# Аутоматско читање текстова на оба писма српског језика

мр Милан Сечујски[1], др Владо Делић[1], Дарко Пекар[2], Радован Обрадовић[1]
[1]Факултет техничких наука, Универзитет у Новом Саду; [2]АлфаНум д.о.о, Нови Сад

Саопштење прочитано на стручнонаучном скупу
Савремене информационе технологије – Интернет и ћирилица
(Српски језик, писмо и култура у савременим информационим технологијама)
Бијељина, РС/БиХ, 25. новембра 2003. године

## 1. УВОД

Аутоматско читање текста од стране машине представља проблем који већ вековима заокупља истраживаче широм света, а појавом савремених рачунара дошло се и до првих применљивих резултата у тој области. Реч је, очигледно, о технологији која је у великој мери зависна од конкретног језика. Због мултидисциплинарности читавог проблема, чије успешно решавање захтева познавање наука као што су фонетика, лингвистика, математика, акустика, али и техничких области попут дигиталне обраде сигнала и програмирања, овим проблемима се у свету баве велики тимови стручњака, и проблем се у овом тренутку може сматрати успешно решеним само за ограничен број светских језика, при чему се успешност огледа у разумљивости синтетизованог говора, али и у његовој природности. У овом раду описан је систем за синтезу говора на основу текста на српском језику, реализован на Факултету техничких наука у Новом Саду, у оквиру пројекта АлфаНум. Перформансе овог система су сасвим упоредиве са системима реализованим за највеће светске језике, а у току његове израде водило се рачуна о бројним специфичностима српског језика, укључујући и чињеницу да је реч о једном од ретких језика који паралелно користи два писма – ћирилицу и латиницу.

## 2. О СИНТЕЗИ ГОВОРА

Развој синтетизатора говора генерално се одвијао у два правца. Једну велику групу чине синтетизатори говора на основу скупа одређених правила која математички описују поједине гласове у језику, као и транзиције између њих. Међу овим синтетизаторима најзначајнији су тзв. *формантни синтетизатори*. Довољно флексибилан формантни синтетизатор у стању је да уз задавање релативно малог броја параметара генерише звучни сигнал чије се временско-спектралне карактеристике могу мењати у широким границама. Захваљујући томе, могуће је, поред осталог, синтетизовати и гласове различитих говорника, под условом да су њихове репрезентативне карактеристике познате. Формантна синтеза, међутим, захтева детаљно познавање фонетике, тако да је за реализовање система који синтетизује говор прихватљивог квалитета, налик природном, потребно уложити веома велик напор. Такви системи се у данашње време радије користе за изучавање перцепције говора, патологије говора и сл.

Другу велику групу чине системи који синтезу говора врше повезивањем сегмената унапред снимљеног говора. Реч је о говорном материјалу који је изговорио један говорник, тако да се од ових система не може очекивати да на основу исте базе снимљеног материјала генеришу глас произвољног говорника. Додуше, савремене методе дигиталне обраде сигнала омогућују одређене модификације у гласу говорника, тако да се ствара утисак да је реч о новом говорнику, чак и супротног пола. Поред тога, овај тип синтетизатора је везан за одређени језик, самим тим што је материјал у оквиру говорне базе везан за одређени језик. Међутим, под условом да један језик има све фонеме које има и други, да се одговарајући фонеми оба језика изговарају на довољно сличан начин, као и да нема битнијих одступања у статистикама појаве фонема у тим језицима, говорна база снимљена на једном језику може се успешно искористити за реализацију синтезе на другом језику.

