

RAZVOJ GOVORNIH ASISTENATA I NJIHOVE PRIMENE U PAMETNIM KUĆAMA I PAMETNIM GRADOVIMA

Vlado Delić

Univerzitet u Novom Sadu – Fakultet tehničkih nauka
Trg Dositeja Obradovića 6, Novi Sad
E-mail: vlado.delic@uns.ac.rs

Sažetak

Veliki jezički modeli koji se razvijaju na bazi veštačke inteligencije predstavljaju najbrže prihvaćenu internet tehnologiju. Za godinu dana dostigli su impresivno vladanje prirodnim jezikom, prvo na engleskom, a ubrzo i na mnogim drugim jezicima. Istovremeno su značajno unapređene mogućnosti virtuelnih govornih asistenata, kao i njihova primena u pametnim kućama i pametnim gradovima. U radu je ukratko prikazana retrospektiva razvoja govornih tehnologija za srpski i srodne južnoslovenske jezike, moguće primene sa posebnim osvrtom na prednosti i opasnosti primena govornih tehnologija. Objašnjena je suština izazova u razvoju automatskog prepoznavanja govora i sinteze govora na osnovu teksta, njihova interdisciplinarnost i potreba za prikupljanjem i obradom ogromnih govornih i jezičkih resursa. Predstavljene su prve primene govornih tehnologija na srpskom i srodnim jezicima na pametnim telefonima i na pojedinim internet portalima, kao i perspektive njihove primene u pametnim kućama i u pametnim gradovima.

Ključne reči: Obrada prirodnog jezika (NLP), veliki jezički modeli (LLMs), govorne tehnologije: automatsko prepoznavanje govora (ASR) i sinteza govora na osnovu teksta (TTS), govorna komunikacija čovek-mašina, primene govornih tehnologija na pametnim uređajima i sistemima.

1. Uvod

Veliki jezički modeli (eng. *Large Language Models* – LLMs) su se u široj javnosti pojavili krajem 2022. godine i brzo su postali najbrže prihvaćena tehnologija koja je svima dostupna preko interneta. Kompanija OpenAI je zvanično lansirala ChatGPT još pre par godina, a trenutno je aktuelna verzija 4.0 koja se pojavila u martu 2023. godine. Do sada su sa ChatGPT diskutovali milioni ljudi širom sveta. ChatGPT baziran je na velikom jezičkom modelu GPT (eng. *Generative Pre-trained Transformer*), a postoje i mnogi drugi LLMs koji se koriste za različite svrhe u domenu obrade prirodnog jezika (eng. *Natural Language Processing* – NLP). Njihove mogućnosti se često ažuriraju i unapređuju kako se tehnologija razvija. Nakon uvodnog poglavlja ćemo pomenuti nekoliko LLMs i posebno one koji se koriste u virtuelnim govornim asistentima.

Paralelno sa razvojem LLMs došlo je i do ključnog progressa u razvoju govornih tehnologija, takođe zahvaljujući primeni veštačke inteligencije u širem smislu, a posebno mašinskog i dubokog učenja. Govorne tehnologije omogućuju čoveku da komunicira govorom s pametnim uređajima i sistemima kao što su roboti, aparati u domaćinstvu, instrumenti u automobilima, personalni asistenti na računarima, mobilnim telefonima, pametnim zvučnicima itd. Iako to mnogima izgleda kao samo još jedan način komunikacije i starije osobe neće lako steći naviku da se govorom obraćaju uređajima, mladi odrastaju uz pametne igračke, pametne telefone, tablete i laptop računare – koji sa njima pričaju.

Govorna komunikacija čovek-mašina u osnovi se bazira na govornim tehnologijama kao što su automatsko prepoznavanje govora (eng. *Automatic Speech Recognition* – ASR) i sinteza govora na osnovu teksta (eng. *Text-to-Speech* – TTS). Za razliku od drugih novih tehnologija, govorne tehnologije su manje ili više jezički zavisne, što znači da se u izvesnoj meri moraju razvijati zasebno za svaki jezik. U tome se najviše odmaklo za jezike koje širom sveta govore stotine ili bar desetine miliona ljudi, zato što je razvoj govornih tehnologija skup proces jer zahteva angažovanje skupih stručnjaka, vreme za sakupljanje i obradu jezičkih i govornih resursa, kao i eksploataisanje značajnih procesorskih resursa. Zahvaljujući nizu projekata koje je u poslednje dve decenije vodio autor ovog rada sa timovima istraživača na Fakultetu tehničkih nauka i inženjera u preduzeću AlfaNum, razvijene

su govorne tehnologije ASR i TTS za srpski, a potom i za srodne južnoslovenske jezike koji se govore u Hrvatskoj, Bosni i Hercegovini, kao i Crnoj Gori.

Primena govornih tehnologija pojačava epitet „pametni“ kako aparatima u pametnim kućama, tako i servisima u pametnim gradovima – a to je i centralna tema ovog rada, koji je organizovan na sledeći način. U drugom poglavlju, ukratko je predstavljeno nekoliko velikih jezičkih modela koji su u osnovi obrade prirodnog jezika, sa posebnim osvrtom na njihovu primenu u pametnim kućama i pametnim gradovima. U trećem poglavlju sagledana je problematika automatskog prepoznavanja i sinteze govora kroz kraću retrospektivu razvoja govornih tehnologija za srpski jezik i posebno kroz progres koji je napravljen na AI projektu S-ADAPT.⁵ U četvrtom poglavlju ukratko su predstavljene neke od prvih primena govornih tehnologija u pomenutim ex-YU republikama, sa posebnim osvrtom na primene u pametnim kućama i u pametnim gradovima. U zaključku se izdvaja nekoliko ključnih detalja i činjenica i sagledavaju nove perspektive koje donosi dalji razvoj govornih tehnologija, a koji će u Srbiji biti baziran na novom AI projektu AI-SPEAK⁶ kod Fonda za Nauku. U spisku referenci nalaze se samo novije publikacije autora ovog rada i njihovih saradnika na projektima razvoja i primene govornih tehnologija.

2. Veliki jezički modeli i njihova primena u virtuelnim govornim asistentima

ChatGPT se zasniva na danas najpoznatijem i najmoćnijem velikom jezičkom modelu GPT-3 (eng. *Generative Pre-trained Transformer 3*) koji je razvila kompanija OpenAI. Veliki jezički model GPT-3 ima 175 milijardi parametara i može da generiše tekstove visokog kvaliteta. Njegov prethodnik GPT-2 je imao samo 1,5 milijardu parametara i već je imao sposobnost generisanja koherentnih i kreativnih tekstova. Google je razvio veliki jezički model BERT (eng. *Bidirectional Encoder Representations from Transformers*) koji koristi transformersku arhitekturu i obučava se da razume kontekstualno značenje reči u rečenici, što ga čini veoma korisnim za zadatke kao što su pretraživanje i razumevanje jezika. Veliki jezički model T5 (eng. *Text-to-Text Transfer Transformer*) kompanije Google Brain, koji koristi jedinstven pristup pretvaranja različitih jezičkih zadataka u univerzalni tekstualni format, pa se trenira za različite zadatke prevođenja, generisanja teksta, odgovaranja na pitanja itd. Pomenućemo ovde i pretraživač Bing kompanije Microsoft, koji koristi sopstveni jezički model za razumevanje i obradu tekstualnih upita korisnika. Specifični detalji jezičkog modela u pretraživaču Bing nisu javno dostupni, ali Microsoft neprestano radi na poboljšanju razumevanja korisničkih upita i pretrage, koristeći tehnike mašinskog učenja i dubokog učenja.

