

Pomen strojnega sluha pri razvoju strojne inteligence

Simon Dobrišek

Katedra za sisteme, avtomatiko in kibernetiko

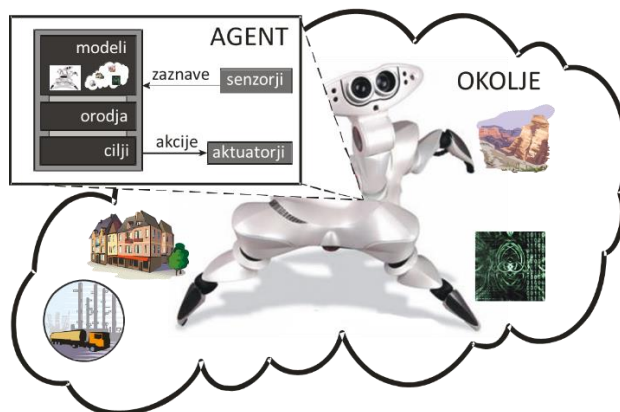
Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška c. 25, 1000 Ljubljana, Slovenija

E-pošta: simon.dobrisek@fe.uni-lj.si

Strojni sluh se nanaša na zmožnost avtonomnih inteligentnih sistemov, da zaznavajo in razpoznavajo različne zvoke v okolju, v katerem izvajajo zastavljene naloge in poskušajo dosegati zastavljene cilje. Strojni sluh temelji na posnemanju procesa slušnega zaznavanja pri ljudeh in drugih živih bitjih. Razvija in uporablja se za različne nadzorne in diagnostične naloge, ki vključujejo analizo in razpoznavanje okoljskih zvokov in zvokov, ki jih oddajo različna živa bitja, proizvodni stroji, vozila in druge naprave. Ena od najpomembnejših uporab strojnega sluha pa je razvoj sistemov za samodejno razpoznavanje in tvorjenje govora. Tovrstne govorne tehnologije namreč omogočajo razvoj sistemov za govorno komunikacijo človek-stroj, ki jo zmorejo izvajati najnaprednejši inteligentni sistemi.

Strojni sluh in umetni inteligentni sistemi

Umetne inteligentne sisteme se na področju avtomatike in kibernetike obravnava kot računalniško avtomatizirane avtonomne sisteme, ki so umeščeni v neko okolje, ki ga zaznavajo, v katerem delujejo in izvajajo zastavljene naloge oziroma poskušajo dosegati zastavljene cilje. Pri tovrstnih sodobnih umetnih inteligentnih sistemih njihovo delovanje temelji na računalniških algoritmih strojnega učenja in drugih metodah umetne inteligence, ki takšnim sistemom omogočajo samodejno pridobivanje znanja in izkušenj iz okolja, v katerem delujejo, ter uporabo tega znanja pri čim bolj učinkovitem izvajanju zastavljenih nalog in doseganju zastavljenih ciljev. Sposobnost prilagajanja okolju omogoča umetnim inteligentnim sistemom, da učinkovito rešujejo vse bolj kompleksne naloge in se dinamično prilagajajo spremembam v okolju. Tovrstne sisteme se obravnava tudi kot t. i. *avtonomne inteligentne agente*. To so polodprti sistemi, ki so umeščeni v nek okoljski sistem, s katerim so v svojem odprtem delu povezani prek senzorjev in aktuatorjev, kot je ponazorjeno na sliki 1. V notranjem delu tega sistema se tvori model znanja, ki ga orodja na osnovi zaznav osvežujejo in upoštevajo pri izbiri akcij za doseganje zastavljenih in spreminjajočih se ciljev.



Slika 1: Ponazoritev inteligentnega avtonomnega agenta, ki prek senzorjev zaznava in razpoznavajo stanje okolja, v katero je umeščen in v katerem izvaja zastavljene naloge in zasleduje zastavljene cilje.