Синтетизатори који повезују постојеће говорне сегменте су одувек представљали једноставну алтернативу формантној синтези, али је тек развојем рачунарске технологије (снажнији процесори, брже и веће меморије) омогућена синтеза говора високог квалитета, и то спајањем сегмената који се бирају из говорне базе у току саме синтезе, односно нису унапред одређени. Овим је омогућено систему да у датом тренутку за синтезу одабере сегменте чији је садржај (фонетски и прозодијски) најприближнији садржају тражене говорне целине. Поред тога, води се рачуна да узастопни одабрани сегменти не буду сувише различити један од другог на месту будућег споја, како би се тај прелаз после примене посебних техника уједначавања што мање чуо. Оваква претрага усложњава читав систем, а с обзиром да се спроводи током саме синтезе, и успорава га. Међутим, савремени рачунари су довољно брзи да успешно могу да превазиђу тај проблем. Овакви системи нису довољно флексибилни да би се могли користити у истраживачке сврхе, и што се тиче флексибилности, највише што се од њих може очекивати јесу могућности промене висине и боје гласа, брзине читања и томе слично. Међутим, у данашње време, ови системи генеришу сасвим разумљив и релативно природан говор, што је, поред њихове једноставности, главни разлог што најшире коришћени синтетизатори говора у свету припадају управо овој групи, што је случај и са АлфаНум синтетизатором.

## 3. АЛФАНУМ СИНТЕТИЗАТОР ГОВОРА

Први проблем који успешан синтетизатор говора треба да реши је да извуче из текста све информације потребне за генерисање говорног сигнала који звучи природно, у оквиру модула за *језичку обраду текста*, приказаног у оквиру принципске шеме синтетизатора говора на слици 1. Примера ради, у тексту по правилу није обележена акцентуација, а она, поред тога што је од суштинског значаја за природност говора, доприноси ефикаснијем растављању говорног тока на речи од стране слушаоца. То нам потврђују тешкоће са којима се сусрећемо ако покушамо да слушамо говор генерисан са потпуно константном висином гласа. У многим језицима постоји и проблем фонетизације, односно, оно што је написано може знатно да одступа од онога што се изговара. У српском језику је тај проблем маргиналан, и своди се на регистровање појава као што су једначење по звучности ("с девојком" → [здевојком]), као и неки типични случајеви асимилације фонема ("дванае**ст** сати" → [дванаесати]). Међутим, с обзиром да се у текстовима које треба прочитати често налазе и страни изрази написани у оригиналу, синтетизатор говора морао би успешно да прочита и то.

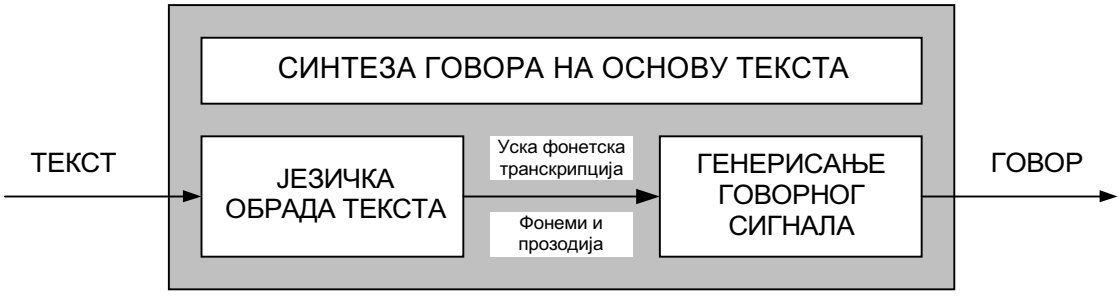
Специфичности српског језика, пре свега сложеност његовог акценатског система и немогућност предвиђања акцентуације речи на основу њеног записа, налажу да се ови проблеми реше помоћу свеобухватног акценатско-морфолошког речника, који би поред податка о акцентуацији сваке речи, садржао и податак о њеној граматичкој категорији (врсти речи), као и о вредностима њених морфолошких категорија (нпр. род, падеж и број код именица). Ове додатне информације су неопходне у случајевима када акцентуација речи није једнозначна. Наиме, чести су случајеви кад нисмо у стању да одредимо акцентуацију речи док не познајемо конкретно окружење у ком се реч налази. Једна реч може се понекад акцентовати на више начина, често у зависности од тога која је лексичка реч у питању ("сѐдети" или "сéдети"). Могуће је и да те две речи не припадају истим граматичким категоријама (глагол "рáдио" или именица "рâдио"), али је могуће и да се ради о истој лексичкој речи са различитим вредностима морфолошких категорија (генитив једнине "днêвника" или генитив множине "днêвникā").

