

# METODI I SISTEMI ZA AUTOMATSKO PREPOZNAVANJE GOVORNIH SEKVENCI

Vladan Vu-kovi<sup>1</sup>  
<sup>1</sup>Elektronski fakultet u Nišu

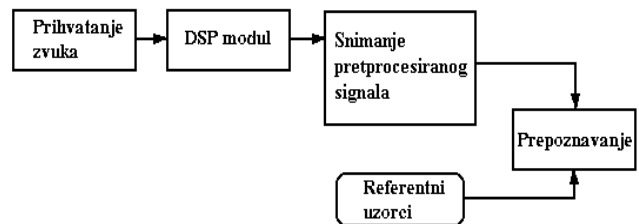
## I UVOD

Ma{insko prepoznavanje govora, uz ostale oblasti ve{ta-ke inteligencije predstavljaju najpropulzivnije pravce razvoja ra-unarske tehnike. Osnovni ciljevi istra`ivanja u ovim oblastima su pobolj{anje interfejsa ra-unara prema ljudskom operateru i omogu}avanje direktne komunikacije sa ra-unarom putem govornog jezika. Dostizanje ovih ciljeva bi omogu}ilo jo{ ve}u ekspanziju ra-unara u mnoge oblasti koje su trenutno nedostupne kao i omogu}avanje automatizovanih usluga onim segmentima ljudske populacije koji su zbog raznoraznih razloga u nemogu}nosti da ve} danas koriste sve prednosti informati-ke revolucije.

Automatsko prepoznavanje govornih sekvenci predstavlja najvi{i mogu}i vid obrade digitalizovanog govornog signala u -iju je realizaciju potrebno uklju-iti niz slo`enih matemati-kih modela i aparata. Problem efikasnog prepoznavanja kontinualnog govora jo{ uvek nije re{en na zadovoljavaju}i na-in, ali zato postoji niz metoda i sistema koji daju veoma zadovoljavaju}e rezultate u re{avanju nekih jednostavnijih sistema za prepoznavanje. Tako|e, postoji i znatan broj komercijalnih sistema baziranih na automatskom prepoznavanju govora koji ve} opslu`uju milione korisnika. Gledano sa dana{njeg stanovi{ta, najinteresantnije primene automatskog prepoznavanja govora su u oblastima *daljinske komunikacije sa ra-unarom (putem telefonske mre`e)* kao i u oblasti *razvoja interfejsa* u smislu uvo|enja automatskog prepoznavanja i generisanja govornih sekvenci kao jednog od osnovnih parametara komunikacije [1],[2]. I ostali jednostavniji vidovi obrade zvuka kao {to su snimanje i reprodukcija digitalizovanog govornog signala, telemarketing i jednostavni sintetizatori govora imaju tako|e brojne primene.

## II OSNOVNE KARAKTERISTIKE SISTEMA ZA AUTOMATSKO PREPOZNAVANJE GOVORA

Osnovne komponente i organizacije sistema za prepoznavanje mo`e se prikazati na slede}oj slici [3]:



Slika 1. Osnovne komponente sistema za prepoznavanje.