Zaznavanje okolja pri tovrstnih sistemih temelji na posnemanju procesa zaznavanja okolja pri ljudeh in drugih živih bitjih, ki vključuje delovanje njihovih različnih čutil in celotnega živčnega sistema. Zaznavanje okolja s čutili je pri živih bitjih proces, ki jim omogoča, da sprejemajo informacijo iz svojega okolja in jo prenašajo v svoj živčni sistem, ki se na te zaznave potem ustrezno odziva. Pri živih bitjih obstaja pet glavnih vrst čutil, ki jim omogočajo zaznavanje okolja, to so: vid, sluh, voh, okus in tip. Pri zaznavanju okolja s čutili se najprej pojavijo dražljaji iz okolja, kot so svetloba, zvok, vonj, okus in dotik. Ti dražljaji sprožijo receptorje v čutilih, ki so v očesu, ušesu, nosu, ustih in koži. Receptorji so občutljivi za določene vrste dražljajev, kot so svetloba, zvok in kemikalije na stični površini ali v zraku. Ko se receptorji aktivirajo, pošljejo svojevrstni signal v živčni sistem, kjer se ta pretvori v električne impulze, ki se kot dražljaji prenašajo po živčnih vlaknih do možganov.

Sluh je pri človekovem razvoju izjemno pomemben, saj predstavlja enega od glavnih vhodnih kanalov za informacijo o svetu okoli njega. Že od zgodnjega otroštva se otrok namreč uči prepoznavati različne zvoke in njihov pomen ter jih povezovati z drugimi zaznavami, kar spodbudi razvoj njegovega govora, posledično pa tudi razvoj komunikacijskega jezika in razuma oziroma inteligence. Primerjalne študije razvoja možganov kažejo, da se možgani otrok, ki odraščajo v zvočno bogatem okolju, razvijajo hitreje in bolje kot možgani otrok, ki živijo v manj zvočno bogatem okolju. Slušno zaznavanje igra tudi ključno vlogo pri razvoju jezikovne komunikacije, saj otrok prek sluha doživlja govorno komunikacijo staršev in drugih govorcev ter se tako nauči govora in jezikovne komunikacije. Čeprav ima vid zagotovo tudi zelo pomembno vlogo pri človekovem razvoju in zaznavanju sveta okoli njega, se izkaže, da je slušno zaznavanje pri razvoju razuma in inteligence celo pomembnejše. Od rojstva gluhi ljudje, ki niso uporabljali nadomestnih slušnih pripomočkov, navadno namreč precej težje razvijajo vse umske sposobnosti.

Zaradi zelo pomembne vloge slušnega zaznavanja pri razvoju razuma in inteligence se tako tudi pri razvoju umetnih inteligentnih sistemov že od nekdaj razvija t. i. strojni sluh. Strojni sluh se nanaša na računalniško vodene sisteme, ki analizirajo, razpoznavajo in se odzivajo na zvoke, ki jih zaznavajo s pomočjo mikrofonov ali drugih zvočnih in vibracijskih senzorjev. Pri razvoju umetnih inteligentnih sistemov se strojni sluh tako uporablja za razpoznavanje različnih zvokov, ki se lahko pojavljajo v njihovem delovnem okolju. Tovrstni sistemi pa se razvijajo tudi za razpoznavanje človekovega govora ter za razpoznavanje govorcev in njihovih psihofizičnih in čustvenih stanj. Strojni sluh se v zadnjih letih zelo hitro razvija, saj sta se z razvojem računske zmogljivosti strojne opreme in kompleksnosti algoritmov izboljšali njegova natančnost in učinkovitost. Uporaba strojnega sluha pri razvoju umetnih inteligentnih sistemov omogoča vse boljšo interakcijo med človekom in strojem ter tudi različne načine analize in razumevanja zaznanih zvokov, kar ima številne možnosti uporabe.

V nadaljevanju so najprej predstavljene možnosti uporabe strojnega sluha za zaznavanje in razpoznavanje različnih zvokov z različnimi nameni v danih okoljih. To so lahko naravno okolje, človekovo bivalno okolje, prometno okolje, proizvodno industrijsko okolje in druga okolja, v katerih je smiselno samodejno razpoznavati različne zvoke.