Popularni govorni asistenti *Alexa (Amazon)*, *Siri (Apple)*, *Cortana (Microsoft)* i *Google Assistant* koriste sopstvene jezičke modele za obradu i razumevanje korisničkih zahteva. Konkretno govorni asistent *Alexa* kompanije *Amazon*, oslanja se na više jezičkih modela za obradu narudžbi korisnika kako bi prepoznali i razumeli upite korisnika na prirodnom jeziku. Govorni asistent *Siri* koristi jezički model razvijen od strane kompanije *Apple* za koji se zna da koristi tehnike dubokog učenja sa rekurentnim neuronskim mrežama i drugim algoritmima kako bi prepoznao i razumeo jezičke upite korisnika. Govorni asistent *Cortana* kompanije *Microsoft* koristi njene tehnologije i jezičke modele o kojima konkretni detalji nisu javno dostupni, ali su izvesno zasnovani na tehnologijama kao što su mašinsko učenje, duboko učenje i različite varijante transformerskih arhitektura. Konačno, *Google Assistant* koristi veliki jezički model BERT i još neke interne algoritme koji su takođe bazirani na dubokom učenju. Ovi modeli se treniraju na ogromnim skupovima podataka kako bi omogućili razumevanje i odgovor na korisničke upite na prirodnom jeziku.

U pametnim kućama se virtuelni govorni asistenti koriste ili preko pametnih telefona ili preko pametnih zvučnika sa mikrofonskim nizovima koji su bežično povezani na uređaje i sisteme u pametnim kućama. Svaki proizvođač kreirao je pametne zvučnike za svog govornog asistenta. Tako su na tržištu dostupni pametni zvučnici *Amazon Echo*, *Google Home*, *Apple HomePod*, *Microsoft's*

⁵ S-ADAPT – “*Speaker/Style Adaptation for Digital Voice Assistants Based on Image Processing Methods*” (2020-2023) je projekat iz oblasti veštačke inteligencije koji je podržan od Fonda za nauku Republike Srbije u okviru granta #6524560.

⁶ AI-SPEAK – “*Multimodal multilingual human-machine speech communication*” (2023-2026) je projekat iz oblasti veštačke inteligencije koji je podržan od Fonda za nauku Republike Srbije u okviru granta #7449.

Cortana, kao i neki koji podržavaju i po dva virtuelna govorna asistenta. Postoje i odgovarajuće aplikacije koje bežično povezuju pametne zvučnike sa uređajima u pametnoj kući kao što su Alexa app, Google Home app, Apple HomeKit, pri čemu su pojedini uređaji kompatibilni samo sa nekim od tih sistema. Sada se pokušava sa uvođenjem novih protokola koji će obezbediti da se isti uređaji u pametnoj kući mogu povezati sa različitim pametnim sistemima i njihovim govornim asistentima (protokol *Matter*). Primena ovih protokola koji će obezbediti izvesnu kompatibilnost i komunikaciju uređaja i sistema u pametnim kućama još uvek nisu zaživeli u većoj meri, pa to znači da se izborom virtuelnog govornog asistenta (softver) praktično opredeljujemo za tip uređaja (pametni zvučnik) i aplikacije za povezivanje sa uređajima i sistemima u pametnoj kući.

Ciljevi dijaloških sistema su:

1. Da prepoznaju potrebu i razumeju nameru, pa i stavove čoveka kao sagovornika.
2. Da odgovore na sva njegova pitanja.
3. Da izvrše zahteve korisnika.
4. Da vode prirodan i zanimljiv razgovor.

Dijaloški sistemi komuniciraju pomoću teksta, govora ili koristeći više modaliteta – multimodalno. Najčešće komuniciraju na engleskom, ali danas i na više desetina različitih jezika. Opšti dijaloški sistemi (npr. *Siri*) mogu da komuniciraju na najrazličitije teme, a postoje i dijaloški sistemi koji su usmereni na specifičnu oblast ili temu (npr. *MEDICTA* – za diktiranje medicinskih nalaza na srpskom uključujući latinske izraze i skraćenice koje se tu koriste; ili *Axon Voice Assistant (AVA)* koji se koristi na mobilnim telefonima pre svega za pozivanje po imenu i čiji je najveći deo rečnika zapravo telefonski imenik datog korisnika). Jednostavniji dijaloški sistemi fokusiraju dijalog na popunjavanje šablona i rade po određenim pravilima, a najnoviji dijaloški sistemi svoje odgovore kreiraju na bazi prethodne obuke sa velikom količinom podataka – koncept LLMs.

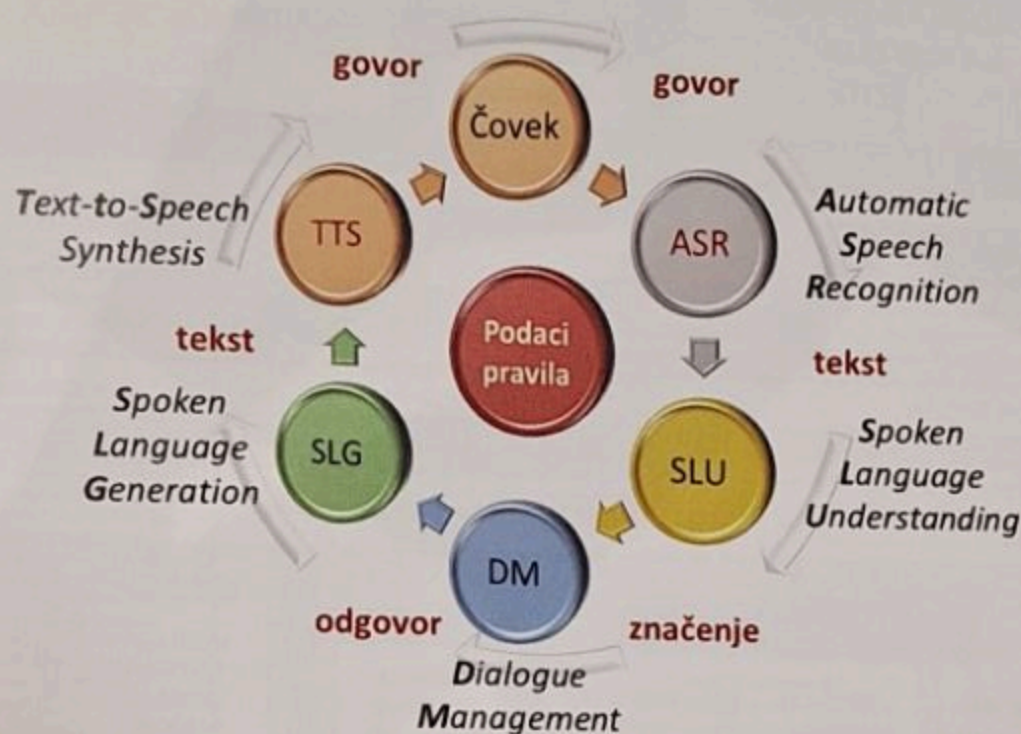
U dijalogu čovek-mašina poruke se najčešće razmenjuju u vidu teksta ili govora. Kao što je prikazano na slici 1, na putu tekstualne poruke od čoveka ka mašini, za svaki dijaloški čin mora da se interpretira tekst poruke, tj. da se utvrdi njeno značenje – i to u kontekstu dijaloga, spram određenih pravila i podataka sa kojima raspolaže NLP sistem; to radi modul za razumevanje prirodnog jezika (eng. *Natural Language Understanding* – NLU). Mašina, odnosno modul za upravljanje dijalogom (eng. *Dialogue Management* – DM) na osnovu interpretacije značenja, pa i prepoznavanja namera i stavova čoveka – inicira odgovarajuću akciju (npr. izvršenje primljene komande u pametnoj kući) i kreira odgovor koji se u modulu za generisanje prirodnog jezika (eng. *Natural Language Generation* – NLG) formuliše u vidu sintaksno i semantički ispravnog teksta za čoveka/sagovornika. Svi ovi NLP moduli su bazirani na određenim pravilima (npr. šablonima) i podacima (npr. velikim jezičkim modelima).



