvišji nivo te hierarhične strukture modelira dani govorni jezik, vmesni nivo slovar izgovarjav danih besed in najnižji nivo akustične uresničitve posameznih glasov danega govornega jezika.

Celotno strukturo govornega modela se lahko obravnava kot strukturirani končni pretvornik (angl. Finite State Transducer), ki v najnižjem nivoju vsebuje stanja prikritih Markovovih modelov (Jelinek, 1998; Mohri et al., 2008). Takšen govorni model se je izkazal kot zelo prilagodljiv različnim področjem uporabe. Za sisteme avdio-nadzora prostora je potrebnega še nekaj razvojnega in raziskovalnega dela za zagotavljanje večje robustnosti delovanja v zahtevnih akustičnih okoljih (ulični hrup ipd), ter za boljše izrabo metod časovne in prostorske lokalizacije govornih zvočnih virov v prostoru.

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 avdio-nadzorne 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 prirejanjem 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 (UBM), ki iz več sto ur posnetkov govora velikega števila različnih govorcev strne večino pomembnih akustičnih lastnosti govorcev v relativno majhno množico parametrov statističnega modela mešanice Gaussovih porazdelitev (GMM) (Reynolds et al., 2000). Uvedba modela UBM pa je poleg številnih drugih prednosti preko postopka največje posteriorne verjetnosti (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 predlagali številne rešitve, s katerimi so poskušali izničiti ali vsaj zmanjšati vpliv sejne spremenljivosti. Med vsemi predlaganimi rešitvami je bila največje pozornosti deležna analiza vezanih faktorjev (JFA) (Kenny et al., 2007; Dehak et al. 2011), s katero se govorne posnetke danega govornika da pretvoriti v nizko-razsežen vektor značilik (poimenovan i-vektor), ki ohrani večino diskriminatorne informacije, ki jo potrebujemo za ločevanje med različnimi govorniki.

3.4. Razpoznavanje psihofizičnega stanja govorcev

Tehnologija razpoznavanja psihofizičnega stanja govorcev je v osnovi precej podobna tehnologiji razpoznavanja govorcev (uporaba UBM-MAP modela itd), pri čemer se govorne posnetke razvršča namesto v razrede govorcev v nekaj razredov obravnavanih psihofizičnih stanj (Gajšek et al., 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.

Precej spodbude razvoju na tem področju so dala tekmovanja, ki so organizirana v okviru serije največjih mednarodnih konferenc na področju tehnologij govornega jezika Interspeech (Schuller et al., 2009). Tovrstne sisteme bi se dalo neposredno uporabiti za razpoznavanje agresivnega obnašanja ali prestrašenosti posameznikov, ki so prisotni v varovanem prostoru.

4. Pravna in etična vprašanja

Predstavljene tehnologije in njihova uporaba v različnih varnostno-nadzornih scenarijih odpirajo precej pravnih in etičnih vprašanj, še posebej v Evropi, kjer se daje precej poudarka človekovim pravicam do varovanja osebnih podatkov in pravic do zasebnosti. Veljavna zakonodaja v Evropi je precej nedorečena glede vprašanja varovanja teh pravic pri samodejni obdelavi in izmenjavi varnostno-nadzornih podatkov (Cannataci, 2010). Veljavna zakonodaja tako praktično onemogoča uporabo pametnih avdio-nadzornih 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 se moralo v obravnavane tehnologije vgraditi systemske varovalke, ki bi onemogočile zlorabo v druge namene, kot je bilo prvotno zamišljeno (denimo, onemogočanje nepotrebne shranjevanja zajetih govornih posnetkov ipd).

Sodelavci našega laboratorija sodelujemo pri evropskih projektih, ki naslavljajo ta vprašanja in katerih cilj je podpora modernizaciji in izboljšanju učinkovitosti sredstev in delovanja organov kazenskega pregona ter izmenjavi informacij na tem področju za izmenjavo dobrih praks ter pripravo smernic in modelnih zakonov, ki bi vsebovali primerne zaščitne ukrepe za državljane pri implementaciji pametnih nadzornih tehnologij. Pridobljeno znanje v okviru teh projektov v laboratoriju upoštevamo tudi pri svojem razvojno-raziskovalnem delu, ki ga izvajamo na tem področju.

