TECHNICS AND INFORMATICS IN EDUCATION
6th International Conference, Faculty of Technical Sciences, Čačak, Serbia, 28–29th May 2016
TEHNIKA I INFORMATIKA U OBRAZOVANJU
6. međunarodna konferencija, Fakultet tehničkih nauka, Čačak, Srbija, 28–29. maj 2016.

UDK: 004.6:811.163.41`342.4 Stručni rad

Srpska Govorna Baza "Phonemes_1.0": Dizajn i Primena

Branko Marković[1], Vladimir Milićević[1], Dragana Petrović[1], Dejan Nešković[1] i Gordana Marković[2]
[1]Visoka Škola Tehničkih Strukovnih Studija Čačak, Čačak, Srbija
[2]Tehnička Škola, Čačak, Srbija
e-mail brankomarko@yahoo.com

Rezime: *U ovom radu smo opisali kako se kreira srpska govorna baza "Phonemes_1.0" i kako se koristi za poređenje govornih uzoraka. Ova baza pokriva listu od 30 fonema koje sadrži srpski jezik i koja se zove "Azbuka". Baza je podeljena na dva dela: deo koji sadrži vokale i deo koji sadrži konsonante. Za vokale je primenjen inicijalni DTW algoritam radi poređenja.*

Ključne reči: *Srpska govorna baza; vokali; konsonanti; DTW algoritam.*

1. UVOD

Sistemi za automatsko prepoznavanje govora (ASR - Automatic Speech Recognition) su danas vrlo popularni. Oni se baziraju na različitim pristupima. Neki od njih su namenjeni za izolovane foneme, neki za slogove ili reci, a neki za kontinualni govor. Takođe oni su podeljeni na sisteme nezavisne i zavisne od govornika.

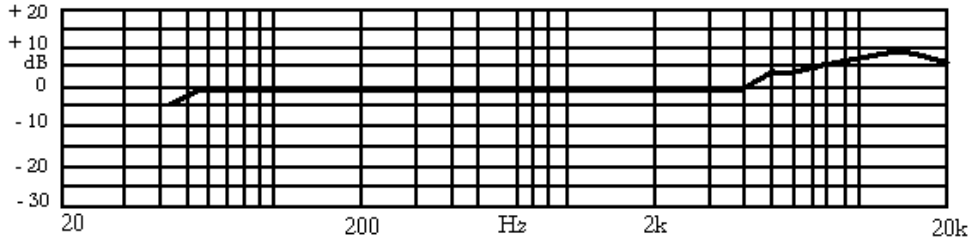
Da bi se vršilo poređenje govornih uzoraka moraju postojati referentni podaci. Stoga je ovaj rad namenjen problemu kako da se kreira baza govornih uzoraka[1] [2] koja bi u tu svrhu poslužila. U ovom slučaju pitanje je: "Kako prikupiti foneme srpskog govornog jezika i kako ih organizovati u odgovarajuću bazu podataka?".

Foneme srpske "Azbuke" su snimane u specijalnoj akustičkoj prostoriji koja omogućava potiskivanje pozadinskog šuma. U ovaj projekat bilo je uključeno 20 volontera (studenti Visoke škole tehničkih strukovnih studija Čačak). Sva snimanja su obeležena na specifičan način tako da ih je moguće kasnije koristiti. Na određen boj elemenat ove baze (vokale) je primenjen inicijalni DTW (Dynamic Time Warping) test i odgovarajući rezultati su prezentovani u ovom radu.

Ovaj rad je organizovan na sledeći način: Sekcija 2 objašnjava kako su podaci snimani i koja vrsta opreme je korišćena. Sekcija 3 objašnjava kako su podaci obeležavani i kako su smeštani u bazu "Phonemes_1.0". U sekciji 4 prezentovali smo inicijalni test za prepoznavanje vokala baziran na tehnologiji poređenja uzoraka. Zadnja sekcija je Zaključak gde su dati sumarni rezultati vezani za ovaj rad.

2. SNIMANJE GOVORA

Baza "Phonemes_1.0" je snimana u tihoj laboratorijskoj sobi korišćenjem Optimus omni-direkcionalnog mikrofona sa dobrom frekvencijskom karakteristikom u oblasti do 16kHz. (Slika 1) i lap-top raučunar Fujitsu-Siemens Esprimo Mobile sa Adobe Audition 1.5 softverskim paketom za snimanje govora.



Slika 1: *Frekvencijska karakteristika Optimus mikrofona*

Mikrofon je bio na udaljenosti od oko 25cm od usta govornika. Govor je digitalizovan korišćenjem frekvencije odmeravanja od 22.050Hz, 16 bita po odmerku, jedan kanal, i smeštan u formi Windows PCM wave fajlova.