Slika 1. Moduli obrade prirodnog jezika (NLP) u dijalogu čovek-mašina razmenom tekstualnih poruka

3. Razvoj govornih tehnologija – šta su izazovi i koji su resursi neophodni

Da bi se dijalog čovek-mašina sa nivoa razmene tekstualnih poruka (dopisivanje ili četovanje, sl. 1) podigao na nivo razmene govornih poruka (razgovor, sl. 2), prethodno opisani NLP sistem mora da se nadogradi govornim tehnologijama koje prevode pisani u govorni jezik i obratno. Govor sadrži daleko više informacija od teksta; u govoru se može razaznati ne samo šta je rečeno (tekst, kao rezultat reverberacija). Automatsko prepoznavanje govora, glasa govornika i stila govora (ambijentalni šum, dinamika) samo po sebi ima elemente veštačke inteligencije, a dijaloški sistem čovek-mašina pored govornih tehnologija obuhvata i jezičke tehnologije. Moduli ASR i TTS govornih tehnologija sa jedne strane imaju govor, a sa druge tekst kao svoj ulaz, odnosno izlaz. Prepoznati tekst kao pisani vid *Spoken Language Generation* – SLG) povezuje sa njihovim značenjem sa kojim barata modul za upravljanje dijalogom (DM), slika 2. Govor koji čovek artikuliše i percipira u jednom odnosno drugom smeru govorne komunikacije – predstavlja fizički nivo govorne komunikacije s mašinom. Odgovarajući tekst u datom govornom činu predstavlja pisanu verziju trenutne komunikacije, dok se interpretacije značenja tekstualnih poruka odnose na kognitivni nivo govorne komunikacije svojstven ljudima.

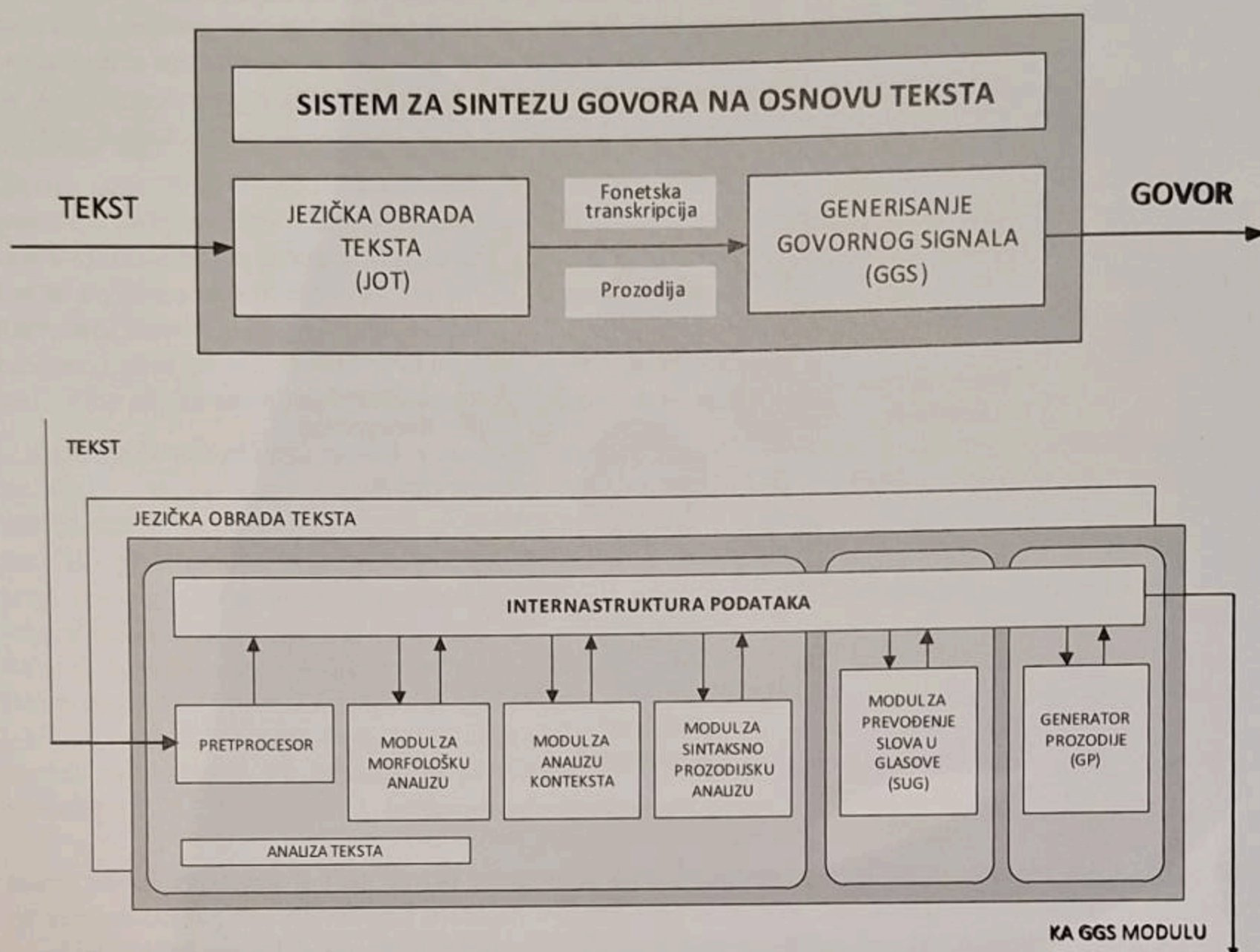


Slika 2. Govorna komunikacija čovek-mašina

Jedna od razlika između pisanog i govornog jezika je u tome što su reči u pisanom jeziku jasno razdvojene praznim mestom dok se u kontinualnom govoru najčešće izgovaraju bez pauza; zastane se obično na kraju rečenice, da govornik udahne vazduh i/ili razmisli šta će dalje da kaže. Dakle, u govoru se reči nižu u kontinuitetu, a čovek ih jasno uočava i razdvaja, ako je govor na njemu poznatom jeziku (na kom su mu te reči poznate) i ako prati kontekst tj. sluša šta se govori i sklapa zvučnu sliku u mozgu gde dolazi do konačnog razumevanja onog što se govori. Da bi čovek mogao da razaznaje reči u kontinualnom sintetizovanom govoru, one moraju zvučati prirodno kao da ih je čovek izgovorio. Ukoliko u sintetizovanom govoru nema dovoljno elemenata prirodne intonacije, otežano je razumevanje sintetizovanog govora. Pogrešna intonacija reči u kontinualnom govoru može dovesti do toga da čak ni jezik na kom se "govori" ne bude prepoznat, a pogrešno intonirane/akcentovane pojedine reči mogu da imaju drugačije značenje, te se menja smisao te reči pa i rečenice u celini. Ovo je posebno izraženo u tzv. tonalnim jezicima u koje spada i srpski, gde isto napisana reč ima različita značenja u zavisnosti od akcentovanja, npr. reči "kosa" i "grad".

Računarski generisana simulacija ljudskog govora predstavlja računarski ekvivalent ljudskog procesa čitanja naglas. Osnovni zadatak TTS sistema jeste da ispravno izgovara pojedine foneme, slogove, reči i rečenice, a to nije jednoznačan zadatak – isti tekst se može pročitati sa raznim inonacijama, sa stavljanjem naglaska na različite reči u istoj rečenici. Problem je što u pisanom tekstu nisu označeni

akcenti, kao ni trajanje ni način izgovora pojedinih fonema, slogova i reči. TTS sistem mora na osnovu analize "golog" teksta da zaključi kako će da ga izgovori. Pri analizi teksta, prvo se vrši predobrada u kojoj, između ostalog, TTS sistem zaključuje kako će da izgovori niz cifara jer nije sve jedno da li je to datum ili novčani iznos, redni ili običan broj. Npr. "2 dečaka i 2 devojčice" treba da se pročita kao "dva dečaka i dve devojčice". Brojni ovakvi i slični problemi se rešavaju u toku predobrade teksta – pretprocesor na sl. 3. Nakon kompletne jezičke obrade teksta (*front end* TTS) dobija se tzv. uska fonetska transkripcija koja vokoderu (*back end* TTS) daje više informacija o tome kako treba izgovoriti dati fonem, slog, reč i rečenicu – sa kakvim trajanjem, intonacijom i drugim tzv. prozodijskim elementima, slika 3. Pod prozodijom možemo da podvedemo sve ono čega u tekstu nema a u govoru postoji. Front end (JOT) baziran je na NLP resursima (rečnici, govorne baze) i znatno je više jezički zavisen od modula za generisanje govornog signala (GSS) na osnovu uske fonetske transkripcije.

