

(19) REPUBLIKA SRBIJA

(12) Patentni spis

(11) 50395 B



ZAVOD ZA
INTELEKTUALNU SVOJINU
BEOGRAD

(51) Int. Cl.⁽⁸⁾
G 10 L 11/00 (2008.04)

(21) Broj prijave: **P-2007/0515**
(22) Datum podnošenja prijave: **28.12.2007.**
(43) Datum objavljivanja prijave: **06.05.2009.**
(45) Datum objavljivanja patenta: **31.12.2009.**
(30) Međunarodno pravo prvenstva:

(61) Dopunski patent uz osnovni patent broj:
(62) Izdvojen patent iz prvobitne prijave broj:

(73) Nosilac patenta:
**CENTAR ZA UNAPREĐENJE
ŽIVOTNIH AKTIVNOSTI**
Gospodar Jovanova 35,
11000 Beograd, RS
(72) Pronalazači:
JOVIČIĆ, S.;
ŠARIĆ, Z.;
FURUNDŽIĆ, D.;
SUBOTIĆ, M.
(74) Zastupnik:

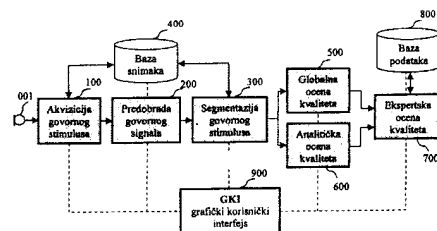
(54) Naziv: **SISTEM I POSTUPAK ZA
PROCENU KVALITETA ARTIKULACIJE**

(51) Int. Cl.⁽⁸⁾
G 10 L 11/00 (2008.04)

(57) Apstrakt:

Pronalazak se odnosi na sistem i postupak za objektivnu procenu kvaliteta artikulacije kod izgovora srpskih fonema. Zasnovan je na računarskoj tehnologiji koja omogućava snimanje izgovora ispitanika pod testom, analizu njegovog glasa i ekspertsku ocenu kvaliteta artikulacije. Sistem se sastoji od mikrofona za snimanje glasa, računara koji vrši analizu snimljenog govornog signala, baze arhiviranih govornih snimaka, baze podataka o pacijentima i njihovim nalazima i grafičkog korisničkog interfejsa putem koga korisnik sistema komunicira sa sistemom u cilju izbora modaliteta operativnog funkcionisanja sistema, kontrole postupka snimanja govornih stimulusa, vizuelizacije rezultata analize i kontrole postupaka: arhiviranja snimaka, održavanja baza sistema, treninga i obuke sistema i izrade izveštaja. Postupak obrade i analize govornog signala se sastoji u primeni neuralne mreže za globalnu klasifikaciju pravilna/nepravilna artikulacija i u primeni statističkih modela za analitičku ocenu kvaliteta

artikulacije na bazi određivanja distanci između skupa akustičkih obeležja tipične/atipične realizacije foneme.



RS 50395 B

OBLAST TEHNIKE NA KOJU SE PRONALAZAK ODNOSI

Pronalazak pripada oblasti obrade govornog signala, ili preciznije, metodama i sistemima snimanja, analize i objektivnog ocenjivanja kvaliteta artikulacije zasnovanim na računarskoj tehnici.

TEHNIČKI PROBLEM

Prepoznavanje govora i prepoznavanje govornika su problemi kojima se naučnici i stručnjaci intenzivno bave poslednjih 30-tak godina. U poslednje vreme analiza glasa i govora se sve više koristi u medicini kao dijagnostičko sredstvo koje može dati pouzdane rezultate u otkrivanju i lečenju različitih obolenja. Jedna od primena je i detekcija patoloških ispoljavanja govora i jezika. Različiti oblici govorne patologije imaju za posledicu nepravilan govor odnosno uočavaju se kao odstupanja od tipičnih izgovora u okviru normi datog jezika. Ova odstupanja, zavisno od težine, mogu biti na nivou pojedinačnih izgovornih glasova, grupe glasova, ritma, tempa i melodije rečenice. Automatsko prepoznavanje odstupanja u kvalitetu izgovora glasova predstavlja složen problem koji u sebi objedinjava elemente prepoznavanja govora i prepoznavanja govornika. Sistemi za prepoznavanje govora koriste različite modele i algoritme kako bi u govornom signalu prepoznali sa određenom verovatnoćom konstitutivne elemente datog jezika. Kod prepoznavanja govornika teži se otkrivanju onih obeležja u govornom signalu koji su vezani za njegove anatomno funkcionalne karakteristike. Detekcija odstupanja u izgovoru zasniva se na definisanju postupaka i procedura koji imaju mogućnost određivanja distance između tipičnog i atipičnog (do patološkog) izgovora.

Svaki izgovorni glas može se predstaviti skupom akustičkih obeležja koja ga definišu. U višedimenzionom prostoru parametara (obeležja) svi varijeteti izgovora nalaze se unutar ograničene hiperdimenzione oblasti. Granice ove oblasti, vezane su za tipične izgovore u okviru datog jezika i određene su lingvističkim i paralingvističkim faktorima. Uticaj lingvističkih faktora se ogleda kroz uticaj semantike, sintakse,

gramatike i leksike dok pol, uzrast, poreklo, psihofiziološki status (emotivno stanje, bolest, itd.) kao paralingvistički faktori imaju značajan uticaj na izgovor. Samim tim broj faktora koji utiču na granice varijacionog polja akustičkog obeležja je veliki.

Odstupanja od tipičnog izgovora mogu se kretati od lakih (kod kojih akustička obeležja malo izlaze iz oblasti dozvoljenog varijacionog polja) preko teških patoloških formi gde je izgovorni glas toliko izmenjen da se veoma teško percipira kao određena fonema, pa do najtežih oblika kod kojih dolazi do supstitucije ili omisije izgovornih glasova.

Sistem za objektivnu procenu kvaliteta izgovora treba da razreši nekoliko problema: definisanje skupa akustičkih obeležja koji odgovaraju svakoj fonemi datog jezika, definisanje funkcije uticaja pojedinih lingvističkih i paralingvističkih parametara na varijaciono polje akustičkog obeležja, iznalaženje funkcije zavisnosti između pojedinih parametara, iznalaženje funkcije zavisnosti između pojedinih akustičkih obeležja, definisanje granice višedimenzionog prostora u kome se kreće tipična realizacija foneme, definisanje mere distance u višedimenzionom prostoru između tipične realizacije foneme i njenog odstupanja.

STANJE TEHNIKE

Analiza govornog signala u cilju identifikovanja artikulacionih odstupanja u izgovoru ili patologija u glasu jeste problem koji je počeo intenzivno da se istražuje još sredinom prošlog veka (na primer: P. Liebermann "Perturbations in vocal pitch" *J. Acoust. Soc. Am.*, Vol. 33, No. 5, 1961). Sa razvojem računarske tehnike, metoda digitalne obrade govornog signala i matematičkih algoritama u prepoznavanju oblika, govora i govornika, razvijali su se i postupci u analizi kvaliteta govornog signala, identifikovanja kvaliteta artikulacije i patologije u govornom izrazu. Na primer: G. Llorente, S. Navarro et al, "On The Selection of Meaningful Speech Parameters Used Pathologic/Nonpathologic Voice Register Classifier", *Eurospeech '99*, Volume 1, pp. 563-566, 1997; D. Michaelis, M. Frohlich, H.W. Strube, "Selection and combination of acoustic features for the description of pathologic voices", *J. Acoust. Soc. Am.*, Vol. 103, No. 3, pp 1628-1639, 1998; M. E. Cesar, R. L. Hugo, "Acoustic Analysis of Speech for Detection of Laryngeal Pathologies", *Proc. 22nd Annual EMBS Int. Conf.*, pp. 2369-2372, July 2000; C. Maguire, P. De Chazal, R. B. Reilly, P. Lacy; „Automatic classification of voice pathology using speech analysis“, *Word Congress on Biomedical*

Engineering and Medical Physics, Sydney, August 2003; C. Maguire, P. De Chazal, R. B. Reilly, P. Lacy; „Identification of voice pathology using automated speech analysis“, *Proc. Of the 3rd International Workshop on Models and Analysis of Voice Emission for Biomedical Applications*, Florence, December 2003; J. I. Godino-Llorente, P. Gomez-Vilda; „Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors“, *IEEE Transaction on Biomedical Engineering*, Vol. 51, No. 2, pp. 380-384, February 2004; R. Moran, R. B. Reilly, P. de Chazal, P. Lacy; “Telephone based voice pathology assessment using automated speech analysis and VoiceXML”, ISSC 2004, Belfast, June 2004; itd. Suština svih ovih istraživanja jeste u izdvajanju vektora obeležja u govornom signalu normalnog i patološkog glasa i prepoznavanja odstupanja na bazi različitih kriterijuma. Pristupi su veoma različiti počev od izbora govornog stimulusa za analizu, do izbora obeležja za analizu, upotrebe matematičkih algoritama za diskriminaciju, interpretacije rezultata, itd. Takođe, još uvek nije dovoljno istražena korelacija objektivnih akustičkih obeležja u govornom signalu sa perceptivnim opservacijama terapeuta-logopeda o artikulacionim odstupanjima.

