SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1338

Automatski sustav za poboljšavanje izgovora

Sandra Selinger

Zagreb, veljača 2017.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA
ODBOR ZA DIPLOMSKI RAD PROFILA

Zagreb, 7. listopada 2016.

DIPLOMSKI ZADATAK br. 1338

Pristupnik: **Sandra Selinger (0036463083)**
Studij: Računarstvo
Profil: Računarska znanost

Zadatak: **Automatski sustav za poboljšanje izgovora**

Opis zadatka:

Sustavi za automatsko prepoznavanje govora ili za identifikaciju govornika vrlo su rašireni. Dobro poznata programska rješenja koriste blokovsku predobradu govornih signala te ekstrakciju vektora značajki (MFCC - Mel frekvencijski kepstar). Izgrađene su akustičke baze u kojima su grozdovi vektora modelirani kao gausovske mješavine. Izgovoreni trifoni (slijedovi tri glasa - fonema) odgovaraju putanjama kroz vektorski prostor, a za prepoznavanje najvjerojatnije putanje (izgovorenog trifona) koriste se skriveni Markovljevi modeli.
U okviru diplomskog zadatka, potrebno je razviti automatizirani sustav za poboljšanje izgovora riječi. Potrebno je redefinirati problem i prilagoditi poznate algoritme za prepoznavanje govora sa svrhom ocjene i korekcije izgovora. Prikupit će se govorna baza autistične djece s raznim govornim poteškoćama, koja kroz igru nastoje savladati osnovne riječi. Definirat će se kriteriji ocjene izgovorene riječi, uzevši u obzir česte pogreške (izostanak ili zamjena redoslijeda fonema). Kreirat će se jednostavna računalna igra u kojoj će dijete biti vođeno na način koji stimulira ispravan izgovor.
U svezi dobivanja detaljnih informacija obratiti se mentoru.

Zadatak uručen pristupniku: 14. listopada 2016.
Rok za predaju rada: 3. veljače 2017.

Mentor:

Prof. dr. sc. Damir Seršić

Djelovođa:

Doc. dr. sc. Tomislav Hrkać

Predsjednik odbora za diplomski rad profila:

Prof. dr. sc. Siniša Srbljić

Hvala mag. rehab. educ. Ivani Jandroković bez čijeg žara za istraživanjem autizma, stručnosti i poklonjenog vremena ovaj rad ne bi bio napravljen.

Veliko hvala mentoru dr. sc. Damiru Seršiću za cjelokupno vodstvo kroz diplomski studij, entuzijazam oko ovog projekta, svo uloženo vrijeme i konstantnu podršku.

Zahvaljujem i dragoj učiteljici Davorini Bakota za sve divne izviđačke uspomene, a i administrativno-logističku pomoć prilikom prikupljanja snimki.

Hvala OŠ Antuna Mihanovića za susretljivost, roditeljima za pristanak da se snimi razgovor s njihovim djetetom, a i djeci za suradnju i vrelo pozitivne energije koju šire.

Hvala svim prijateljima, kolegama i profesorima koji su mi dijeljenjem znanja i otvaranjem vidika studiranje učinili tako lijepim periodom života.

Posebno zahvaljujem mami Tanji, tati Darku i sestri Andrei za bezuvjetnu ljubav i podršku kroz cijeli život. Bez vas ne bih uspjela izrasti u osobu koja danas jesam.

Konačno, hvala i Franji. Hvala ti što me činiš sretnom svaki dan.

SADRŽAJ

1. Uvod

Ovaj diplomski rad bavi se tehničkom izvedbom sustava za raspoznavanje i poticanje govora kod djece s poremećajem iz autističnog spektra (PAS). Napravljen je akustički model hrvatskog jezika za djecu s PAS i računalna igra *Brbljalica* koja ih kroz igru nastoji potaknuti na govor. Za izgradnju akustičkog modela odabrane su specifične riječi bitne za komunikaciju koje djeca usvajaju u ranom stadiju razvitka govora. Nakon toga je snimljena akustička baza dječjeg izgovora tih riječi nad kojom se izgradio akustički model jezika. Model je iteracijski Konačno, taj model se koristi u *Brbljalici* koja prema bihevioralnim principima učenja navodi djecu s PAS na govor.

Ideja je nastala u razgovoru s prijateljicom magistrom edukacijske rehabilitacije na području autizma koja je trenutno na doktoratu neuroznanosti Medicinskog fakulteta u Zagrebu. Pri upitu postoji li neka tehnička stvar koja bi joj olakšala rehabilitaciju djece s autizmom, a koja trenutno ne postoji, odgovor je bio odlučno potvrdan. Rekla je kako bi bilo idealno kada bi postojala računalna igra koja bi poticala dijete na govor i svaku točno upotrijebljenu riječ nagradila pozitivnim potkrjepljenjem prema bihevioralnim principima učenja. Računalne igre i aplikacije imaju veliki potencijal za poticanje razvoja komunikacije kod djece s poremećajem iz autističnog spektra ako su stručno napravljene jer ta djeca pokazuju velik interes za svim elektroničkim napravama (računalima, tabletima, telefonima), dok ih međuljudska socijalna interakcija uopće ne zanima (to je jedna od glavnih odlika autizma). Zato računalni programi razvijeni u suradnji sa stručnim osobama imaju potencijala pomoći djeci s PAS da bolje razviju svoje komunikacijske vještine.

Do sada su već i u Hrvatskoj napravljeni projekti koji mogu djeci s poremećajem iz autističnog spektra olakšati komunikaciju. Jedan od njih je projekt *ADORE* Šimleša et al. (2016) u sklopu kojeg humanoidni robot može pomagati pri dijagnozi i rehabilitaciji djece prateći njihovo ponašanje. Drugi primjer je projekt *Komunikator* Klopotan et al. (2014) koji može pomagati u razvoju komunikacije bazirane na vizualnom uparivanju. Još jedna aplikacija za vizualno uparivanje je *AuThink*. Postoje i strane aplikacije koje olakšavaju komunikaciju i izražavanje djetetovih želja preko sli-

čica. Ti sustavi mogu pomoći u dijagnozi, rehabilitaciji i komunikaciji, no niti jedna od njih ne potiče govor izravno. Govor kao viša kognitivna funkcija jedna je najtežih, ali i ključnih funkcija za usvajanje kod sve djece, a naročito kod one s poremećajima iz autističnog spektra. Kao odgovor na potrebu za aplikacijom koja se bavi govorom kod djece s PAS proizašla je web računalna igra *Brbljalica* čiji se razvoj prati kroz ovaj diplomski rad.

U poglavlju 2 se opisuje autizam kao poremećaj iz autističnog spektra, njegove značajke i karakteristike bitne za fokus ovog diplomskog rada. Poglavlje 3 se bavi akustičkim modelom. Prvo je opisana matematička podloga prepoznavanja govora, a kasnije prati proces izgradnje konačnog akustičkog modela. Izgradnja modela započinje opisom automatizacije razvijene u sklopu rada Dropuljić (2008) koja se nakon toga koristi na obrađenim podacima prikupljene akustičke baze. U poglavlju broj 4 se prikazuju i analiziraju rezultati različitih načina izrade modela za raspoznavanja govora. Poglavlje 5 sadrži tehničke i funkcijske detalje razvijene računalne igre.

2. Autizam

Širok zbir tema o poremećajima iz autističnog spektra na hrvatskom jeziku i relevantne reference na originalnu literaturu mogu se pronaći u knjizi i sveučilišnom udžbeniku Bujas-Petković et al. (2010). U daljnjem tekstu će se na bazi knjige razraditi samo općenite karakteristike autizma i dva područja bitna za razumijevanje ovog rada - razvijanje govora kod djece s PAS i biheviоralni tretmani.

2.1. Općeniti podaci o autizmu

2.1.1. Klasifikacija i definicija

Autizam (kao i poremećaj iz autističnog spektra) je razvojni poremećaj koji se javlja u ranom djetinjstvu (do treće godine) i traje do kraja života. Prema trenutno važećoj američkoj klasifikaciji i priručniku DSM-V (engl. *Diagnostical and Statistical Manual of Mental Disorders*) Američka Psihijatrijska Udruga (2014) simptomi se dijele u dvije kategorije. Prvu kategoriju čine odstupanja u području socijalne komunikacije: manjak komunikacije, socio-emocionalne recipročnosti i razumijevanja te održavanja odnosa. Druga kategorija opisuje ponašanja i interese: stereotipna, repetitivna i rutinska ponašanja, veoma usko područje interesa u kojem je prisutna abnormalno visoka razina koncentracije te senzorna hipoosjetljivost ili hiperosjetljivost.

Dakle u osnovi autizam je poremećaj komunikacije i ponašanja. U komunikacijom području autistična djeca ne ostvaruju kontakt pogledom, u pravilu se ne služe govorom, ne započinju komunikaciju, ne zanima ih sklapanje prijateljstava i imaju teškoće u prepoznavanju tuđih i vlastitih emocionalnih stanja. U samoj definiciji govor nije eksplicitno naveden jer neka djeca razviju određenu razinu govora, dok druga djeca uopće ne govore, pa se ne može točno reći kakav govor mora biti prisutan da bi se dijagnosticirao poremećaj iz autističnog spektra. No ono što se može reći je da su govorne poteškoće najtipičnije i najučestalije za poremećaj te postoje istraživanja koje ukazuju na to da djeca koja ranije razviju govor kasnije imaju bolje rezultate. Upravo

na tome se temelji smisao ovog rada, a više o autizmu i govoru će se reći u poglavlju 2.2.

2.1.2. Prevalencija i razlika među spolovima

Učestalost broja djece s poremećajem iz autističnog spektra je dosta teško precizno odrediti jer se i sami kriteriji za dijagnozu često mijenjaju. Jedno od istraživanja s najvećim uzorkom je Croen et al. (2002) u kojem se temeljem prebrojavanja u populaciji od 4,5 milijuna djece došlo do prevalencije od 11 slučajeva autizma na 10 000 djece.

Zanimljivo je da većinu dijagnosticirane djece čine dječaci. Jedna od najvećih studija Chakrabarti i Fombonne (2001) pokazuje da je udio dječaka među dijagnosticiranom djecom iz spektra čak 79%. Druge velike studije također daju slične brojke. Ne zna se točan razlog zašto prevladavaju dječaci, no postoje različita objašnjenja. Jedno popularno objašnjenje je objavljeno u istraživanju Head et al. (2014) gdje autori smatraju da je moguće da moguće da djevojčice općenito imaju bolje razvijenu komunikaciju u toj dobi, a da su dijagnostički alati napravljeni po uzoru na komunikaciju koju razvijaju dječaci, pa se poremećaj kod djevojčica lako može previdjeti. No bez obzira na pravi razlog, zbog značajnog prevladavanja broja dječaka u ovom radu je akustička baza napravljena isključivo na snimkama dječaka.

2.1.3. Dugoročne prognoze

Potkraj prošlog stoljeća se nekolicina znanstvenika počela baviti prognozama životnog vijeka i kvalitete života osoba s PAS, no trenutno još nije prošlo dovoljno vremena da bi se dobili potpuno vjerodostojni rezultati. Najpoznatije istraživanje Lotter (1978) koje prikupljanjem podataka 25 istraživanja koja ukupno prate preko 1000 djece s autizmom zaključuje da $62\% - 74\%$ djece ima lošu prognozu te su ovisni o tuđoj pomoći, dok samo $5\% - 17\%$ djece dobiva dobru prognozu što znači da žive normalno ili gotovo normalno (pohađaju školu, rade itd.).

U kasnijem istraživanju Gillberg (1991) autor zaključuje da su govor i kvocijent inteligencije (IQ) dobre značajke za davanje prognoze. Rezultati pokazuju da djeca čiji je IQ iznad 70 i koja s pet godina imaju imalo razvijen govor te već u tim godinama pokazuju napredak imaju najbolju prognozu. Ta djeca pokazuju najbolje rezultate bez obzira o terapiji koju dobivaju. Ipak, većina djece (nekih 60%) je i nakon odrastanja ovisna o tuđoj pomoći. Najslabije prognoze dobivaju petogodišnjaci s IQ kvocijentom ispod 50 i nerazvijenim govorom.

Iz ovih podataka se može vidjeti da je razvitak govora prije pete godine uz IQ jedan od najvećih pokazatelja kasnijeg samostalnog i normalnog života. Zato je roditeljima i terapeutima od velike važnosti što raniji razvitak govora i upravo se zato ovaj rad bavi poticanjem govora kod djece s poremećajem iz autističnog spektra.

2.2. Autizam i govor

U ovom odjeljku se pažnja posvećuje isključivo govoru — njegovoj definiciji, normalnom razvitku govora te karakteristikama govora kod djece s poremećajem iz autističnog spektra. Na kraju odjeljka se razjašnjava pitanje izgovora i govornih mana kod autistične djece koje se spominju u diplomskom zadatku te način redefiniranja problema nakon što je područje podrobnije proučeno i nakon što su prikupljene snimke djece s PAS.

2.2.1. Definicija i razvoj govora

Govor i jezik su sastavni dijelovi složenog procesa međuljudske komunikacije, a pravilno korištenje govora uključuje efikasnu interpersonalnu i intrapersonalnu komunikaciju. S obzirom na to da govor spada među više kognitivne sposobnosti, poteškoće kod savladavanja govora i jezika se očituju u mnogim poremećajima uključujući PAS i djecu sa sniženim intelektualnim sposobnostima. Problema kod komunikacije mogu imati i gluha djeca, no za razliku od gluhe djece i djece sa sniženim intelektualnim sposobnostima, jedino djeca s PAS ne mogu pročitati osnovne emocije drugih osoba (je li osoba sretna, tužna, ljuta i sl.) niti ne razviju neki drugačiji način komunikacije gestama ili mimikama.

Da bi komunikacija bila uspješna bitno je da svi dijelovi komunikacijskog lanca budu uspješni. Prvi dio komunikacije podrazumijeva receptivni govor, a iza njega dolazi ekspresivni govor. Receptivni govor uključuje slušanje i prepoznavanje riječi. Ekspresivni govor je artikulacija i vokalizacija odgovora na situaciju. Ako je jedna karika tog lanca poremećena, tada je i cijela komunikacija poremećena. Većina autistične djece ima velike poteškoće u receptivnom govoru, a ekspresivni govor se često uopće ne razvije.

Dijete normalnog razvoja s okolinom komunicira već s nekoliko mjeseci smiješkom i podizanjem ruku, a kasnije počinje pokazivati prema predmetima, počinje govoriti potkraj prve godine te s pet godina već slaže gramatički točne rečenice i razumije apstraktne pojmove. Dijete s PAS ima poteškoće u cjelokupnoj komunikaciji još od

najranije dobi.

2.2.2. Karakteristike govora autistične djece

U upitniku napravljenom 1980. godine za utvrđivanje autizma kod djece od dvije do četiri godine postoje četiri kategorije, od kojih jedna odgovara procjeni jezičnih problema. Njeni dijelovi su:

- teškoće u razumijevanju govora,
- usporen i abnormalan govorni razvoj,
- teškoće u uporabi i razumijevanju gesta, mimike, facijalne ekspresije i posture tijela.