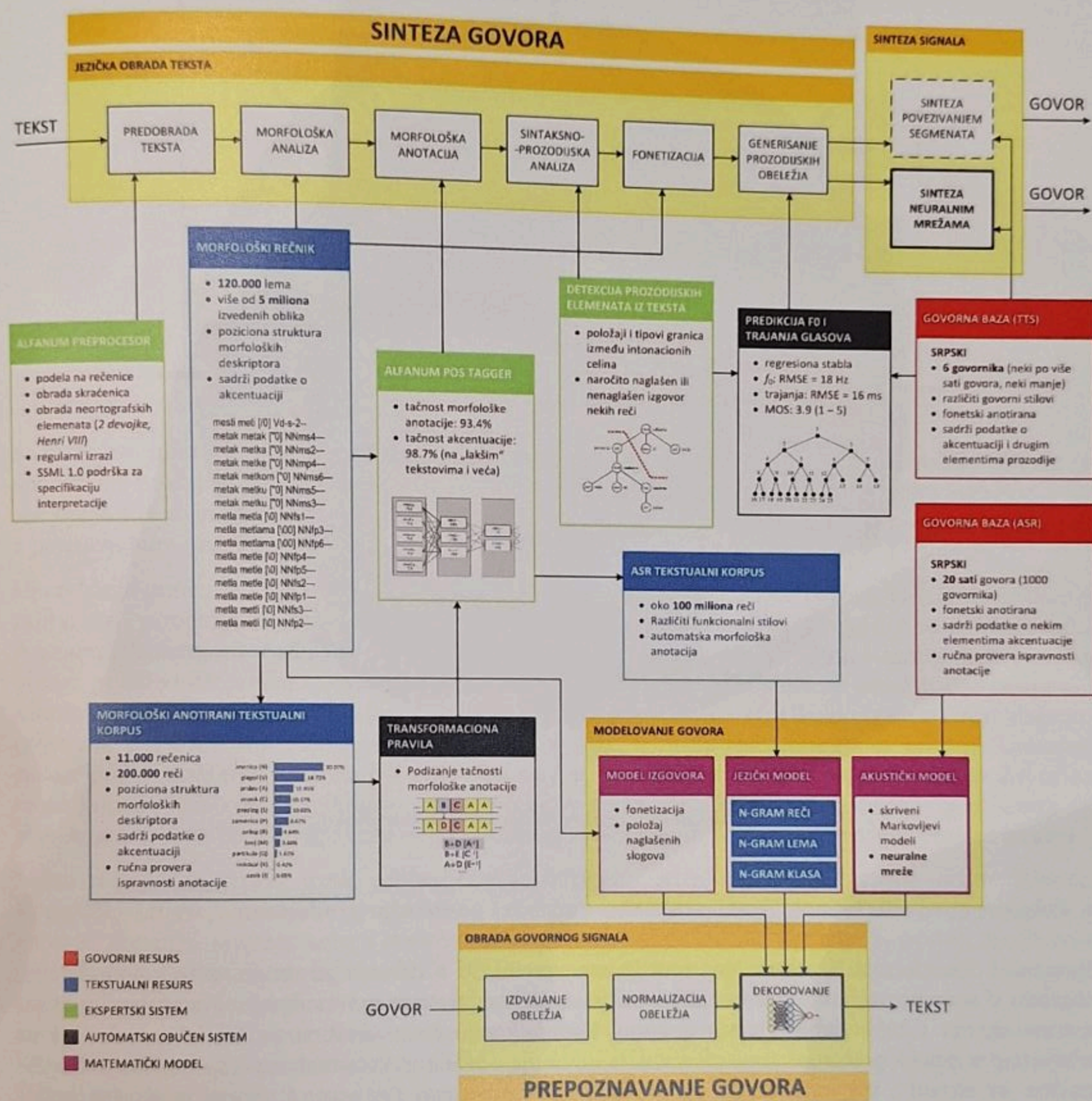


Slika 3. Struktura TTS sistema sa više detalja koji se odnose na jezičku obradu teksta

Problematika ASR sistema je kompletno drugačija. Kod TTS-a je bio zadatak da se na goli tekst doda mnoštvo prozdijskih informacija koje će se naći u govoru. Kod ASR potrebno je iz mnoštva informacija u govornom signalu eliminisati sve, a ostaviti samo niz fonema tj. slova uklopljenih u reči datog jezičkog modela. Razne varijabilnosti u govoru posledica su različitog glasa pojedinih govornika, pa čak se i istom govorniku menja glas u zavisnosti od raspoloženja, umora, sa starenjem i sl. Takođe, ambijentalna buka i reverberacija mogu da budu veoma različitog nivoa. Uticaj pozicije i kvaliteta mikrofona, komunikacionog kanala (da li je npr. govor prvo prenet telefonskom vezom ili je direktno sa mikrofona dospeo do ASR sistema). Sve ove varijabilnosti čine automatsko prepoznavanje govora jednim od najtežih zadataka sa kojim su se suočili istraživači. Za ASR jedino su relevantne varijabilnosti i akustičke razlike u izgovoru pojedinih fonema, a sve ostale pomenute varijabilnosti samo otežavaju zadatak ASR-a. Dugi niz godina ASR sistemi su bazirani na stohastičkim modelima kao što su skriveni Markovljevi modeli, ali ključni progres u razvoju govornih tehnologija baziran je

na paradigmi dubokih neuronskih mreža (eng. *Deep Neural Networks* – DNN) koje su u osnovi većine danas korišćenih ASR i TTS sistema [1]. Preduslov za ovaj progres u tačnosti i robusnosti ASR treniranje DNN obezbede ogromne baze anotiranih govornih i jezičkih resursa za dati jezik, kao i teorija NN bila poznata decenijama ranije.

Da bi se razvile govorne tehnologije za jedan jezik, neophodno je razviti brojne NLP resurse, među kojima su anotirane govorne baze, tekstualni resursi kao što je akcenatsko-morfološki rečnik i vrsta reči (morfologija i akcentovanje) i detekciju prozodijskih elemenata iz teksta, pa automatski prepoznavanje govora – akustički, leksički i jezički model, sl. 4.