Zbog široke lepeze odstupanja u artikulaciji do sada nije pronađeno jedinstveno rešenje sistema koji bi na automatizovan način mogao identifikovati odstupanje i pravilno ga klasifikovati. Parcijalna rešenja su patentirana kao na primer u: U.S. patentu br. US 6,714,911 B2, prihvaćen 30. marta 2004., sa naslovom „Speech transcription and analysis system and method“, koji daje jedno rešenje metodološkog pristupa u proceni korektne/nekorektne produkcije foneme primenom računarske tehnologije ali uz neophodno učešće terapeuta, ili na primer u: U.S. patentnoj prijavi br. US 2007/0005357 A1, prijavljena 29. juna 2005., sa naslovom „Telephone pathology assessment“, koja prikazuje rešenje koje procenjuje postojanje patologije u glasu udaljenih govornika preko telefonske mreže na bazi merenja većeg skupa parametara u govornom signalu i klasifikacije normalan/patološki govor primenom LDA (*Linear Discriminant Analysis*) i HMM (*Hidden Markov Model*).

Rešenje, izloženo u ovom patentu, bazira se takođe na vektoru obeležja detektovanih u analiziranom govornom stimulusu koji se neuralnom mrežom analizira u cilju dobijanja globalne procene kvaliteta artikulacije ali i posebnom analizom pojedinih obeležja koja su tipična za specifično artikulaciono odstupanje karakteristično za svaki izgovorni glas, za svaku fonemu.

IZLAGANJE SUŠTINE PRONALASKA

Predmet ovog patenta je sistem i postupak za objektivnu procenu kvaliteta artikulacije pomoću računara. Računar, akvizicioni sistem za reprodukciju i snimanje i programski paket za snimanje, skladištenje podataka i specijalizovanu obradu govornog signala čine elemente predloženog patentnog rešenja.

Sistemom, koji je predmet pronalaska, definiše se skup test stimulusa koji se koriste kao modeli pravilnog izgovora trideset fonema srpskog jezika, način snimanja i skladištenja izgovora ispitanika, i algoritmi i softverski moduli za obradu i procenu kvaliteta artikulacije.

Suština pronalaska jeste u specifičnoj obradi govornog signala koji se snima u akustičkom ambijentu prostorije u kojoj se nalazi i sistem i govornik. Izgovor test stimulusa od strane ispitanika snima se u digitalnom obliku (i ostaje sačuvan u računaru) i služi kao ulazni set podataka za procenu kvaliteta artikulacije. Dodatni parametri bitni za adekvatnu procenu kvaliteta artikulacije su pol i starost ispitanika koji se unose u bazu podataka o ispitanicima. Ovaj set podataka ulazi u module za globalnu i analitičku procenu kvaliteta artikulacije, koji vrše suštinsku analizu govornog signala.

Specifičan aspekt pronalaska jeste realizacija sistema na bazi dva testa: testa globalne procene kvaliteta artikulacije i testa analitičke ocene kvaliteta artikulacije svih 30 fonema u srpskom jeziku. Testom globalne procene kvaliteta artikulacije dobija se osnovni podatak o kvalitetu artikulacije: da li je artikulacija dobra, da li je artikulacija loša ili da li je artikulacija na granici dobre/loše artikulacije. Test analitičke ocene kvaliteta artikulacije odnosi se na procenu položaja svih akustičkih obeležja u granicama varijacionog polja za posmatranu fonemu. Pojedina od ovih obeležja direktno koreliraju sa anatomo funkcionalnim karakteristikama artikulacionog aparata ispitanika. Time je omogućena precizna dijagnostika na nivou pojedinih elemenata artikulacionog aparata.

Procena globalnog kvaliteta artikulacije zasniva se na akumuliranom ekspertskom znanju ocenjivanja artikulacije "Globalnim artikulacionim testom" (GAT). GAT (Kostić, Đ., Vladisavljević, S., & Popović, M., *Testovi za ispitivanje govora i jezika*. Zavod za udžbenike i nastavna sredstva, Beograd, 1983) je test koji se sastoji od trideset stimulus reči koje ispitanik ponavlja (imitira pravilan izgovor ispitivača) i na osnovu audio vizuelne procene ekspert daje ocenu o kvalitetu artikulacije svakog

izgovornog glasa. Test reči su uglavnom dvosložne reči (najčešće sastavljene od tri tj. četiri foneme). Kod test reči u kojima se ispituje artikulacija vokala, vokal se nalazi u interkonsonantskom položaju, a u test rečima za ispitivanje artikulacije suglasnika, suglasnici se nalaze u inicijalnoj poziciji u reči. Glasovno okruženje ispitivane foneme birano je tako da je uticaj koartikulacije minimalan na položaj artikulacionih organa pri izgovoru ispitivanog glasa. Način izvođenja testa omogućava da se izbegnu dijalekatski varijeteti i uticaj dijalekta na artikulaciju ispitivanog glasa u svakoj od stimulus reči. Testom se ocenjuje samo izgovor ispitivanog glasa, ne i ostalih glasova u test rečima. Ocena predstavlja globalni akustički i vizuelni utisak o kvalitetu izgovora i položaju artikulacionih organa u toku artikulacije za svaki izgovorni glas. Ocene na GAT-u su date na dva nivoa: procena artikulacije svakog izgovornog glasa (ocena od 1 do 7 ili (+), (+-), (-)), i kvalitet izgovora u odnosu na hronološki uzrast. U ovom patentu korišćene su ocene (+) dobra artikulacija, (-) loša artikulacija i (+-) granična artikulacija.

Analitička ocena se odnosi na mogućnost procene odstupanja akustičkih obeležja foneme u odnosu na tipično varijaciono polje datog obeležja u okviru posmatrane foneme. U okviru analitičkog testa za procenu kvaliteta izgovora fonema srpskog jezika (Kostić, D., Vladislavljević, S., & Popović, M., *Testovi za ispitivanje govora i jezika*. Zavod za udžbenike i nastavna sredstva, Beograd, 1983) definisan je skup artikulaciono akustičkih obeležja za svaki izgovorni glas u okviru glasovne grupe kojoj pripada. Svako od ovih obeležja doprinosi globalnoj predstavi artikulacije ispitivanog glasa. Skup obeležja definisan u analitičkom testu iskorišćen je za procenu kvaliteta izgovora glasova u globalnom artikulacionom testu posebno na nivou akustičkih obeležja datog glasa. Neka od ovih obeležja su binarna odnosno jedna isključuju druga dok su druga u obliku gradacije (slaba, jaka, ima nema).

Test stimulusi su reči i oni čine sledeći specifičan aspekt pronalaska. Reči su izabrane tako da zadovolje nekoliko kriterijuma: da je njihovo značenje poznato najširoj populaciji govornika (pre svega dečijoj), da se testirane foneme nalaze u takvom koartikulacionom okruženju u kome se najlakše izgovaraju i da se lako percipiraju. Stimuluse čini skup od 30 reči sa značenjem. Test stimulusi su memorisani u računaru u digitalnom obliku i služe kao model tipičnog izgovora koji se pušta ispitaniku sa ciljem da ispitanik reprodukuje test reč koju je čuo. Test stimulusi su snimljeni na bazi izgovora terapeuta-logopeda kao tipičnog govornika sa pravilnom artikulacijom svih izgovornih glasova srpskog jezika.

Sledeću specifičnost pronalaska čini moduo za globalnu procenu artikulacije koji je zasnovan na modelu neuralnih mreža. On na osnovu prethodnog znanja klasifikuje izgovor foneme u kategoriju dobre, loše ili granične artikulacije. Znanje koje neuralna mreža koristi pri donošenju odluka zavisi od pola i starosti ispitanika tako da su parametri mreže različiti za osobe muškog i ženskog pola, decu, mlade i stare. Prethodno znanje se obezbeđuje obučavanjem neuralne mreže na ciljano struktuiranom uzorku izgovora ocenjenih od strane eksperata-logopeda.

