

TTS pristup makedonskom jeziku

Slavcho Chungurski, Ivan Kraljevski, Gorgi Kakasevski, Dragan Mihajlov

Sadržaj — U ovom radu je postavljen problem koji se treba rešiti kako bi se dobio visoko-kvalitetni govorni sintetizator na makedonskom jeziku. Problem je lingvističke prirode, pa ovaj rad obuhvata komponente makedonskog jezika koje bi se trebalo dobro proučiti pre početka tehničke implementacije same govorne sinteze.

Pristup makedonskom jeziku izložen u ovom radu iskorišten je za razvoj TTS-MK koji predstavlja prvi sintetizator govora na makedonskom jeziku, baziran na snimljenim govornim segmentima na makedonskom jeziku.

Glavne reči — Govorna sinteza, Makedonski jezik, Prozodija, TTS, TTS-MK

I. UVOD

SINTEZA govora kod sintetizatora TTS-MK izvodi se u tri faze. U prvoj fazi dobija se govor bez prozodičnih elemenata - kontinuiran govor kod kojeg se teško mogu ustanoviti granice između reči. Druga faza obuhvata transformaciju dobivenog signala što rezultira akcentiranim govorom kod kojeg se lako mogu prepoznati granice između reči zahvaljujući implementiranim pravilima akcentuacije. Kod rezultatnog govora druge faze ne mogu se prepoznati granice između rečenica. Ovo je predmet treće faze sinteze u kojoj u sintetizovanom govoru primenjuje se intonacija rezultirajući pravilnim izgovorom.

Sledeći deo ovog rada obuhvata tri dela koja odgovaraju svakoj fazi sinteze govora na makedonskom jeziku. Sve karakteristike jezika koje su obuhvaćene ovim radom utiču na pravilan izgovor sintetizovanog govora na makedonskom jeziku, čime je moguće i njihovo modeliranje [1].

II. LINGVISTIČKI PRISTUP MAKEDONSKOM JEZIKU

A. Fonološki inventar makedonskog jezika

U svakom jeziku reči se upotrebljavaju kao elementi koji imaju svoju formu i značaj. Sastavljeni su od glasova koji čine njihovu zvučnu, materijalnu stranu - formu.

Čovek može proizvesti mnogo glasova, ali samo jedan

S. Čungurski, Faculty of Information and Communication Technologies, FON University, bul. Vojvodina bb, 1000 Skopje, Macedonia (phone: +389-2-2445555, e-mail: chungurski@fon.edu.mk)

I. Kraljevski, Faculty of Information and Communication Technologies, FON University, bul. Vojvodina bb, 1000 Skopje, Macedonia (phone: +389-2-2445555, e-mail: chungurski@fon.edu.mk)

Đ. Kakaševski, Faculty of Information and Communication Technologies, FON University, bul. Vojvodina bb, 1000 Skopje, Macedonia (phone: +389-2-2445555, e-mail: chungurski@fon.edu.mk)

D. Mihajlov, Faculty of Electrical Engineering and Information Technologies, Ss. Cyrilus and Methodius University, 1000 Skopje (phone: +389-2-3099100, e-mail: dragan@feit.ukim.edu.mk)

deo njih koristi se za stvaranje reči. Obično kod jezika upotrebljavaju se 25 do 40 osnovnih govornih glasova. Oni formiraju glasovni sistem jednog jezika. Makedonski standardan jezik ima 31 glas.

Izgovor glasa ima različite nijanse koje zavise pre svega od položaja glasa u jednoj reči. Primera radi, postoje više varijante glasa *k* u rečima *kruna*, *maka*, *kelner*, *kitka*. Kod *kruna* i *maka*, *k* se izgovara tvrdo, a kod *kelner* i *kitka* njegov izgovor je mekši. Ova pojava je rezultat činjenice da se glasovi u govoru ne izgovaraju izdvojeno nego povezano što znači da njihov izgovor ovisi o susednim glasovima govornog niza. Ove nijanse glasa *k* ne unose promenu u značenju navedenih reči jer te nijanse nisu toliko odlučujuće za dobijanje nove reči sa novim značenjem. Sa druge strane, ako se uporedi izgovor reči *kuki* i *kući*, ili izgovor reči *lagi* i *lađi* lako se može primetiti njihova sličnost. Ta sličnost ipak ne utiče u izjednačavanju dve reči u jednu, zato što neka izgovorna svojstva glasa, imaju posebno važnu ulogu kod njegovog formiranja kao posebni jezični element koji vrši razlikovanje značenja dve reči.