Намећу се два закључка – први је да речник мора да обухвати не само сваку лексичку реч, већ и сваку реч у свим њеним појавним облицима, што намеће питање обима оваквог речника и времена потребног за његову реализацију. Други закључак је да чак ни са таквим речником проблем није решен, јер је потребно обезбедити и технике којима се одређује која је од постојећих могућности исправна. Ово неће бити могуће увек, пошто је некада чак и човек у недоумици кад треба да одреди синтаксу дате реченице, али се коришћењем аутоматске синтаксне анализе реченице могућност грешке ипак значајно умањује. Осим тога, на тај начин је могуће решити и проблем преласка акцента на клитике ("мòгу" → "нѐ могу"). У оквиру АлфаНум синтетизатора говора састављен је речник описаних карактеристика, а имплементиране су и одређене технике одређивања исправне акцентуације на нивоу реченице, засноване махом на диграмима (поређењу граматичких и морфолошких категорија суседних речи и утврђивању степена њиховог слагања). Свеобухватна синтаксна анализа реченице заснована на формалним граматикама била би још успешније средство за елиминацију грешака у акцентуацији, али се сасвим задовољавајући резултати могу постићи и на овај начин.

Под условом да су ови проблеми успешно решени, односно да су нам фонетски и прозодијски садржај реченице познати, остаје да се реши питање избора сегмената из говорне базе (дефинисање критеријума и алгоритми ефикасне претраге базе), као и конкретних алгоритама прилагођавања пронађених сегмената траженом садржају и њиховог повезивања, уз што мању чујност прелаза. Постоји неколико стандардних техника које се за то користе, међу којима треба поменути технику синтезе преклапањем фрејмова полазног сигнала у временском домену – TD-PSOLA (енгл. *Time-Domain Pitch Synchronous OverLap and Add*), употребљену код АлфаНум синтетизатора, као и технику синтезе на основу хибридног хармонијско-стохастичког модела говора. Детаљнији опис ових техника излази ван оквира овог рада.

## 4. ПРОБЛЕМ ДВА ПИСМА

Проблем читања ћириличних текстова у великој мери се преклапа са проблемом формата улазног текста. Читање текстова на српском језику, на било ком писму, могуће је и употребом система кодовања карактера који користе један бајт по карактеру, али су нас масован прелазак на Unicode кодни распоред и напуштање превазиђеног YUscii распореда определили да учинимо АлфаНум синтетизатор говора Unicode компатибилним. Улазни текст је стога у Unicode формату, који предвиђа коришћење два бајта по карактеру и покрива и ћирилицу и латиницу. Поред тога, на овај начин је омогућено и читање текстова са мешовитим писмом (као што је и овај). Овакви текстови су у данашње време честа појава, поготово у штампи.




*Слика 1.* Основна структура синтетизатора говора

Једина разлика између ћириличног и латиничног писма са становишта система за синтезу говора на основу текста јесте у третману диграма "lj", "nj" и "dž", који се јављају само у латиничном писму. Проблем лежи у томе што у неким речима, ти диграми не представљају један фонем већ два ("i**nj**екција", "na**dž**iveti"), и то је чињеница о којој би теоретски требало водити рачуна и при састављању речника. У ћириличном писму тај проблем не постоји (правилно је написати "инјекција", а не "ињекција"). Сматрали смо да овај проблем није потребно посебно решавати, и то из неколико разлога. Разлике у изговору две варијанте критичних диграма (као један фонем или као два) су незнатне, а и речи у којима се они јављају као репрезент два фонема су изразито ретке. Најбитнији разлог је, међутим, било то што се кренуло од претпоставке да је кориснику система најважније да свака реч буде прочитана онако како би то учинио човек, чак и у случају да није правилно написана, већ је учињена нека релативно честа правописна грешка. Акценатски речник, који је из историјских разлога урађен у латиничној варијанти, стога не води рачуна о овом проблему и сваку од комбинација "lj", "nj" и "dž" третира као један фонем. Као што је већ напоменуто, због специфичне природе ових гласова, разлике у изговору су незнатне, и у великом броју случајева се и не могу регистровати.