Specifični aspekt pronalaska jeste i moduo za analitičku ocenu kvaliteta artikulacije. On je zasnovan na statističkim modelima koji za procenu pojedinih akustičkih obeležja foneme koriste prethodno znanje dobijeno obučavanjem. Za svaku fonemu postoji definisan skup akustičkih obeležja kao i algoritmi za izdvajanje i procenu kvaliteta samog obeležja. Za obuku se koristi isti uzorak kao i kod neuralnih mreža. I u ovom modulu parametri za statističko određivanje distance zavisni su od pola i uzrasta ispitanika.

Dobijeni rezultati globalne i analitičke ocene kvaliteta artikulacije, svih trideset fonema, ulazni su podaci modula za dijagnostičku procenu kvaliteta izgovora. Ovo je posebna specifičnost pronalaska. U ovom modulu ugrađeno je lingvističko, medicinsko i defektološko znanje na osnovu koga se procenjuje kvalitet izgovora. Za dete uzrasta od 2 do 5 godina poznato je koji glasovi i u kojoj starosnoj dobi mogu da odstupaju od pravilnog izgovora. Odstupanje jednog akustičkog obeležja u više fonema jasno ukazuje na poremećaj specifičnog dela artikulacionog aparata itd.

Inventivnost u ovom pronalasku se nalazi u algoritamskoj realizaciji svake od navedenih specifičnosti, koje su zasnovane na rezultatima sopstvenih istraživanja u domenu srpskog jezika, ali i u postupku integrisanja svih algoritama u jedinstvenu celinu koja funkcioniše stabilno i kvalitetno pri čemu su mnoge rutine optimizirane korišćenjem zajedničkih resursa.

Ovi i drugi aspekti, specifičnosti i benefiti ovog pronalaska biće očigledniji nakon uvida u detaljan opis pronalaska, patentne zahteve i pripadajuće crteže.

KRATAK OPIS SLIKA I NACRTA

Slika 1 – prikazuje osnovne elemente sistema za procenu kvaliteta artikulacije.

Slika 2 – prikazuje detaljnu blok šemu akvizicionog modula.

Slika 3 – prikazuje postupak predobrade govornog signala.

Slika 4 – prikazuje blok šemu modula u kome se odvija proces segmentacije ulaznog stimulusa.

Slika 5 – prikazuje: a) strukturu neuronske mreže **NM**, b) konfiguraciju jednog neurona i c) prenosnu funkciju jednog neurona.

Slika 6 – prikazuje: a) strukturu ulaznog sloja neurona, b) strukturu skrivenog sloja neurona i c) strukturu izlaznog sloja neurona.

Slika 7 – prikazuje detaljnu blok šemu modula za analitičku ocenu kvaliteta artikulacije.

Slika 8 – prikazuje strukturu grafičkog korisničkog interfejsa **GKI**.

DETALJAN OPIS PRONALASKA

Ovaj pronalazak opisuje sistem i postupak obrade govornog signala za objektivnu ocenu kvaliteta artikulacije ili konkretnije, za utvrđivanje stepena odstupanja pojedinog akustičkog obeležja u okviru posmatrane foneme.

Na slici 1 prikazana je blok šema osnovne strukture sistema. Ispitanik izgovara stimuluse prema specifikaciji testa implementiranog u sistem a izgovor se snima putem mikrofona **001** i memoriše u digitalizovanoj formi u računar. Proces akvizicije govornog signala kontroliše modul **100**, koji je u neposrednoj vezi sa bazom snimaka, modul **400**, i grafičkim korisničkim interfejsom **GKI**, modul **900**. Baza snimaka **400** ima dvojaku ulogu: prvo, pri upotrebi sistema svaki prihvaćeni stimulus može se memorisati čime se baza snimaka proširuje i drugo, izabrani snimci iz baze koriste se u fazi akvizicije ulaznog stimulusa (kasnije će biti detaljnije objašnjena uloga baze). Snimljen i memorisan stimulus prosleđuje se u modul **200** gde se započinje proces obrade govornog signala. Ovaj postupak podrazumeva primenu seta algoritama pomoću kojih se iz govornog signala određuje skup parametara koji karakterišu govorni signal u frekvencijskom domenu. Dalje se i govorni signal i određeni parametri prosleđuju modulu **300** u kome se vrši segmentacija ulaznog stimulusa. Segmentacija obuhvata određivanje inicijalne i finalne granice stimulusa, granice fonema u stimulusu i granice određenih subfonemskih segmenata. U ovom bloku se vrši i estimacija dodatnog skupa statističkih obeležja potrebnih za funkcionisanje blokova **500** i **600**.

U bloku **500** vrši se globalna ocena kvaliteta artikulacije primenom neuralne mreže (**NM**) na bazi skupa estimiranih obeležja u ulaznom stimulusu. Zadatak **NM** je da izvrši klasifikaciju (ocenu) artikulacije u tri nivoa: korektna, nekorektna i granična. Parametri **NM** su optimizirani za svaki stimulus prema Globalnom artikulacionom testu. Paralelno se vrši analiza ulaznog stimulusa u bloku **600** gde se dobija analitička ocena kvaliteta artikulacije. U ovom bloku se vrši analiza različitog skupa akustičkih obeležja izdvojenih za svaki fonem u srpskom jeziku a koji su karakteristični indikatori za pojedine oblike odstupanja u artikulaciji. Sve informacije iz blokova **500** i **600** koriste se u bloku **700** za formiranje konačne ekspertske ocene kvaliteta artikulacije. Ova ocena se prezentuje ispitaniku i ekspertu (logopedu) vizuelno na ekranu računara ili u formi izveštaja na posebnom formularu, a istovremeno se i arhivira u bazi podataka **800**. Ova baza je organizovana tako da se može pretraživati (koristiti) prema različitim zahtevima (kriterijumima) kao što su: pacijent, vrsta stimulusa, vrsta odstupanja, itd.

Komunikacija sa sistemom od strane korisnika (ispitanik i/ili logoped) vrši se preko grafičkog korisničkog interfejsa **GKI**, blok **900**. On je u vezi sa svim ostalim blokovima u sistemu i obezbeđuje prenos podataka od sistema ka korisniku i obrnuto. Time je omogućena potpuna i optimalna kontrola funkcionisanja sistema ali i potpuna informisanost korisnika sistema.

Na slici 2 prikazana je detaljna blok šema akvizicionog modula, moduo **100**. Govorni signal iz mikrofona **001** ulazi u blok **101** koji vrši detekciju aktivnosti govornog signala, **VAD** (*voice activated detector*). Naime, kada se preko **GKI**, blok **900**, uputi zahtev sistemu za snimanje stimulusa, sistem se uključuje i u stanju je čekanja za izgovorom. Nakon izgovorenog stimulusa sistem se isključuje posle izvesnog vremena a **VAD** detektor detektuje efektivan stimulus i obezbeđuje da se optimalno memoriše samo govorni signal, bez pauza ispred ili iza stimulusa ili eventualnih smetnji u snimku. Takođe, u ovom bloku se vrši analiza kvaliteta govornog signala na bazi estimacije odnosa signal/šum (S/N) i u slučaju prisustva neprihvatljivog nivoa smetnji, loš odnos S/N , sistem blokira snimanje takvog stimulusa i preko bloka **GKI**, blok **900**, informiše o tome korisnika. U bloku **102** vrši se poređenje snimljenog stimulusa sa referentnim stimulusom iz baze snimaka, blok **400**, i na bazi mere odstojanja donosi se odluka o prepoznatom stimulusu. Ovo je neophodna aktivnost sistema kako bi se sprečio prihvatanje neadekvatno (pogrešno) izgovorenog stimulusa. U bloku **103** se vrši odluka o ulaznom stimulusu, koja nije jednostavna obzirom na

varijabilnosti u izgovoru kao i moguću patologiju. Odluka može biti pozitivna, negativna ili neizvesna. U slučaju negativne odluke šalje se informacija korisniku, preko **GKI**, da ponovi izgovor. U slučaju neizvesne odluke sistema pruža se mogućnost korisniku da odluči da li će da prihvati ili odbaci snimljeni stimulus. Pozitivnom odlukom prihvata se stimulus i sa jedne strane memoriše se u bazi snimaka **400**, dok se na drugoj strani upućuje na dalju analizu.

Baza snimaka **400** je neophodna iz nekoliko razloga; prvo, arhivirani snimak može naknadno da se analizira i time dijahrono prati proces rehabilitacije pacijenta; drugo, baza snimaka je neophodna za dodatnu obuku delova sistema (neuralne mreže i podsistema za segmentaciju) i treće, baza snimaka može da se koristi za istraživačke namene.