Kod nauke jezika glasovi koji unose razlike u značenju samih reči nazivaju se fonemi koji udruživanjem stvaraju veće jezične jedinice – morfeme, reči.

Za svaki govorni glas kod makedonskog jezika postoji posebno slovo, tj. za 31 glas upotrebljavaju se isto toliko slova. Na osnovu oblika i položaja govornih organa kod artikulacije i na osnovu različitih vibracija vazdušne struje, glasovi su podeljeni na više grupe i podgrupe. Najopštije deljenje glasova je na vokale i konsonante.

Vokali. Postoje 5 vokala u makedonskom jeziku: *a*, *e*, *i*, *o* i *u*. Postoje tri klasifikacije vokala i to na osnovu pomeranja jezika napred-nazad, visine jezika i položaja usta. Sve tri klasifikacije su predstavljene u tabeli 1.

TABELA 1: VOKALI.

		pomeranje jezika napred-nazad		
		prednji	srednji	zadnji
visina jezika	visoki	i		u
	srednji	e		o
	niski		a	
		nezaokružena		zaokružena
		oblik usta		

Kod makedonskog jezika postoji mogućnost da slovo *r* u nekim slučajevima igra ulogu vokala, kao na primer kad je između dva konsonanta (*drvo*, *crkva*), kad je na početku reči ispred konsonanta (*'rbet*, *'rž*) i kad je u reči sa prefiksom koja završava vokalom (*za 'rža*, *za 'rti*).

Konsonante. Makedonski jezik sadrži 26 suglasnika koji se mogu klasifikovati po nekoliko kriterijuma, a najčešće po mestu i načinu artikulacije i po zvučnosti suglasnika. Ove klasifikacije se nalaze u tabeli 2.

TABELA 2: KONSONANTE.

	okluzivni	frikativni	afrikati
labijalni	b, p, m	v, f	
dentalno-alveolarni	d, t, n	z, s, l	dz, c
alveolarni		ž, š, lj, r	dž, č
palatalni	nj	j	ć, đ
velarni	g, k	h	

(a) artikulacija

zvučni	bezvučni	sonant
b	p	r
v	f	l
g	k	lj
d	t	m
đ	ć	n
ž	š	nj
z	s	j
dz	c	
dž	č	
	h	

(b) zvučnost

B. Ortoepija – pravogovor makedonskog jezika

Svaki jezik ima svoje pravopisne norme koje se moraju ispoštovati. Pored njih, postoje i pravogovorne norme koje se moraju primenjivati. Kod nekih jezika, koji nisu fonetski, postoji razlika između slova, kojim se označuje glas, i samog izgovorenog glasa. Kao primer se može uzeti da jedan govorni glas može biti označen sa dva ili tri slova. Kod makedonskog jezika, koji je fonetski, vladanje normativnim izgovorom ne treba biti teško, međutim ovde se javljaju neka odstupanja između napisanog i izgovorenog glasa.

Izgovor vokala. Pri pojavi duplog vokala između prefiksa ili partikule i osnovne reči izgovara se sa osetnom granicom između ta dva susedna vokala, kao na primer reč *voobrazi* se izgovara sa osetnom granicom između dva *o*. Nedoslednosti pri izgovoru vokala se javljaju kad se nalaze dva vokala jedan do drugog u situacijama kad se radi o pojavi duplog vokala u sredini ili na kraju jedne reči. U ovim situacijama ovaj dupli vokal izgovara se kao jedan, ali malo produžen, primer *vakuum – vakūm*.

Izgovor vokalnog r. Glas *r* spada u red sonanta i odlikuje se velikom zvučnošću. U slučaju pojave glasa *r* u ulozi samoglasnika u izgovoru ispred *r* čuje se jedan tamni prizvuk koji je potpun glas. Ovaj tamni prizvuk se izgovara isključivo ispred glasa *r*. Na primer, reč *vrba* se izgovara kao *v'rba*.