Sesije za snimanje su organizovane četiri puta tako da se sakupi dovoljan broj kvalitetnih uzoraka (neki su eliminisani). Tokom pojedinačne sesije govornici su imali da pročitaju spisak od 30 fonema srpske "Azbuke" po dva puta. Zatim je čitav set snimaka ručno segmentiran i nad dobijenim fonemama je vršena kontrola kvaliteta. Ako su ispitivani uzorci dobri oni su označavani i smeštani u bazu "Phonemes_1.0"; u protivnom su eliminisani. Na ovaj način generisano je više od 1200 fonema, ali je samo 1200 najboljih smešteno u bazu "Phonemes_1.0".

Kontrola kvaliteta prilikom snimanja je otkrila različite vrste grešaka. Neke od njih su bile vezane za pogrešnu artikulaciju, neke za pogrešan izgovor, neke za duvanje u mikrofon i slično. Više novih snimaka je urađeno da bi se eliminisali ovi problemi.

Svi uzorci u bazi su podeljeni na osnovu kategorija u dve grupe: vokali (5 tipova vokala) i konsonante (25 tipova konsonati) [4]. Oni su prikazani u Tabeli 1 sa IPA (International Phonetic Alphabet) notacijom za svaki od njih.

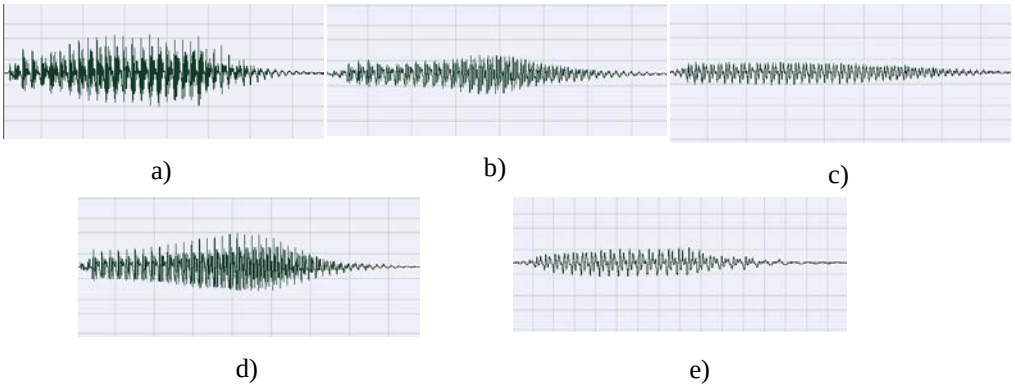
Tabela 1. *Foneme smeštene u "Phonemes_1.0" bazi (sa IPA notacijom)*

Tip	Fonema	IPA	Tip	Fonema	IPA
vokal	/a/	/a/	kons.	/љ/	/ ʎ /
vokal	/e/	/e/	kons.	/м/	/m/
vokal	/и/	/i/	kons.	/н/	/n/
vokal	/o/	/o/	kons.	/њ/	/ ɲ /
vokal	/у/	/u/	kons.	/п/	/p/
kons.	/б/	/b/	kons.	/р/	/r/
kons.	/в/	/v/	kons.	/с/	/s/
kons.	/г/	/g/	kons.	/т/	/t/
kons.	/д/	/d/	kons.	/ћ/	/ t͡ɕ /

kons.	/ђ/	/ dʑ /	kons.	/ф/	/f/
kons.	/ж/	/ ʒ /	kons.	/х/	/h/
kons.	/з/	/z/	kons.	/ц/	/ ts /
kons.	/j/	/j/	kons.	/ч/	/ tʃ/
kons.	/к/	/k/	kons.	/џ/	/ dʒ /
kons.	/л/	/l/	kons.	/ш/	/ ʃ/

Sa aspekta prepoznavanja govora vokali su mnogo interesantniji od konsonanti zato što se češće pojavljuju u govoru, a takođe mogu da postoje samostalno.

Na Slici 2 prikazani su talasni oblici za svaki od pet vokala za jednog govornika.



Slika 2: *Talasni oblici za vokale a) za /a/, b) za /e/, c) za /i/, d) za /o/ i e) za /u/*

Sa ove slike može se videti da je većina talasnih oblika za vokale slična. Ali kada se primene metodi za spektralnu analizu pokazuje se njihova spektralna različitost.

3. OZNAČAVANJE U BAZI

Da bi se vršilo lako i automatizovano testiranje podataka koji su smešteni u bazu potrebno je odgovarajuće označavanje (labeliranje). Oznake se biraju tako da same sebe objašnjavaju. Stoga, za označavanje vokala, svi fajlovi koji ih predstavljaju su označeni na sledeći način: ***vn_m_p.wav***. Slovo“***v***” označava vokal, a ***“n”***, “***m”*** i “***p”*** su brojevi sa sledećim značenjem:

- ***“n”*** je broj koji označava koji je vokal po redu (1 - znači vokal /a/, 2 - znači vokal /e/ itd.)
- ***“m”*** je broj koji označava govornika (1 - znači prvog govornika, 2 - znači drugog govornika itd.)
- ***“p”*** je broj koji označava redni broj izgovora od istog govornika (1 - znači prvi izgovor, 2 – znači drugi izgovor itd.)