Na slici 3 prikazan je postupak predobrade govornog signala koji se izvršava u bloku **200**. On podrazumeva korišćenje više standardizovanih postupaka kao što su: podela govornog signala na frejmove na kojima se vrši dalja obrada i analiza, primena preemfazisa, primena Hanningovog prozora, određivanje brze Fourijeve transformacije (FFT), određivanje magnitude spektra, primena banke mel-filtara i konačno, određivanje kepstra i mel-kepstralnih koeficijenata **MFCC** (*Mel Frequency Cepstral Coefficients*). U daljoj analizi koristi se prvih 12 **MFCC** koeficijenata.

Slika 4 prikazuje blok šemu modula **300** u kome se odvija proces segmentacije ulaznog stimulusa. Izgovoreni stimulus kao jedna reč deli se segmentacijom na fonemske i subfonemske segmente. Postupak koji se primenjuje je tzv. uslovna segmentacija, odnosno deoba reči na akustičke celine kada se unapred zna koja je reč izgovorena. Ovakav pristup omogućava znatno precizniju segmentaciju u odnosu na pristup gde se ne uzima u obzir koja je reč izgovorena.

Segmentacija se ostvaruje u bloku **301** postupkom dinamičkog vremenskog usklađivanja (DTW) ulazne reči izgovorene od strane ispitanika sa *modelima* iste reči koje su nalaze memorisane u bazi modela reči, blok **302**. Baza *modela* reči formirana je od uzoraka normalnog i patološkog izgovora posmatrane reči od strane ispitanika oba pola i različitog doba starosti, koji su ocenjeni i klasifikovani od strane eksperta-ocenjivača prema tipu i stepenu patologije. Svi uzorci reči su segmentirani ručno zbog veće preciznosti. Segmentne celine uzoraka reči su parametrizovane u cilju sažetog opisa statističkih karakteristika segmentnih celina. Parametarski opis sadrži sledeće: a)

segmentnu strukturu reči, b) opis spektra segmenata na bazi mel-kepstralnih koeficijenata (**MFCC**), c) vremensko ograničenje svakog segmenta i d) simbolički opis trenda snage unutar pojedinih segmenata. U cilju redukovanja broja modela u bazi izvršeno je klasterovanje modela, uz minimalan gubitak polazne informacije, tako da baza sadrži **M** modela reči.

Postupak automatske segmentacije reči može se opisati kroz sledeće algoritamske korake:

Za svaki od modela reči iz baze **302**, bilo da je on patološki ili normalan, primenjuje se **DTW** postupak dinamičkog vremenskog usklađivanja u modulu **301**. Rezultat ovog postupka je vrednost mere odstojanja između ulazne reči i svakog od modela reči iz baze **302**. Za meru odstojanja upotrebljena je Euklidska mera odstojanja **MFCC** koeficijenata definisana relacijom

$$d(i, j) = \sum_{k=1}^{12} (C_k^i - C_k^j)^2,$$

gde je $d(i, j)$ mera odstojanja i -tog segmenta analizirane (test) reči i j -tog segmenta modela (referentne) reči. Veličine C_k^i i C_k^j su k -ti **MFCC** koeficijenti redom i -tog test segmenta i j -tog referentnog segmenta. Favorizovanje prolaska **DTW** putanje kroz određene tačke od interesa na grafu povezanosti, kao što su na primer trenuci nagle promene stacionarnosti signala, ostvaruje se dopunom odstojanja $d(i, j)$ novim članovima prema relaciji

$$\tilde{d}(i, j) = d(i, j) + d^1(i, j) + \dots + d^k(i, j),$$

gde su $d^1(i, j), \dots, d^k(i, j)$ članovi koji opisuju dopunsko znanje bilo da su zasnovani na apriornom znanju o granicama segmenata ili da je to znanje dobijeno na osnovu dopunske analize.

Od **M** kandidata modela reči iz bloka **302** vrši se izbor 10 modela sa najmanjom distancom u bloku **303** a njihove distance se usrednjavaju u bloku **304**. Time je dobijena optimalna procena segmenata ulazne reči. Konačno, posle izvršene segmentacije ulazne reči određuju se sledeća akustička obeležja u analiziranom govornom signalu u bloku **305** koja ulaze u sadržaj ulaznog vektora za neuralnu mrežu **NM**: veličina trajanja foneme koja se analizira Globalnim i Artikulacionim testom T_f , veličina trajanja cele

analizirane reči T_r , odnos ove dve veličine T_f/T_r , srednja snaga signala foneme P_{srf} , standardna devijacija kratkovremene snage foneme STD_{Pr} , standardna devijacija kratkovremene snage reči STD_{Pr} i odnos P_{srf}/STD_{Pr} .

Na slici 5a prikazana je struktura neuronske mreže **NM** koja je bazirana na višeslojnom Perceptronu. Sastoji se iz 3 sloja: ulazni sloj koji sadrži 3 neurona, skriveni sloj sa 10 neurona i izlazni sloj sa 3 neurona. Struktura jednog neurona data je na slici 5b, dok je prenosna funkcija neurona prikazana na slici 5c, u formi funkcije $Y = \tanh(X)$. Ulazni vektor obeležja, koji dolazi na ulazni sloj neurona, sadrži 19 obeležja: 12 MFCC koeficijenta i 7 obeležja određena u bloku **305**. Svako obeležje iz ulaznog vektora (slika 5a) dolazi na svaki ulazni neuron preko matrice težinskih koeficijenata W_{Pi} ($P = 1, \dots, 19$; $i = 1, 2, 3$). Detaljan opis ulaznog sloja neurona prikazan je na slici 6a. Izlaz ulaznog sloja neurona je vektor $V1t$. On se pojavljuje kao ulaz skrivenom sloju neurona preko matrice $W2j$ ($j = 1, \dots, 10$) (slika 6b), dok je izlaz ovog sloja vektor $V2t$. I konačno, izlazni sloj neurona prihvata vektor $V2t$ preko matrice $W3k$ ($k = 1, 2, 3$) i nakon transformacije dobija se izlazni vektor neuronske mreže Y , sa svoja tri kodovana izlaza. Vektorom $[1 \ 0 \ 0]$ označena je pravilna artikulacija, vektorom $[0 \ 0 \ 1]$ označena je nepravilna artikulacija, a vektorom $[0 \ 1 \ 0]$ označena je granična artikulacija. Znanje obučene neuronske mreže **NM** sadržano je u matricama težinskih koeficijenata W_{Pi} dimenzija 19×3 , $W2j$ dimenzija 3×10 i $W3k$ dimenzija 10×3 . Finalna globalna ocena kvaliteta artikulacije analizirane foneme dobija se kao srednja vrednost ocena po frejmovima unutar granica segmentacije za datu fonemu.

U fazi obuke neuronskoj mreži **NM** dovode se sinhrono parovi ulaznih i izlaznih vektora u cilju minimizacije srednje kvadratne greške izlaza u odnosu na željeni izlaz. Procedura traje dok model ne postigne željene performanse koristeći u toku obuke druge ulazno – uzlazne parove za validaciju sopstvene sposobnosti generalizacije odnosno za predupređenje *overfitting* situacije. Posredstvom nelinearne funkcije preslikavanja i paralelne distribucije rešenja, a u skladu sa teoremom Kolmogorova, greška pri obuci konvergira postavljenoj zahtevanoj vrednosti. Povećanjem prezentovanog skupa ispitanika u toku obuke povećava se sposobnost generalizacije i tačnost evaluacije kvaliteta izgovora (ocene) nepoznatih ispitanika.

Na slici 7 prikazana je detaljna struktura modula za analitičku ocenu kvaliteta artikulacije. Ovoj oceni se pristupa nakon detekcije prisustva patologije u govoru u bloku **500**. Najpre se u bloku **601** vrši izdvajanje u ulaznom govornom signalu dela

signala sa fonemom koja je predmet analize a na bazi informacija o granicama segmenata iz bloka **300**. U blokovima 602 i 603 vrši se vremensko-frekvencijska analiza izdvojene foneme i to u bloku 602 spektar signala se ocenjuje DFT transformacijom, dok se u bloku 603 spektar signala ocenjuje Burg-ovim postupkom modeliranja signala AR modelom. U oba slučaja rezultat je 2D matrica kojom se opisuje vremensko-frekvencijska zavisnost logaritma spektra snage signala. Dalje se u bloku **604** vrši izdvajanje trajektorije snage signala na unapred definisanim frekvencijskim podopsezima i u bloku **605** lociranje u vremenu energetskih koncentrata koji će u narednim koracima obrade biti podvrgnuti analizi u odnosu na moguće prisustvo patologije govora.