5. Sklep

V članku je obravnavana problematika uporabe tehnologij govornega jezika v pametnih nadzornih sistemih. To področje ponuja precej priložnosti za bolj intenzivno razvojno in raziskovalno delo tudi pri nas, saj so te tehnologije v veliki meri odvisne od govornega jezika in ni pričakovati, da bodo tuji razvijalci v kratkem razvili tovrstne sisteme, ki bodo uspešno delovali tudi za sloven-

sko govorno področje. V laboratoriju imamo dolgoletne izkušnje z razvojem vseh omenjenih tehnologij za slovenski govorni jezik, vključno z razvojem emocionalnega sintetizatorja govora. V prihodnosti se imamo zato namen bolj posvetiti razvoju teh tehnologij tudi za uporabo v izbranih primernih varnostno-nadzornih scenarijev.

6. Zahvala

Delo, predstavljeno v tem prispevku, je bilo deloma podprto s financiranjem iz Sedmega okvirnega programa Evropske unije (FP7-SEC-2010-1) na podlagi sporazuma o financiranju številka 261727. 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.

7. Literatura

- Clavel, C., Vasilescu, I., Devillers, L., Ehrette, T. in Richard, G., 2006, The SAFE Corpus: fear-type emotions detection for surveillance applications. In Proc. LREC'06, Genoa, Italy, pp. 1099–1104.
- Cannataci, J. A., 2010, Squaring the circle of smart surveillance and privacy, Fourth Inter. Conf. on Digital Society, DOI 10.1109/ICDS.2010.55, pp. 323-328.
- Dehak, N., Kenny, P., Dehak, R. in Dumouchel, P., 2011, »Front-End Factor Analysis for Speaker Verification«. IEEE Trans. Audio, Speech, Lang. Process., zv. 19, št. 4, pp. 788-798.
- Gajšek, R., Mihelič, F., Dobrišek, S., 2012, "Speaker state recognition using an HMM-based feature extraction method", Computer Speech & Language, DOI 10.1016/j.csl.2012.01.007.
- Jelinek, F., 1998. Statistical Methods for Speech Recognition. MIT Press, Cambridge, MA, USA.
- Kenny, P., Boulianne, G., Ouellet, P. in Dumouchel, P., 2007, »Speaker and Session Variability in GMM-Based Speaker Verification«. IEEE Trans. on Audio, Speech, and Language Processing, 15 (4), pp. 1448–1460.
- Keyrouz, F., Diepold, K., Keyrouz, S., 2007, »High performance 3D sound localization for surveillance applications«, In Proc. IEEE AVSS '07, London, UK, pp. 563-566.
- Mohri, M., Pereira, F. C. N. in Riley, M., 2008. Speech recognition with weighted finite-state transducers, pogl. Part E: Speech recognition. Springer-Verlag, Germany.
- Onut, I.V., Aldridge, D., Mondel, M. in Perelgut, S., 2011. 2nd Workshop on Smart Surveillance System Applications. In Proc. CASCON '11, IBM Corp., Riverton, NJ, USA, pp. 382-384.
- Piciarelli, C. in Foresti, G. L., 2011, "Surveillance-oriented event detection in video streams", IEEE Intelligent Systems, vol. 26, issue 3, pp. 32–41.
- Reynolds, D. A., Quatieri, T. F. in Dunn, R. B., 2000, »Speaker Verification using Adapted Gaussian Mixture Models«. Digital Signal Processing, 10, pp. 19–41.
- Schuller, B., Steidl, S., Batliner, A., 2009. The Interspeech 2009 emotion challenge. In: Proc. Interspeech 2009, ISCA, Brighton, UK, pp. 627: 312–315