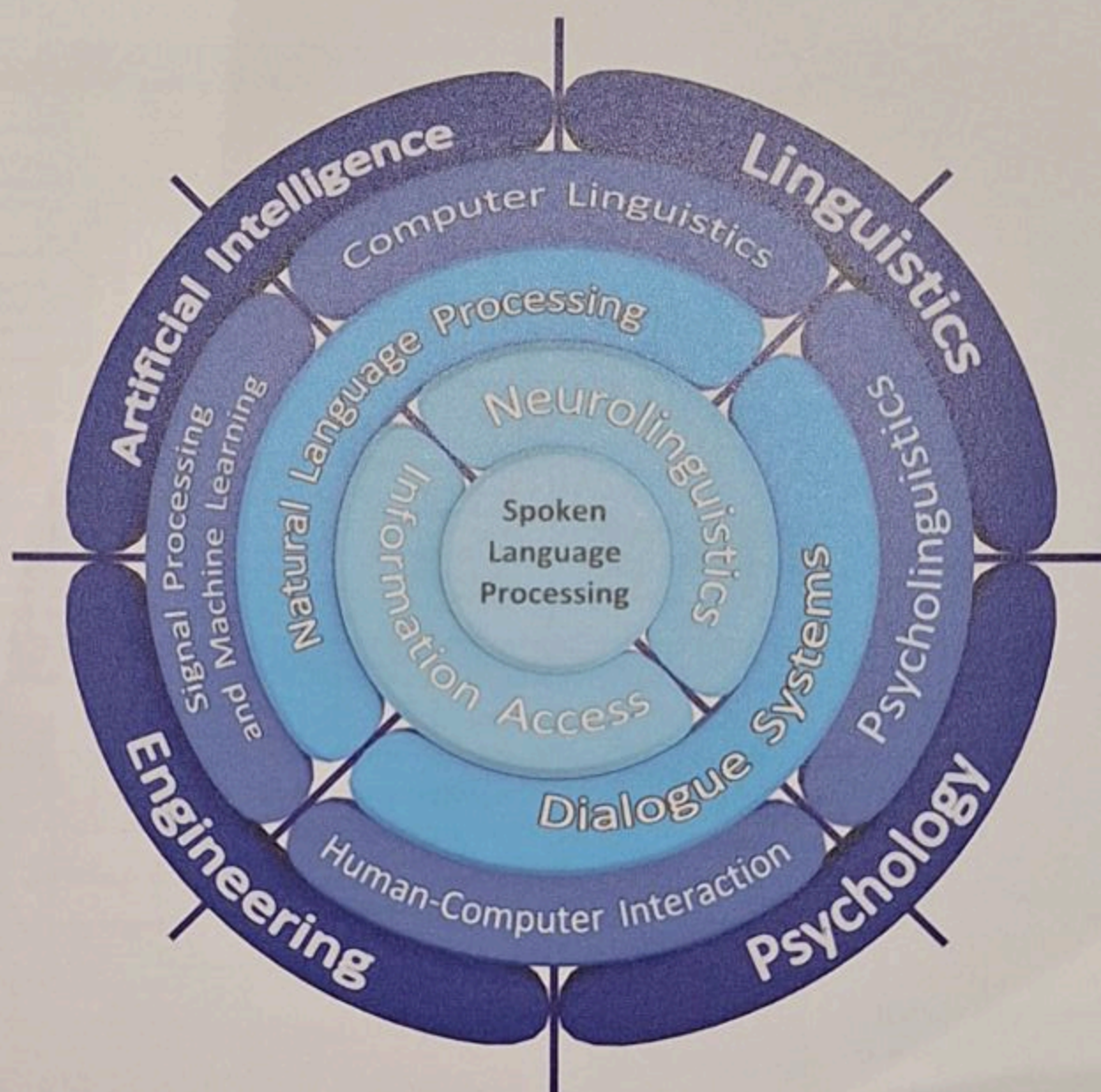


Slika 4. NLP resursi i moduli u prepoznavanju (ASR) i sintezi (TTS) govora

Zahvaljujući nizu projekata na FTN i u preduzeću AlfaNum u Novom Sadu, zaključno sa projektom S-ADAPT (2020-2023), razvijene su visokokvalitetne govorne tehnologije ASR i TTS za srpski jezik.

Na AI projektu S-ADAPT kod Fonda za nauku unapređena je tačnost i robusnost ASR, kao i prirodnost i fleksibilnost TTS – dobrim delom na bazi adaptacije izabranih algoritama koji su originalno uspešno primenjivani za obradu slike. Naredni korak – projekat AI-SPEAK (2023-2026) – usmeren je na razvoj višejezičkih multimodalnih dijaloških sistema, koji će pored govora i teksta koristiti i druge modalitete (slika, video, dodir) kako bi se dalje unapredila govorna komunikacija čovek-mašina.

Istraživanje i razvoj govornih tehnologija zaista su interdisciplinarna oblast. Obrada prirodnog govora je u preseku oblasti lingvistike, psihologije, inženjerstva i veštačke inteligencije, sl. 5. Sa jedne strane obrada prirodnog jezika (NLP) kao oblast veštačke inteligencije bazira se i na lingvistici i na inženjerstvu, a sa druge strane dijaloški sistemi koji omogućuju interakciju čoveka i mašine zadiru u oblast psiholingvistike i inženjerskih disciplina.



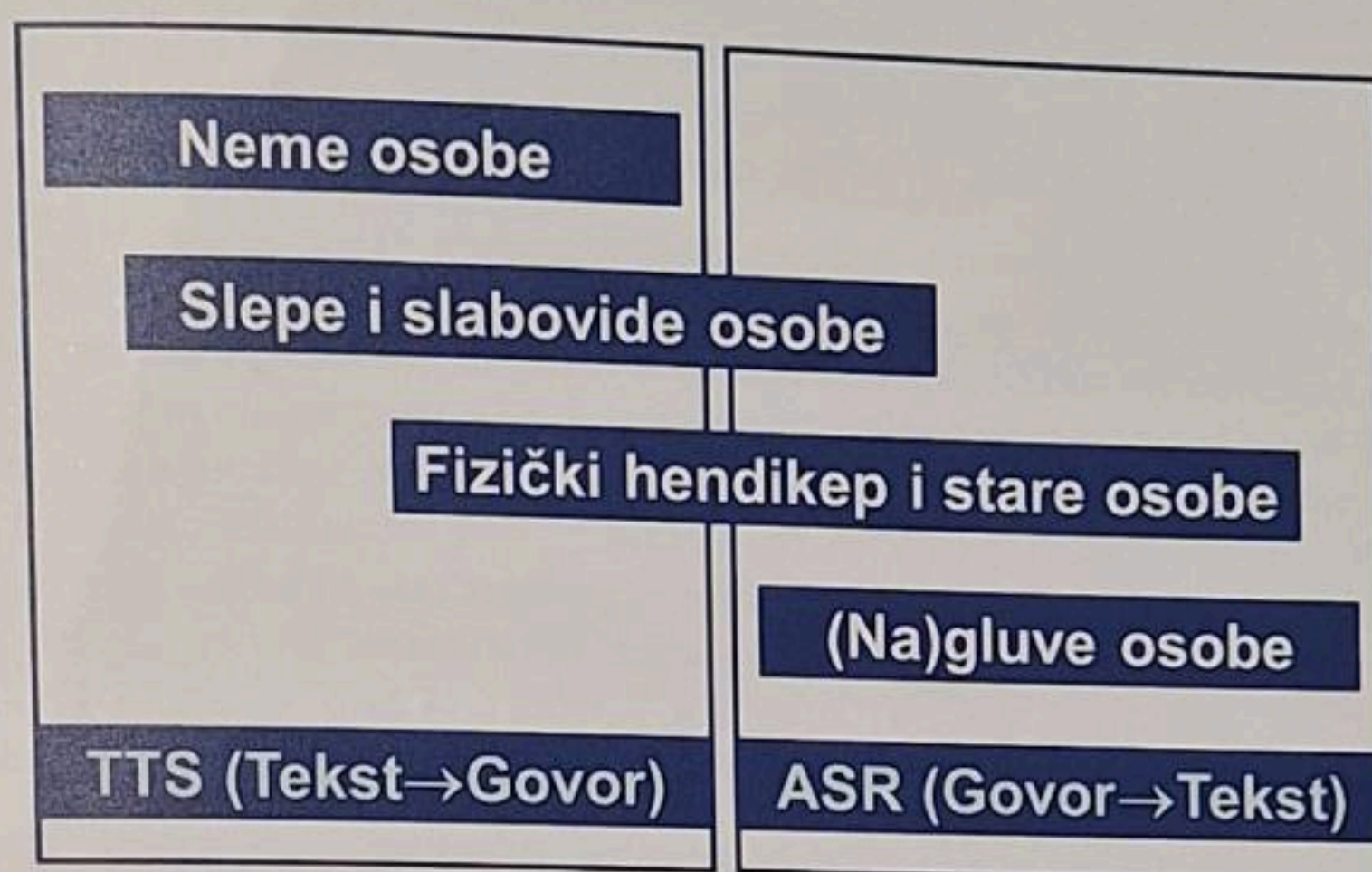
Slika 5. Interdisciplinarni karakter govornih tehnologija

4. Primene govornih tehnologija u pametnim kućama i pametnim gradovima

Govorna komunikacija čovek-mašina ima izvesne prednosti u odnosu na druge vrste komunikacije, posebno u situacijama u kojima su ruke i oči zauzete; za govornu komunikaciju čovek koristi samo govorni aparat i čulo sluha. Uz to, govorna komunikacija čovek-mašina može da se odvija i sa udaljenog mesta – preko telefona. Razvoj tehnologije ide na ruku govornoj komunikaciji čovek-mašina jer ekrani i tastature ne mogu biti mnogo manji nego što jesu, a procesori su sve brži i memorija je sve veća, a i komunikacioni linkovi su sve brži – sve ono što je potrebno za implementaciju računarski zahtevnih govornih tehnologija kako bi funkcionisale u realnom vremenu.