Simptomi tih jezičnih problema navedeni su u Bujas-Petković et al. (2010) te uključuju potpun izostanak razvoja govora, dijete se ne odaziva na svoje ime, ne započinje razgovor s drugima, koristi čudan ritam i intonaciju, ponavlja riječi, krivo koristi zamjenice "ti" i "ja" jer ponavlja tuđe fraze i pogrešno upotrebljava fraze (recimo kada želi čokoladu dijete može reći "Hoćeš čokolade?"). No u pravilu djeca ne koriste govor te često niti odrasle osobe s PAS koje imaju razvijen govor ga ne koriste za komunikaciju.

Razina govora djeteta se može podijeliti u četiri stupnja:

1. Ne govori,
2. Eholaličan govor (ponavljanje riječi),
3. Nepravilno koristi zamjenice i stvara bizarne konstrukcije,
4. Gramatički ispravan govor, postavlja pitanja.

Na cijelom autističnom spektru poremećaja zadnji se (četvrti) stupanj javlja samo kod visoko-funkcionalnog autizma i osoba s Aspergerovim sindromom, dok se kod klasičnog autizma uglavnom radi o prva tri stupnja.

Cilj ovog rada je potaknuti dijete s PAS na govor i vježbanje pravilne upotrebe jednostavnih riječi i fraza, pa se razvijenom igrom mogu služiti djeca na drugom i trećem stupnju razvoja, dok je kod djece koja su potpuno neverbalna teško postići bilo kakav govor, pa se tu trenutno može pričati samo o poticanju na govor zbog njihovog interesa prema tehnici.

2.2.3. Izgovor i redefinicija problema

Osobe s PAS tijekom govora (eholaličnog ili spontanog) često griješe u ritmu, intonaciji, visini, naglasku i sadržaju. Njihov govor je lišen emocija i prepoznatljiv, različit je od govora djece sa smanjenim intelektualnim sposobnostima i djece urednog razvoja.

Greške u sadržaju se često svode na gramatički neispravne konstrukcije ili poteškoće pri upotrebi zamjenica i priloga. Izvorna ideja prilagodbe sustava djeci s PAS je bila modelirati njihov krivi izgovor tretirajući ih kao govorne mane. Djeca s PAS mogu imati i neke od govornih mana koje se pojavljuju i kod djece urednog razvoja, no one nisu karakteristične za poremećaj. Postoji desetak vrsta standardiziranih govornih mana. Njihovi nazivi dolaze od grčkih naziva glasova koje predstavljaju te su najčešće govorne mane sljedeće:

- Rotacizam - glas R,
- Sigmatizam - grupe glasova (C, Z, S), (Č, Ž, Š), (Ć, Đ, DŽ),
- Lambdacizam - glasovi L i LJ.

Svaka od tih vrsta se dalje dijeli na tri načina izvedbe:

1. Izostanak glasa,
2. Supstitucija drugim glasovima (primjerice u rotacizmu najčešće glasovima J i L),
3. Distorzija glasa.

Međutim, nakon prikupljanja snimki govora djece s autizmom vidjelo se da unutar seta zapravo uopće nema dovoljno primjera u kojima se provodi neki način standardne govorne mane osim rotacizma kod dijela djece i veoma mali broj primjera sigmatizma i lambdacizma. Velika većina krivo izgovorenih riječi je uključivala naizgled nasumične pogreške koje ne odgovaraju niti jednoj opisanoj govornoj mani koja se može predvidjeti i sustavno modelirati. Zato je odlučeno da će se problem redefinirati te se u okviru ovog diplomskog rada neće modelirati dio vezan uz modeliranje govornih mana koji uključuje izostanak glasova i zamjenu glasova niti graditi sustav koji bi te greške ispravljao.

Ono što je u zamjenu uključeno u sustav je definiranje razreda riječi gdje se sve riječi unutar jednog razreda smatraju jednako vrijednima. Kod imenica se radi o svim oblicima riječi u padežima, a kod glagola su to prezent, perfekt i ponekad još neki vrlo čest oblik. Osim toga se uvodi i pravilo da se prijedlozi mogu po volji izbaciti ili ubaciti u izgovor. Prikaz korištenih razreda riječi može se vidjeti na slici 3.12. U

svakom retku se nalazi po jedan razred riječi, dok se u zadnjem retku nalaze svi prijedlozi. Svrha ovih preinaka jest prepoznati i gramatički netočan, ali logički jasan govor kao prihvatljiv odgovor. S obzirom na to da osobe s PAS koje razviju spontani govor najčešće imaju problema s gramatikom, pretpostavka je da će oslabljivanje gramatičkih pravila povoljno utjecati na ocjenu modela raspoznavanja govora. Djeca s PAS u promatranoj dobi (pet do sedam godina) i kada govore ne govore puno, tj. govor se svodi na svega nekoliko riječi, pa i u ovom radu najdulja fraza za prepoznavanje ima svega tri riječi.

2.3. Bihevioralni tretmani

Bihevioralni tretmani proizašli su iz primjene analize ponašanja (engl. *Applied behavior analysis - ABA*) te se široko koriste u radu s osobama s PAS (Schreibman i Ingersoll (2005)). Bihevioralne metode intenzivno se razvijaju već gotovo pola stoljeća, kao i sama dijagnoza.

Kod tretmana gdje sudjeluju terapeut i dijete postoje strukturirane bihevioralne intervencije i bihevioralne intervencije u prirodnom kontekstu. Strukturirana bihevioralna intervencija je vrsta tretmana primijenjene analize ponašanja u kojoj je okolina strogo definirana. To znači da terapeut stvara visoko strukturiranu okolinu u kojoj se provodi podučavanje i rastavlja zadatke na manje dijelove koji se serijalno podučavaju. Terapeut također bira sredstva i inicira podučavanje. Za razliku od toga u zadnje vrijeme je razvijena i intervencija u prirodnom kontekstu (Ingersoll i Schreibman (2006), Prizant et al. (2003)) gdje je nema stroge okoline, a dijete samo bira sredstva i inicira podučavanje pokazujući interes za neki predmet ili radnju. Takvi tretmani su bliže stvarnom svijetu te se trenutno smatraju sveobuhvatnim terapijskim pristupom. S obzirom na to da se *Brbljalica* nalazi na računalu, tabletu ili mobitelu te je unaprijed vrlo određena programiranjem, ona je bliža strukturiranoj bihevioralnoj intervenciji. Naravno, cilj igre nije zamijeniti terapeuta niti biti kompletno rješenje problema, već samo nadomjestak svakodnevnom podučavanju i vježbi.

2.3.1. Primjena strukturirane bihevioralne intervencije

Strukturirana bihevioralna terapija se provodi u sklopu rane intervencije kod djece s PAS, a za njenu primjenu se koristi podučavanje diskriminativnim nalozima — PDN-ima (engl. *discrete trial instructions — DTI*).

Svrha PDN-a je podučavanje djeteta na vrlo jasan, sažet i strukturiran način. Uvi-

jek se koriste unaprijed određeni materijali, podučavanje se provodi na istom mjestu i dijetu se pruža podrška pri obavljanju zadatka. Postoje različite faze učenja zadatka PDN-om, počevši od faze gdje dijete treba usvojiti novi zadatak, a u krajnjoj fazi dijete može obaviti zadatak i razlikovati ga od drugih zadataka. Faze su sljedeće:

1. Izolacija: ponavljanje podražaja nekoliko puta uzastopce uz podršku kako bi dijete naučilo što treba napraviti (primjerice pokazivanje slike kao diskriminativni podražaj, a dijete na njega treba odgovoriti imenovanjem stvari sa slike),

2. Diskriminacija: kada dijete može u izolaciji obaviti zadatak bez podrške, tada se u okolinu ubacuju drugi predmeti nevezani uz zadatak te se ponavlja podražaj uz podršku kako bi dijete razlikovalo novi zadatak od onih koje je prethodno naučilo s ostalim predmetima,

3. Nasumična rotacija: kada dijete može diskriminirati i uspješno obaviti zadatak bez podrške tada se djetetu nasumično prezentiraju podražaji više predmeta u okolini te mu se daje podrška da u pravom trenutku obavi odgovarajući zadatak.

Pokazano je da je PDN učinkovit u podučavanju osoba s autizmom još u 60-im godinama prošlog stoljeća Baer et al. (1967), a uspjeh je potvrđen u još mnogo istraživanja nakon toga od kojih je jedno i Smith et al. (2000). Razvoj igre rabi pristup sličan PDN-u, a detaljnije je opisan u poglavlju 5.

3. Prepoznavanje govora

3.1. Teorijska podloga

Iz teorijskog dijela prepoznavanja govora izdvajamo:

- Mel-frekvencijski kepstar opisan u odjeljku 3.1.1 kao način računalne reprezentacije govornog signala,
- Skrivene Markovljeve modele (engl. *Hidden Markov Model — HMM*) zbog toga što su oni nositelji modeliranja sustava za prepoznavanje govora. Skriveni Markovljevi modeli i osnovni algoritmi nad njima su opisani u odjeljku 3.1.2,
- Gaussove mješavine kao način reprezentacije stanja HMM-a opisane su u odjeljku 3.1.3.

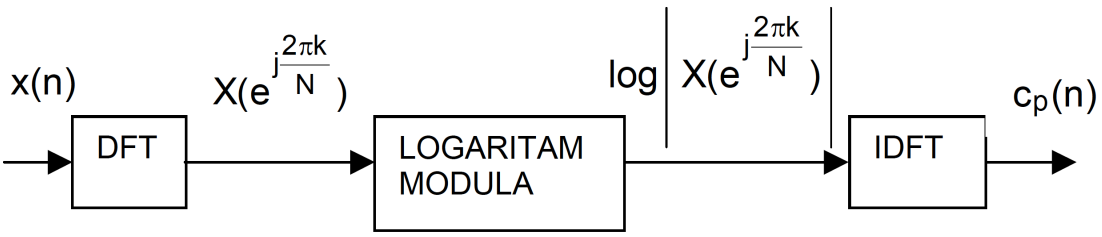
Ti koncepti su nužni za razumijevanje rada sustava za raspoznavanje govora, a dodatni elementi koji se grade prilikom razvoja sustava proizlaze iz njih.

3.1.1. Mel-frekvencijski kepstar

Mel-frekvencijski kepstar služi za pretvorbu snimljenog govornog signala u niz brojki (vektor) s kojima računalo može raditi. U snimljenom govornom signalu slušanjem možemo s velikom vjerojatnošću zaključiti mnogo toga kao na primjer govornikov spol, starost, osjećaje (je li govornik sretan, tužan, ljut itd.) i izgovorene riječi. Cilj je iz snimke odstraniti sve dijelove osim izgovorenih riječi. Ukupnu snimku možemo stoga promatrati kao kombinaciju signala i šuma, gdje šum čine sve ostale komponente osim onih odgovornih za prijenos informacije o izgovorenim riječima. Trenutno ne postoji postupak kojim bi se u potpunosti isključile sve ostale komponente, no Mel-frekvencijski kepstar je jedan od najboljih pokušaja.

Običan kepstar

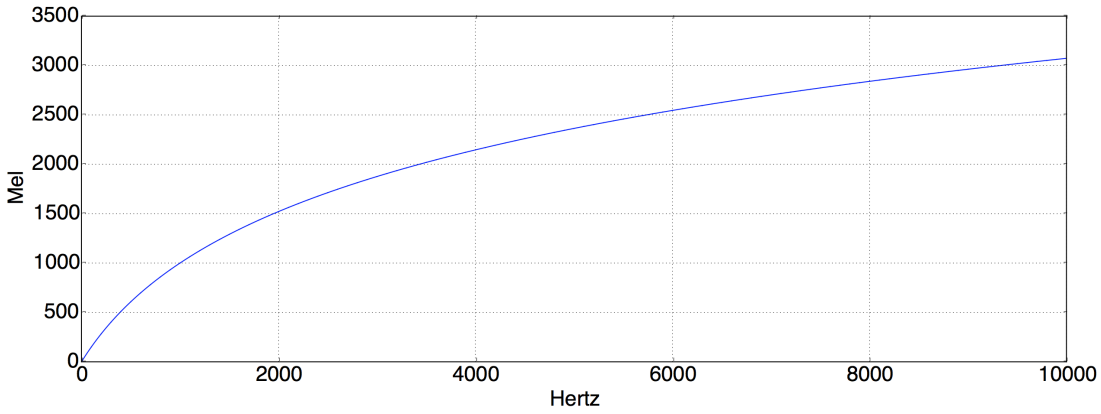
Običan kepstar dobiva se kao logaritam amplitudnog spektra signala nad kojim je obavljena diskretna Fourierova transformacija. To je zato što se signal dijeli na kratke dijelove u kojima smatramo da je sustav linearan, te u njemu znamo da je zbroj dvaju signala (u našem slučaju signala i šuma) na uzlazu linearno nepromjenjivog sustava jednak konvoluciji tih dvaju signala na izlazu. Dakle kako bi dvojili komponente signala moramo napraviti dekonvoluciju. Signal diskretnom Fourierovom transformacijom prevedemo u Z-domenu gdje konvolucija signala postaje umnožak. Kako bi se umnožak pretvorio u zbroj uzmemo logaritam modula. Sada imamo zbroj komponenti od kojih nas zanimaju samo neke. Zato na tom signalu radimo još jednu Fourierovu transformaciju (IDFT - inverznu diskretnu Fourierovu transformaciju) i biramo koeficijente za koje znamo da dobro opisuju ono što nam treba — prvih 13 koeficijenata isključujući nulti koeficijent jer za njega znamo da je proporcionalan energiji sustava. Ukupni prikaz sustava dan je na slici 3.1. Slika je preuzeta iz Petrinović (2010) gdje se može pronaći i detaljniji pregled informacija i intuicije iza računanja kepstra.



Slika 3.1: Proces dobivanja kepstra

Mel-frekvencijski kepstar

Upravo je objašnjen nastanak kepstra, no u prepoznavanju govora se gotovo uvijek koristi Mel-frekvencijski kepstar. Motivacija iza toga leži u činjenici da je ljudsko uho puno osjetljivije na podražaje nižih frekvencija nego viših. To znači da ljudsko uho bolje prepoznaje frekvencijske razlike između dva tona niže frekvencije nego dva tona iste frekvencijske razlike koja su puno većih apsolutnih frekvencija. Na primjer, ljudsko uho će percipirati razliku u visini tona između 100 Hz i 200 Hz kao mnogo veću nego razliku između frekvencija 10100 Hz i 10200 Hz. Melova funkcija radi mapiranje između skale u Melima i skale u Hertzima te je odabrana tako da će slušatelj svaku jednaku razliku na Melovoj skali uvijek subjektivno odrediti kao jednaku razliku u visini tona. Melova funkcija ima logaritamski oblik te je njena formula dana jednadžbom



Slika 3.2: Graf konverzije frekvencije u Hertzima i Melima

3.1, dok je njen graf dan na slici 3.2. Skaliranje frekvencija se događa odmah nakon prve diskretne Fourierove transformacije jer je tada sustav u frekvencijskoj domeni, pa se frekvencije dijele u preklapajuće pojaseve jednake na Melovoj ljestvici te se na tim pojasevima dalje radi logaritam modula i dalje se postupa kao i kod običnog kepstra.

$$m(f) = 2595 \cdot \log_{10}(1 + \frac{f}{700}) \tag{3.1}$$

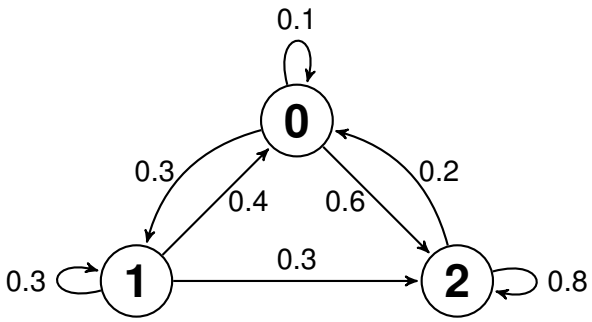
3.1.2. Skriveni Markovljevi modeli

Markovljevi lanci

Skriveni Markovljevi modeli su nadogradnja Markovljevih lanaca. Markovljevi lanci su vjerojatnosni model u kojem za svako stanje u lancu postoji vjerojatnost da u idućem trenutku sustav pređe u neko drugo stanje. U takvom sustavu iduće stanje uvijek ovisi samo o trenutnom stanju, a ne ovisi o stanjima prije trenutnog. Primjer Markovljevog lanca s vjerojatnostima prijelaza između stanja može se vidjeti na slici 3.3. Stanja $0 - 3$ su prikazana kao vrhovi grafa, a vjerojatnosti prijelaza iz jednog u drugo stanje se nalaze na lukovima grafa.