Izgovor konsonanta. Odstupanje napisanog i izgovorenog konsonanta se javlja kad se stvara jednačjenje suglasnika po zvučnosti. Kao što se vidi u tabeli 2-b, za svaki zvučni konsonant, postoji bezvučni par. Jednačjenje suglasnika po zvučnosti znači „obezvučavanje“ zvučnog suglasnika na kraju jedne reči, tj. zvučni suglasnik na kraju reči se menja njegovim bezvučnim parom. Evo nekoliko primera napisane i izgovorene reči: *grob* se čita *grop* (*b-p*), *gluv* se čita *gluf* (*v-f*), *bod* se čita *bot* (*d-t*), *oriz* se čita *oris* (*z-s*), *jadež* se čita *jadeš* (*ž-š*), *breg* se čita *brek* (*g-k*) itd.

C. Akcenat makedonskog jezika

Makedonski akcenat se nalazi na tačno utvrđenoj poziciji u jednoj reči. Bez razlike na to dali se dodaju prefiksi ili sufiksi na jednu osnovnu reč, akcenat uvek se nalazi na treći slog, računajući unatrag. Primera radi, akcenat u reč *pišuvanje* pada na slog *-šu*, međutim, ako se doda prefiks ili sufiks, akcenat se pomera u zavisnost njegovog dodatka, kao na primer akcenat na reči *potpišuvanje* ili *pišuvanjetto*, pada na *-šu*, tj. *-va*.

Ovo pravilo važi za reči sa više od dva sloga. Kod jednosložnih reči, akcenat pada na jedini slog, a kod dvosložnih reči akcenat pada na prvi slog.

U malo slučajeva ima odstupanja trećosložnog akcentiranja, kao što su slučajevi kad se radi o rečima koji su prihvaćeni iz stranih jezika (*foajè, genèza* itd), kao i slučajevi kad ima pojava akcentske celine (*nè_saka, kàj_beše, pòd_brada* itd).

D. Intonacija makedonskog jezika

Deo makedonskog pisma su i interpunkcijski znaci [2], i to: tačka, zapeta, upitnik, uzvičnik, tačka i zapeta, tri tačke, zagrade, crta, navodnici i polunavodnici. Ovi interpunkcijski znaci u napisanom tekstu imaju značajnu ulogu u procesu implementacije prozodije kod sinteze govora. Pored njih ima i karakteristične reči koje utiču u deljenju rečenice na intonacijske fraze (*i, a, ili, ama* i *no*).

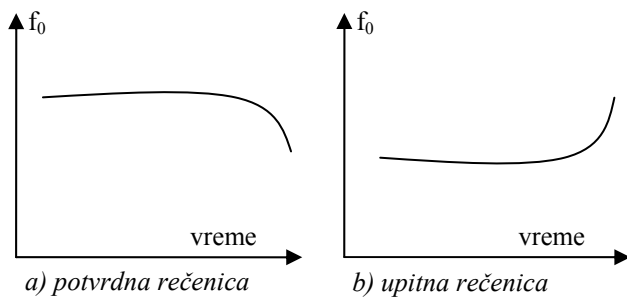
Interpunkcijski znaci i karakteristične reči su oznaka za odgovarajuću intonaciju reči ispred njih. Njihovom detekcijom rečenica se može podeliti na intonacijske grupe reči, koje predstavljaju intonacijske fraze kojima se aplicira odgovarajuća prozodija [3].

Interpunkcijski znaci i karakteristične reči se mogu podeliti u tri grupe:

- Slabi interpunkcijski znaci: (,), (:), *i, ili, a, ama, no*
- Jaki interpunkcijski znaci: (.), (:)
- Upitnik: (?)