Prve primene govornih tehnologija u Srbiji i regionu bile su vezane za osobe sa invaliditetom. Prvo su slepe osobe pre dvadesetak godina počele da koriste AlfaNum TTS za srpski dok je još bio tzv.

konkatenativni sintetizator sa elementima prirodne intonacije, ali se teško menjao i unapređivao. Danas se AlfaNum TTS bazira na parametarskoj sintezi pomoću DNN što je značajno olakšalo promenu glasa i stila govora, unapredilo prirodnost i proširilo opseg primene – od ozvučenih web sajtova na kojima se dati tekstovi mogu preslušavati sintetizovanim govorom (npr. RTS), do primena u izdavanju audio-knjiga i kao pomagala za širi krug osoba sa invaliditetom: TTS za osobe koje ne vide, ne mogu da govore, ne mogu da drže knjige u ruci, ili ASR za osobe koje govornim komandama upravljaju uređajima u svom okruženju, osobe koje ne čuju itd, sl. 6. Govorne tehnologije omogućuju osobama sa invaliditetom brže i jednostavnije obavljanje kako svakodnevnih poslova, tako i komplikovanih zadataka čime se značajno podiže kvalitet njihovih života.



Slika 6. Primene govornih kao asistivnih tehnologija

Sa razvojem informaciono-komunikacionih tehnologija nastaju mogućnosti implementacije sistema za upravljanje uređajima u poslovnim okruženjima kao i u domaćinstvima, za međusobnu interakciju tih uređaja, a sve sa ciljem efikasnijeg obavljanja svakodnevnih poslova. U nastavku ovog rada biće reči o potencijalnim primenama govornih tehnologija u pametnim kućama i pametnim gradovima.

Upravljanje uređajima u pametnim kućama glasom danas se najčešće vrši preko virtuelnih govornih asistenata i pametnih zvučnika kao što su pomenuti proizvodi kompanija Google, Apple, Microsoft i Amazon. Pametni zvučnici (Apple HomePod, Amazon Echo, Google Home) sadrže mikrofonske nizove preko kojih primaju govorne komande (ASR), a sam zvučnik služi za saopštavanje povratnih informacija (TTS) ili reprodukciju tražene muzike, čitanje vesti i sl. Ukoliko nije na raspolaganju prijemnik govornih komandi u vidu mikrofonskog niza, moguće rešenje je da se govorne komande zadaju preko mobilnog telefona čiji mikrofon nam je uvek pri ruci i na raspolaganju. Svakako se ne može koristiti samo jedan mikrofon, udaljen od govornika, jer u njega stiže znatno više refleksija sa svih strana tj. reverberantnog zvuka nego direktnog zvuka koji treba da bude predmet prepoznavanja.

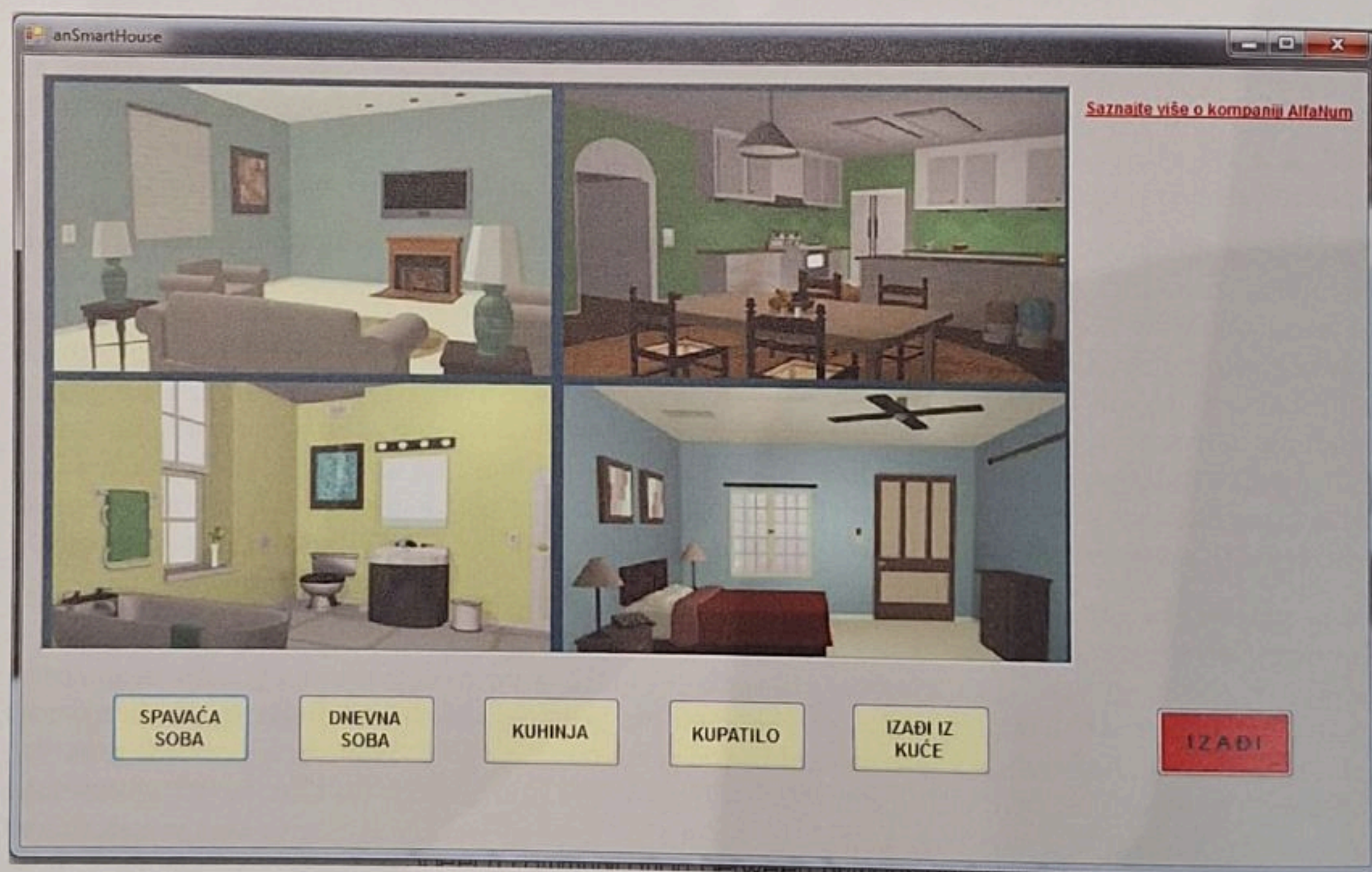
Jedno rešenje za srpski jezik napravljeno je u okviru AlfaNum demo sistema anSmartHouse. Korisnički interfejs sastoji se iz grafičkog prikaza prostorija jednog tipičnog stambenog objekta – kuhinje, kupatila, spavaće sobe i dnevne sobe, sl. 7. Za svaku od ovih prostorija postoji mogućnost upravljanja određenim uređajima, sl. 8. Aplikacija omogućuje jednostavno dodavanje i uklanjanje pojedinih uređaja iz skupa sa kojim se upravlja u centralnom sistemu. Upravljanje se vrši pomoću kontrolnih tastera ili izgovaranjem odgovarajućih govornih komandi (predviđen je niz sinonima). Nakon izvršenja bilo koje komande grafički se prikazuju izmene nastale u prostoriji koja je trenutno u fokusu grafičkog korisničkog interfejsa.

Moguće je vršiti podešavanja i u prostorijama koje su trenutno van fokusa, odnosno koje nisu prikazane pomoću grafičkog interfejsa u datom trenutku (naravno, u takvim slučajevima potrebno je u govornoj komandi naglasiti u kojoj se prostoriji nalazi uređaj kojim upravljamo). Promena fokusa na novu prostoriju se takođe vrši kontrolnim tasterima ili odgovarajućim govornim komandama.