Skriveni Markovljevi modeli

Skriveni Markovljevi modeli se od Markovljevih lanaca razlikuju po tome što se ne zna u kojem se stanju u nekom trenutku sustav nalazi, već imamo samo opservacije koje su posljedice sustava u nekom stanju. Opservacija je vektorski niz mel-frekvencijskih značajki signala kratke vremenske duljine. Za svako stanje postoji funkcija koja određenoj opaženoj opservaciji pridodaje vjerojatnost da je ona posljedica upravo tog sta-



Slika 3.3: Graf jednog Markovljevog lanca

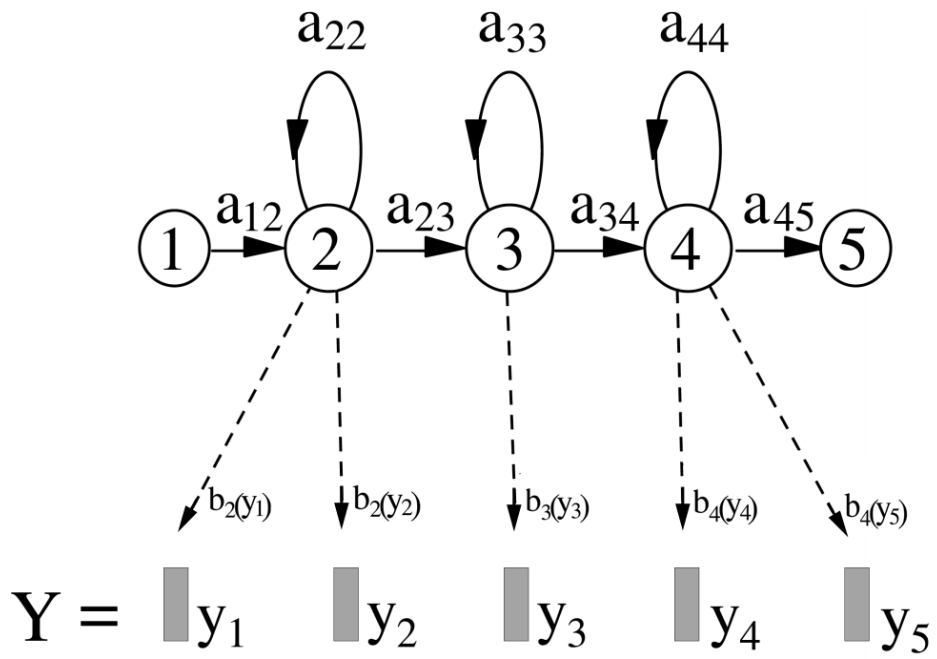
nja. Slikovni opis modela je na slici 3.4 koja dolazi iz članka Gales i Young (2008) koji se bavi pregledom korištenja HMM-a za raspoznavanje govora. Na slici notacija a_{ij} označava vjerojatnost prelaska u idućem vremenskom trenutku iz stanja i u stanje j. Međutim, za razliku od Markovljevih lanaca postoji i funkcija b_s pridružena svakom stanju s koja za određeni niz značajki $\boldsymbol{y_t}$ iz vremenskog trenutka t određuje vjerojatnost da je taj niz značajki posljedica toga da je model u stanju s. Formalno, $b_{s,t} = P(\boldsymbol{y_t}|s)$.

Jedan ovakav model može predstavljati jednu riječ ili jedan monofon/difon/trifon. Kada je u pitanju 1 model = 1 riječ tada se radi o prepoznavanju izoliranih riječi. To znači da se svakoj riječi u rječniku pripisuje jedan model te se na kraju traži najvjerojatniji model. Automatizacija korištena u ovom radu koristi monofone, difone i trifone koji onda čine prepoznavanje slijednog govora. Difon je slijed od dva monofona u riječi, dok je trifon slijed od tri monofona: lijevog konteksta, centralnog monofona i desnog konteksta.

Kod HMM-a postoje tri glavna problema: evaluacija, dekodiranje i treniranje. Za evaluaciju se koristi običan unaprijedni ili unatražni prolaz, dekodiranje se izvodi Viterbijevim algoritmom, a treniranje Baum-Welch algoritmom. Sva tri algoritma se koriste u automatizaciji sustava za prepoznavanje govora uz napomenu da se uzima njihov logaritamski oblik. To znači da se uzima logaritam cijele jednadžbe kako bi se uzastopno množenje malenih vjerojatnosti s kojima se barata pretvorilo u zbroj. U idućim odjeljcima će se proći kroz algoritme za sva tri osnovna problema HMM-a.

Evaluacija

Recimo da je model na slici 3.4 model jednog trifona ili riječi s ulaznim i izlaznim ne-emitirajućim stanjima i tri emitirajuća stanja: 2, 3 i 4. Tada prilikom računanja vjerojatnosti da je taj model $\boldsymbol{M}$ generirao slijed vektora $\boldsymbol{Y}$ dok je prolazio kroz stanja



Slika 3.4: Slika modela HMM-a za raspoznavanje govora

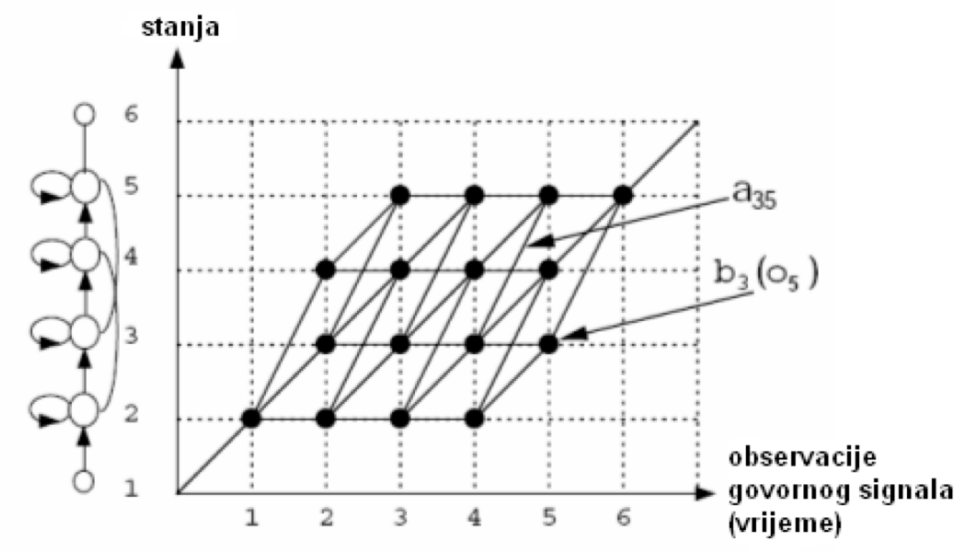
$\boldsymbol{S} = [1, 2, 2, 3, 4, 4, 5]$ jednostavno evaluira se formula 3.2. Ta formula prolazi unaprijed u vremenu, a može se i zapisati obratno tako da se kreće od kraja, no formula ima isto značenje i uporabu.

$$P(\boldsymbol{S}, \boldsymbol{Y}|\boldsymbol{M}) = a_{s_0 s_1} \prod_{t=s_1}^{s_6} b_{s_t}(\boldsymbol{y_t}) \cdot a_{s_t s_{t+1}} \tag{3.2}$$

Dekodiranje

Dekodiranje kod prepoznavanja govora znači otkrivanje izrečene fraze na temelju opservacija (slijeda značajki Mel-frekvencijskog kepstra) i za to se koristi Viterbijev algoritam. Drugim riječima, Viterbijev algoritam se koristi za pronalazak najvjerojatnijeg slijeda stanja u HMM-u na temelju danih opservacija $\boldsymbol{Y}$. Kod prepoznavanja slijednog govora temeljenog na trifonima kao što je to ovdje slučaj to znači da će Viterbi algoritam pronaći najvjerojatniju izgovorenu riječ ili frazu slaganjem modela trifona koji čine riječi iz rječnika maksimiziranjem vjerojatnosti da upravo taj slijed trifona generira opažene opservacije $\boldsymbol{Y}$. Formalno rečeno, želimo maksimizirati Bayesovu vjerojatnost sekvence ostvarenih stvarnih stanja $\boldsymbol{S}$ znajući njihovo opažanje $\boldsymbol{Y}$. Ako sve moguće sekvence skrivenih stanja označimo s $\boldsymbol{S_{svi}}$, onda zapravo tražimo $\max_{\boldsymbol{S} \in \boldsymbol{S_{svi}}} P(\boldsymbol{S}|\boldsymbol{Y})$.

Viterbijev algoritam je dinamički algoritam, što znači da prvo izračuna i zapamti



Slika 3.5: Viterbi algoritam

rezultate manjih problema koji mu kasnije služe za brži izračun rezultata većeg problema. Algoritam se može prikazati formulama 3.3 – 3.4, gdje π_k predstavlja vjerojatnost da se model na početku nalazi u stanju k, a $s \in S$ predstavlja element skupa skrivenih stanja.

$$V_{1,k} = b_k(\boldsymbol{y}_1) \cdot \pi_k \tag{3.3}$$

$$V_{t,k} = \max_{s \in S}(b_k(\boldsymbol{y}_t) \cdot a_{s,k} \cdot V_{t-1,s}) \tag{3.4}$$

Primjer slikovnog prikaza algoritma iz Dropuljić (2008) nalazi se na slici 3.5. U horizontalnom smjeru su označeni vremenski trenuci kojima pripadaju opservacije s Mel-frekvencijskim značajkama, dok su vertikalno poredana sva moguća stanja. Crnom bojom prikazani su mogući putevi kroz stanja. Prijelazi bez označenog puta odgovaraju prijelazima vjerojatnosti jednakoj nuli.

Maksimalna vrijednost u zadnjoj iteraciji daje kraj puta najveće vjerojatnosti. Algoritam dan prethodnim formulama se može nadograditi tako da na kraju ispiše kompletnu sekvencu stanja $\boldsymbol{S}$ kroz koju je model prošao, tj. trifone koji tvore izgovorenu frazu.

Trening

Kod HMM-a trening služi za estimaciju unutrašnjih parametara modela. Kod prepoznavanja govora i HTK paketa koji se koristi za izvedbu HMM-a trening se izvodi

Baum-Welch algoritmom. Baum-Welch algoritam vrsta je EM algoritma (engl. *Expectation Maximization algorithm*). Koristeći znanje od prije HMM možemo definirati kao trojku funkcija $\boldsymbol{Y} = (A, B, \pi)$. Baum-Welch algoritam pronalazi lokalni maksimum $\boldsymbol{Y}_{max} = \text{argmax}_{\boldsymbol{M}} P(\boldsymbol{Y}|\boldsymbol{M})$. Dakle algoritam postavlja unutarnje parametre tako da se maksimizira vjerojatnost opažanja $\boldsymbol{Y}$.

Algoritam se odvija u tri faze:

1. *Unaprijedni prolaz*: računanje vjerojatnosti da je sustav u vremenskom trenutku t u stanju i i da je prvih t opservacija upravo prvih t vektora od $\boldsymbol{Y}$ — označimo to $\boldsymbol{Y}[1:t]$. Formalno, računa se $\alpha_i(t) = P(\boldsymbol{Y}[1:t], s_t = i|\boldsymbol{M})$ i dobiva se rekurzivno jednadžbama 3.5 - 3.6.

$$\alpha_i(1) = \pi_i b_i(\boldsymbol{y}_1) \tag{3.5}$$

$$\alpha_j(t+1) = b_j(\boldsymbol{y}_{t+1}) \sum_{i=1}^{|S|} \alpha_i(t) a_{ij} \tag{3.6}$$

2. *Unatražni prolaz*: računanje vjerojatnosti da je sustav u vremenskom trenutku t u stanju i te da je ostatak opservacija $\boldsymbol{Y}[t+1:T]$. Dakle računa se $\beta_i(t) = P(\boldsymbol{Y}[t+1:T], s_t = i|\boldsymbol{M})$, a ta funkcija se dobiva rekurzivno jednadžbama 3.7 - 3.8.

$$\beta_i(T) = 1 \tag{3.7}$$

$$\beta_i(t) = \sum_{j=1}^{|S|} \beta_j(t+1) a_{ij} b_j(\boldsymbol{y}_{t+1}) \tag{3.8}$$

3. *Osvježavanje parametara*: općenita formula za vjerojatnost da je model u stanju i u trenutku t je dana jednadžbama 3.9 – 3.10. Nakon toga možemo izračunati ukupnu vjerojatno da je model u stanju i u trenutku t i da radi prijelaz u stanje j u trenutku $t+1$ prema jednadžbi 3.11. Tada znamo je očekivanje broja prijelaza iz s_i jednako $\sum_{t=1}^{T} \gamma_i(t)$, a očekivani broj prijelaza iz s_i u s_j jednak $\sum_{t=1}^{T} \xi_t(i,j)$. Konačno, sada se pomoću njih može napraviti korak maksimizacije, tj. osvježavanje parametara a i b prema jednadžbama 3.12 – 3.14 uz napomenu da je vrijednost funkcije $1_{\boldsymbol{y}_t = \boldsymbol{y}_k}$ jednaka 1 ako je $x = y$, a u suprotnom iznosi 0.

$$
\begin{aligned}
P(s_t = i, \boldsymbol{Y}|\boldsymbol{M}) &= P(\boldsymbol{Y}[1:t], s_t = i|\boldsymbol{M}) \cdot P(\boldsymbol{Y}[t+1:T]|s_t = i, \boldsymbol{M}) \\
&= \alpha_i(t)\beta_i(t)
\end{aligned}
\tag{3.9}
$$

$$
\begin{aligned}
P(s_t = i|\boldsymbol{Y}, \boldsymbol{M}) &= \frac{P(s_t = i, \boldsymbol{Y}|\boldsymbol{M})}{P(\boldsymbol{Y}|\boldsymbol{M})} \\
&= \gamma_i(t)
\end{aligned}
\tag{3.10}
$$

$$
\begin{aligned}
\xi_t(i,j) &= P(s_t = i, s_{t+1} = j|\boldsymbol{Y}, \boldsymbol{M}) \\
&= \frac{\alpha_i(t) \cdot a_{i,j} \cdot b_{s_j}(\boldsymbol{y}_{t+1}) \cdot \beta_j(t+1)}{P(\boldsymbol{Y}|\boldsymbol{M})}
\end{aligned}
\tag{3.11}
$$

$$
\pi_i^* = \gamma_i(1) \tag{3.12}
$$

$$
a_{i,j}^* = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_i(t)} \tag{3.13}
$$

$$
b_i^*(\boldsymbol{y}_k) = \frac{\sum_{t=1}^{T} 1_{\boldsymbol{y}_t = \boldsymbol{y}_k} \gamma_i(t)}{\sum_{t=1}^{T} \gamma_i(t)} \tag{3.14}
$$

3.1.3. Gaussove mješavine

Nakon objašnjenja kako reprezentiramo govor računalu i kako koristimo model za prepoznavanje govora, još ostaje opisati od čega se zapravo sastoje ti modeli, tj. koje informacije čuva svako stanje skrivenog Markovljevog modela. Knjiga Yu i Deng (2014) na sistematičan i matematički način obrađuje širok spektar tema o prepoznavanju govora, a ovdje ćemo se koncentrirati samo na njeno poglavlje o Gaussovim mješavinama.