Koristeći vremensko-frekvencijsku predstavu analiziranog fonema, energetske trajektorije, položaje energetskih koncentrata reči, kao i delove reči koji nisu neposredno obuhvaćeni analizom, ali koji na posredan način predstavljaju relativne pokazatelje artikulacionih odstupanja, u bloku **606** se računaju obeležja koja najbolje opisuju pojedine tipove artikulacionih odstupanja ili su u nekoj korelaciji sa njima. Na osnovu izračunatih pokazatelja artikulacionih odstupanja, u bloku **607** se formira multimodalna funkcija odluke koja ima onoliko izlaza koliko postoji poznatih oblika artikulacionih odstupanja za pojedine foneme koji su predmet analize. Izlazi iz multimodalne funkcije odluke prolaze kroz nelinearne funkcije kojima se realizuje meka odluka o prisustvu određenog artikulacionog odstupanja.

Na slici 8 prikazana je organizacija grafičkog korisničkog interfejsa **GKI**, blok **900**. On sadrži četiri osnovne celine koje omogućavaju optimalnu komunikaciju čovek-sistem i maksimalno korišćenje mogućnosti sistema, to su: analiza ulaznih stimulusa (blok **901**), izrada izveštaja (blok **902**), održavanje baze snimaka (blok **903**) i trening i obuka kako elemenata sistema tako i korisnika sistema (blok **904**).

Izbor i prezentacija stimulusa, blok **905**, je procedura u kojoj se iz baze podataka bira ispitanik ili ako on ne postoji unose se podaci o novom ispitaniku (koji se smeštju u bazu podataka) i formira se radno okruženje za snimanje i procenu kvaliteta izgovora u kome je omogućen izbor jedne od 30 stimulus reči, njena prezentacija ispitaniku (maksimalno do 3 puta) snimanje izgovora ispitanika (maksimalno tri izgovora), itd.

Snimanje, blok **906**, je procedura u kojoj je moguće preslušati snimljeni izgovor ispitanika i odabrati onaj koji je najbolji reprezent njegovog izgovora (ovaj snimak se trajno čuva u bazi podataka).

Predobrada govornog signala, blok **907**, je procedura u kojoj se normalizuje snimljeni materijal (intenzitet, granice itd) automatski segmentira i formira se set podataka (uzrast, pol, granice segmenata, digitalni zapis snimka).

Obrada, blok **908**, je procedura u kojoj se paralelno ocenjuje kvalitet artikulacije na globalnom i analitičkom nivou. **GKI** omogućava da se formira okruženje za optimalan rad sistema prema okolnostima u kojima se koristi. Na primer: parametri neuronske mreže, odnosno predznanje koje koristi, zavise od foneme, pola i starosti ispitanika. Sa druge strane, odstupanja u izgovoru su arikulacione i akustičke prirode i postoji određena korelacija između pojedinih obeležja u ova dva domena. Takođe pojedina artikulaciona odstupanja je moguće uočiti samo vizuelno jer govornik stvara kompenzatorne mehanizme tako da izgovor iako artikulaciono neadekvatan auditivno zvuči korektno. Deo problema koji je rešavan u okviru sistema je iznalaženje veze između artikulacionih obeležja i njihove akustičke manifestacije, kao i definisanje algoritama za ekstrakciju akustičkih obeležja iz govornog signala. Izborom određenih parametara koje **GKI** ponudi korisniku sistema formira se radni režim sistema.

Smeštanje rezultata i snimaka u bazu podataka, blok **909**, je procedura koja omogućava trajno čuvanje dobijenih rezultata globalne i analitičke ocene izgovora svake foneme, podataka o datumu snimanja, dodatnim komentarima o toku snimanja i pacijentu i snimaka izgovora u bazu podataka.

Izveštaji, blok **902**, predstavljaju deo programa i **GKI** u kome se na osnovu globalne i analitičkih ocena za svaku fonemu i ekspertskog znanja o pravilnim i patološkim tokovima razvoja govora i jezika vezanim za uzrasne norme donose zaključci o kvalitetu izgovora ispitanika. Na primer, za ispitanika starosti dve godine koji ima odstupanja u izgovoru 15 izgovornih glasova (ako pripadaju glasovnim grupama za koje je dozvoljeno odstupanje u tom uzrastu) izveštaj će pored klasifikacije globalnih i analitičkih ocena ukazati da je kvalitet izgovora u skladu sa normalnim tokovima razvoja govora i jezika. Međutim ako se kod ispitanik starosti dve godine pojave odstupanja u izgovoru onih fonema koje bi morale biti usvojene i pravilno artikulisane za taj uzrast tada se kao rezultat dobija da izgovor odstupa od normalnih tokova razvoja govora. Da bi se dobio izveštaj potrebno je iz baze podataka odabrati ispitanika, snimanje za koje se želi izveštaj (pošto za jednog ispitanika može biti više snimaka) i slično (blok **910**), vrstu izveštaja (globalni, analitički ili zbirni) (blok **911**) i na kraju izraditi izveštaj (blok **912**).

Održavanje baze, blok **903**, je deo programa u kome se nalaze rutine za održavanje baze kao što su: sažimanje, brisanje podataka, odlaganje podataka, itd., dok **GKI** putem menija omogućava korisniku upravljanje ovim delom programa.

Trening i obuka, blok **904**, je deo programa u kome se po određenim kriterijuma biraju i reprodukuju snimci izgovora pojedinih fonema, koji imaju određeni tip artikulacionog odstupanja, ili se mogu preslušavati izgovori jednog ispitanika kako bi se uočile razlike u kvalitetu izgovora u određenom vremenskom periodu i slično. Posebne procedure su predviđene za trening delova sistema, neuronske mreže i modula za segmentaciju, sa novom i/ili proširenom bazom pacijenata u cilju optimizacije sistema ili njegove adaptacije na novu klasu pacijenata.

U ovom pronalasku opisan je postupak obrade govornog signala u sistemu za objektivnu procenu kvaliteta artikulacije. Postupci i tehnike obrade govornog signala u ovom pronalasku ne mogu se generalno primeniti na bilo koji jezik. Naime, globalni artikulacioni test pripada klasi prepoznavanja oblika (poređenje pravilno i nepravilno izgovorenog stimulusa) i on se generalno može primeniti na stimulusima bilo kog jezika. Međutim, analitički artikulacioni test se neposredno odnosi na analizu fonema u okvirima fonetske i artikulacione baze datog jezika i zbog toga je ovo rešenje primenljivo samo za srpski jezik.

Postupci i tehnike obrade govornog signala u ovom pronalasku se nalaze pod kontrolom većeg broja parametara koji omogućavaju optimalno funkcionisanje sistema u celini. Konceptija rešenja je fleksibilna u smislu ažuriranja baze snimaka stimulusa i njenog proširivanja, što omogućava dodatnu, naknadnu, obuku sistema na proširenoj bazi atipičnih stimulusa i uvećanja pouzdanosti funkcionisanja sistema.

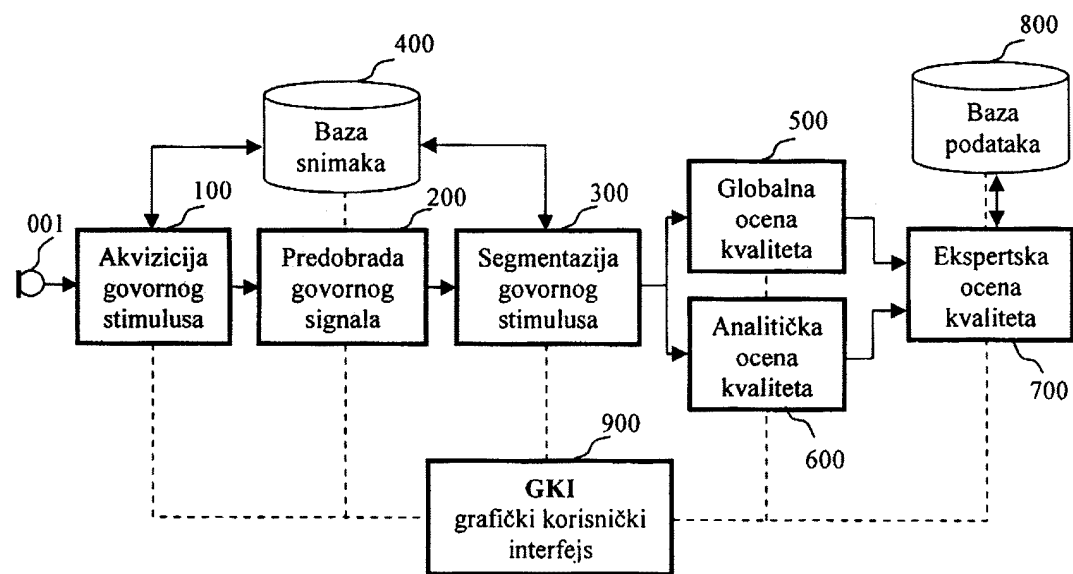
Detalji ovog pronalaska opisani ovde omogućavaju bilo kom stručnjaku u ovoj oblasti da generičke principe ovog pronalaska može implementirati u drugim sistemima za druge jezike.