- ♦ **Prihvatanje zvuka** - U prvoj fazi vr{i se prihvatanje i digitalizacija govornog signala uz pomo} mikrofona i A/D konvertora. Naj-e{e se koriste kondenzatorski mikrofoni koji imaju najbolje prenosne i spektralne karakteristike. A/D konvertori su standardni sa 8, 12 ili 16 bita konverzionom rezolucijom. Poja-ava-ki stepen je linearan. Naj-e{e se u ovaj stepen ugra|uje i automatska kontrola nivoa (AGC).
- ♦ **DSP modul** - Ovaj modul ima zadatak da obavi osnovnu obradu nad digitalizovanim govornim signalom. Obrada podrazumeva odre|ivanje frekvencijskog domena, odre|ivanje granica govornog signala, razdvajanje korisnog signala od signala {uma, skaliranje, filtriranje, sekvenciranje, podelu na *Hamming*-ove prozore ... Cilj ove faze je izdvajanje samo onih parametara koji su korisni u fazi prepoznavanja. ^esto se u sisteme za prepoznavanje ugra|uju specijalizovani DSP procesori visokih performansi.
- ♦ **Snimanje preprocesiranog signala** - Imaju zadatak da predprocesirani signal iz predhodne faze snimi u *buffer* memoriju i tako pripremi digitalizovan signal za zavr{nu fazu.
- ♦ **Referentni uzorci** - Skup referentnih uzoraka na osnovu kojih se vr{i prepoznavanje.
- ♦ **Prepoznavanje** (*pattern matching*) - Algoritam za prepoznavanje na osnovu digitalizovanog zapisa signala sme{tenog u *buffer* memoriji i tabele referentnih uzoraka odlu-uje koji je uzorak iz tabele najsliniji uzorku koji se prepoznaje.

Ve}ina metoda za prepoznavanje koriste dva osnovna pristupa:

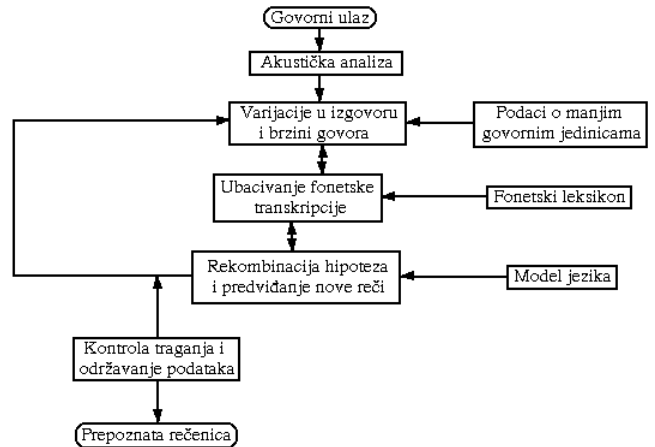
- prepoznavanje pretra`ivanjem odre|enog skupa referentnih uzoraka i nala`enjem referentnog uzorka koji ima minimalno udaljenje od tra`enog,
- prepoznavanje tra`enjem optimalnog puta kroz mre`u kona-nog automata - pristup koji korespondira sa metodom skrivenih Markovljevih modela.

Brojni algoritmi koji su razvijeni za re{avanje problema prepoznavanja predstavljaju zapravo samo varijacije osnovnih metoda sa raznim unapre|enjima u pogledu efikasnosti i brzine prepoznavanja. Osnovni koraci kod ve}ine algoritama za prepoznavanje su:

- Detekcija kraja govornog signala,
- Nelinearno tra`enje u skupu referentnih uzoraka,
- Klasifikacija.

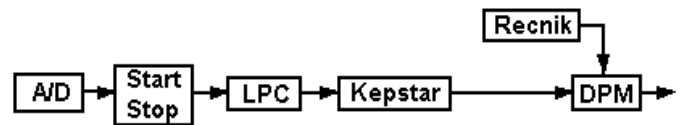
Linearno traganje kod slo`enih skupova referentnih uzoraka usled uticaja kombinatorne eksplozije ima samo teoretski zna-aj. Zbog toga se primenjuju razne metode za smanjivanje broja mogu}ih puteva. Dva -esto kori}ena metoda su *dinami-ko programiranje* i *heuristi-ko tra`enje*. Uop{teno uzev{i, heuristike predstavljaju odre|ena pravila dobijena konsultovanjem znanja eksperata i odnose se na smanjivanje broja kombinacija odnosno na usmeravanje toka traganja u odre|enom pravcu. Primena heuristika omogu}ava zna-ajno smanjivanje obima traganja ali sa druge strane donosi optere}enje u vremenu procesiranja i potrebnoj memoriji. Uporedna istra`ivanja pokazuju da dinami-ko programiranje sa implementiranim *beam-search* metodom ima su{tinske prednosti u odnosu na heuristi-ki pristup [4].