Započnimo s nekom kontinuiranom nasumičnom varijablom x. Svaka kontinuirana nasumična varijabla ima dvije funkcije - funkciju gustoće vjerojatnosti p i razdiobe P te su prikazane formulama 3.15 – 3.16 uz $x = a$. Za vektore slučajnih kontinuiranih varijabli te se funkcije definiraju na sličan način.

$$
p(a) = \lim_{\Delta a \to 0} \frac{P(a - \Delta a < x \leq a)}{\Delta a} \tag{3.15}
$$

$$P(a) = P(x \leq a) = \int_{-\infty}^{a} p(x)dx \tag{3.16}$$

Slučajna varijabla koja prati Gaussovu distribuciju ima gustoću vjerojatnosti danu formulom 3.17, te se označuje kao $x \sim \mathcal{N}(\mu, \sigma^2)$. Ovdje vrijede jednostavni oblici za očekivanje i varijancu: $E(x) = \mu$, te $var(x) = \sigma^2$.

$$p(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{3.17}$$

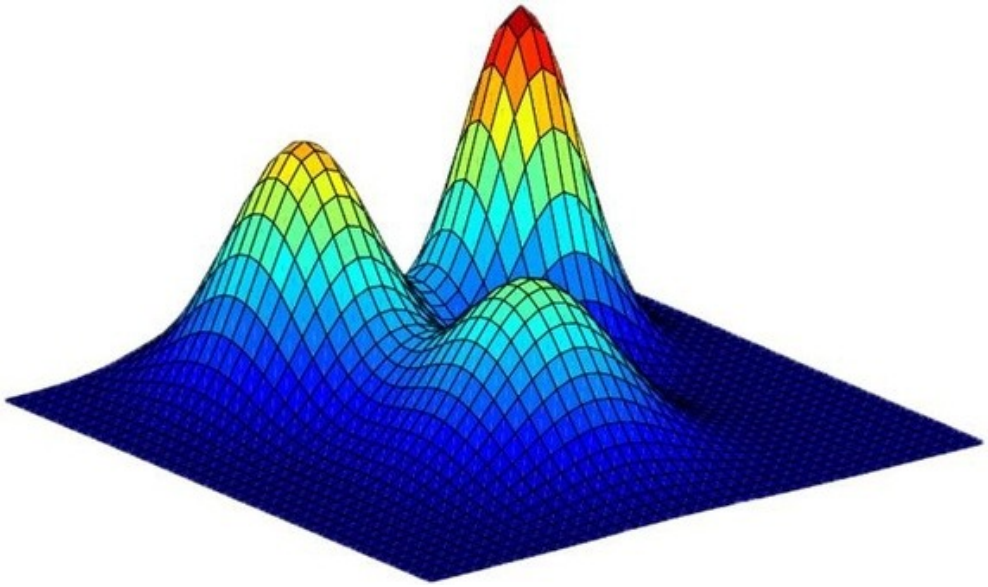
Ako se umjesto obične kontinuirane slučajne varijable koristi vektor-stupac $\boldsymbol{x}$ kontinuiranih slučajnih varijabli dimenzije D onda funkcija gustoće vjerojatnosti prelazi u 3.18 i vektor označujemo s $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu} \in \mathbb{R}^D, \boldsymbol{\Sigma} \in \mathbb{R}^{DxD})$. Očekivanje i varijanca prelaze u svoje vektorske oblike: $E(\boldsymbol{x}) = \boldsymbol{\mu}$ i $E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T] = \boldsymbol{\Sigma}$.

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right] \tag{3.18}$$

Ovi opisani modeli su unimodalni, tj. slučajna varijabla (dimenzije D) opisana je jednom Gaussovom razdiobom. Međutim, pokazalo se da su potrebe prepoznavanja govora korisniji multimodalni modeli zbog više komponenti koje čine podatke, pa tada svaka može biti uzrok jedne Gaussove komponente. Tada je slučajna varijabla definirana kao linearna kombinacija M Gaussovih razdioba, a njena gustoća vjerojatnosti je jednaka linearnoj kombinaciji gustoća vjerojatnosti pripadajućih Gaussovih razdioba. Taj oblik se zove Gaussova mješavina te joj je gustoća vjerojatnosti dana jednadžbom 3.19. Svaka težina c_m je pozitivna te je njihov zbroj jednak jedinici, tj. $\sum_{m=1}^{M} c_m = 1$. Tada je očekivanje te varijable jednostavno $E(\boldsymbol{x}) = \sum_{m=1}^{M} c_m\boldsymbol{\mu}_m$. Promatranjem formule za očekivanje vidljivo je da ovakav model jedino ima smisla ako su srednje vrijednosti bliske jedna drugoj. Radi vizualizacije krajnjeg produkta, na slici 3.6 je prikaz dvodimenzionalne Gaussove mješavine s tri komponente.

$$\begin{aligned} p(\boldsymbol{x}) &= \sum_{m=1}^{M} \frac{c_m}{(2\pi)^{D/2}|\boldsymbol{\Sigma_m}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_m})^T{\boldsymbol{\Sigma_m}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_m})\right] \\ &= \sum_{m=1}^{M} c_m\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu_m}, \boldsymbol{\Sigma_m}) \end{aligned} \tag{3.19}$$

Važno je još spomenuti da je kod većih dimenzija (primjerice u ovom radu je to 39) nepraktično spremati cijelu matricu $\boldsymbol{\Sigma}$ jer je ona veličine 39x39 elemenata, a u samo jednoj mješavini postoji M takvih matrica. Zato se često uzima samo dijagonala



Slika 3.6: Gaussova mješavina

kovarijantne matrice Σ koja onda znatno smanjuje memorijske zahtjeve, a i računanje postaje jednostavnije. Takva matrica i dalje čuva najbitnije informacije jer se u njoj i dalje nalaze varijance svih varijabli, pa se često koristi u praksi, a koristi se i u automatizaciji korištenoj u ovom radu. Njene srednje vrijednosti i varijance su upravo parametri koji se optimiziraju korištenjem Baum-Welch algoritma.

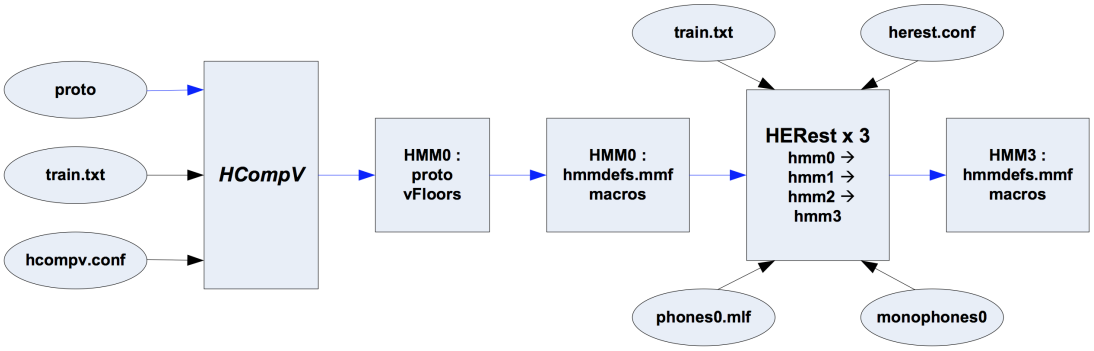
3.2. Automatizacija prepoznavanja govora

U ovom radu se koristi automatizacija napravljena u sklopu diplomskog rada Dropuljić (2008). U ovom će se poglavlju ukratko opisati proces koji obavlja automatizacija (treniranje i testiranje modela) i njeno korištenje, a detalji se mogu pronaći u izvorniku Dropuljić (2008).

Preduvjeti za korištenje sustava su instaliran HTK toolkit Young i Young (1993) (i dodan u varijablu okoline), instaliran MATLAB i mogućnost izvođenja Perl i Python programa.

3.2.1. Proces izgradnje akustičkog modela

Proces izgradnje akustičkog modela opisat će se u tri dijela pomoću grafova procesa koji se nalaze na slikama 3.7 - 3.9. Elipse predstavljaju datoteke koje se koriste, a u pravokutnicima se nalaze imena korištenih HTK funkcija. `HMM0–HMM13` su mape u kojima se gradi sustav, a istrenirani model se na kraju nalazi u mapi `HMM13`. Na po-



Slika 3.7: Inicijalizacija automatizacije

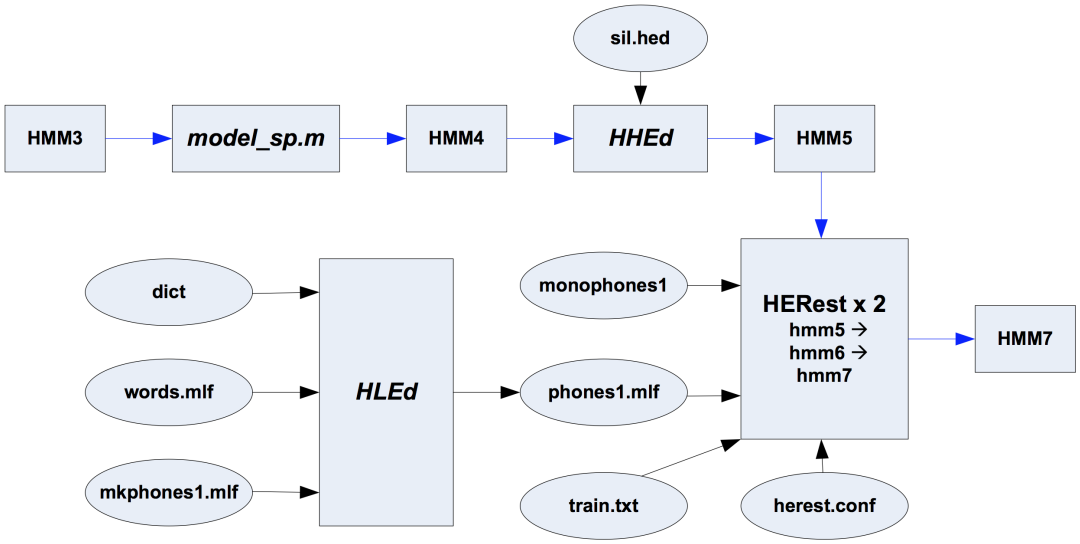
četku imamo datoteku rječnika `dict`, konfiguracijske datoteke, rečenice izgovorene u uzorcima skupa za treniranje `train.txt` i njihove transkripcije `words.mlf`. Njihovo dobivanje opisano je u poglavlju 3.4.

Prvi dio automatizacije se nalazi na slici 3.7 i pokazuje dio koji se tiče inicijalizacije cijelog procesa. Početna mapa `HMM0` dobiva se zvanjem inicijalizacijske HTK funkcije `HCompV` koja radi *flat-start* inicijalizaciju, tj. sve srednje vrijednosti i varijance postavlja na globalnu srednju vrijednost i varijancu. Nakon toga se parametri reestimiraju tri puta uporabom funkcije `HERest` koja izvodi Baum-Welch algoritam te se pune mape `HMM1` –`HMM3` . Važno je napomenuti da se ovdje radi o monofonima — svaki uzorak iz seta za treniranje je raspodijeljen na monofone pomoću transkripcije u datoteci `phones0.mlf`.

Na slici 3.8 prikazan je daljnji nastavak razvijanja akustičkog modela u kojem se uvodi model kratke pauze između riječi `sp`. Kroz MATLAB skriptu `model_sp.m` stvara se model kratke tišine koji je prva nadogradnja u poboljšavanju osnovnog sustava. Nakon toga se pomoću HTK-ove funkcije `HHed` centralno stanje modela `sp` povezuje s centralnim stanjem model tišine `sil`. Rezultat odlazi u mapu `HMM5`.

Funkcija `HLEd` radi novu datoteku `phones1.mlf` koja se od transkripcije `phones0.mlf` razlikuje jedino u tome što ima i labelu `sp` između riječi. Nad novim sustavom sa `sp` modelom se izvodi dvostruka reestimacija parametara, a njen izlaz se sprema u mapu `HMM7`.

Završni, a i najveći dio automatizacije prikazan je na slici 3.9. Ovaj dio treninga uključuje kreiranje kontekstno ovisnih trifonskih modela. Do sada su se koristili isključivo monofoni uz pretpostavku da svaki monofon uvijek isto zvuči, no izgovor monofona ovisi i o položaju fonema u riječi i fonemu prije i iza spomenutog fonema. Primjerice, u riječi *banka* glas *n* je nosnozubnik, a u riječi *danas* je mekonepčanik. Takvi različiti izgovori istog fonema se zovu alofoni. Pretpostavka je da modeliranje



Slika 3.8: Uvođenje modela `sp`

trifona pokriva alofone jer tada svaki centralni fonem u trifonu ima svoj lijevi i desni kontekst, a njihovi parametri se estimiraju zasebno.