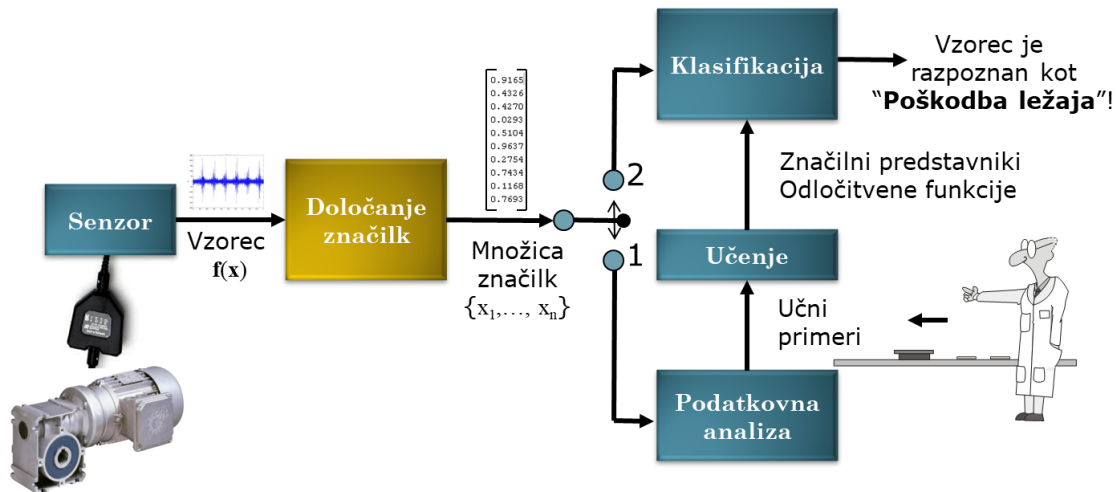
Uporaba tehnologij strojnega sluha za razpoznavanje različnih zvokov

Tehnologije strojnega sluha se lahko uporablja za razpoznavanje različnih zvokov v naravnem okolju in drugih okoljih, kjer se v umetnih inteligentnih sistemih tovrstno pridobljeno informacijo smiselno ter za izpolnjevanje zastavljenih nalog in doseganje zastavljenih ciljev koristno uporabi. Namen sistemov za samodejno razpoznavanje različnih vrst zvokov, ki se jih je do neke mere že razvilo, se nanaša predvsem na naslednje naloge:

- odzivanje na različne varnostno sumljive zvoke, kot so razbijanja, kričanje, streljanje in eksplozije;
- nadzorovanje kakovosti izvajanja proizvodnih procesov pri izdelovanju in sestavljanju izdelkov;
- razpoznavanje različnih živalskih vrst na osnovi zvokov, ki jih oddajajo;
- razpoznavanje različnih vrst vozil, naprav in strojev, ki oddajajo zvoke;
- diagnostika okvar strojev, vozil in drugih naprav, ki se odražajo v nenavadnih zvokih;
- zaznavanje živalskih škodljivcev in nepooblaščenih del na kmetijskih poljih;
- zaznavanje prostorskega stanja kopenskega, zračnega ali vodnega prometa;
- razpoznavanje zvokov, ki izhajajo iz teles ljudi in živali, kot so dihanje, kašljanje in srčni utrip; ter

- razpoznavanje katerikoli zvokov v okolju, ki bi lahko imeli pomembno vlogo pri delovanju in odločanju avtonomnih inteligentnih sistemov.

Zmožnost zaznavanja in razpoznavanja različnih zvokov omogoča inteligentnim sistemom, da avtonomno sprejemajo odločitve na podlagi zaznanih in razpoznanih zvokov, kar jim omogoča, da se čim učinkoviteje prilagajajo spremenjenim okoliščinam pri izvajanju zastavljenih nalog.



Slika 2: Struktura razpoznavalnika vzorcev, ki ima fazo strojnega učenja (1) in fazo samodejnega razpoznavanja vhodnih vzorcev (2). Ponazorjen je primer sistema za samodejno slušno diagnosticiranje napak pri delovanju motorja.

Za avtomatizirano razpoznavanje zvokov se razvija in preizkuša različne razpoznavalnike vzorcev, ki že dosegajo sprejemljivo zanesljivost in točnost razpoznavanja relevantnih zvokov. Nekaj primerov tovrstnih sistemov je podanih v nadaljevanju. Na sliki 2 je prikazana običajna struktura razpoznavalnikov vzorcev, pri čemer se pri razpoznavanju zvokov kratkočasovne značilke izsekov zvočnega signala običajno določa kot logaritme amplitud melodičnega frekvenčnega spektra, kar se zgleduje po domnevnem delovanju človekovega slušnega organa in sluha.