Neka pobolj{anja vezana za skrivene Markovljeve modele su kori}enje ekvalizatora prvog reda za smanjivanje uticaja {uma do nivoa 15-20 dB [5] i primena SCHMM-a {to mo`e da donese zna-ajna pove}anja u faktoru uspe{nog prepoznavanja u odnosu na DHMM i CHMM (30% i 20% respektivno) [6]. Kori}enje *Bayesian*-ovog pristupa u klasifikaciji [7] tako je donosi pove}anju faktora uspe{nog prepoznavanja. Upotreba neuronskih mre`a za samostalno prepoznavanje govornog signala je jo{ uvek u ranim fazama razvoja ali je njihovo kori}enje u sprezi sa HMM radi boljeg odre|ivanja emisionih verovatno}a dalo pozitivne rezultate [8]. U cilju re{avanja problema kod raspoznavanja kontinualnog govora u realnom vremenu koji se javljaju usled potrebe za brzom obradom velikih re-nika i baza znanja koriste se efikasniji metodi traganja kao {to je *metoda traganja vo|ena podacima*. Primer sistema za prepoznavanje kontinualnog govora koji koristi izvore znanja dat je na slede}oj slici [9]:



**Slika 2.** Arhitektura sistema za prepoznavanje kontinualnog govora i interakcija izvora znanja.

Prikazani sistem koristi znanje na tri kognitivna nivoa i radi na principu postavljanja i dokazivanja hipoteza. Sli-nu strukturu poseduje i ekspertni sistem HEARSAY [10]. Kori}enje povratnih petlji zna-ajno smanjuje brzinu prepoznavanja ali se navedenim sistemom posti`e maksimalna pouzdanost u prepoznavanju kontinualnog govora. Visoke performanse mogu}e je posti}i i jednostavnijim metodama bez kori}enja specijalizovanih baza znanja upotrebom spektralnih koeficijenata i LPC (*linear-predictive coefficients*) analize. Blok {ema sistema za prepoznavanje na osnovu LPC analize data je na slede}oj slici [11]:



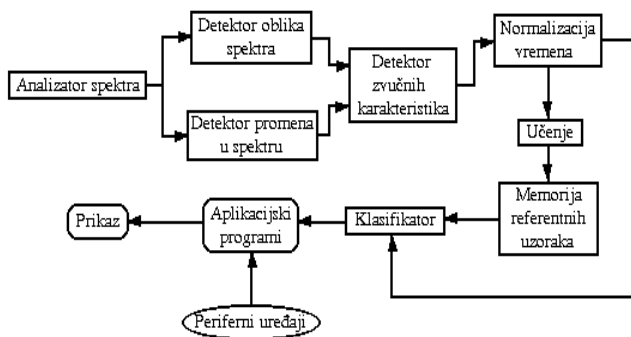
**Slika 3.** Blok {ema sistema za prepoznavanje kontinualnog govora na osnovu LPC analize.

Sistem funkcioni}e na slede}i na-in - posle analogno/digitalne konverzije vr{i se odre|ivanje po-etka i kraja digitalizovanog govornog signala. Slede}i korak je obavljanje LPC analize i odre|ivanje *kepstar* -nih koeficijenata. Na osnovu re-nika i dinami-ke analize ( **DPM** - *Dynamic Programming Matching*) odre|uje se referentni uzorak sa najslabijom strukturom kepstralnih koeficijenata koji predstavlja rezultat prepoznavanja. Za govornika koji je obu-io bazu sistem posti`e faktor pogodaka od 95%. Kori}enje znanja kroz ekspertski zasnovanu fokusiranu spektralnu analizu tako je daje dobre rezultate (*APG<sub>L</sub> sistem*) [12]. Segmentacija govornih sekvenci na manje govorne jedinice predstavlja tako je te`ak problem koji jo{ uvek nije u potpunosti re{en. Neki od pristupa su kori}enje *BOX-counting* metode i *teorije haosa (teorija fraktala)* [13] kao i kori}enje sistema baziranih na znanju eksperata koje je

ekstraktovano u fazi ru-ne segmentacije govora [14]. Ovi pristupi u segmentaciji daju slabe rezultate tako da je za sada daleko bolja alternativa koristiti standardni pristup preko HMM ili preko ostalih statisti-kih metoda [15].