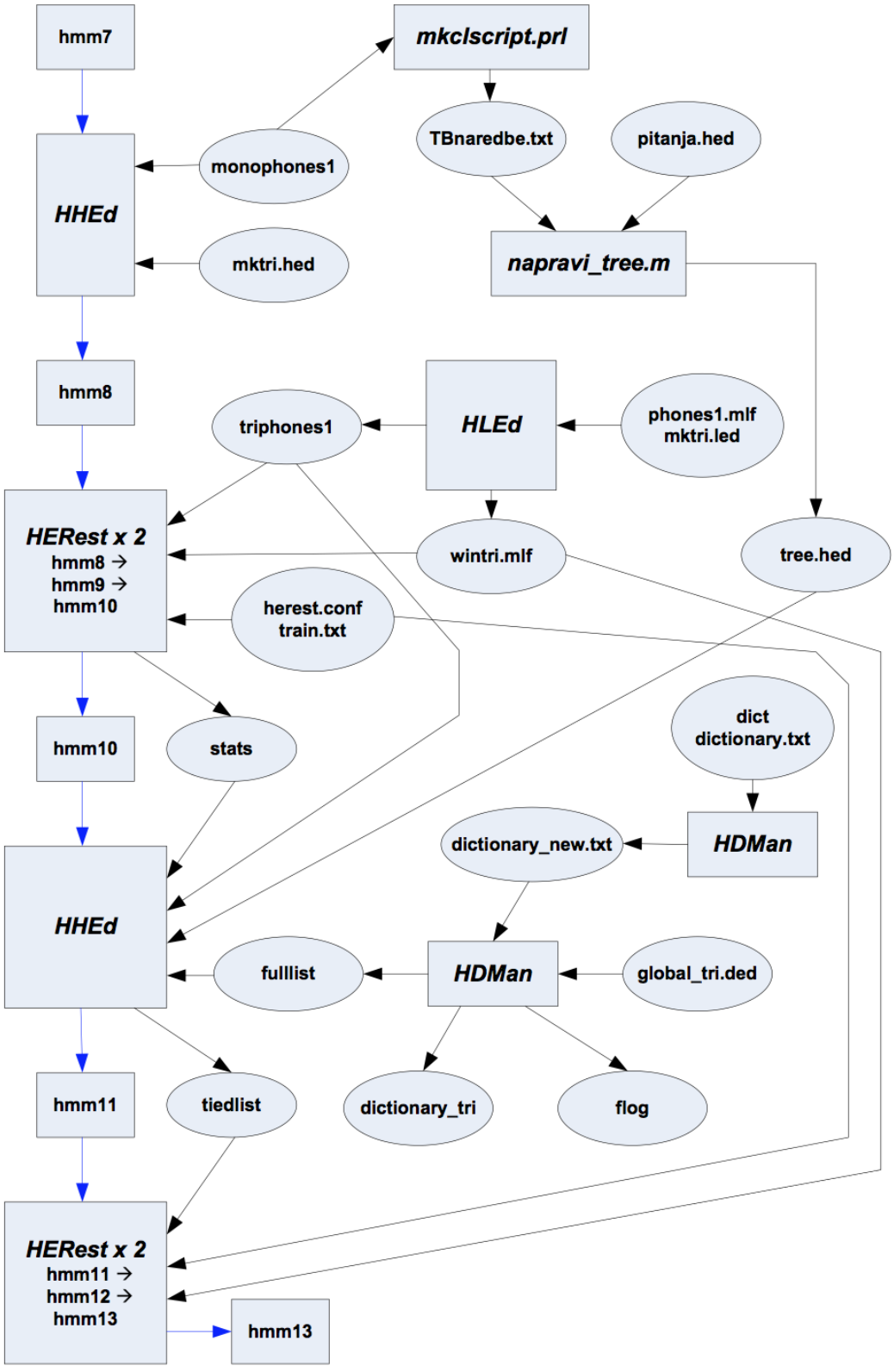
HTK funkcija `HLEd` iz monofonskih stvara trifonske transkripcije uzoraka `wintri.mlf` i njihovu listu `triphones1`. Osim trifonske transkripcije potrebni su i trifonski modeli koje radi poziv HTK funkcije `HHEd`, a datoteka `mktri.hed` joj govori da trifonske modele napravi pomoću kopija monofonskih modela. Sada se trifonski model može ponovo dva puta reestimirati naredbom `HERest` te rezultat odlazi u mapu `HMM10`.

Sada HMM model sadrži sve trifone koje se pojavljuju u uzorcima za trening i test, no ako se želi postići prepoznavanje govora cijelog jezika potrebno je modelirati sve trifone jezika. To se radi iz rječnika `dictionary.txt` u kojem se nalazi najčešćih 15000 riječi hrvatskog jezika te se pomoću funkcije `HDMan` stvara lista trifona `fulllist` za koju se smatra da opisuje čitav jezik.

Još jedino preostaje izgradnja modela novih trifona. To se obavlja pozivanjem funkcije `HHEd` koja to obavlja stapanjem nekih modela u zajedničke razrede i pripisivanje tih modela preostalim trifonima iz `fulllist`. Konačno, vrši se zadnja dvostruka reestimacija modela i tu je izgradnja akustičkog modela gotova.

3.2.2. Korištenje automatizacije

Korištenje automatizacije je vrlo jednostavno. Razvijeno je grafičko korisničko sučelje u kojem je potrebno samo odabrati treći mod rada, direktorije za trening i test i pokrenuti program. Može se odabrati i parametar penalizacije svake nove riječi p, no



Slika 3.9: Završni dio automatizacije

o tome, kao i o stvaranju direktorija za trening i test će se više reći u idućem odjeljku.

3.3. Podaci akustičke baze

Za prikupljanje podataka za izradu akustičkog modela hrvatskog jezika imali smo pristup snimkama bihevioralnih terapija petero dječaka s dijagnosticiranim poremećajem iz autističnog spektra. Međutim, razvijena igra treba generalizirati prepoznavanje i na novo dijete čije snimke nisu uključene u set za treniranje, a za to nije dovoljno imati petero govornika. Uz to, broj izrečenih riječi koji se može snimiti u jednom tretmanu je relativno malen zbog njihove same dijagnoze, pa ukupan broj snimljenih fraza nije prevelik iako je snimljeno barem nekoliko tretmana za svako dijete. Zato se odlučilo nadopuniti set za treniranje dodatnim snimkama. Dodatne snimke za učenje moraju biti dječje jer se vokalni trakt i izgovor odraslih osoba znatno razlikuju od dječjeg. S obzirom na to da se fizionomija vokalnog trakta djece s PAS ne razlikuje od djece urednog razvoja, dio snimki je zamijenjen govorom djece urednog razvoja. Kako spol i uzrast djece također igraju ulogu, a kod snimljene djece s PAS se radi o dječacima u dobi od 5 do 7 godina, dječaci učenici prvog razreda su idealni kandidati za popunjavanje seta za treniranje. Tako je snimljen govor još 17-ero dječaka iz dva prva razreda osnovne škole. To je ukupno 22 sudionika što je dovoljno za sustav koji može generalizirati. Sve snimke su prikupljene uz suglasnost roditelja.

Riječi koje se koriste u ovom radu su one koje djeca nauče u ranom stadiju razvoja govora. Tako je uzeto ukupno 42 riječi i fraza. Za imenice su još snimljeni oblici u svim padežima, a kod glagola su snimljeni prezent i perfekt uz eventualno još neki česti oblik. Uz odvojene riječi snimljeno je i nekoliko osnovnih fraza koje povezuju dvije ili tri riječi s popisa. Popis riječi nalazi se na slici 3.10. Uz padeže i vremena ukupan se broj različitih riječi penje na 139. Taj popis riječi se nalazi u datoteci `rjecnik_odabran.txt` direktorija `lib` opisanog u poglavlju 3.4.

Kod osnovnoškolske djece su snimljene sve riječi, fraze i gramatički oblici kroz razgovor i ponavljanje za mnom ili čitanje, pa se među uzorcima za svako dijete nalaze skoro sve riječi osim manjeg broja slučajeva gdje se prilikom preslušavanja utvrdilo da se neki uzorak mora odbaciti. S druge strane, kod djece s poremećajem iz autističnog spektra glavna tema svake snimljene terapije je bila ponavljanje osnovnih 42 riječi za terapeutkinjom jer je njima već i tih 40-ak riječi teško reći ili ponoviti unutar jednog tretmana.

Treba još napomenuti da su djeca znala izreći i riječi i fraze koje nisu na popisu, a neke su riječi s popisa izrekli više puta. Svaka se dodatna riječ ili fraza unijela u

1. auto
2. boja
3. bojati
4. bojica
5. čitati
6. čokolada
7. dječak
8. gleda
9. jakna
10. jede
11. kuća
12. kula
13. lopta
14. mač
15. mačka
16. medo
17. mlijeko
18. nos
19. pas
20. pije
21. pijesak
22. plava
23. pleše
24. pliva
25. šuma
26. vozi
27. zelena
28. žuta
29. bolestan je
30. dječak čita
31. dječak gleda
32. dječak ima mač
33. dječak jede
34. ispod stola
35. kula od pijeska
36. medo je u šumi
37. medo vozi auto
38. mačka pije mlijeko
39. nema više
40. otišao je
41. oblači se
42. oblači jaknu

Slika 3.10: Riječi i fraze korištene za prepoznavanje govora

uzorke kako se ne bi gubili uzorci koji će možda jednog dana zatrebati. Oni se u ovom radu koriste za oplemenjivanje akustičkog modela, no uzorci koji nisu s popisa se ne koriste za testiranje.

Ukupno je snimljeno 3934 uzorka. Tablica uzoraka u kojoj su zapisani podaci o svakom uzorku ima sedam stupaca:

1. `ID`: ID snimke
2. `Govornik`: Ime djeteta koje je izgovorilo uzorak
3. `Datum`: Datum snimke
4. `Rijec`: Riječ ili fraza koja se izgovara na snimci
5. `IzrecenoDrugo`: Ako riječ nije pravilno izgovorena, ovdje se zapisuje što je zapravo izrečeno
6. `OcjenaIzgovora`: Ocjena 1 – 3 za ocjenu izgovora (1 - veće nepravilnosti u izgovoru, 2 - manje nepravilnosti u izgovoru, 3 - uredno izgovoreno)
7. `VanjskiSum`: Ocjena vanjskog šuma 0 – 2 (0 - nema šuma, 1 - prisutan je manji šum, 2 - šum može smetati pri prepoznavanju izrečene riječi ili fraze)

Svi uzorci pohranjeni su u *wav* formatu s frekvencijom otipkavanja od 48000 Hz i zapisom od 16 bita po uzorku. Broj uzoraka po djetetu dan je u tablici 3.1. Prvih pet govornika su djeca s poremećajem iz autističnog spektra, dok su preostalih 17 djeca urednog razvoja.

Tablica 3.1: Broj uzoraka po djetetu

Govornik	Broj uzoraka	Govornik	Broj uzoraka
Filip	126	Jakov2	218
Ilija	261	Josip	205
Ivor	53	Leon	177
Lovro	129	Marko	207
Lukas	220	Mihovil	169
Alan	162	Rafael	182
Eli	180	Roko	196
Filip2	196	Roko2	166
Filip3	190	Vid	177
Gabrijel	166	Vinko	178
Jakov	188	Zvonimir	185

Pažljivijim promatranjem broja uzoraka iz tablice 3.1 može se vidjeti da nedostaju tri uzorka. Prilikom izgradnje akustičkog modela moraju biti prisutni svi fonemi jezika, no u cijelom setu za treniranje nije bilo fonema *đ*, *dž* i *f* jer se ti fonemi ne javljaju na popisu niti su ih djeca ikada sama izrekla. Zato su za ispunjavanje uvjeta dodana još tri dodatna uzorka s tim fonemima, no oni se u sustavu i dalje efektivno ne koriste jer se u rječniku za test također ne pojavljuju ta tri fonema.

3.4. Ocjena točnosti modela i automatizacija pripreme podataka

3.4.1. Ocjena točnosti modela

Skupovi za treniranje i testiranje se mogu napraviti na više načina. U odjeljku 3.5 se prati njihov razvoj. U svakom načinu rada metoda evaluacije modela je ista - modificirana 22-dijelna krosvalidacija. S obzirom na to da model treba generalizirati prepoznavanje govora i na snimke djece koja nisu dio istreniranog modela, model se trenira na snimkama 21-og djeteta, a testira se na 22-om i tako 22 puta. Za svaki model ocjenjuje se točnost prepoznavanja po riječima i po frazama. Zbog toga što djeca imaju međusobno različit broj uzoraka za testiranje, ukupna točnost modela je težinski prosjek točnosti po djetetu gdje je težina udio testnih uzoraka tog djeteta među svim testnim

uzorcima. Točnost t jednog dijela krosvalidacije dana formulom 3.20 gdje je n ukupan broj primjera za test u tom dijelu, d je broj izgubljenih riječi (engl. *deletions*), a s je broj riječi koje su zamijenjene nekom drugom (engl. *supstitutions*). Postoji i metrika i koja označuje dodane riječi (engl. *insertions*), no ovdje se ne uzima u obzir jer u konačnici nije bitno je li neko dijete dodalo još neku riječ u frazu ako je izrečeno ono što je trebalo izreći. Dodatno, oznaka h označava broj točno prepoznatih fraza, a koristi se kasnije u analizi rezultata. Ukupna točnost T dana je formulom 3.21.

$$t = \frac{n - d - s}{n} \cdot 100\% \tag{3.20}$$

$$T = \frac{\sum_{i=1}^{22} n_i * t_i}{N} \tag{3.21}$$

Ova metoda se razlikuje od klasične krosvalidacije po tome što se uzorci za testiranje ne biraju slučajnim odabirom, već su unaprijed određeni i odvojeni po djetetu. Razlog tome je važnost generalizacije sustava na novo dijete. Takva podjela još povlači i razliku u veličini dijelova za test kod svakog djeteta, dok se u formalnoj N-dijelnoj krosvalidaciji set uzoraka dijeli na N jednakih dijelova. Taj je problem riješen uzimanjem težinske srednje vrijednosti, za razliku od običnog prosjeka točnosti svakog modela kod klasične krosvalidacije. Važno je još napomenuti da je za krajnji akustički model koji se koristi u računalnoj igri korišten cijeli set uzoraka, bez ostavljanja dijela za test.

3.4.2. Princip rada automatizacije

U ovom radu se koristi nekoliko načina izrade skupova za treniranje i testiranje. Kako bi se olakšao proces izrade svih potrebnih datoteka i raspodijele uzoraka po skupovima napravljena je automatizacija tog procesa. Ako se koristi automatizacija stvaranja akustičkog modela iz odjeljka 3.2 mnogi dijelovi su već na mjestu, no prije pokretanja ih je potrebno nadopuniti podacima specifičnim za način stvaranja modela kojeg želimo. Koraci potrebni za uspješno pripremanje podataka za trening i test su:

1. Izrada rječnika koji će se koristiti za testiranje
2. Izrada gramatike koju fraze moraju pratiti i koja se koristi prilikom prepoznavanja govora
3. Podjela uzoraka na uzorke za treniranje i testiranje

4. Spajanje transkripcija s uzorcima za treniranje i testiranje

5. Vađenje Mel-frekvencijskih značajki iz uzoraka

6. Stvaranje grafa iz gramatike

Automatska priprema podataka za treniranje i testiranje se odvija u dva dijela. Prvi korištenjem Python skripte stvara i priprema sve potrebne datoteke (koraci 1–4), a drugi dio korištenjem MATLAB-ovih funkcija računa Mel-frekvencijske značajke i gradi graf gramatike (koraci 5–6).

Python dio

Dio pripreme koji se izvodi u Pythonu započinje s izradom rječnika i gramatike. Rječnik za testiranje se gradi ili od svih riječi koje su izrečene ili samo iz različitih oblika riječi iz liste sa slike 3.10, ovisno o načinu rada. Popis svih riječi se jednostavno dobiva iz tablice uzoraka i njenog stupca `Rijec`, a lista izabranih riječi je unaprijed određena datotekom `rjecnik_odabran.txt`. Posao izbora i sastavljanja rječnika obavlja modul `napravirjecnik.py`.

Nadalje, u ovom radu je korištena i jednostavna gramatika - tzv. *nulta gramatika*. To znači da nema određenog nužnog slijeda riječi već je svaki redoslijed moguć. Gramatika se radi na temelju gotovog rječnika i sprema se u datoteku `grammar.txt`, a time se bavi modul `napravigrammar.py`.