Sistemi za nadzorovanje proizvodnih procesov in ugotavljanje okvar strojev

Razvitih je bilo že več sistemov za zvočno nadzorovanje delovanja strojev in proizvodnih procesov. Ti sistemi se imenujejo akustični sistemi za nadzorovanje in diagnostiko (ASND), akustični diagnostični sistemi (ADS) ali akustični nadzorni sistemi (ANS).

Vsi ti sistemi uporabljajo mikrofone in druge vibracijske senzorje za zajem zvokov, ki se nato obdelujejo z algoritmi za razpoznavanje zvoka z uporabo akustičnih modelov, tvorjenih z metodami strojnega učenja. S temi algoritmi se potem samodejno zaznajo spremembe v zvokih, ki jih povzročajo delovanje proizvodnih strojev. Spremembe v zvoku lahko kažejo na različne vrste okvar, napak in nepravilnosti, kot so nesprejemljivo trenje, nenavadne vibracije, pretoki in izpusti tekočin ali plinov, obraba strojnih komponent in drugo.

Sistemi ADS lahko pomagajo pri razpoznavanju zgodnjih znakov okvar, kar omogoča hitro ukrepanje in preprečevanje nadaljnjih poškodb ali okvar strojev. Ti sistemi lahko preprečujejo daljše izpade proizvodnih procesov ter zmanjšajo izgube zaradi proizvodnih zastojev in podaljšajo življenjsko dobo proizvodnih strojev.

Tovrstne sisteme ponuja že več proizvajalcev. Na razpolago so predvsem komponente, ki omogočajo večkanalno zajemanje zvočnih in vibracijskih signalov, ki se potem obdelujejo v izbranem računalniškem sistemu. Primeri takšnih sistemov so:

- Strojna in programska oprema podjetja Brüel & Kjær Sound & Vibration Measurement (<https://www.bksv.com/>), kot so sistemi SoundPower, PULSE, LAN-XI DAS in drugi.

- Strojna in programska oprema podjetja Mitras, kot so sistemi za nadzor zvočnih emisij Acoustic Emission Technology (<https://www.physicalacoustics.com/ae-technology>).
- Strojna in programska oprema podjetja Siemens, kot je akustični sistem SITRANS DA 400 (<https://siemens.polinski.com.pl/sitrans-da-400>), ki temelji na akustičnem spremljanju delovanja strojev in zgodnjem odkrivanju okvar na osnovi analize zvočnega šuma.

Zvočno nadzorovanje delovanja strojev in proizvodnih procesov navadno temelji tudi na metodah strojnega učenja, kjer se sistem iz primerov že razvrščenih in označenih zvočnih posnetkov, ki odražajo pravilno ali nepravilno delovanje naprav, sam nauči zaznavati in razpoznavati značilne zvoke, ki odražajo stanje teh nadzorovanih naprav in celotnih proizvodnih procesov.

Uporaba tehnologij strojnega sluha pri diagnosticiranju zdravstvenih in psihičnih težav

Pri diagnosticiranju zdravstvenih težav ljudi se je razvilo sisteme za razpoznavanje različnih značilnih zvokov, kot so zvoki, ki nastajajo med dihanjem, zvoki, ki jih oddaja srce, črevesni zvoki, zvoki sklepov in drugi zvoki, ki odražajo delovanje človekovih organov in celotnega telesa oziroma okostja. Z različnimi senzorskimi napravami se zajema tovrstne vibracije in zvoke, nato pa se te zaznave uporabi za analizo in diagnosticiranje zdravstvenih težav. Na primer, z analizo in razpoznavanjem zvokov, ki se jih lahko pridobi z avskultacijo (poslušanjem, oslušovanjem) srca, se lahko odkrije različne srčne nepravilnosti, kot so šumenje srca, nepravilnosti v ritmu ter druge srčne nepravilnosti in težave.

Samodejno se lahko obdeluje, analizira in razpozna tudi zvoke, ki nastajajo z dihanjem, kašljem ali drugimi telesnimi funkcijami. Pridobljene spektralne značilnosti teh zvokov se nato lahko uporabi za diagnosticiranje bolezni, kot so astma, bronhitis, pljučnica in druge bolezni pljuč.

