

ПРОЈЕКТНИ ЗАДАТАК

Развој говорних технологија које омогућују говорну комуникацију човек-машина на српском језику:

- ASR – аутоматско препознавање говора
- TTS – синтеза говора на основу текста

ЦИЉЕВИ ПРОЈЕКТА:

- Повећавати квалитет TTS на српском језику
- Повећавати тачност ASR на српском језику
- Проширити развојне потенцијале и ресурсе за развој и примену говорних технологија у конкурентним иновативним производима.
- Учврстити лидерску позицију у развоју и примени говорних технологија у региону и омогућити прве примене ASR и TTS.

УКРАТКО О ГОВОРНИМ ТЕХНОЛОГИЈАМА

ASR – говорна технологија помоћу које рачунар разуме говорне команде, а у развијенијој форми може да преводи говор у текст. TTS – говорна технологија помоћу које рачунар може усмено да одговара човеку претварањем текста у синтетизовани говор. Говорне технологије се морају посебно развијати за различите језике. То је до сада успешно урађено за мали број језика у свету. Ми смо до сада постигли најзначајније резултате на подручју јужнословенских језика. Резултатима на овом пројекту омогућили смо развој иновативних и корисних производа и на српском говорном подручју, почев од помагала за особе са инвалидитетом, преко телефонских говорних портала, уређаја за претрагу аудио снимака, па до аутоматског превођења говора.



Основни проблеми препознавања говора јесу велика варијабилност говорног сигнала: различити говорници, различит ниво амбијенталне буке, различит положај и квалитет микрофона, разни комуникациони канали. Због свега тога говорни сигнал који долази до рачунара може да изгледа сасвим различито за исте изговорене речи и реченице. Поред тога што је речено (лингвистичке информације) могу се уочити карактеристике гласа – ко је и како то изговорио. Ту су могућности идентификације и верификације гласа говорника, као и његових емоција (да ли је задовољан или збуњен и сл.). Све ове информације могу се искористити за унапређење квалитета дијалога човек-машина. Други смер говорне комуникације човек-машина базиран је на TTS синтези говора. Основни проблеми синтезе говора састоје се у томе што у тексту који треба претворити у синтетизован говор нема информација о томе како треба изговорити поједине гласове, речи и реченице. За сваки глас треба аутоматски одредити колико треба да траје при изговору, да ли је у наглашеном слогу или не, како га интонирати у току непознатог трајања, па сви ти тзв. прозодијски елементи на нивоу гласа, слога, речи и реченице. Свака грешка у препознавању говора може да одведе дијалог човек-машина у погрешном смеру, а погрешно интониране речи од стране синтетизатора говора могу да збуне слушаоца.

ASR и TTS спадају у најтеже мултидисциплинарне проблеме са којима се истраживачи данас суочавају. Захваљујући овом пројекту српски језик је сврстан у узак круг језика за које су ASR и TTS достигли висок степен развоја и проналазе прве примене.