Modul koji radi podjele uzoraka na skupove za trening i test se zove `pripremi_uzorke_trening_test.py`. Kao što je objašnjeno ranije, točnost modela se dobiva preko točnosti njegovih 22 komponenti. Zato se u direktorijima za treniranje i testiranje mora nalaziti po 22 direktorija. To je izvedeno tako da za jednu komponentu postoji po jedan direktorij u direktoriju za treniranje i jedan za testiranje s imenom govornika (djeteta) kojem ta komponenta pripada. Ovaj modul iz tablice uzoraka izvlači sve govornike (u ovom slučaju je to 22 djece) i njihove uzorke zajedno s izgovorenim riječima. Zatim za svakog govornika njegove uzorke s transkripcijama preimenuje u uzorke $1 - n$ njegovog direktorija za testiranje, a sve ostale uzorke i pripadne transkripcije također preimenuje i stavlja u njegov direktorij za treniranje. Kod različitih načina rada ovdje se može odvijati i filtriranje uzoraka o kojem će više riječi biti u odjeljku 3.5.

Ova tri modula su povezana u `main.py` skriptu koja prima samo jedan argument — način (mod) rada, a iz nje se dalje pozivaju i grade datoteke i strukture potrebne za treniranje i testiranje akustičkog modela.

MATLAB dio

MATLAB je također potrebno inicijalizirati za prepoznavanje pokretanjem skripte `napravi_code_wordnet_wordlist.m` koja spaja nekoliko dijelova automatizacije iz Dropuljić (2008). Ona prvo vadi Mel-frekvencijske značajke iz svih uzoraka za treniranje i testiranje te stvara datoteku koja spaja računalne puteve uzoraka i njihovih značajki, a zatim na temelju gramatike radi njen graf i listu stanja.

3.4.3. Korištenje automatizacije

Priprema podataka koja prethodi gradnji modela se provodi u nekoliko jednostavnih koraka:

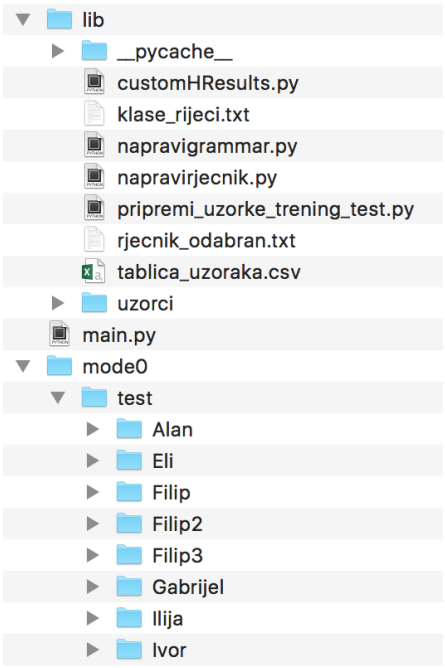
1. Pozivanje `main.py` skripte s argumentom načina rada
2. Kopiranje direktorija za trening i direktorija za test željenog načina rada u radni direktorij `HTK_uzorci` koji se nalazi u glavnom MATLAB direktoriju — `prepoznavanje_govora`
3. Iz direktorija `prepoznavanje_govora` pozvati MATLAB funkciju `napravi_code_wordnet_wordlist.m`

Struktura direktorija `prepoznavanje_govora` preuzeta je iz Dropuljić (2008), dok je struktura pripremnog direktorija djelomično prikazana slikom 3.11. U njoj se nalazi glavna skripta `main.py` koja se služi podacima iz subdirektorija `lib`, a rezultate spremna u `trening` i `test` subdirektorije direktorija `mode`n kada se radi o načinu rada n.

3.5. Načini izrade skupova za treniranje i testiranje

3.5.1. Osnovni način rada: mod 0

U osnovnom načinu rada, ovdje vođenom pod brojkom 0 je izrada trening i test seta od svih uzoraka. Ti uzorci uključuju i riječi koje se nalaze na popisu riječi i njihovih oblika, ali i ostatak riječi i fraza koje su djeca spontano izrekla. Rezultati ovog pristupa su referentna točka koja služi za ocjenjivanje poboljšanja ostalih načina rada. Rezultati su dani u tablici 4.1.



Slika 3.11: Struktura pripremnog direktorija

3.5.2. Selekcija uzoraka za test: mod 1

U prvom poboljšanju, tj. prvom načinu rada su za izgradnju skupa za treniranje korišteni svi uzorci osim onih koji dolaze od govornika na kojem se testira. Međutim, u skupu za testiranje su odabrani isključivo uzorci čije se riječi i fraze nalaze na popisu odabranih riječi i fraza i njihovih oblika. Takav model daje realističniju procjenu kvalitete modela jer će se model ionako koristiti za točno te fraze. Rezultati tog načina rada su dani u tablici 4.2.

3.5.3. Optimizacija parametra p: mod 2

Način rada broj 2 podrazumijeva način rada 1 i na takvim skupovima za treniranje i testiranje optimizira parametar p. Parametar p se koristi prilikom testiranja — traženja najvjerojatnije izrečene fraze i predstavlja penalizaciju koju svaki mogući put kroz modele dobiva na prijelazu riječi, tj. prilikom dodavanja nove riječi u frazu. Kada je parametar p jednak nuli, onda nema penalizacije, a što je parametar manji od nule, to je penalizacija veća. Rezultati su prikazani u tablici 4.3 i grafički na slici 4.1.

1. auta auto autom autu
2. boja bojama boje bojom
3. bojanje bojenje
4. bojao bojati boja
5. bojica bojicama bojice bojici bojicom bojicu
6. bolestan
7. čita čitamo čitao čitat čitati
8. čokolada čokolade čokoladi čokoladom čokoladu
9. dječak dječaka dječakom dječaku
10. gleda gledao
11. ide otišao
12. ima
13. ispod
14. jakna jakne jakni jaknom jaknu
15. je bio
16. jede jeo
17. kuća kuće kući kućom kuću
18. kula kule kuli kulom kulu
19. lopta lopte lopti loptom loptu
20. mač mača mačem mačom maču
21. mačka mačke mački mačkom mačku
22. mede medi medo medom medu
23. mlijeka mlijeko mlijekom mlijeku
24. nema
25. nos nosa nosom nosu
26. oblači obukao obukla
27. od
28. pas psa psom psu
29. pije pila
30. pijesak pijeska pijeskom pijesku
31. plava plave plavo plavoj plavom plavu
32. plesale plesali pleše plešu
33. pliva plivao
34. s sa
35. se
36. stola
37. su
38. šuma šume šumi šumom šumu
39. u
40. više
41. vozi vozio
42. zelena zelene zeleno zelenoj zelenom zelenu
43. žuta žute žuto žutoj žutom žutu
44. je od s sa se su u

Slika 3.12: Razredi riječi

3.5.4. Priznavanje različitih oblika riječi: mod 3

U ovom načinu rada dodaju se razredi riječi. Razredi riječi su skupine riječi koje u ovom radu uzimamo kao jednakovrijedne. Za imenice to su različiti padeži, a za glagole prezent, perfekt i poneka veoma česta drugačija forma. Razlog dodavanja razreda riječi objašnjen je u odjeljku 2.2.3. Popis razreda riječi nalazi se na slici 3.12, a u sustavu se nalazi u transkripcijskom obliku u datoteci `klase_rijeci.txt`. U zadnjem retku se nalazi i popis prijedloga koji se mogu, ali i ne moraju nalaziti u izrečenoj frazi, tj. ne smatra se problematičnim ako dijete upotrijebi pogrešan prijedlog ili on izostane.

Prilagođena verzija ocjene modela se pokreće nakon što je pronađena najvjerojatnija fraza korištenjem Viterbi algoritma. Umjesto korištenja uobičajene HTK metode `HResults`, poziva se Python skripta `customHResults.py` koja dopušta da se umjesto neke riječi može izreći i bilo koja druga riječ iz istog razreda. Rezultati se mogu pogledati u tablici 4.4.

3.5.5. Eliminacija lošije izgovorenih uzoraka iz skupa za treniranje: mod 4

Prilikom unošenja uzoraka u sustav za svaki je uzorak zapisana i kvaliteta izgovora podijeljena na tri razine: 1 (veće nepravilnosti u izgovoru), 2 (manje nepravilnosti

u izgovoru) i 3 (dobro izgovoreno). Moguće je da pojedini lošije izgovoreni uzorci smanjuju kvalitetu modela prilikom treniranja. Zato je u ovom načinu rada isprobano treniranje samo na uzorcima koji su ocijenjeni ocjenama 2 i 3. Testiranje se i dalje odvija na svim primjerima jer se ne može očekivati da će dijete sve uvijek skroz točno izgovoriti. Rezultati ovog načina rada prikazani su u tablici 4.5.

3.5.6. Treniranje konačnog modela: mod 5

Na kraju, napravljen je i način rada za pripremu skupa za treniranje završnog modela. U njemu se koriste uzorci za treniranje svih govornika bez ostavljanja uzoraka za testiranje. Taj model se ne ocjenjuje, pa za njega nema rezultata, ali se koristi u računalnoj igri.

4. Rezultati i analiza akustičkog modela

4.1. Osnovni način rada: mod 0

Rezultati osnovnog načina rada dani su u tablici 4.1. Ukupna točnost modela je 59.32% na razini riječi, a tek 54.10% na razini cijele fraze. Ako se uzme u obzir i to da su u prvih pet redaka zapisani rezultati djece s poremećajem iz autističnog spektra koji su ispod 50% može se reći da sustav koji ovako radi nije jako pouzdan, tj. griješit će u više od 50% slučajeva kada dijete izgovori točnu riječ ili frazu.

Međutim, dubljim promišljanjem se vidi kako je ovo pesimistična procjena modela. U kontekstu računalne igre se koristi fiksan vokabular, dok se ovdje testira i na riječima koje su daleko od svih ostalih primjera. S obzirom na to da su sva djeca izgovorila riječi iz zadanog vokabulara postoji mnogo primjera riječi i trifona koji se nalaze u njima, ali ostatak riječi koje su bile spontani govor djece se može uvelike razlikovati od zadanog vokabulara. Moguće je da za neke foneme u tim riječima postoji tek jedan primjer u cijelom skupu za treniranje, pa nije realno za očekivati da će se te riječi prepoznati. Zato se kao prva izmjena radi filtracija skupa za treniranje.

4.2. Selekcija uzoraka za test: mod 1

Ovdje se radi izmjena skupa za testiranje, dok skup za treniranje ostaje nepromijenjen. Prilikom stvaranja skupa za testiranje izvodi se filtriranje fraza koje sadrže riječi izvan zadanog vokabulara. Time se skup za treniranje smanjuje, no sličniji je kontekstu unutar kojeg će se model koristiti. Rezultati se mogu vidjeti u tablici 4.2.

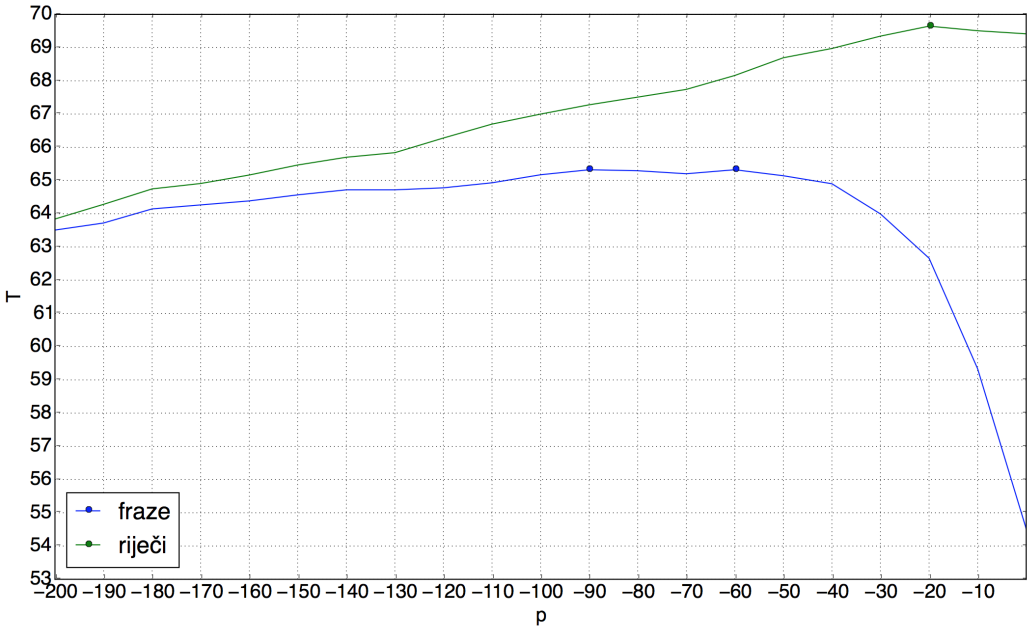
U ovoj verziji se točnost povećala za 10-ak % i kod fraza i kod riječi. Sada realnija procjena modela ukazuje na točnost modela od 69.34% na bazi riječi te 63.99% na bazi fraza što znači da model griješi otprilike pri svakom trećem točnom izgovoru fraze. To

Tablica 4.1: Rezultati osnovnog načina rada

	FRAZA				RIJEČ					
Dijete	T	H	S	N	T	H	D	S	I	N
Filip	42.06	53	73	126	46.58	68	4	74	38	146
Ilija	32.57	85	176	261	46.73	157	5	174	147	336
Ivor	45.28	24	29	53	42.19	27	2	35	17	64
Lovro	33.33	43	86	129	37.04	60	4	98	71	162
Lukas	14.55	32	188	220	18.32	48	0	214	142	262
Alan	66.67	108	54	162	66.22	149	12	64	8	225
Eli	51.11	92	88	180	56.39	150	20	96	28	266
Filip2	58.67	115	81	196	68.22	176	8	74	30	258
Filip3	65.26	124	66	190	66.32	191	15	82	12	288
Gabrijel	41.57	69	97	166	46.46	105	7	114	22	226
Jakov	62.77	118	70	188	62.87	171	14	87	31	272
Jakov2	70.18	153	65	218	74.28	231	11	69	21	311
Josip	55.12	113	92	205	57.19	159	10	109	23	278
Leon	50.28	89	88	177	56.71	131	7	93	16	231
Marko	61.35	127	80	207	63.19	182	12	94	19	288
Mihovil	62.72	106	63	169	68.10	158	11	63	12	232
Rafael	50.55	92	90	182	52.92	127	18	95	10	240
Roko	57.65	113	83	196	66.55	189	6	89	51	284
Roko2	68.07	113	53	166	70.56	163	9	59	8	231
Vid	63.84	113	64	177	64.06	164	13	79	19	256
Vinko	75.84	135	43	178	78.05	192	11	43	12	246
Zvonimir	59.46	110	75	185	70.04	187	10	70	56	267
Ukupno	**54.10**				**59.32**					

Tablica 4.2: Rezultati načina rada s filtriranjem testnih uzoraka

	FRAZA				RIJEČ					
Dijete	T	H	S	N	T	H	D	S	I	N
Filip	67.12	49	24	73	72.84	59	2	20	9	81
Ilija	52.32	79	72	151	65.05	121	3	62	53	186
Ivor	63.64	21	12	33	72.22	26	1	9	5	36
Lovro	53.73	36	31	67	66.67	52	1	25	16	78
Lukas	28.46	35	88	123	35.62	52	0	94	64	146
Alan	72.19	109	42	151	72.50	145	15	40	3	200
Eli	58.43	97	69	166	62.61	144	12	74	12	230
Filip2	65.73	117	61	178	73.01	165	6	55	11	226
Filip3	73.05	122	45	167	77.19	176	11	41	3	228
Gabrijel	46.00	69	81	150	52.91	100	8	81	16	189
Jakov	72.99	127	47	174	68.94	162	15	58	3	235
Jakov2	80.73	155	37	192	85.16	218	9	29	7	256
Josip	60.42	116	76	192	62.60	159	10	85	12	254
Leon	52.66	89	80	169	58.64	129	7	84	4	220
Marko	66.48	119	60	179	68.22	161	13	62	7	236
Mihovil	67.72	107	51	158	72.04	152	10	49	4	211
Rafael	53.18	92	81	173	55.11	124	21	80	3	225
Roko	67.90	110	52	162	82.27	167	5	31	28	203
Roko2	71.90	110	43	153	76.10	156	10	39	9	205
Vid	71.95	118	46	164	74.66	165	11	45	10	221
Vinko	80.12	133	33	166	81.82	180	10	30	5	220
Zvonimir	64.02	105	59	164	78.64	173	5	42	37	220
Ukupno	**63.99**				**69.34**					



Slika 4.1: Ovisnost penalizacije dodavanja nove riječi p i ukupne točnosti T

je nešto bolji rezultat. Gledajući rezultate djece s poremećajem iz autističnog spektra vidi se da su i ti rezultati slični ukupnoj točnosti, pa je model upotrebljiv. Ovakav način filtriranja skupa za testiranje se primjenjuje i na ostalim načinima rada.