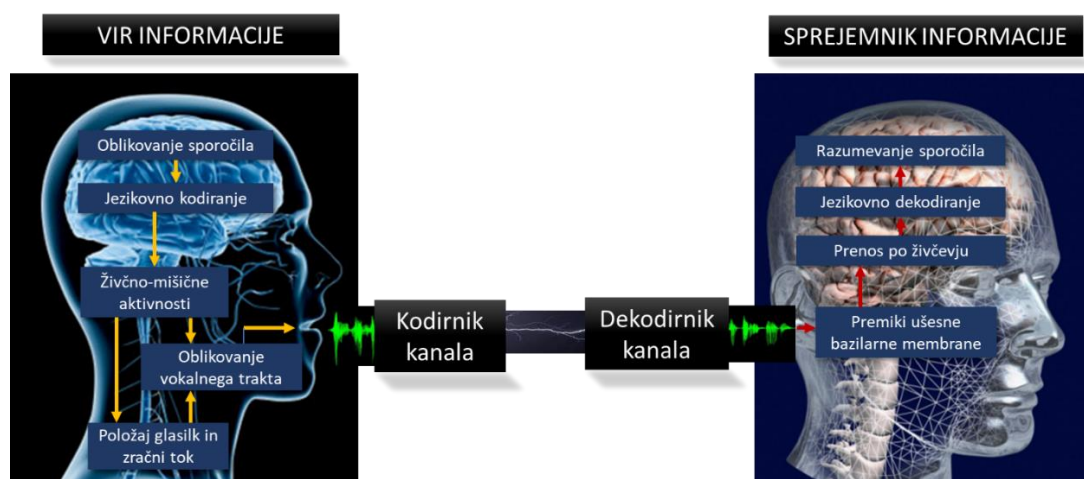
Z analizo akustičnih in jezikovnih značilnost človekovega govora pa se lahko samodejno razpozna in diagnosticira tudi različna psihofizična stanja ljudi, kot so:

- afazija, pri kateri gre za izgubo sposobnosti izražanja in razumevanja jezika, ki se lahko pojavi zaradi možganske poškodbe ali bolezni, kar se odraža v akustičnih in jezikovnih značilnosti govora;
- demenca, pri kateri gre za degenerativno bolezen možganov, ki lahko privede do izgube spomina, kognitivnih funkcij in govornih zmožnosti, kar se prav tako odraža v značilnostih govora;
- parkinsonova bolezen, nevrološka bolezen, ki lahko vpliva na nadzor nad mišicami, vključno z mišicami, ki se v govorilih uporabljajo za tvorjenje govora – akustična in jezikovna analiza govora lahko tako pomaga pri diagnosticiranju in spremljanju napredovanja te bolezni;
- depresija, duševna motnja, ki lahko vpliva na razpoloženje, čustva in vedenje, kar se odraža v akustičnih in jezikovnih značilnosti govora, kot so upočasnen govor, monoton glas, zmanjšano število različnih besed in druge jezikovne značilnosti govora;
- shizofrenija, duševna motnja, ki lahko privede do izkrivljene percepcije resničnosti, ki se odraža v disfunkcionalnem govoru, nenavadni govorni intonaciji, neobičajnih besednih zvezah in drugih jezikovnih značilnosti govora; ter
- avtizem, pri katerem gre za motnjo, ki lahko vpliva na socialne interakcije in komunikacijo z drugimi ljudmi ter se prav tako odraža v akustičnih in jezikovnih značilnostih govora.