У случају да се користи ћирилица, елиминисан је још један проблем – наиме, код писања електронске поште латиничним писмом, чест је обичај пошиљаоца да не користи слова са дијакритичким знацима ("ć", "č","đ","š" и "ž"), већ да их изрази на разне друге начине (нпр. "c" или "ch" уместо "č"). Уколико би улаз у синтетизатор говора представљао текст у том облику, слушалац би имао великих проблема у разумевању поруке, јер такав текст не би било могуће ни акцентовати на адекватан начин. У том случају би био задатак синтетизатора говора да одреди како треба прочитати коју реч (АлфаНум синтетизатор говора обухвата ову функционалност). Међутим, за разлику од грешака у акцентуацији, грешка у интерпретацији речи написаних у овом облику била би недопустива, и корисник би с правом могао да се запита због чега му систем не прочита текст барем онако како је написан, кад већ не уме правилно да га интерпретира. Због тога је овај проблем посебно осетљив, и његово одсуство представља једну од предности коришћења ћирилице у кореспонденцији електронском поштом.

## 5. ЗАКЉУЧАК

У овом раду је укратко представљен синтетизатор говора реализован на Факултету техничких наука у Новом Саду, у оквиру пројекта АлфаНум. Захваљујући бројним специфичностима српског језика, у реализацији овог система велика пажња је посвећена језичким проблемима, што је, заједно са ефикасном имплементацијом TD-PSOLA методе синтезе говорног сигнала довело до првог система за синтезу говора високог квалитета на српском језику. Овакав систем има веома широко поље примене, почев од пружања разноврсних информација и услуга преко телефона, обезбеђивања приступа Интернету и другим базама података путем телефона (енгл. Voice Portals), вокалног надзора у мерним и управљачким системима, учења страних језика, па до помоћи слепим и слабовидим особама у самосталном раду на рачунару. С обзиром да је реч о технологији до те мере зависној од језика да се резултати не могу очекивати од великих иностраних фирми, већ се до њих мора стићи сопственим снагама, реално је закључити да је реч о пројекту од националног значаја.

## 6. ЛИТЕРАТУРА


[1] T. Dutoit: *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, Dordrecht/ Boston/London, 1997.

[2] M. Beutnagel, M. Mohri, M. Riley: Rapid Unit Selection from a Large Speech Corpus for Concatenative Speech Synthesis, *EUROSPEECH '99*, pp. 607-610, Будимпешта, Мађарска, 1999.

[3] И. Лехисте, П. Ивић: Прозодија речи и реченице у српскохрватском језику, The Massachusetts Institute of Technology, 1986.

[4] В. Делић, С. Крчо, Д. Главатовић: Основни елементи за аутоматско препознавање и синтезу говора из текста за српски језик, *ДОГС*, pp. 32-37, Фрушка Гора, 1998.

[5] М. Сечујски: Синтеза говора на основу текста с осврtом на српски језик (дипломски рад), Факултет техничких наука, Нови Сад, 1999.

[6] М. Сечујски: Акценатски речник српског језика намењен синтези говора на основу текста, *ДОГС*, Бечеј, 2002.

[7] Р. Обрадовић, Д. Пекар: Ц++ библиотека за обраду сигнала – SLIB, *ДОГС,* Нови Сад, 2000.