АУТОМАТСКО ПРЕПОЗНАВАЊЕ ГОВОРА

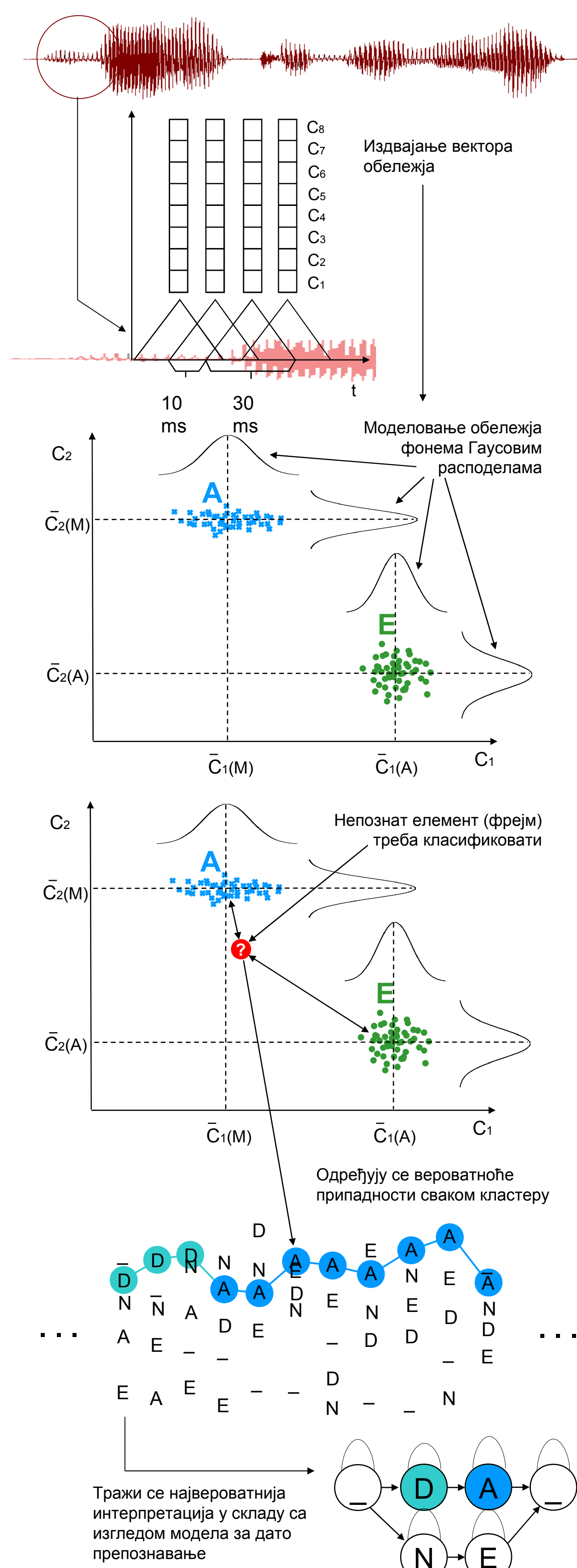
Аутоматско препознавање говора (ASR) спада у највеће техничке изазове савременог доба и већ више од пола века заокупља пажњу истраживача широм света. Задатак аутоматског препознавања говора је да на основу улазних података у виду звучног записа неке говорне целине (речи или реченице) да излаз у виду текстуалног записа. На тај начин се говор практично конвертује у текст, односно, "препознаје се" шта је одређени говорник изговорио.

Аутоматско препознавање говора реализује се тако што се из звучног записа издвајају одређена обележја говорног сигнала, која се затим пореде са унапред припремљеним референтним узорцима или моделима. Већ ту се поставља низ питања: која су то обележја релевантна за препознавање говора, како се она могу издвојити из говорног сигнала, како се затим припремају референтни узорци односно модели, како се врши поређење обележја говорне целине коју треба препознати са референтним моделима односно узорцима. Посебан проблем представља велика варијабилност говора, јер не изговарају све особе исту реч на исти начин, нити је једна иста особа изговара увек на исти начин. Из тих разлога се за препознавање говора користе системи засновани на тзв. скривеним Марковљевим моделима (Hidden Markov models) и/или вештачким неуронским мрежама.

Аутоматско препознавање говора одвија се у две фазе: обучавање и препознавање. Обучавање се врши у припремној фази, помоћу одговарајућих база говорних узорака. Ако систем мора посебно да се обучава за препознавање датог говорника, реч је о препознавању зависном од говорника. Ако, међутим, може успешно да препознаје велики број говорника чији гласови нису коришћени у фази обучавања, реч је о препознавању независном од говорника. Ова подела умногоме одређује и примену система. Наиме, уколико је реч о систему који је намењен препознавању гласова и оних особа које нису учествовале у обуци система, такав систем ће се моћи примењивати у интерактивним говорним сервисима, јер тамо систем мора подједнако добро да препознаје говор без обзира на то ко је позивалац. С друге стране, системи зависни од говорника углавном се примењују за персонализоване намене као што је диктирање текста.

Разликујемо ASR системе који препознају изоловане речи од система који могу да препознају и повезано изговорене речи. У оба случаја речи морају бити из унапред дефинисаног речника. Величина речника утиче на тачност препознавања - што је речник шири, вероватноћа грешке је већа. Посебан проблем јавља се у случају да у речнику постоје речи које се слично изговарају, али ово се при креирању апликације најчешће може избећи.