Evo po jedan primer svih grupa:

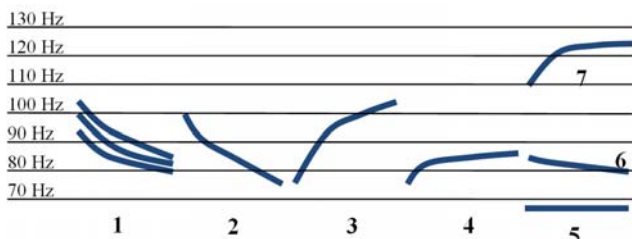
- zapeta, označuje deo rečenice i izgovor se karakterizira blagim povišenjem glasa ispred nje i kratke pauze
- tačka, pretstavlja završetak jedne rečenice. Izgovor ispred tačke je karakterističan po tome što se glas smanjuje, kao na Sl. 1-a, a onda sledi veća pauza usled poslednje reči
- upitnik, označuje upitnu rečenicu i izgovor je karakterističan po tome što se javlja veće pojačavanje glasa ispred njega (kao na Sl. 1-b) i veća pauza iza njega



Sl. 1. Tipovi konture visine glasa

Modeliranjem konture visine glasa na osnovu sloga [4], definisani su sedam tipa slogova:

- 1 običan slog
- 2 preposlednji slog ispred slabog interpunkcijskog znaka
- 3 slog ispred slabog interpunkcijskog znaka
- 4 preposlednji slog ispred jakog interpunkcijskog znaka
- 5 slog ispred jakog interpunkcijskog znaka
- 6 preposlednji slog ispred upitnika
- 7 slog ispred upitnika



Sl. 2. Intonacijski šabloni tipova slogova

Za svaki tip sloga definisani su intonacijski šabloni kao na slici 2. Intonacijski šablon definiše konturu vrednosti visine glasa.

III. ELEMENTI MAKEDONSKOG JEZIKA KOD TTS-MK

Sistem TTS-MK je razvijen kao prvi sistem za automatsku sintezu visoko-kvalitetnog govora na osnovu proizvoljnog teksta za makedonski jezik, baziran na govornom korpusu. Termin visoko-kvalitetni podrazumeva ugradnju prozodičnih karakteristika karakteristične za makedonski jezik koja obuhvataju automatsku akcentuaciju teksta i implementaciju intonacijskih fraza. Termin prvi ne znači da nije bilo i ranijih pokušaja razvitka TTS sistema. Postoje i raniji pokušaji za razvoj TTS sistema na osnovu snimljenih segmenata, kao što je Alfanum sistem [5], ali problem je u tome što je ovaj sistem baziran na govornom korpusu namenjenom za sintezu govora na srpskom jeziku, a zbog sličnosti samih jezika, primenjen za sintezu govora na makedonskom jeziku, što dovodi do sintetizovanog govora na makedonskom jeziku sa srpskim naglaskom.

TTS-MK je zasnovan na konceptima koji omogućuju njegovo daljnje usavršavanje prema krajnjem cilju – dostignuće kvaliteta najprestižnijih TTS sistema.

A. Govorni korpus

Potrebna su nekoliko koraka za definiciju govornog korpusa za sintezu govora na makedonskom jeziku koji se koristi kod TTS-MK:

- selekcija tipova govornih segmenata
- definicija skupa fonema koji bi predstavljao fonološki inventar TTS-MK-a
- definicija skupa govornih segmenata koji sadrže izabrane foneme
- definicija teksta za ekstrakciju izabranih govornih segmenata

Broj govornih segmenata trebao bi biti što manji, dok duži segmenti smanjuju gustinu tački spajanja, što dovodi do boljeg kvaliteta sintetizovanog govora. Ovo utiče na vreme generisanja govora na taj način što u velike kolekcije govornih segmenata, za pretraživanje odgovarajuće jedinice potrebno bi bilo više vremena.

Difoni su svakako najčešće upotrebljavani osnovni segmenti kod sinteze govora [6]. To su govorne jedinice koje sadrže završni deo jednog glasa i početni deo drugoga glasa. Difoni nisu previše brojni, ali česta pojava spajanja osnovnih segmenata dovodi do smanjene tečnosti govora.