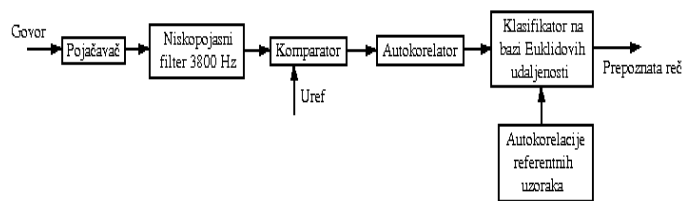
### III SISTEMI ZA PREPOZNAVANJE GOVORNIH SEKVENCI

Jedan od prvih poznatijih komercijalnih sistema za raspoznavanje je **VIP-100** kompanije *Threshold Technology Inc.* koji raspoznaje re-i na osnovu merenja odre|enih zvu-nih karakteristika i njihovim upore|ivanjem sa skupom referentnih vrednosti [10]. Osnovne karakteristike koje sistem koristi u prepoznavanju vezane su za oblik i promene spektralne strukture digitalizovanog signala. Blok {ema VIP-100 sistema prikazana je na slede}oj slici:



Slika 4. Blok {ema sistema VIP-100 za raspoznavanje izolovanih re-i.

Referentne karakteristike dobijaju se u fazi u-enja gde govornik 5-10 puta ponavlja odre|enu re-. Sistem VIP-100 je nai{ao na primenu u oblasti kontrole kvaliteta, kod unosa podataka u inteligentnim terminalima, programiranja ra-unara glasom ... Mnogi komercijalni sistemi za raspoznavanje govora koji su nastali kasnije koristili su iskustva ste-ena u eksploataciji sistema VIP-100 i imali su sli-nu koncepciju i organizaciju osnovnih modula. Zbog te{ko}a u realizaciji spektralne analize primenom Furijeovih transformacija u realnom vremenu, kod mikroprocesorskih sistema za prepoznavanje se koriste jednostavnije i efikasnije transformacije. Tako se na primer -esto koristi pristup preko autokorelacije. Blok {ema jednog mikroprocesorskog sistema za prepoznavanje koji koristi autokorelaciju signala prikazana je na slede}oj slici [10]:



Slika 5. Funkcionalna blok {ema mikroprocesorskog sistema za prepoznavanje.

Govorni signal se preko poja-ava-a dovodi na niskopojasni filter od 3800 Hz koji ima zadatak da elimini{e vi{e harmonike i signal {uma tako da se obezbe|uje vi{i faktor uspe{nog prepoznavanja. Zadatak komparatora je da na osnovu intenziteta ulaznog signala i reference  $U_{ref}$  odredi po-etak i kraj govorne sekvence. Autokorelator izra-unava autokorelaciju ulaznog signala po slede}oj formuli:

$$\Theta_{xx}(\tau) = \lim_{n \rightarrow \infty} (1/N \sum_{k=1, N} x(k)x(k+\tau)) \quad (1)$$

Klasifikator na osnovu vrednosti autokorelacije uzorka i referentnih vrednosti autokorelacija snimljenih u memoriji izra-unava Euklidovo rastojanje za sve referentne uzorke a minimum svih rastojanja predstavlja rezultat prepoznavanja:

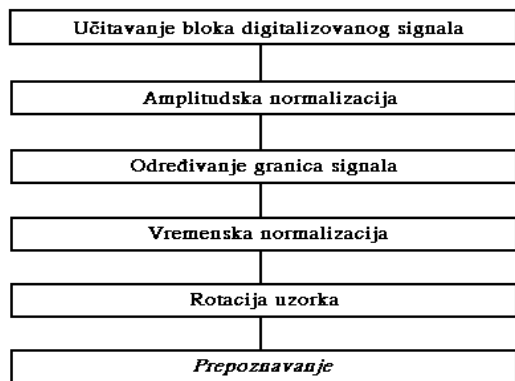
$$D_k = \min D_i, \text{ gde je } D_i = \sum_{k=0, n} |X_k - Y_k| \quad (2)$$

( $X_k$  i  $Y_k$  su autokorelacije snimljenog i referentnog uzorka respektivno). Sistem baziran na Motorolinom procesoru [10] koriste{i prikazan metod izra-unavanja autokorelacije postizao je faktor uspe{nog prepoznavanja od 90% rade{i u realnom vremenu.

### IV REALIZACIJA ALGORITMA ZA PREPOZNAVANJE GOVORNIH SEKVENCI

Kao prakti-ni deo rada autor je razvio hardversko-softverski sistem za prepoznavanje izolovanih govornih sekvenci za PC ra-unare [20].

Prepoznavanje govorne sekvence obavlja se obradom digitalizovane *master.WAV* datoteke. Metoda koja se koristi prilikom realizacije algoritma za prepoznavanje zasniva se na kori{enju karakteristika signala koje se mogu efikasno izdvojiti iz originalnog uzorka signala. U programu koji je implementiran paralelno se koriste dve karakteristike: **energetska** koja se zasniva na izra-unavanju sume intenziteta u nizu sukcesivnih prozora signala (blok u-itanog signala) kao i na **frekvencijskoj karakteristici** implementiranoj kao standardni algoritam za registrovanje broja prolazaka kroz nulu (**zero-crossing**). Pojedina-no kori{enje ovih karakteristika za prepoznavanje govornih sekvenci poznato je u literaturi [3],[10],[16],[17]. Novina koja je uvedena u metodi je njihovo simultano kori{enje u prepoznavanju. Faze u prepoznavanju govorne sekvence predstavljene su slede}im dijagramom toka:

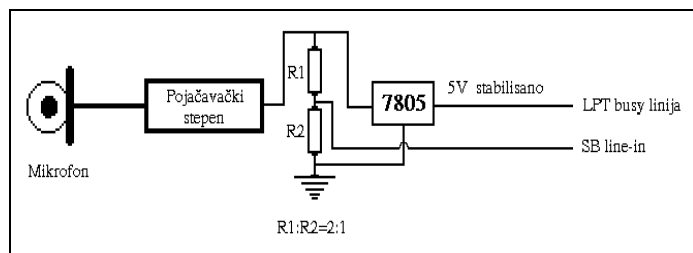


Slika 6. Dijagram toka procedure za prepoznavanje.

Funkcija za prepoznavanje sastoji se od poziva elementarnih procedura koje implementiraju pojedine faze navedene u predhodnom dijagramu toka.

## V REALIZACIJA HARDVERSKOG PODSISTEMA

Hardverski podsistem sistema za prepoznavanje izolovanih govornih sekvenci sastoji se od personalnog računara opremljenog minimalno procesorom **Pentium I 100Mhz** i šestobitnom zvukovnom karticom tipa *Sound Blaster* (Slika 7):

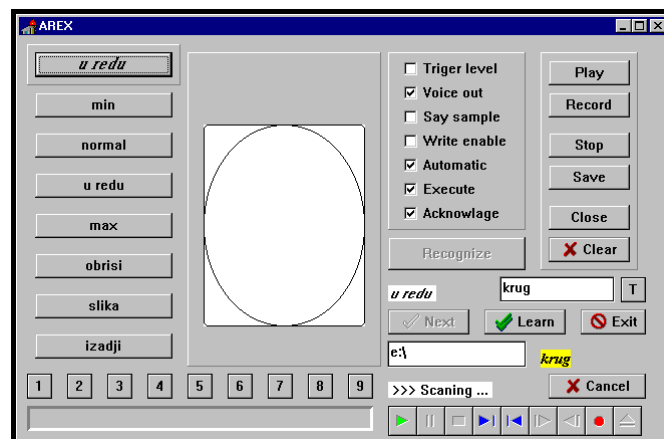


Slika 7. Hardverski podsistem.