Системи до сада развијени у оквиру овог пројекта омогућују препознавање говора телефонског квалитета (у обуци је коришћена база гласова добијена преко јавне телефонске мреже). Препознавање у овим условима је тежи задатак, јер је део спектра говорног сигнала при проласку кроз телефонску мрежу оштећен, а и преостали део је значајно измењен у складу са карактеристикама канала као и других специфичних околности (квалитет микрофона, амбијентални шум...). У овим условима ASR системи развијени у оквиру овог пројекта постижу тачност од око 98% на реченицама од неколико стотина речи, што је за практичне примене и више него довољно.



СИНТЕЗА ГОВОРА НА ОСНОВУ ТЕКСТА

Задатак аутоматске синтезе говора на основу текста је да се текст записан у рачунару конвертује у говорни сигнал. Ова технологија се користи тамо где није могуће или је непрактично да говорник прочита текст и сними садржај.

Конверзија текста у говор спроводи се у две основне фазе: (1) анализа текста и (2) синтеза говорног сигнала. Ово је неопходно јер се у тексту не могу наћи подаци о томе како се која реч наглашава (што је у српском језику нарочито сложен проблем), како се у конкретном говорном сигналу акустички манифестује тај нагласак, како се висина гласа говорника мења на нивоу реченице, те колико треба да траје сваки од говорних сегмената – гласова. Ове податке је неопходно одредити, а законитости које међу њима владају нису довољно изучене ни за велике светске језике. Овде ће уkratко бити изложено решење које се користи у оквиру синтетизатора говора развијеног на Факултету техничких наука.

Први корак је акцентуација реченице. С обзиром на врло комплексну акценатску структуру српског језика, за ово је било потребно развити обиман акценатско-морфолошки речник који садржи преко 3.000.000 милиона изведених облика речи (око 100.000 основних облика). С обзиром на то да је језик жив и да се сваког дана појављују нове речи, овај речник се стално допуњава.