Korišćenjem istog principa obeležili smo i konsonante na jedinstven način. Tako, fajlovi za konsonate imaju oznake sledećeg oblika: ***cn_m_p.wav***. Ovde je slovo ***“c”*** oznaka za konsonatu (eng. consonant). Značenje brojeva ***“n”,”m”*** i “***p”*** je identično kao što je objašnjeno za vokale.

4. INICIJALNI DTW TEST

Da bi se evaluirali podaci u ovoj bazi izvršeni su određeni inicijalni testovi. Cilj ovih testova je se vidi kako kreirana baza može da se koristi za automatsko prepoznavanje govora (sa aspekta fonema) i koja će biti verovatnoća prepoznavanja za vokale.

Kao prednji deo (predobrada) za automtasko prepoznavanje govora korišćene su LPC (Linear prediction coding) osobine [5], gde je za red autokorelacije izabrano ***p=12***. Za zadnji deo (odlučivanje) korišćen je DTW algoritam [6].

DTW algoritam je baziran na dinamičkom programiranju i cilj je naći optimalnu stazu između početnih i završnih tačaka u kojima se poklapaju poređeni govorni uzorci. Govorni uzorci se reprepzentuju skupom vektora koji se dobija tokom predobrade. Prvi skup uzoraka (5 vokala) je korišćen kao referentni, a ostali uzorci (devet skupova, svaki od po 5 vokala) kao test uzorci. Za lokalno ograničenje korišćen je tip I predložen od strane Sakoe i Chiba [7] pri čemu je akcenat stavljen na dijagonalni prelaz. Globalna ograničenja nisu korišćena. Sistem nije treniran.

Rezultati u obliku broja prepoznatih reči (WRR - word recognition rates) su prikazani u Tabeli 2. Dijagonala matrice prikazuje broj uspešno prepoznatih reči (maksimalno je 9).

Tabela 2. *Broj prepoznatih reči za vokale sa matricom konfuzije*

Ref/Test	/a/	/e/	/i/	/o/	/u/
/a/	7				1
/e/	2	5	1		
/i/		4	7	1	
/o/			1	6	
/u/				2	8
Srednje	**77.78**	**55.56**	**77.78**	**66.67**	**88.89**
Ukupno	**73.33**				

Na osnovu Tabele 2 vokali /e/ i /o/ daju najlošije rezultate. Najbolji rezultat je za vokal /u/ . Srednji broj prepoznatih vokala je 73.33%.

5. ZAKLJUČAK

Ovaj rad daje primer kako da se kreira govorna baza za srpski jezik koja je bazirana na fonemama od kojih se sastoji "Azbuka". Korišćenjem odgovarajućih tehnika i označavanja ova baza može biti dobro organizovana, laka za pristup i korišćenje.

Za automatsko prepoznavanje govornih uzoraka različiti algoritmi mogu biti korišćeni. U ovom radu je LPC korišćen za predobradu, a DTW za poređenje. Sa njima je pokazano kako se može izvršiti odgovarajući test i dobiti broj prepoznatih vokala. Sličan scenario može se koristiti za konsonante kao i za reči.

Dalje istraživanje i rad mogu biti usmereni ka ovim oblastima.

REFERENCE

[1] B. Marković, S.T. Jovičić, J. Galić, Đ. Grozdić: "Whispered Speech Database: Design, Processing and Application", 16th International Conference, I. Habernal and V.

Matousek (Eds.): TSD 2013, LNAI 8082, Springer-Verlag Berlin Heidelberg, pp. 591-598. (2013).

[2] S. Itahashi, "A Japanese Language Speech Database", ICASSP 86, Tokzo, pp. 321-324.

[3] L. Rabiner, B-H. Juang, "Fundamentals of speech recognition", (Prentice Hall, New Jersey) (1993).

[4] S. T. Jovičić, "Govorna komunikacija – fiziologija, psihoakustika i percepcija", Nauka, Beograd, 1999.

[5] B. R. Marković and Đ. T. Grozdić, „The LPCC-DTW Analysis for Whispered Speech Recognition", Proceedings of 1st International Conference of Electrical, Electronic and Computer Engineering, IcETRAN 2014, pp. AKI1.1.1-4, Vrnjačka Banja, Serbia, June 2-5, 2014.

[6] G. Marković, B. Marković, "Vizuelni DTW kao nastavno sredstvo za poređenje govornih uzoraka", Tehnika i informatika u obrazovanju, TIO '08, str. 409-415, Tehnički fakultet, Čačak, 9-11. maja.

[7] H. Sakoe and S. Chiba, „Dynamic programming optimization for spoken word recognition", IEEE Trans. Acoustics, Speech, Signal Proc., pp 43-49, 1978.