Kod TTS-MK, ovaj problem je rešen što u govorni korpus su uključeni dvoslozi – govorni segmenti koji počinju u sredini jednog vokala, obuhvataju celu konsonantsku grupu i završavaju u sredini sledećeg vokala. Ovo je dobro rešenje, jer dvoslozi obuhvataju niz difona, ali njihov broj je mnogo veći nego broj difona. Ovo ukazuje da najbolje rešenje za generisanje govornog korpusa za makedonski jezik je rešenje koje uključuje i određene dvosloge koji se u govoru javljaju češće nego drugi. Za taj cilj, u okviru istraživanja obezbeđena je baza tekstova na makedonskom jeziku koja obuhvata celu dnevnu štampu iz Makedonije koja je izašla tekom jedne godine [7], i na osnovu njega izvršena je statistika za dvosloge koji se najčešće upotrebljavaju u makedonskom jeziku.

S obzirom da je skup govornih segmenata relativno mali, no i s obzirom na performanse današnjih računara govorni korpus TTS-MK-a obuhvata još jednu kategoriju govornih segmenata, a to su cele reči. Spomenuta baza je iskorišćena i za izdvajanje najviše korištenih reči u makedonskom jeziku, tako da i govorni segmenti koji obuhvataju cele reči su uključeni u govorni korpus. Ovi govorni segmenti obuhvataju cele reči, sa izuzetkom prve polovine prvog fonema u reči i druge polovine zadnjeg fonema u datoj reči. Na ovaj način se dobijaju dugački segmenti pogodni za spajanje sa drugim segmentima, kao što su difoni i dvoslozi, no i druge cele reči.

Skup fonema i skup govornih segmenata. U izboru skupa fonema koji će biti uključeni u govorni korpus najvažnije je da skup obuhvati sve moguće govorne pojave makedonskog jezika. Ocenjeno je da za makedonski jezik ovaj skup čine svi govorni glasovi, 5 vokala i 26 konsonanta, kao i dva dopunska fonema tamni prizvuk (u pojavi vokalnog *r*) *a* - (*schwa*), koji predstavlja vokal, i alofon *ŋ* suglasnika *n*, zbog njegove specifičnosti kad *n* se nalazi ispred *k* ili *g*. Ovaj alofon se često

upotrebljava i u skupovima fonema i drugih jezika [5]. U skup fonema dodaje se i znak „_“ koji označava tišinu i koji učestvuje u tranzicijama tipa tišina-glas i tipa glas-tišina.

Kvalitetan govorni sintetizator treba proizvesti sintetizovan govor bez obzira na tekstualni ulaz. Ovo znači, da postoje i difonske kombinacije koje se ne pojavljuju u makedonski govor, a koje moraju biti uključene u govorni korpus. Za njihovu ekstrakciju koriste se *logatomi* ili reči bez značenja koje sadrže ove kombinacije.

Spomenuta baza [7] je ponovo upotrebljena za generisanje minimalnog skupa postojećih reči koje pokrivaju što veći broj difonskih kombinacija. Ovaj skup broji tačno 500 reči, a njemu su dodati još dopunskih 258 logatoma kako bi se pokrio svaki teoretski mogući difon, što praktično čini i prvu makedonsku difonsku bazu.

Dopunski koraci za nadogradnju ove difonske baze do govornog korpusa, namenjenog za TTS-MK sistem, su napravljeni dodavanjem dopunskih neelementarnih govornih segmenata. U spomenut tekstualni korpus [7] prebrojani su svi dvoslozi i dobiven je broj 937987. Odlučeno je da se u govorni korpus uključe dvoslozi koji su se pojavili minimum 1000 puta, ili u 0,1% slučajeva. Generisani su tačno 180 dvosloga koji su uključeni u govorni korpus. Za ekstrakciju ovih dvosloga upotrebljene su već postojeće 758 reči, no dodani su i dopunski 46 reči.

Tekstualni korpus uključuje i 392 cele reči koje obuhvataju brojeve, imena meseca i geografske pojmove, da bi se dobio konačni broj od 1196 reči za ekstrakciju tačno 1086 difona, 180 dvosloga i 392 cele reči.

B. Prozodična obeležja.

Akcentat kod TTS-MK zasniva se na pravilu za trećesložno akcentiranje, uključujući rečnik akcentskih izuzetaka koji obuhvata i akcentske celine.