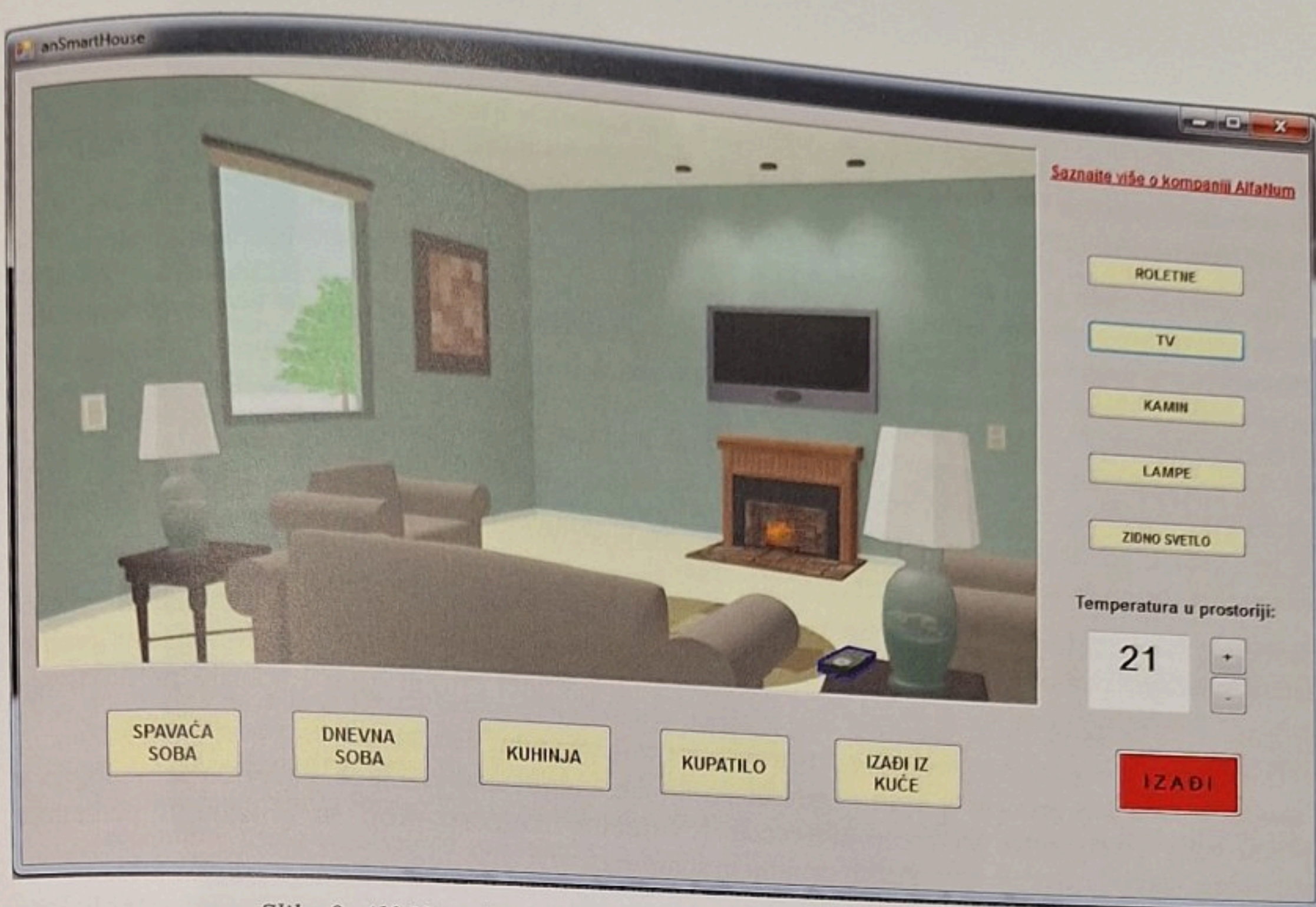
U svakom trenutku korisnik može, pomoću govornih komandi, proveriti stanje bilo kog uređaja u bilo kojoj prostoriji. Takođe, jednom komandom prilikom izlaska iz doma korisnik može isključiti sve uređaje i izvršiti neophodna podešavanja radi bezbednosti. Pored navedenih mogućnosti, postoji i mogućnost adaptacije prostorija za određene potrebe (poput večernjeg opuštanja – kada želimo da imamo spuštene roletne, uključen televizor, upaljenu vatru u kaminu i eventualno neka prigušena svetla) – sve jednom govornom komandom, koja podrazumeva promenu stanja više uređaja odjednom.

U glasu govorne komande može se prepoznati i govornik, pa tako kontrolisati koje komande su dozvoljene deci, odraslima i starim osobama u kući. Uz to, pametni sistem može u glasu govornika razaznati i njegovo raspoloženje i prilagoditi tok dijaloga čovek-mašina, a najnapredniji sistemi mogu da uče navike korisnika pomoću veštačke inteligencije i da svakom izlaze maksimalno u susret. Ove mogućnosti personalizacije posebno su interesantne ukoliko se koriste govorne komande i vrši govorna komunikacija čovek-mašina tj. čovek-pametna kuća.

Pored raznovrsnih primena u pametnim kućama, govorne tehnologije mogu značajno da unaprede brojne servise u pametnim gradovima. Ovde će biti pomenute samo neke mogućnosti povezane s aplikacijama proširene i virtuelne realnosti namenjene turistima. Naime, umesto klasičnih audio-vodiča, tj. preslušavanja unapred pripremljenih i snimljenih audio sadržaja o pojedinim spomenicima, građevinama, eksponatima u muzejima i sl. sada postoje fleksibilne aplikacije koje se adaptiraju na razne korisnike. Na primer, preko aplikacije na mobilnim telefonima, otac, majka i dete mogu da slušaju različitu priču o nekom eksponatu u muzeju – prilagođeno njihovim interesovanjima, sl. 9. Pored obima i nivoa sadržaja lako se može ponuditi sintetizovan govorni sadržaj na različitim jezicima, sa različitim glasovima spikera, stilovima govora... Često su audio i video sadržaji o nečemu u pametnom gradu dostupni na licu mesta preko QR koda koji ih povezuje na odgovarajući web sajt sa tim audio i video sadržajima.



Slika 7. AlfaNum demo anSmartHouse – primer osnovnog ekrana



Slika 8. AlfaNum demo anSmartHouse – dodatne opcije u dnevnoj sobi



Slika 9. Adaptirane primene audio-vodiča kroz muzeje

Napredne aplikacije proširene i virtuelne realnosti omogućuju, na primer, dok gledaju neku građevinu kroz mobilni telefon da turisti vide i virtuelnog viteza koji se pojavljuje između njih i tog građevinskog objekta – vitez im „uživo“ priča i objašnjava šta se tu dešavalo u nekom momentu u istoriji i slično – slika 10.

5. Zaključak

Buran razvoj u oblasti obrade prirodnog jezika zasnovan na velikim jezičkim modelima praćen je i progresom u razvoju i primeni govornih tehnologija. Svi ovi sistemi se razvijaju pomoću algoritama mašinskog učenja, dubokog učenja i naprednih arhitektura i sistema veštačke inteligencije.

Fokus u ovom radu je na mogućnostima primene govornih tehnologija u pametnim kućama. Sa više detalja prikazan je primer jedne takve aplikacije na srpskom jeziku – anSmartHome, koju karakteriše jednostavan i intuitivan dijaloški sistem, visok kvalitet sintetizovanog govora kojim se korisniku saopštavaju tražene informacije, visoka tačnost prepoznavanja govora, kao i jednostavan grafički korisnički interfejs. Govorne tehnologije se mogu iskoristiti i za dodatnu personalizaciju proizvoda, što se pokazalo kao izvanredan marketinški potez kod velikih svetskih kompanija. Naime, uz automatsko prepoznavanje govornih komandi, zahvaljujući veštačkoj inteligenciji ovi sistemi mogu da prepoznaju govornike i njihovo raspoloženje i da u skladu sa tim prilagode tok dijaloga i pruže personalizovane servise.

Osobe sa invaliditetom su populacija kojoj tehnologija pametnih kuća sa govornim tehnologijama donosi značajnu pomoć u savladavanju mnogih svakodnevnih problema, čime se poboljšava kvalitet njihovih života, one postaju u većoj meri nezavisne od pomoći drugih ljudi i postaju produktivniji članovi društva.

Takođe je kroz par primera u radu nagoveštena široka mogućnost primene govornih tehnologija u aplikacijama i servisima proširene stvarnosti i virtuelne realnosti, koji se pružaju u pametnim gradovima, kako turistima, tako i svim drugim građanima.