Električni signal od mikrofona se preko pojačavačkog stepena vodi na razdelnik napona koji ima ulogu atenuatora predpojačanog govornog signala. Srednji izlaz iz razdelnika vodi se direktno na ulaz zvukovne kartice. Sa druge strane, paralelno razdelniku spojen je stabilizator **7805** čiji je zadatak da ograniči napon iz izlaza pojačavača na **5V** i da tako dobijeni signal dovede na jednu od ulaznih linija LPT porta (*busy*) [18]. Na ulazu se govorni signal digitalizuje na dve binarne vrednosti. Na taj način, paralelno sa signalom koji se digitalizuje 16-bitnim odmeravanjem generiše se binarni signal čija je svrha automatsko određivanje početka i kraja govorne sekvence. Skeniranje LPT porta obavlja se procedurom pisanom na asemblerskom jeziku [19] koja ima zadatak da kroz 500 prolaza prebroji promene binarnih vrednosti na LPT *busy* ulazu da bi se zatim na osnovu tih promena automatski odredio početak i kraj govorne sekvence.

## VI AREX - APLIKACIJA ZA PREPOZNAVANJE I IZVRŠAVANJE GOVORNIH KOMANDI

Program **AREX** predstavlja aplikaciju za računarsko prepoznavanje govornih sekvenci i izvršavanje programiranih komandi [20]. Aplikacija je pisana u sastavu praktičnog dela ovog rada u *Delphi V3.0* razvojnom okruženju i radi pod operativnim sistemom *Windows 95-2000*. Vizuelni izgled aplikacije dat je na sledećoj slici:



Slika 8. Vizuelni izgled aplikacije **AREX for Windows**.

Funkcionisanje pod operativnim sistemom *Windows* donosi neka značajna kvalitativna poboljšanja u odnosu na starije sisteme: bolji pregled i pristup opcijama programa, dinamičnost, mogućnost izvršavanja određenih programiranih funkcija u zavisnosti od rezultata prepoznavanja, automatsko određivanje početka i kraja govorne sekvence...

## VII ZAKLJUČAK

U radu su prikazane neke osnovne karakteristike sistema za prepoznavanje kontinualnog govora kao i za prepoznavanje govornih sekvenci kao jednostavniji vid tog problema. Kao praktični deo rada, prikazan je sistem AREX v1.0. Ova aplikacija, razvijena od strane autora, obuhvata hardverski i softverski podsistem i predstavlja komandni procesor upravljani govornim komandama.

## LITERATURA

- [1] Carl R. Strathmeyer "Voice in Computing: An Overview of Available Technologies", IEEE Computer, Vol. 23, No. 8, str. 10-16, August 1990.
- [2] Sadaoki Furui "Digital Speech Processing, Synthesis and Recognition", III version, NTT Human Interface Laboratories, Nippon Telegraph and Telephone Corporation, Tokio, Japan.
- [3] Richard D. Peacocke, Daryl H. Grat "An Introduction to Speech and Speaker Recognition", IEEE Computer, Vol. 23, No. 8, str. 26-34, August 1990.