NLP modul [1] vrši analizu ulaznog teksta i utvrđuje eventualne pojave vokalnog r , na koje se vrši umetanje znaka a , i utvrđuje eventualne pojave fonema η . U NLP fazi paralelno se generiše niz difona, kao i binarni niz jednako dugačak kao niz difona gde 1 označava potrebu transformacije, za dobijanje akcenta. Pored ovog niza, ulazni tekst se deli na intonacijske fraze i generiše se dopunski niz cifri od 1 do 7 koji označava eventualne transformacije na osnovu intonacijskih šablona sa Sl. 2.

Kao ilustracija može se razgledati sledeći primer. Na ulazni tekst „Kisela voda e edna od prvite naselbi vo Skopje“ NLP modul generiše sledeće nizove:

```
_kisElavOdaEEdnaOdp@rvitEnasElbivOskOpjE_
0000010000110001001000000100000010010000
111111111111111111111111111111111111111111111111111111444555
```

U ovaj primer se mora istaći da rečnik akcentskih izuzetaka sadrži informaciju da je „Kisela voda“ akcentska celina, tj. ove dve reči se izgovaraju kao jedna na koju se primenjuje pravilo za trećesložno akcentiranje (*Kiselāvoda*).

Dobijena tri niza predstavljaju dovoljnu informaciju za

prelaz u DSP fazu sinteze govora kod TTS-MK.

IV. ZAKLJUČAK

U ovom radu su predstavljena istraživanja u polju govorne sinteze na makedonskom jeziku. Prezentirani su lingvistički elementi makedonskog jezika potrebni za visoko-kvalitetnu govornu sintezu i predstavljen je model ovih elemenata u implementaciju TTS-MK. Važno je izvesti dobro procesiranje ulaznog teksta što omogućava pravilnu akcentuaciju i prozodične transformacije, vodeći prema razumljivijoj sintezi govora.

Iduća istraživanja na ovom polju biće u pravcu daljnjih poboljšanja TTS-MK, kao i njegova primena.

LITERATURA

- [1] S.Chungurski, I.Kraljevski, D.Mihajlov, S.Arsenovski, "Concatenative Speech Synthesizers and Speech Corpus for Macedonian Language", 30th International Conference ITI Cavtat/Dubrovnik, Croatia, June 23-26, 2008, pp. 669-674
- [2] B. Vidoeski, T. Dimitrovski, K. Koneski, R. Ugrinova-Skalovska, "Pravopis na makedonskiot literaturni jazik", Prosvetno delo, Skopje, 1999
- [3] M. Zrmanovska, „Kompjuterska sinteza govora na makedonskom jeziku sa prozodijom i emocije“, Magistarski rad, Elektrotehnički fakultet, Skopje, 2005
- [4] S. Kula, P. Dymarski, A. Janicki, C. Jobin, P.B. De Mareuil, "Prosody Control in Diphone-Based Speech Synthesis System for Polish", Institute of Telecommunications, Warsaw University of Technology, Prosody Workshop: Speech Recognition and Synthesis, Krakow, 2000
- [5] M. Secujski, V. Delic, "Snimanje govorne baze na srpskom jeziku namenje sintezi govora na osnovu teksta", DOGS, Novi Sad, 2000
- [6] T. Dutoit, „High Quality Text-To-Speech Synthesis of the French Language“, Ph.D. dissertation, Faculté Polytechnique de Mons, October 1993
- [7] S. Chungurski, I. Kraljevski, S. Arsenovski, S. Samardziska, "Information system for newspaper clipping", Naucno-strucni simpozijum Infoteh, Jahorina, BiH 2007

ABSTRACT

This paper deals with the problem which should be solved for achievement of high-quality speech synthesizer for Macedonian language. The problem has linguistic nature, so this paper includes components of Macedonian language which should be well known before technical implementation of the speech synthesis.

The approach to Macedonian language expressed in this paper is used for development of TTS-MK which is the first speech synthesizer for Macedonian language, based on recorded speech segments on Macedonian language.

TTS APPROACH TO MACEDONIAN LANGUAGE

S. Chungurski, I. Kraljevski, G. Kakasevski, D. Mihajlov