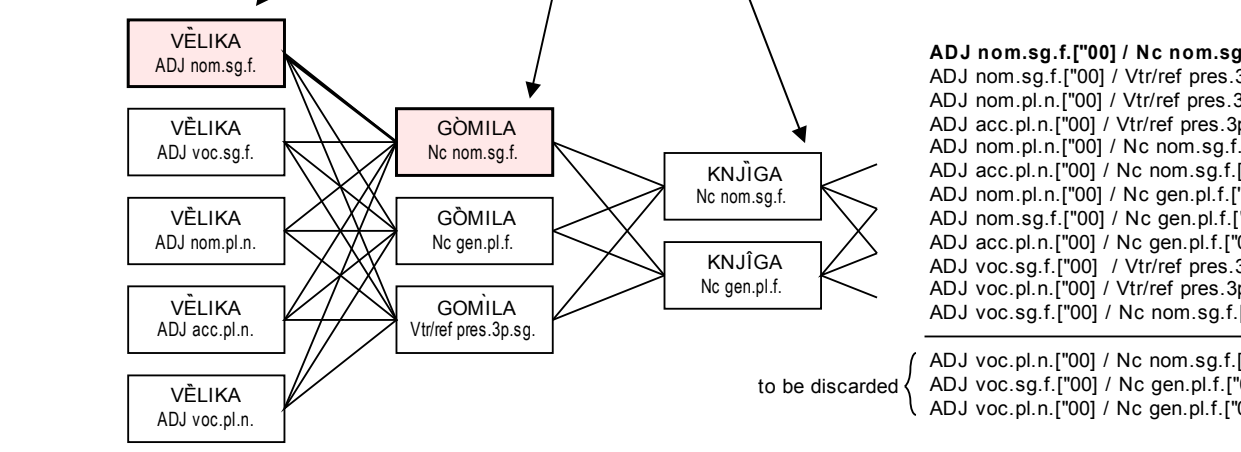
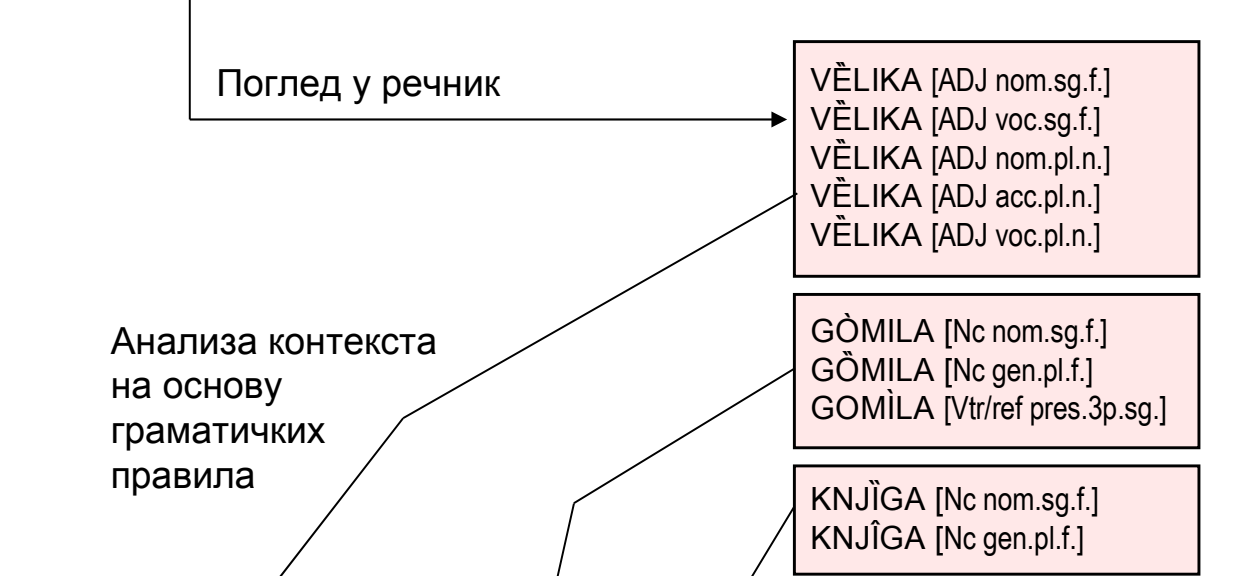
Неке речи се могу акцентовати на више начина, па је неопходно развити алгоритме анализе контекста, чиме би се утврдило која је од могућих варијаната она права. У приказаном примеру, реч "гомила" могла би бити именица, али и глагол. На граматичким правилима је да у оквиру имплементираних стек-алгоритама утврде да се у овој реченици ради о именици номинатива јединине, па да, према томе, она има краткоулазни нагласак на првом слогу. На исти начин се одређује акцентуација свих речи у реченици, и долази се до акценатске структуре читаве реченице, а мало детаљнијом анализом синтаксе и до података као што су места где би требало направити краћу паузу у говору.

Следећи корак је на основу акценатске конфигурације доћи до њене акустичке реализације. Ово је урађено под претпоставком да акцентуација доминантно утиче на кретање висине гласа. За акустичку манифестацију сваког типа акцента узима се реализација која је присутна у снимљеном говору, а за коју је утврђено да се може сматрати типичном. Добијене криве висине гласа за сваку реч се спајају, а затим се примењује нагиб у складу са типом реченице (нпр. константно опадање висине гласа у изјавној реченици).