- [4] Hermann Ney "A Comparative Study of Two Search Strategies for Connected Word Recognition: Dynamic Programming and Heuristic Search" , IEEE Transactions on pattern analysis and machine intelligence, vol. 14, No. 5, **str. 586-595**, May 1992.
- [5] Biing-Hwang Juang, Kuldip K. Paliwal "Hidden Markov Models with First-Order Equalization for Noisy Speech Recognition" , IEEE Transactions on signal processing, Vol. 40, No. 9, **str. 2136-2143**, September 1992.
- [6] X.D.Huang "Phoneme Classification Using Semicontinuous Hidden Markov Models" , IEEE Transactions on signal processing, Vol. 40, No. 5, **str. 1062-1067**, May 1992.
- [7] Neri Merhov, Yariv Ephraim "A Bayesian Classification Approach with Application to Speech Recognition" , IEEE Transactions on signal processing, Vol. 39, No. 10, **str. 2157-2166**, October 1991.
- [8] Radoslav Brki} "Prepoznavanje kontinualnog govora" , Diplomski rad, Elektronski fakultet u Ni{u.
- [9] Hermann Ney, Dieter Mergel, Andreas Noe, Annedore Peaseler "Data Driven Search Organization for Continous Speech Recognition" , IEEE Transactions on signal processing, Vol. 40, No. 2, **str. 272-281**, February 1992.
- [10] Ludvik Gyergyek, Nikola Pave{i}, Slobodan Ribari} "Uvod u raspoznavanje uzoraka" , Tehni-ka knjiga, Zagreb, Septembar 1988.
- [11] S.M.Kr-o, V.D.Deli}, V.S.Milo{evi} "Realizacija sistema za prepoznavanje izolovano izgovorenih re-i" , XXXVIII Konferencija za ETRAN, Sveska II, Komisija za Akustiku, **str. 195-196**, Ni{, 7-9 juna 1994.
- [12] Goran Jovanovi} "APG<sub>L</sub> sistem za prepoznavanje govora na bazi ekspertske zasnovane fokusirane strukturalne analize" , XXXVIII Konferencija za ETRAN, Sveska II, Komisija za Akustiku, **str. 205-206**, Ni{, 7-9 juna 1994.
- [13] Z.Uro{evi}, Milan D.Savi}, S.Ili}, B.Savi} "Primena BOX counting metode za segmentaciju govora" , XXXVIII Konferencija za ETRAN, Sveska II, Komisija za Akustiku, **str. 197-198**, Ni{, 7-9 juna 1994.
- [14] Milan D. Savi}, Z. Uro{evi}, B. Savi} "Segmentacija izolovano izgovorenih re-i na manje govorne jedinice" , XXXVIII Konferencija za ETRAN, Sveska II, Komisija za Akustiku, **str. 199-200**, Ni{, 7-9 juna 1994.
- [15] Yariv Ephraim "Statistical-Model-Based Speech Enhancement Systems" , Proceedings of the IEEE, Vol. 80, No. 10, **str. 1526-1558** , October 1992.
- [16] R. Rabiner, R. W. Schafer "Digital Processing of Speech Signals", Bell Laboratories, Prentice-Hall, Inc. , U.S.A. , 1978.
- [17] James W. Pitton, Kuansan Wang, Biing-Hwang Juang "Time-Frequency Analyses and Auditory Modeling for Automatic Recognition of Speech" , Invited Paper, Proceedings of the IEEE, Vol. 85, No.9, **str. 1199-1215**, September 1996.
- [18] Te{evi} "PC/XT Hardware" , Tehni-ka knjiga, Beograd, 1992.
- [19] Borland International "Turbo Assembler User's Guide" , Scotts Valley , U.S.A. , 1988.
- [20] Vladan Vu-kovi} "Digitalna obrada i ma{insko prepoznavanje izolovanih govornih sekvenci" , magistarska teza, Elektronski fakultet u Ni{u, maj 1997.

**Abstract** - This paper is concerned with some basic characteristics of the continual speech recognition as well as isolated speech sequences recognition as the simpler variance of that problem. As the practical contribution, author has developed the AREX v1.0 system, including hardware and software subsystem, representing voice command processor.

## THE METHODS AND SYSTEMS FOR THE AUTOMATIC ISOLATED SPEECH SEQUENCE RECOGNITION, Vladan Vu-kovi}