Razvoj govornih tehnologija za srpski i širenje njihove primene predstavlja rezultat od nacionalnog značaja jer predstavlja doprinos očuvanju srpskog jezika u eri sve češće govorne komunikacije sa uređajima i servisima. Smanjenje tehnološke zavisnosti od globalnih AI kompanija i raspolaganje *in-home* rešenjima posebno je značajno za državnu upravu, sudstvo i zdravstvo, ali i za privatne pametne kuće jer neko ne želi da snimak onog što govori ide na internet da bi tamo bio prepoznat i pretvoren u tekst.



Slika 10. Govorne tehnologije u aplikacijama virtuelne realnosti na mobilnim telefonima

Literatura

- [1] Vlado Delić, Zoran Perić, Milan Sečujski, Nikša Jakovljević, Jelena Nikolić, Dragiša Mišković, Nikola Simić, Siniša Suzić, Tijana Delić: "Speech technology progress based on new machine learning and Adaptive Learning Methods", *Computational Intelligence and Neuroscience*, Special Issue "Advanced Signal Processing June 2019", DOI: 10.1155/2019/4368036, Hindawi, Vol. 2019, Article ID 4368036, 19 pages,
- [2] Edvin Pakoci, Branislav Popović, Darko Pekar: "Using Morphological Data in Language Modeling for Serbian Large Vocabulary Speech Recognition", *Computational Intelligence and Neuroscience*, ISSN: 1687-5265, Hindawi, Vol. 2019, Article ID 5072918, 8 pages, 2019, DOI: 10.1155/2019/5072918
- [3] Tijana Nosek, Siniša Suzić, Darko Pekar, Radovan Obradović, Milan Sečujski, Vlado Delić: "Cross-lingual neural network speech synthesis based on multiple embeddings", *The Int. Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI*, pp. 110-120, 2021, DOI: 10.9781/ijimai.2021.11.005
- [4] Branislav Popović, Lenka Cepova, Robert Cep, Marko Janev, Lidija Krstanović: "Measure of Similarity between GMMs by Embedding of the Parameter Space That Preserves KL Divergence", *Mathematics*, ISSN: 2227-7390, Vol 9, Issue 9, 957; 21 pages, 25 April 2021, DOI: 10.3390/math9090957
- [5] Milan Sečujski, Darko Pekar, Siniša Suzić, Anton Smirnov, Tijana Nosek: "Speaker/style-dependent neural network speech synthesis based on speaker/style embedding", *Journal of Universal Computer Science*, vol. 26, no. 4, pp. 434-453, 2020
- [6] Siniša Suzić, Tijana Delić, Darko Pekar, Vlado Delić, Milan Sečujski: "Style transplantation in neural network based speech synthesis", *Acta Polytechnica Hungarica, Journal of Applied Sciences*, 16(6):171-189, September 2019, DOI: 10.12700/APH.16.6.2019.6.11
- [7] Nikola Simić, Siniša Suzić, Tijana Nosek, Mia Vujović, Zoran Perić, Milan Savić, Vlado Delić: "Speaker recognition using constrained convolutional neural networks in emotional speech", *Entropy – an Open Access Journal by MDPI*, Special Issue "Methods in Artificial Intelligence and Information Processing", ISSN: 1099-4300, 24(3), Article ID 414, 16.03.2022. DOI: 10.3390/e24030414
- [8] Lidija Krstanović, Branislav Popović, Marko Janev, Branko Brkljač: "Bootstrapped SSL CycleGAN for Asymmetric Domain Transfer", *Applied Sciences*, Special Issue "Computer Vision-Based Intelligent Systems: Challenges and Approaches", António J. R. Neves (Ed.), ISSN: 2076-3417, MDPI, Vol. 12, No. 7: 3411, pp. 1-17, 2022, DOI: 10.3390/app12073411
- [9] Branislav Popović, Marko Janev, Lidija Krstanović, Nikola Simić, Vlado Delić: "Measure of Similarity between GMMs Based on Geometry-Aware Dimensionality Reduction", *Mathematics*, Special Issue "Artificial Intelligence and Mathematical Methods", ISSN: 2227-7390, Vol 11, Issue 1, 175, 22 pages, 29 December 2022, DOI: 10.3390/math11010175
- [10] Lidija Krstanović, Branislav Popović, Marko Janev, Branko Brkljač: "Feature Map Regularized CycleGAN for Domain Transfer", *Mathematics*, Special Issue "Artificial Intelligence and Mathematical Methods", ISSN: 2227-7390, Vol 11, Issue 2, 372, 21 pages, 10 January 2023, DOI: 10.3390/math11020372

Zahvalnica: Najzaslužniji članovi istraživačko-razvojnog tima su doktorirali u oblasti govornih tehnologija i dalje su na FTN-UNS ili u preduzeću AlfaNum: dr Milan Sečujski, dr Branislav Popović, dr Darko Pekar, dr Nikša Jakovljević, dr Dragiša Mišković, dr Edvin Pakoci, dr Siniša Suzić, dr Nikola Simić i dr Tijana Nosek.



Vlado Delić je doktorirao 1997. godine na Univerzitetu u Novom Sadu gde je na Fakultetu tehničkih nauka okupio tim istraživača koji je izgradio poziciju lidera u razvoju govornih tehnologija za srpski i srodne južnoslovenske jezike. Iz tog tima je poniklo preduzeće AlfaNum – regionalni lider u primeni ASR i TTS. Prof. Delić je obogatio studentske programe na FTN predmetima iz oblasti akustike, obrade audio signala, kao i govornih tehnologija. Naučno-istraživački rad praćen je sa preko 350 publikacija i na desetine tehničkih rešenja i патената. Kroz niz projekata koje je vodio uspostavio je saradnju u regionu i šire u Evropi. Doprinos naučnoj i društvenoj zajednici (organizacije konferencija, časopisa, projekata za osobe sa invaliditetom i sl.) praćen je sa većim brojem priznanja.

**SERIJA STRUČNIH PREDAVANJA
POSVEĆENIH UNAPREĐENJU
PROJEKTOVANJA TELEKOMUNIKACIONIH
MREŽA I SISTEMA**

ZBORNİK RADOVA 2022/2023

**Urednik
Mirjana Jarić-Ćirić**



FTTH udruženje Srbije

Urednik

Mirjana Jarić - Ćirić

Uređivački odbor

Prof.dr Slobodan Jovičić
Prof. dr Vlado Delić
Prof dr Nedžad Hadžiefendić
Prof. dr Predrag Milošević
Prof. SS dr Goran Zajić
Prof.SS dr Vladimir Stanojević
Mr Milenko Nikolić
Mr Marija Zajeganović
Mr Vladan Cvejić
Mr Dragorad Milovanović
Mr Dejan Todorović
Mr Milan Perać
Bojan Papić
Ivana Lacković
Danijela Đokić
Goran Medic
Dragan Vukojević
Predrag Bogdanović
Nebojša Filipović

Recenzenti

Prof. dr Isak Karabegović, akademik ANU BiH
Prof. dr Nikša Jakovljević
Prof. SS dr Sanja Jevtić

Izdavač

FTTH Udruženje Srbije
Zdravka Čelara 16, 11000 Beograd

Štampa

JP Službeni Glasnik
Jovana Ristića 1, 11000 Beograd

Tiraž

100 primeraka

ISBN – 978-86-905152-1-9

Izdavač zadržava sva prava.
Reprodukcija pojedinih delova ili celine ove publikacije nije dozvoljena.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

621.39(082)
004.7(082)

ZBORNİK radova : 2022/2023 : ciklus serije stručnih predavanja posvećenih
unapređenju projektovanja telekomunikacionih mreža i sistema / urednik
Mirjana Jarić-Ćirić. - Beograd : Udruženje FTTH Srbije, 2024 (Beograd : Službeni
glasnik). - XIV, 213 str. : ilustr. ; 30 cm + 1 elektronski optički disk (CD-ROM ; 12 cm)

Tiraž 100. - Reč urednika: str. V-VIII. - Str. IX-XIII: Reč recenzenata / Isak
Karabegović, Nikša Jakovljević, Sanja Jevtić. - Napomene i bibliografske reference
uz radove. - Bibliografija uz većinu radova.

ISBN 978-86-905152-1-9

a) Телекомуникациони системи -- Зборници b) Телекомуникационе мреже
-- Зборници v) Рачунарске мреже -- Зборници

COBISS.SR-ID 144153609