4.3. Optimizacija parametra p: mod 2

Daljnjim promišljanjem o modelu vidljivo je da je korišten predodređen parametar penalizacije na prijelazu riječi p jednak -30 koji ne mora nužno biti najbolji za naše potrebe. Rečenice/fraze ovog sustava su veoma kratke, pa se može očekivati da bi veća penalizacija dodavanja riječi tijekom prepoznavanja mogla poboljšati sustav. Upravo to je zadatak načina rada broj 2. Parametar p varira od vrijednosti 0 (nema penalizacije) do -200 (velika penalizacija) u razmacima vrijednosti 10. Rezultati točnosti na razini riječi i fraze su vidljivi u tabličnom obliku u tablici 4.3 i u grafičkom obliku na slici 4.1. U tablici su maksimalne vrijednosti točnosti modela označene podebljanim brojkama, a na grafu su te točke označene malim krugom.

Vidljivo je da prepoznavanje na razini riječi postiže svoj maksimum za vrijednost parametra p jednakoj -20, a prepoznavanje na razini fraze postiže maksimalnu točnost kada je p jednak -60 i -90. Dvije vrijednosti maksimalne točnosti na razini fraze se razlikuju za 1 na trećoj decimali, no ta razlika je premalena da bi se smatralo da je jedan model zaista bolji od drugoga, pa se može reći da obje vrijednosti daju maksimalnu

Tablica 4.3: Rezultati variranja parametra p

p	FRAZE	RIJEČI
0	54.52	69.42
-10	59.33	69.51
-20	62.66	**69.65**
-30	63.99	69.34
-40	64.90	68.97
-50	65.14	68.70
-60	**65.32**	68.16
-70	65.20	67.74
-80	65.30	67.51
-90	**65.32**	67.28
-100	65.17	67.00
-110	64.93	66.70
-120	64.78	66.28
-130	64.72	65.84
-140	64.72	65.70
-150	64.57	65.47
-160	64.39	65.17
-170	64.27	64.91
-180	64.15	64.75
-190	63.72	64.28
-200	63.51	63.84

točnost.

Ovdje je maksimalna točnost na razini riječi jednaka 69.65%, a na razini fraze 65.32% što čini malenu razliku u odnosu na prethodne rezultate gdje je korišten p jednak -30. Ipak, u daljnjoj analizi će se koristiti optimalan parametar p za točnost na razini fraza, jer su fraze ono što je u konačnici prepoznaje. S obzirom na to da ovdje imamo dvije vrijednosti u kojima se postiže maksimum, u idućem koraku će se isprobati obje mogućnosti i izabrati konačan parametar p. Valja još napomenuti da su obje vrijednosti parametra p veće od predodređene vrijednosti zbog veoma kratkih fraza sustava.

4.4. Priznavanje različitih oblika riječi: mod 3

Kao što je objašnjeno u odjeljku 2.2.3, osobe s poremećajem iz autističnog spektra često griješe u gramatici. U ovom sustavu to ne želimo smatrati greškom, pa se uvode jednakovrijedni razredi riječi opisani i navedeni u odjeljku 3.5.4. Njihova primjena je temelj trećeg načina rada. Ocjena modela provedena je za obje vrijednosti parametra p koje maksimiziraju točnost modela, a rezultati se nalaze u tablici 4.4. Rezultati sadrže samo ocjenu modela na temelju fraza jer se u ovom načinu rada uz zamjenu riječi unutar razreda može i izostaviti neki prijedlog, pa analiza na temelju riječi više nema smisla.

Rezultati ove metode pokazuju drastično povećanje u točnosti modela. Ukupna točnost modela s p jednakim -60 je 86.90%, dok je za vrijednost -90 točnost jednaka 87.47%. Razlika između te dvije vrijednosti nije velika, no ipak će se nadalje koristiti vrijednost p jednaka -90. Točnost od 87.47% je povećanje za više od 20% u odnosu na metodu bez priznavanja jednakovrijednih riječi čija točnost iznosi 65.32%. To znači da je model u preko 20% slučajeva bio *blizu* pogađanja točnog izraza, ali je pogriješio u padežu i sl. Može se argumentirano reći da ova metoda kao takva zapravo daje prostora modelu za prihvaćanje sličnih fraza, pa su rezultati bolji. No ovakav sustav i djetetu daje slobodu da može izgovoriti gramatički nepravilne konstrukcije s obzirom na to da je model i prije radio s dosta dobrim učinkom. Osim toga, ovakav sustav bi mogao pomoći i u problemu kada dijete iskrivi zadnji glas ili zadnji dio riječi što je uočeno da postoji prilikom obrade uzoraka i kod djece s PAS i one urednog razvoja. Kako je već rečeno, ovdje greške u gramatici, a niti izgovoru zaista nisu bitne, pa se korištenje ovakvog modela testiranja koristi za daljnji razvoj.

U tablici 4.4 se još mogu pogledati rezultati točnosti kod djece s poremećajem iz autističnog spektra (prvih pet redaka) u usporedbi s djecom urednog razvoja (ostali

Tablica 4.4: Rezultati za način rada 3 (p = -60 i p= -90)

	P=-60			P=-90		
Dijete	H	N	T	H	N	T
Filip	56	73	76.71	58	73	79.45
Ilija	110	151	72.85	116	151	76.82
Ivor	27	33	81.82	27	33	81.81
Lovro	51	67	76.12	53	67	79.10
Lukas	53	123	43.09	58	123	47.15
Alan	140	151	92.72	138	151	91.39
Eli	137	166	82.53	138	166	83.13
Filip2	166	178	93.26	168	178	94.38
Filip3	154	167	92.22	153	167	91.61
Gabrijel	130	150	86.67	131	150	87.33
Jakov	152	174	87.36	154	174	88.50
Jakov2	185	192	96.35	184	192	95.83
Josip	168	192	87.50	170	192	88.54
Leon	153	169	90.53	152	169	89.94
Marko	158	179	88.27	159	179	88.82
Mihovil	144	158	91.14	142	158	89.87
Rafael	157	173	90.75	156	173	90.17
Roko	146	162	90.12	147	162	90.74
Roko2	137	153	89.54	138	153	90.19
Vid	143	164	87.20	144	164	87.80
Vinko	157	166	94.58	156	166	93.97
Zvonimir	148	164	90.24	149	164	90.85
Ukupno		3305	**86.90**		3305	**87.47**

reci). Može se vidjeti da je točnost prepoznavanja oko 80% što je sasvim prihvatljiva brojka za sustav za raspoznavanje govora. Iznimka je jedno dijete čije je prepoznavanje ispod 50% kroz cijelu izgradnju sustava. Na snimkama se može čuti da su te snimke često dosta nerazgovijetne, pa je računalnom sustavu teško prepoznati pravu izgovorenu riječ ili frazu. Razmotreno je da se ti uzorci u potpunosti izbace, no s obzirom na malen broj uključene djece s PAS odlučeno je ostaviti sve uzorke.

Zanimljivo je još napomenuti da su rezultati ostalih četvero djece s PAS obrnuto proporcionalni s njihovim stupnjem razvoja govora. Kod djeteta na međusobno najvišem stupnju razvoja govora je prepoznavanje najslabije (76.82%), dok se kod ostale djece ta brojka penje do djeteta na međusobno najnižem stupnju razvoja govora (prepoznavanje od 81.81%). Jedno moguće objašnjenje je da djeca na nižem stupnju razvoja govora češće koriste eholaličan govor (ponavljaju za terapeutom), pa se zapravo radi o njihovoj mogućnosti imitacije glasova, dok djeca koja se češće koriste spontanim govorom unose više svojih izmjena u izgovor, pa je prepoznavanje malo slabije. U svakom slučaju oba postotka su sasvim dovoljna za korištenje u računalnoj igri *Brbljalica* čiji se razvitak prati kroz iduće poglavlje.

4.5. Eliminacija lošije izgovorenih uzoraka iz skupa za treniranje: mod 4

Prilikom unosa uzoraka u sustav obavljena je i ocjena izgovora od strane stručne osobe ocjenom od 1 do 3, kao što je predstavljeno u odjeljku 3.3. S obzirom na to da su neki uzorci dosta loše izgovoreni, moguće je da su oni negativno utjecali na parametre modela i otežali prepoznavanje ostalih uzoraka. U načinu rada broj 4 se ispituje utjecaj tih loše izgovorenih uzoraka tako da se oni filtriraju prilikom izrade skupa za treniranje. Iste se snimke i dalje koriste u skupovima za testiranje jer se može očekivati da će dijete isto ponekad lošije izgovoriti frazu koju bi svejedno voljeli prepoznati. Rezultati ovog pristupa dani su u tablici 4.5.

Rezultati pokazuju ukupnu točnost od 87.87%. Razlika od prethodnog načina rada koji je davao točnost od 87.47% je oko 0.4% što znači da ovaj postupak ne doprinosi značajno ukupnoj točnosti modela. No ono što je bitno uočiti je da je u ovim rezultatima razlika točnosti prepoznavanja među djecom s PAS veća nego u prošlom primjeru. Nama bolje odgovara imati ujednačenije rezultate s obzirom na to da je ta razina točnosti prihvatljiva, pa se ovaj način kreiranja skupa za treniranje neće koristiti pri izradi završnog modela.

Tablica 4.5: Rezultati dobiveni eliminacijom loše izgovorenih uzoraka

Dijete	H	N	T
Filip	59	73	80.82
Ilija	111	151	73.51
Ivor	28	33	84.85
Lovro	52	67	77.61
Lukas	57	123	46.34
Alan	136	151	90.07
Eli	139	166	83.73
Filip2	169	178	94.94
Filip3	154	167	92.22
Gabrijel	134	150	89.33
Jakov	155	174	89.08
Jakov2	187	192	97.40
Josip	169	192	88.02
Leon	156	169	92.31
Marko	161	179	89.94
Mihovil	145	158	91.77
Rafael	156	173	90.17
Roko	148	162	91.36
Roko2	139	153	90.85
Vid	142	164	86.59
Vinko	157	166	94.58
Zvonimir	150	164	91.46
Ukupno		3305	**87.87**

5. Računalna igra *Brbljalica*

Brbljalica je računalna igra namijenjena poticanju govora kod djece s poremećajem iz autističnog spektra kao dodatak bihevioralnim tretmanima. Početni ekran igre se može vidjeti na slici 5.1. Igra se temelji na principima podučavanja diskriminativnim nalozima — PDN-ima objašnjenim u odjeljku 2.3. U ovom poglavlju će se podrobnije opisati detalji izvedbe i korištenja igre.

5.1. Način rada igre

Igra se sastoji od niza zadataka (situacija) u kojima dijete treba dati prihvatljiv govorni odgovor na vizualan i auditivni podražaj. Prođimo kroz princip igre na primjeru prvog zadatka kroz koji se potiče izgovor riječi *kuća*.

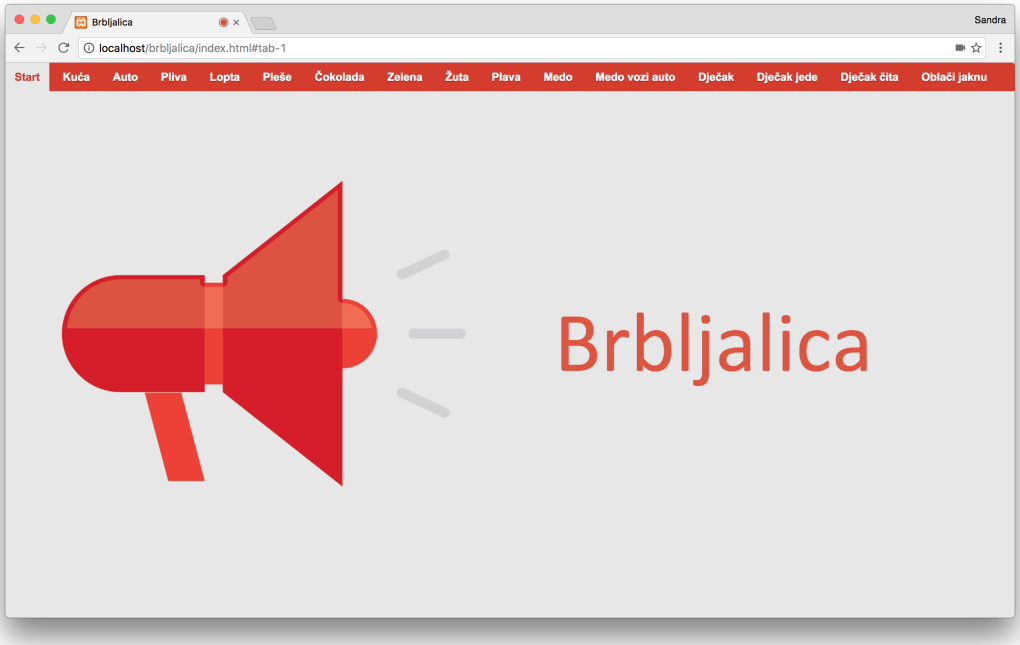
Na početku zadatka se na ekranu prikazuje vizualan podražaj u obliku slike koja predstavlja neku od unaprijed definiranih riječi s liste 3.10, u ovom slučaju je to kuća. Primjer izgleda prozora u tom trenutku je prikazan na slici 5.2. Kao auditivni podražaj upućuje se pitanje na koje je ispravan odgovor upravo ono što je na slici. U ovom slučaju, pitanje je: *Što je ovo?*. Nakon toga se snima zvuk djetetove reakcije. Ako je dijete odgovorilo točnom riječju daje se nagrada (pozitivna potkrjepa), te se prelazi na idući zadatak. U suprotnom, podražaj se ponavlja. Ako niti tada nema odgovora, onda se daje govorno rješenje koje dijete može ponoviti.