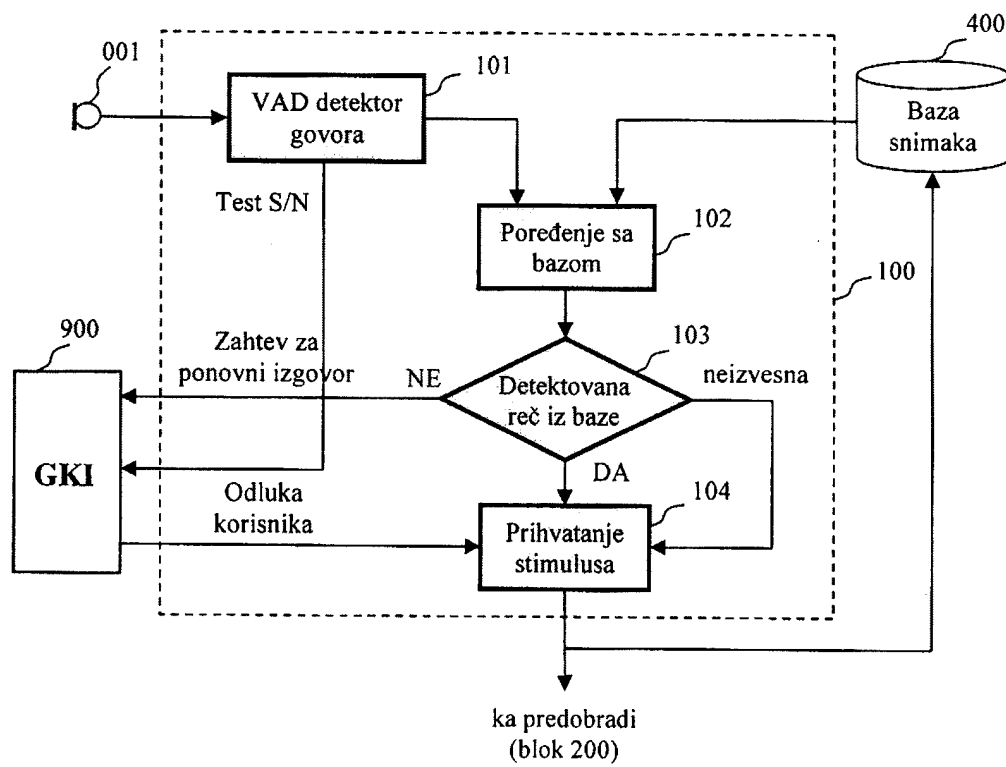
PATENTNI ZAHTEVI

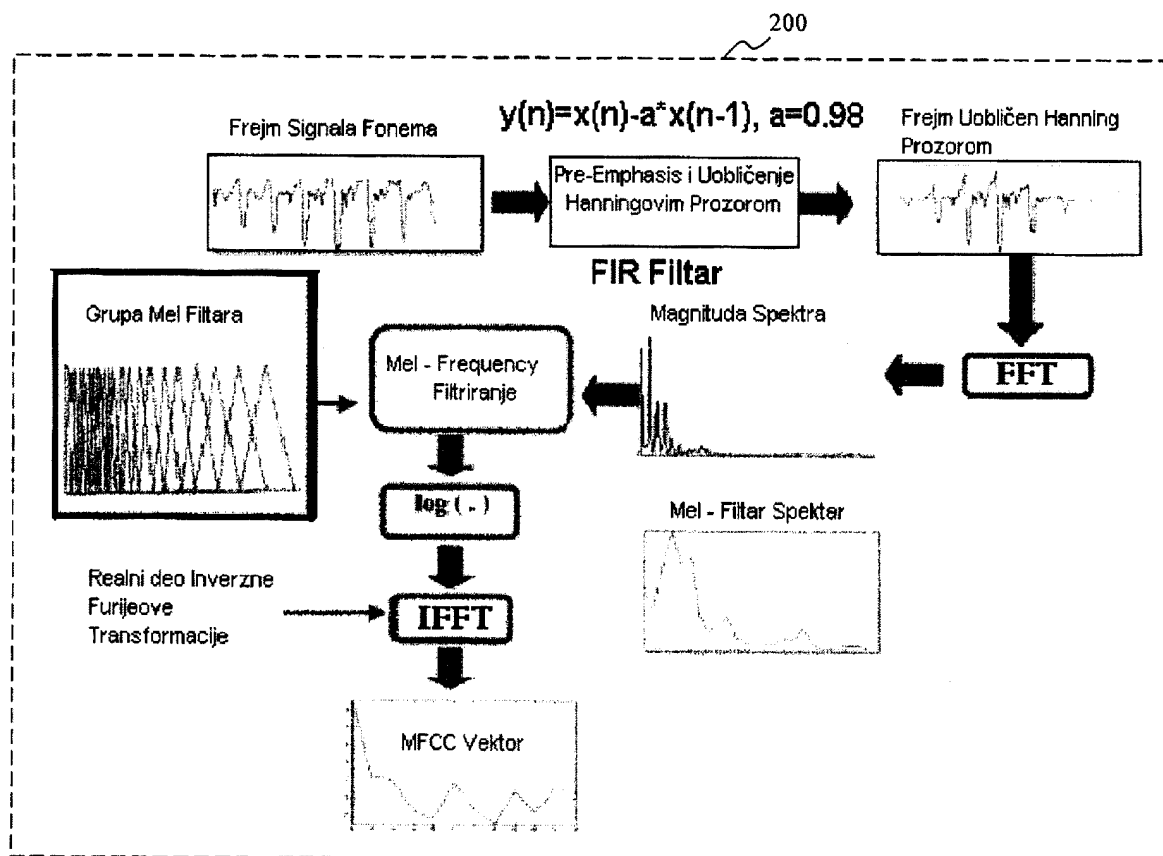
1. Sistem za objektivnu procenu kvaliteta artikulacije **karakterisan time** što sadrži mikrofona za snimanje govornog stimulusa; što sadrži računar sa odgovarajućim softverom koji vrši akviziciju govornog signala, zatim njegovu obradu u cilju prepoznavanja i identifikovanja kvaliteta artikulacije, i koji putem grafičkog korisničkog interfejsa u potpunosti komunicira sa korisnikom sistema; što sadrži procedure za globalnu i analitičku procenu kvaliteta artikulacije; što sadrži bazu snimaka govornih stimulusa i što sadrži bazu pacijenata.
2. Sistem prema zahtevu 1 **karakterisan time** što se akvizicionim modulom prihvataju korektno snimljeni govorni stimulusi i što se arhiviraju u bazu stimulusa koja se višenamenski koristi za dijahronu analizu rehabilitacije pacijenta ili za trening o obuku delova sistema u cilju poboljšanja kvaliteta funkcionisanja sistema.
3. Sistem prema zahtevu 1 **karakterisan time** što vrši globalnu procenu kvaliteta artikulacije pomoću globalnog artikulacionog testa definisanog na bazi ekspertske logopedске procene odstupanja u artikulaciji na nivou fonema standardne artikulacione baze srpskog jezika.
4. Sistem prema zahtevu 1 **karakterisan time** što vrši analitičku procenu kvaliteta artikulacije pomoću analitičkog artikulacionog testa definisanog na bazi analize akustičkih obeležja u izgovoru svake foneme srpskog jezika.
5. Sistem prema zahtevu 1 **karakterisan time** što sadrži bazu pacijenata koja sadrži sve podatke o pacijentu, koja omogućava izradu izveštaja određenih formi i koja omogućava dijahrono praćenje procesa rehabilitacije artikulacionih poremećaja.
6. Sistem prema bilo kom od prethodnih zahteva, **naznačen time**, što putem grafičkog korisničkog interfejsa omogućava potpunu komunikaciju korisnik-sistem, što omogućava optimalnu kontrolu analitičkog procesa i što omogućava fleksibilnu vizuelnu i operativnu interpretaciju rezultata analize.
7. Postupak za objektivnu procenu kvaliteta artikulacije, **karakterisan time**, što obradu govornog signala vrši u vremenskom, intenzitetskom, frekvencijskom i parametarskom domenu, što obradu govornog signala fokusira na foneme srpskog jezika, što vrši segmentaciju analizirane foneme u izgovorenom stimulusu, što vrši globalnu procenu kvaliteta artikulacije pomoću neuronske

mreže, i što analitičku procenu kvaliteta artikulacije vrši na bazi algoritama estimacije akustičkih obeležja karakterističnih za pojedino odstupanje u artikulaciji analizirane foneme.