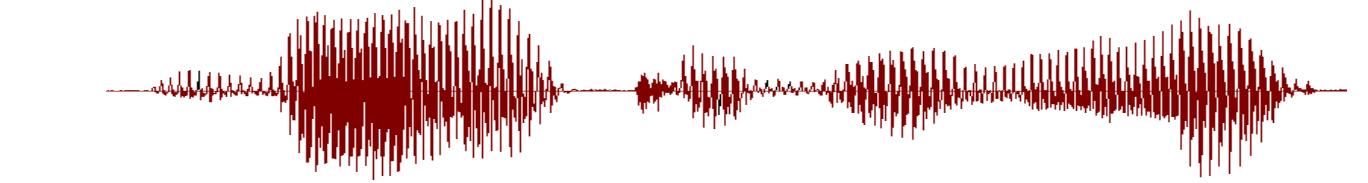
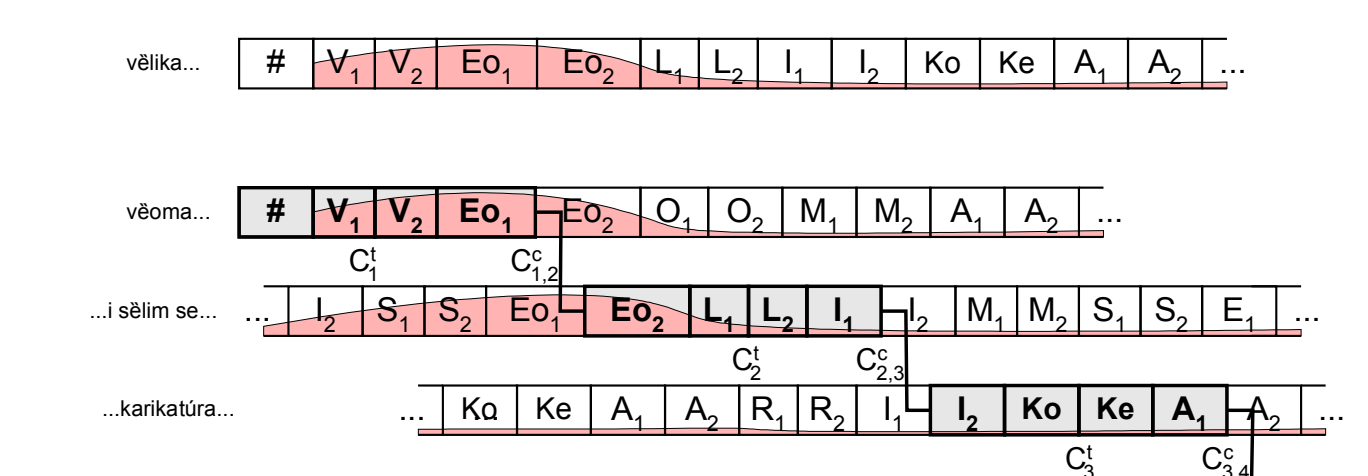
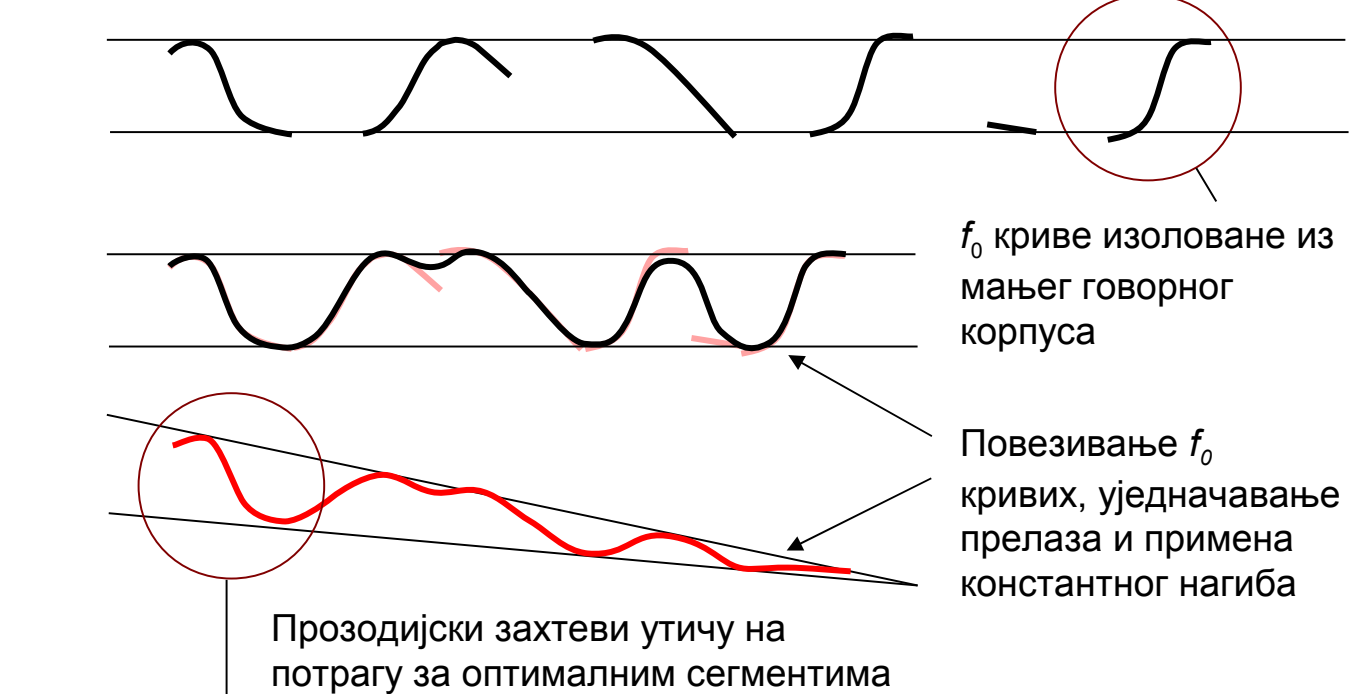
Сам говорни сигнал генерише се спајањем сегмената из говорне базе од око 2 сата снимљеног говора. Ова база је нарочито припремљена за то, сви гласови у њој, њих више десетина хиљада, пажљиво су прегледани и обележени у погледу степена артикулације и типа акцента. Помоћу алгоритма сличног Витербијевом, у бази се идентификују оптимални сегменти (низови гласова) чијим ће се повезивањем добити тражена говорна целина. При томе се настоји да се створи утисак да се ради о континуалном говору, односно, да прелази између појединих сегмената буду нечујни.