Akustična in jezikovna analiza govora torej ne omogoča le izgradnje sistemov za govorno komunikacijo človek-stroj, ampak omogoča tudi svojevrstno medicinsko in psihiatrično diagnostiko. Primarno pa se človekov govor sicer analizira in obdeluje predvsem zaradi razvoja umetnih inteligentnih sistemov, ki podpirajo govorno komunikacijo človek-stroj. Za tvorjenje tovrstnih sistemov je treba razviti predvsem tehnologije za zanesljivo samodejno razpoznavanje in tvorjenje govora. Tovrstne govorne tehnologije se razvijajo že desetletja in najnovejši sistemi pri razpoznavanju in tvorjenju govora že dosegajo zmožnosti, ki jih imajo pri govornem komuniciranju ljudje. Sistemi za samodejno razpoznavanje in tvorjenje govora omogočajo razvoj različnih sistemov za govorno komunikacijo človek-stroj. Taka komunikacija je potrebna in zelo uporabna v vseh okoliščinah, ko imajo ljudje pri svojih delovnih in drugih aktivnostih, ki vključujejo upravljanje različnih naprav ali sodelovanje z njimi, že zasedene roke in oči. Zelo uporabna je tudi v okoliščinah, ko ljudje med delovnimi in drugimi aktivnostmi proizvedejo po informacijah oziroma vnašajo podatke v različne informacijske sisteme.

Sistemi za samodejno razpoznavanje in tvorjenje govora

Sistemi za samodejno razpoznavanje govora pretvarjajo govorne signale v normalizirano besedilo ali v semantični jezikovni zapis izrečenega, ki je osnova za pomensko analizo govora, ki jo izvajajo inteligentni sistemi pri govorni interakciji z ljudmi. Pri samodejnem tvorjenju govora pa gre samo za obratno pretvorbo besedilnega zapisa govora v zvočni signal, ki se nato predvaja poslušalcu. Za samodejno razpoznavanje in tvorjenje govora se pogosto uporabljajo isti akustični in jezikovni modeli, temelječi predvsem na modelih umetnih nevronske omrežij, ki se jih tvori z metodami strojnega učenja. Pri tem se pri tvorjenju govora ti modeli uporabijo na generativni način – tvorijo se zvočni signali, ki najbolj ustrezajo modelu v podanem besedilnem kontekstu. Predvidene možganske aktivnosti in procesi med govorno komunikacijo ljudi, ki se jih z različnimi modeli umetnih nevronske omrežij poskuša simulirati, so ponazorjeni na sliki 3.



Slika 3: Ponazoritev možganskih aktivnosti in procesov med govorno komunikacijo, ki vključujejo tudi pretvarjanja med zvočnimi in električnimi valovanji.

Razvoj govornih tehnologij je zelo pomemben tudi za preživetje obstoječih jezikov in človekove govorne komunikacije v vseprisotnem digitalnem okolju. Pomemben prispevek k preživetju in razvoju slovenščine v digitalnem okolju je tudi projekt RSDO – Razvoj slovenščine v digitalnem okolju (<https://rsdo.slovenscina.eu>), ki sta ga sofinancirali Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj.

V okviru projekta RSDO je bil razvit demonstracijski spletni portal (<https://www.slovenscina.eu>), na katerem je možno preizkusiti razvite govorne in jezikovne tehnologije. Pomemben prispevek projekta so tudi zbrani besedilni korpusi in korpus transkribiranih govornih posnetkov, ki so urejeni v govornem korpusu Artur 1.0 (<https://www.clarin.si/repository/xmlui/handle/11356/1776>). Ti korpusi in podatkovne zbirke omogočajo vsem raziskovalnim in komercialnim razvijalcem razvoj različnih govornih in jezikovnih tehnologij z uveljavljenimi metodami strojnega učenja, ki bodo omogočile in podprle preživetje slovenskega pisanega in govornega jezika v digitalnem okolju.

Simon Dobrišek je leta 2001 doktoriral na Fakulteti za elektrotehniko Univerze v Ljubljani, kjer je trenutno izredni profesor na Katedri za sisteme, avtomatiko in kibernetiko ter predstojnik Laboratorija za strojno inteligenco. Pri svoji raziskovalnih in pedagoških aktivnostih je osredotočen predvsem na področje umetne inteligence, teorije informacij, razpoznavanja vzorcev, biometrije in govornih tehnologij. V zadnjih desetih letih je sodeloval kot ključni raziskovalec pri več nacionalnih in mednarodnih projektih ter kot vodilni raziskovalec pri dveh projektih EU FP7 na področju biometrije, govornih tehnologij, razpoznavanja vzorcev, strojnega učenja in pametnih nadzornih sistemov. Na teh področjih je objavil tudi večje število izvornih znanstvenih člankov in znanstvenih prispevkov na konferencah.