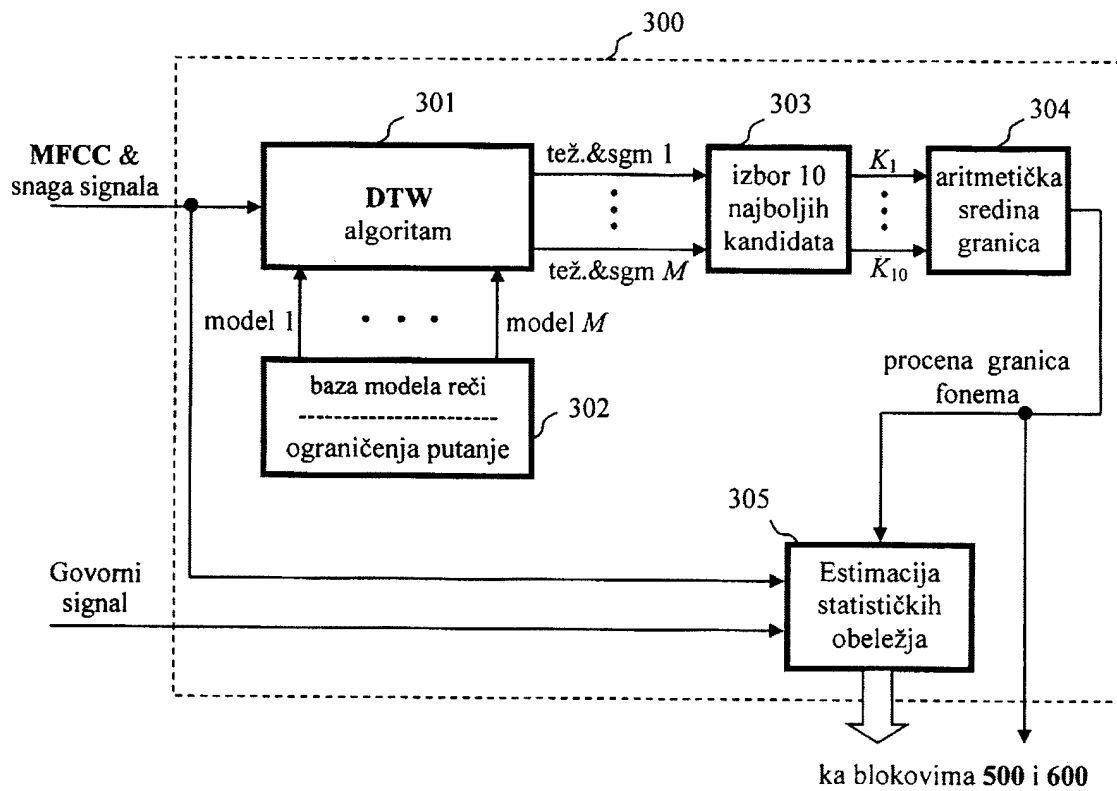
8. Postupak prema zahtevu 7, **karakterisan time**, što se segmentacija foneme u okviru pojedinog stimulusa vrši na bazi prepoznavanja segmenata pomoću DTW algoritma i baze segmentiranih stimulusa sa tipičnom i atipičnom artikulacijom.
9. Postupak prema zahtevu 7, **karakterisan time**, što se globalna procena kvaliteta artikulacije bazira na neuronskoj mreži tipa višeslojnog perceptrona, koja klasifikaciju korektna/nekorektna/granična artikulacija vrši na ulaznom vektoru od 19 parametarskih i akustičkih obeležja detektovanih u segmentiranoj fonemi.
10. Postupak prema zahtevu 7, **karakterisan time**, što se analitička procena kvaliteta artikulacije zasniva na skupu akustičkih obeležja detektovanih u segmentiranoj fonemi a karakterističnih za analizirano odstupanje i što je za svaku fonemu i svako njeno odstupanje definisan algoritam za prepoznavanje odstupanja.

**Slika 1.**

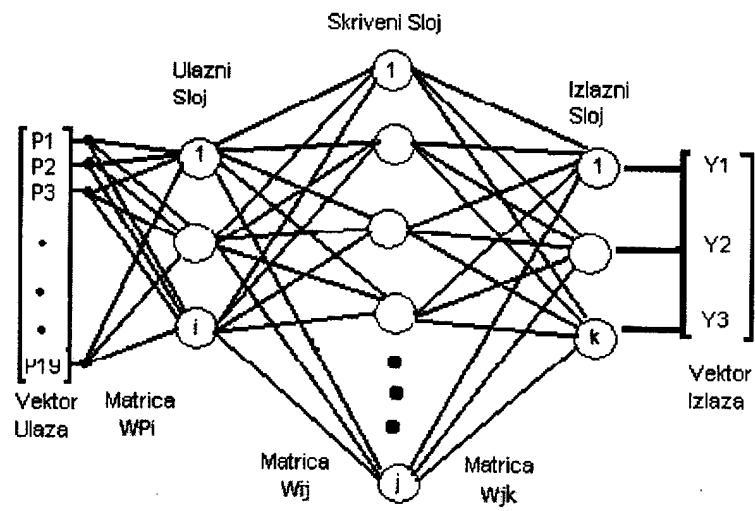
**Slika 2.**



Slika 3.

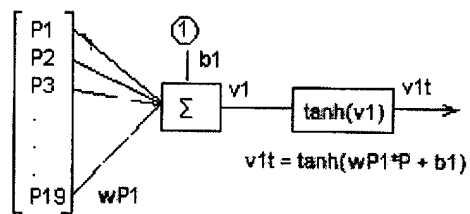


Slika 4.

Arhitektura primenjene
neuronske mreže

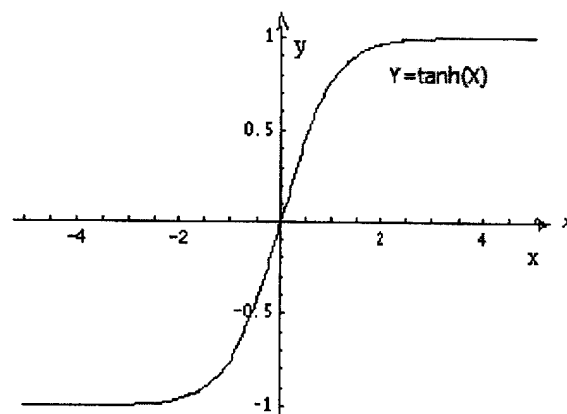
a)

Blok šema neurona



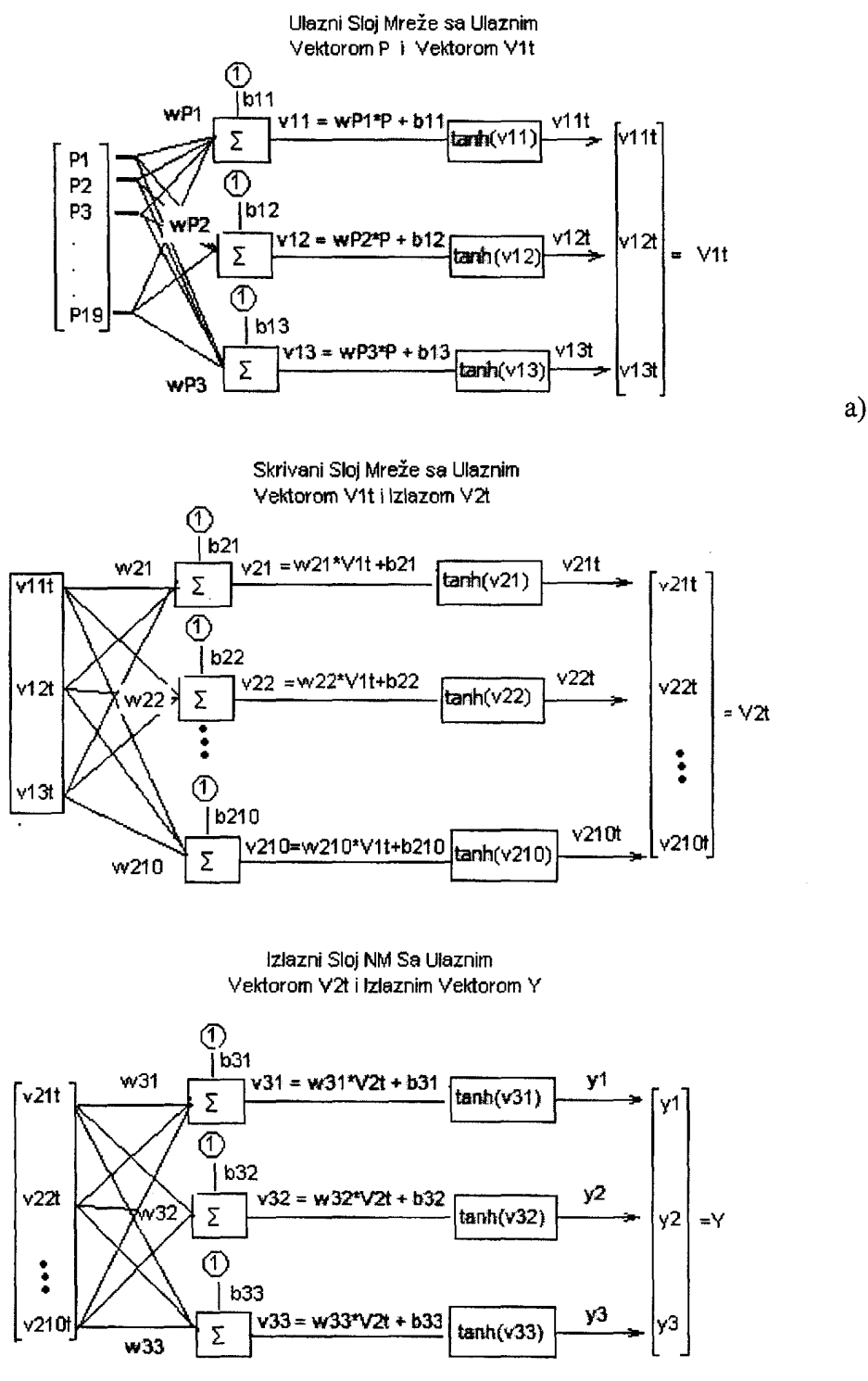
b)