На овај начин ствара се говор на српском језику, сасвим разумљив и релативно природне интонације. Наравно, синтеза говора је и на светским језицима далеко од коначног решења, тако да и у овом систему има места за побољшавање квалитета синтезе. Ипак, за све наведене примене квалитет је сасвим задовољавајући.

Велика гомила књига стоји на столу.



Велика гомила књига стоји на столу.

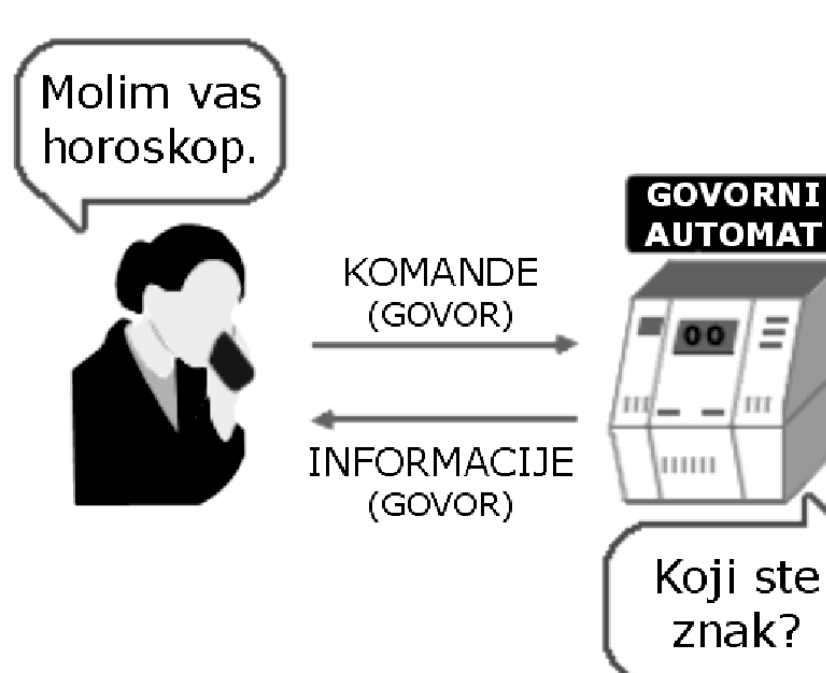


ПРИМЕНЕ И ТЕХНИЧКА РЕШЕЊА

Успешна имплементација обе говорне технологије омогућава успостављање у потпуности аутоматске двосмерне говорне комуникације између човека и машине. На овај начин се омогућује да људи на ефикасан и економичан начин долазе до жељених информација или иницирају одређене акције – слушајући информације саопштене синтетизованим говором, и издвајају одговарајуће команде својим гласом. На овај начин се, примера ради, штеди време оператера у позивним центрима и омогућује се њихово концентрисање на важније послове које није могуће обавити без помоћи оператера. Поред тога, веома значајне примене ових технологија су и помоћ особама са инвалидитетом, као и могућност оспособљавања неких од њих за обављање одређених послова, што овом пројекту осим технолошке и економске, даје и значајну социјалну димензију. Зависност од језика га карактерише као пројекат од националног значаја.

Неке од првих примена ових система које смо имплементирали обухватају следеће:
anReader – говорни софтвер за слепе. Овај софтвер омогућава слепој или слабовидној особи у потпуности самосталан рад на рачунару, чиме се у ери технолошког напретка оваквим особама омогућава да се равноправније информишу и образују, а самим тим оспособљавају за неки посао који би могле да обављају. Слепе особе захваљујући синтези говора имају могућност да им рачунар саопштава битне информације звучним путем, уместо да користе монитор као видеће особе.
Аудио библиотека за слепе и слабовиде. Овај систем представља складиште књига у електронском формату, којима слепе и слабовиде особе могу по потреби приступити и преслушавати их звучним путем, захваљујући интегрисаној могућности синтезе говора на основу текста. Овакав систем је већ имплементиран у неколико школа и организација слепих, а далеко је економичнији од библиотеке књига у Брајеовом запису, чија је штампа веома скупа.
Контакт – говорни портал за слепе. Овај систем користи могућност двосмерне комуникације човек-машина говорним путем, и намењен је пружању информација слепим особама. Реч је о дневним вестима, информацијама о правима слепих, помагалима за слепе, као и о забавним садржајима.
Састанак – говорни портал за упознавање. Овај систем је технолошки напреднија варијанта система за упознавање особа сличних интересовања, која гарантује позиваоцима дискрецију јер се читава комуникација (остављање личних података и претрага туђих профила) врши кроз аутоматску комуникацију с рачунаром, уместо са живим оператером. Један овакав систем већ више од годину дана успешно функционише на јавној телефонској мрежи у Србији.
Advertising Monitor – апликација која омогућује претрагу аудио снимка у потрази за дефинисаним аудио садржајем, као што је рекламни спот. На овај начин могуће је надзор радио-станица и ТВ канала и провера емитовања одређених рекламних блокова, али и других садржаја као што су музички.
Word Spotter – апликација која омогућује претрагу аудио снимка у потрази за кључним речима дефинисаним у текстуалном облику. Примене оваког система су многобројне, једна од најчешћих је press-clipping, али су могуће и примене у безбедносне сврхе.
ITEMA – вишејезични систем за приступ e-mail порука путем телефона, реализован у оквиру ЕУРЕКА пројекта, омогућава велике уштеде времена особама које би могле да преслушају пристигле поруке док се, примера ради, возе до канцеларије.
Говорни портали на аутобуским станицама у Крагујевцу и Бања Луци, као и TTS функционалност на сајту www.rtv.rs су иновације каквих нема у ширем региону. Достигнути квалитет TTS омогућио је реализацију "TTS портала" који ће свим сајтовима на српском језику омогућити једитну и ефикасну реализацију говорног портала. Са друге стране квалитет ASR-а на српском језику омогућује развој бројних иновативних апликација и решења.

Примене ових и сличних система су многобројне, и њихово време у земљама као што је наша свакако тек долази.



НАГРАДЕ И ПРИЗНАЊА

Поред научних резултата верификованих на бројним домаћим и међународним часописима (30) и конференцијама (85), пројекти развоја говорних технологија (ТР6144А и ТР11001) награђени су и бројним наградама:

- ★ Златна медаља "Архимед" за говорне технологије са највеће светске изложбе иновација у Москви (2008)
- ★ Контакт – пројекат социјалне интеграције у региону, награда ЕРСТЕ фондације (2008)
- ★ Повеља школе "Вељко Рамадановић" за допринос образовању ученика оштећеног вида (2007)
- ★ Диплома и златна медаља "Тесла Феста" од Савеза проналазача Војводине (2007)
- ★ Advertising Monitor – најбоља технолошка иновација на такмичењу у организацији Министарства науке (2006)
- ★ Повеља "Миша Анастасјевић" за иновације ТВ Нови Сад, ФТН и Медиа инвент (2005)
- ★ anReader – најбољи информатички производ – Плакета Друштва за информатику Србије (2005)

Поред наведених награда, ту су и бројне награде у периоду пре почетка пројекта технолошког развоја Факултета техничких наука у Новом Саду и партнерских организација уз подршку МНТР.