Prenosna funkcija neurona

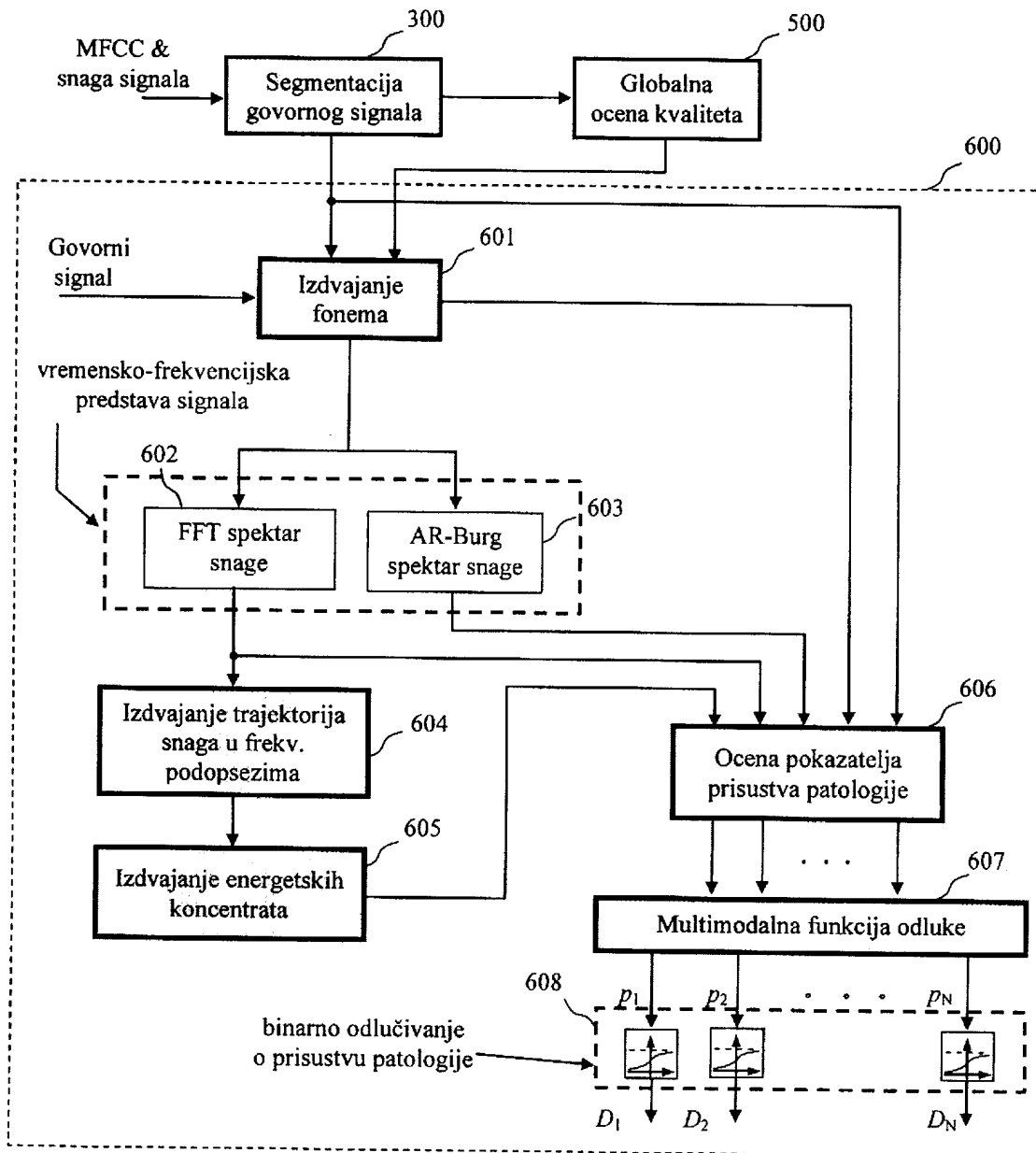


c)

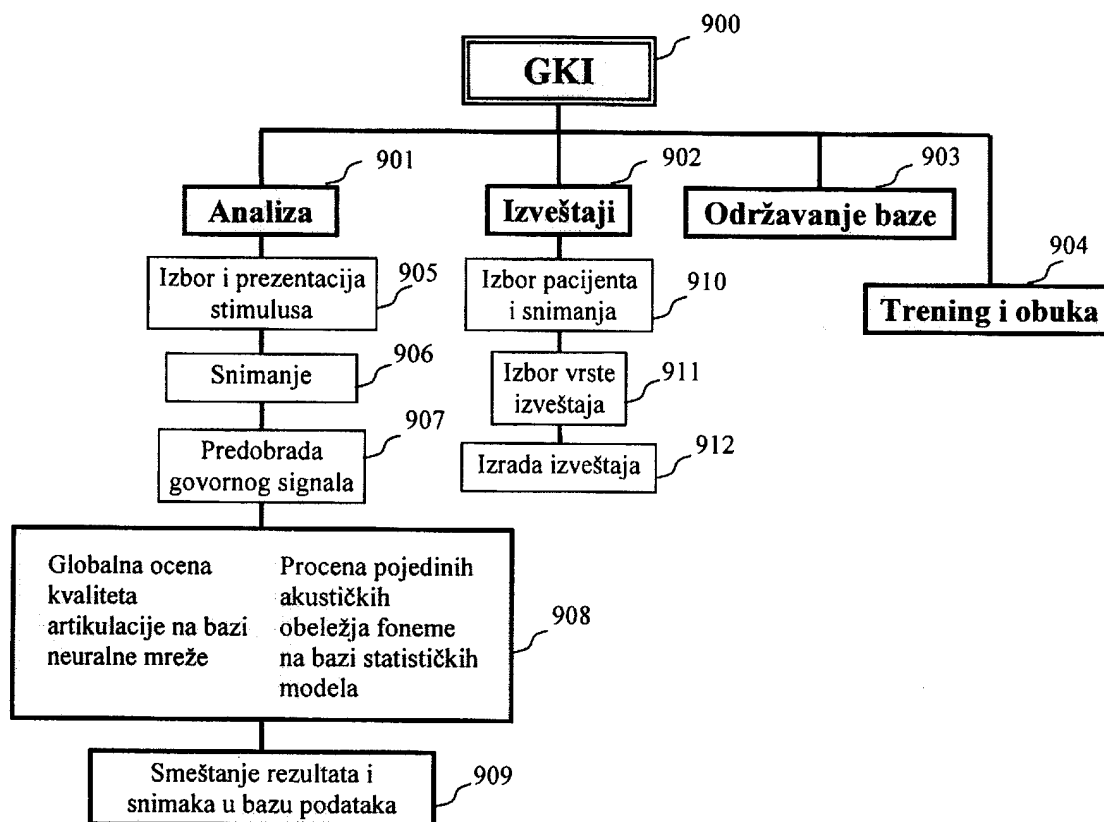
Slika 5.



Slika 6.



Slika 7.



Slika 8.