Preciznije se proces jednog zadatka može definirati ovako:

1. Prikaz vizualnog podražaja u obliku slike i puštanje auditivnog podražaja u obliku pitanja na koje je odgovor ono što je na slici.

2. Sluša se djetetov odgovor u trajanju od pet sekundi jer se odgovor unutar tog roka još može smatrati kognitivnim odgovorom na podražaj, te se vrši prepoznavanje onoga što je izrečeno.

3. Ako je dijete točno izgovorilo traženu riječ ili frazu, onda dobiva vizualnu i auditivnu nagradu. Vizualna nagrada je jedna od nekoliko mogućih slika šarenih



Slika 5.1: Početni ekran igre Brbljalica

balona, a uz nju se još dobiva pljesak i pohvala (*Bravo!*, *Odlično!* ili *Izvrsno!*). Nakon nagrade se izvodi sljedeći zadatak.

Ako dijete nije dalo odgovarajući govorni odgovor, tada se ponavljaju koraci 1 i 2, te se dobiva nagrada ako je ovog puta odgovor bio ispravan.

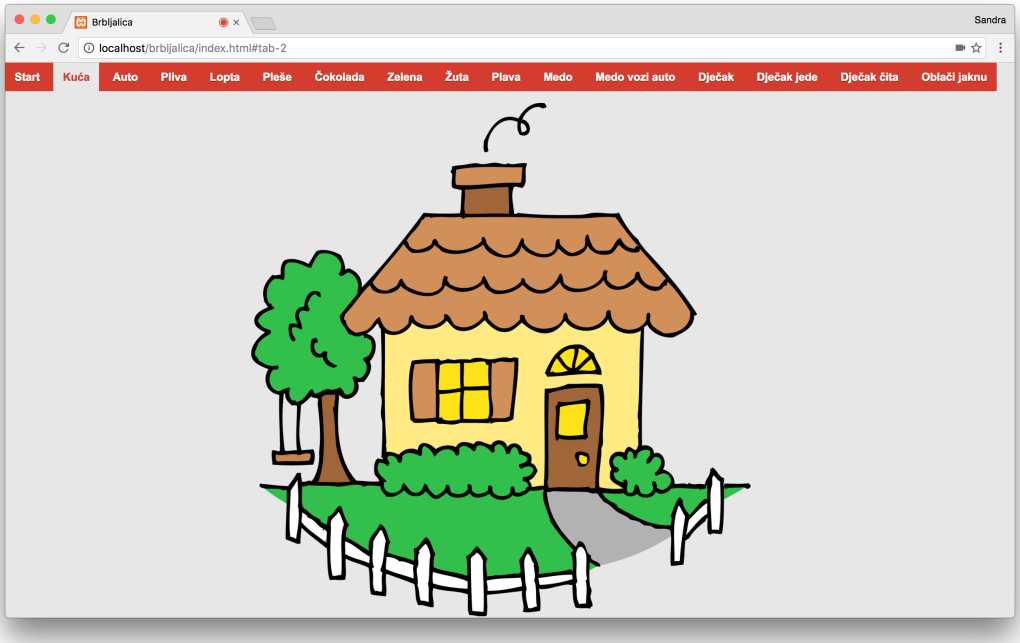
4. U slučaju da niti tada odgovor nije ispravan, djetetu se daje odgovor u obliku izraza *Ponovi za mnom* za kojim slijedi odgovor (npr. *kuća*). Ovakav način podražaja razumiju djeca s razvijenim spontanim govorom, ali i djeca koja se služe eholalijom.

5. Ponovno se snima djetetov odgovor u trajanju od pet sekundi i vrši se prepoznavanje govora.

6. Slično kao u koraku broj 3, dobiva se nagrada za ispravan odgovor, a za neispravan odgovor dolazi do izostanka nagrade i ponavljanjem koraka broj 4 i 5. U slučaju dobro izvršenog zadatka se dobiva nagrada.

7. Ako niti tada zadatak nije uspješno izvršen nagrada se preskače i prelazi se na idući zadatak.

Zadaci se izvode po redoslijedu navođenja na upravljačkoj traci na vrhu prozora: *kuća*, *auto*, *pliva*, *lopta*, *pleše*, *čokolada*, *zelena*, *žuta*, *plava*, *medo*, *medo vozi auto*,



Slika 5.2: Prikaz zadatka "kuća"

dječak, *dječak jede*, *dječak čita*, *oblači jaknu*. Ime svakog zadatka je upravo i odgovor koji se očekuje u tom zadatku. Dobro je još napomenuti da se za fraze od više riječi priznaje i izgovor njihove skraćene verzije. Tako se za *medo vozi auto* i pitanje *Što medo radi?* priznaje i *vozi auto* ili samo *vozi*. Umjesto *dječak jede* i *dječak čita* priznaju se i samo *jede* i *čita*. Umjesto *oblači jaknu* se može reći i samo *oblači se*. Naravno, uz sve ove verzije, priznaju se sve verzije ovih fraza koje se mogu dobiti korištenjem razreda riječi iz odjeljka 3.5.4. Kada su izvršeni svi zadaci igra se vraća na početnu stranicu.

5.2. Korištene tehnologije

Ova igra napravljena s namjerom korištenja u web preglednicima. Igra je testirana na pregledniku Google Chrome.

Statički dio stranice opisan je HTML kodom, stil se konfigurira CSS-om, dok se dinamika ostvaruje Javascript kodom. Za ostvarivanje tabova korišten je Javascript *plug-in* Responsive-Tabs.js [1], a za snimanje sa stranice korišten je AudioRecorder.js [2].

Serversku stranu odrađuje PHP preuzima snimku sa stranice i točan odgovor, pa

[1] https://github.com/jellekralt/Responsive-Tabs commit 3da3531c5bc61a2ad26dc64d619c4a895b8f205b
[2] https://github.com/cwilso/AudioRecorder commit 90a3a08ba4dd4e28879bb1931d9bfcedfc88a173

nakon toga poziva Python skriptu koja odrađuje posao prepoznavanja govora koristeći HTK toolkit [3]. Web server je Apache iz distribucije XAMPP.

5.3. Tehnička izvedba

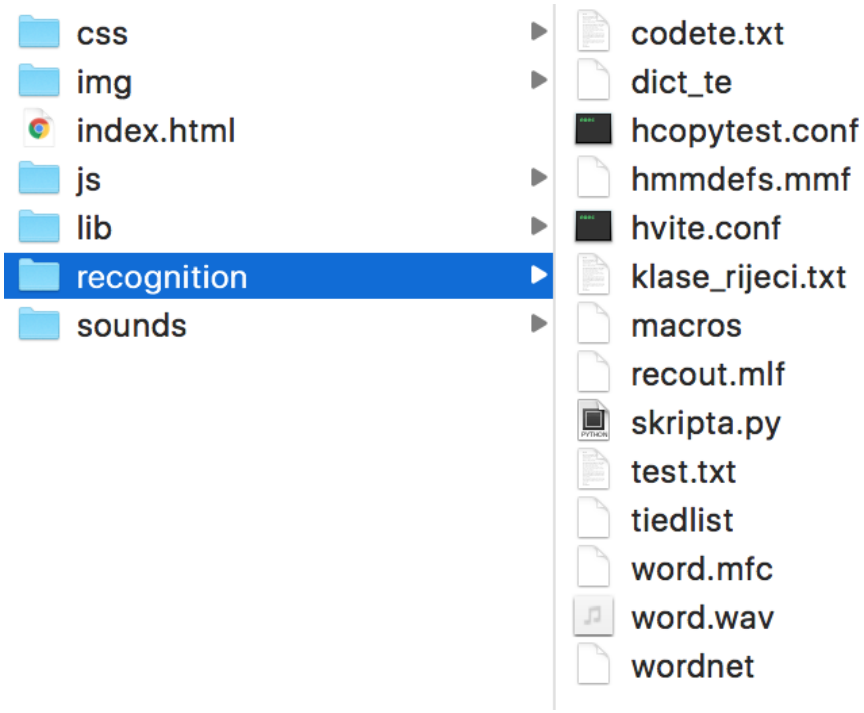
Kod XAMPP web poslužitelja se svaki projekt stavlja u pripadni subdirektorij direktorija `xamppfiles/htdocs/`. U njega se stavi i direktorij `brbljalica` sa strukturom danom na slici 5.3. Direktorij `css` sadrži sve potrebne CSS datoteke koje određuju izgled komponenti. Unutar `img` direktorija se nalaze sve slike korištene na stranici uključujući sliku početne stranice, svih zadataka i svih nagrada. Datoteka `index.html` sadrži HTML kod stranice. U direktorij `js` su stavljena oba Javascript *plug-ina* i Javascript kod za upravljanje radom igre. Direktorij `lib` sadrži skriptu za serversku stranu obrade zahtjeva za prepoznavanjem govora izrečenog unutar igre. U `recognition` direktoriju se nalaze sve datoteke potrebne za prepoznavanje govora uključujući istrenirani model, podatke za prepoznavanje, datoteku `word.wav` u kojoj se nalazi snimka snimljena tijekom igre i Python skripta `skripta.py` koja te sve podatke povezuje i vrši prepoznavanje. Konačno, direktorij `sounds` sadrži sve snimke koje se puštanju na stranici: snimke pitanja, pomoći i nagrade.

5.4. Upute za instalaciju i korištenje

Za pokretanje igre potrebno je imati instaliran neki web preglednik (preporučljivo Google Chrome), XAMPP, HTK i Python. U XAMPP postavkama `php.ini` potrebno je dozvoliti korištenje naredbe `exec` za pozivanje naredbi iz ljuske tako da se postavi `safe_mode_exec_dir=Off`. Ukoliko to već nije tako, potrebno računalnom korisniku koji izvodi kod stranice (u slučaju macOS-a je to `daemon`) dati pravo pristupa datotekama iz direktorija `brbljalica/recognition`.

Sada se za pokretanje igre u web pregledniku može otvara adresa `localhost/brbljalica/index.html`. Potrebno je još jedino prilikom prvog otvaranja stranice web pregledniku odobriti audio snimanje na stranici. Za početak igre treba kliknuti na bilo koji tab zadatka, a od tog trenutka na dalje ih igra vrti sama. Za ponavljanje igre može se ponovno učitati početna stranica, a zatvaranje igre se postiže gašenjem web preglednika.

[3] http://htk.eng.cam.ac.uk/ verzija 3.4.1.



Slika 5.3: Struktura direktorija `brbljalica`

6. Zaključak

Ovaj diplomski rad bavi se tehničkom izvedbom sustava za automatsko raspoznavanje i poticanje govora kod djece s poremećajem iz autističnog spektra. Razvitak govora kod djece s PAS je iznimno bitan jer je upravo to jedan od ključnih pokazatelja kasnije veće kvalitete života.

U sklopu rada je sakupljen materijal za izgradnju akustičke baze govora dječaka u dobi od 5 do 7 godina s poremećajem iz autističnog spektra i onih urednog razvoja. Snimke su obrađene te su za izgradnju modela korištene njihove Mel-frekvencijske značajke. Akustički model se temelji na skrivenim Markovljevim modelima, a za njegovu izgradnju je korištena automatizacija alata HTK iz Dropuljić (2008) koja prepoznaje slijedni govor korištenjem modela trifona. S obzirom na to da osobe s PAS imaju problema s gramatičkim konstrukcijama, u ovom radu je proces prepoznavanja riječi obogaćen uvođenjem razreda jednakovrijednih riječi koje se sastoje od svih padeža neke riječi i drugih čestih oblika kako bi se dozvolilo prepoznavanje gramatički neispravnih, ali logički jasnih izraza i tako olakšala komunikacija. Konačni akustički model hrvatskog jezika prepoznaje govor djece iz ciljane skupine uz točnost od 87.47%.

Razvijena je i računalna igra *Brbljalica* koja djecu s PAS kroz igru nastoji potaknuti na govor i savladavanje jednostavnih fraza. Njen glavni element je upravo model za raspoznavanje govora. Igru je moguće pokrenuti u Web pregledniku, a testirana je na pregledniku Google Chrome. Sama igra je napravljena prema bihevioralnim principima učenja korištenjem pozitivnog potkrjepljenja (nagrade) za svaku dobro izgovorenu i upotrebljenu riječ ili frazu. Tijek igre kreiran je na bazi strukturirane bihevioralne intervencije — sastavljena je od niza unaprijed definiranih zadataka čiji je cilj potaknuti dijete na izgovaranje određene fraze ili rečenice. Ako dijete uspješno izvrši zadatak ono dobije nagradu u obliku slike šarenih balona, pljeska i pohvale.

LITERATURA

Američka Psihijatrijska Udruga. *DSM-5 Dijagnostički i statistički priručnik za duševne poremećaje*. urednici hrvatskog izdanja: Vlado Jukić i Goran Arbanas, Slap, Jastrebarsko, 2014.

Donald M Baer, Robert F Peterson, i James A Sherman. The development of imitation by reinforcing behavioral similarity to a model1. *Journal of the Experimental analysis of Behavior*, 10(5):405–416, 1967.

Z Bujas-Petković, J Frey Škrinjar, D Hranilović, B Divčić, i J Stošić. Poremećaji autističnog spektra. *Školska knjiga, Zagreb*, 2010.

Suniti Chakrabarti i Eric Fombonne. Pervasive developmental disorders in preschool children. *Jama*, 285(24):3093–3099, 2001.

Lisa A Croen, Judith K Grether, Jenny Hoogstrate, i Steve Selvin. The changing prevalence of autism in california. *Journal of autism and developmental disorders*, 32 (3):207–215, 2002.

Branimir Dropuljić. Development of acoustic and lexical model for automatic speech recognition for croatian language. Magistarski rad, Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, 2008.

Mark Gales i Steve Young. The application of hidden markov models in speech recognition. *Foundations and trends in signal processing*, 1(3):195–304, 2008.

Christopher Gillberg. Outcome in autism and autistic-like conditions. *Journal of the American Academy of Child & Adolescent Psychiatry*, 30(3):375–382, 1991.

Alexandra M Head, Jane A McGillivray, i Mark A Stokes. Gender differences in emotionality and sociability in children with autism spectrum disorders. *Molecular autism*, 5(1):19, 2014.

Brooke Ingersoll i Laura Schreibman. Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *Journal of autism and developmental disorders*, 36 (4):487, 2006.

Žarka Klopotan, Radmila Amanović, Umićević Ljiljana, Alen Conjar, Vedran Mornar, Ivica Botički, Ivan Lučin, Mladen Subotić, i Nikola Predovan. Komunikator system for people with the autistic spectrum disorder. U *10. Kongres edukacijskih rehabilitatora*, 2014.

V Lotter. Follow-up studies. U *Autism*, stranice 475–495. Springer, 1978.

Davor Petrinović. Digitalna obradba govora (interna zavodska skripta). unpublished internal learning material, 2010.

Barry M Prizant, Amy M Wetherby, Emily Rubin, i Amy C Laurent. The scerts model: A transactional, family-centered approach to enhancing communication and socioemotional abilities of children with autism spectrum disorder. *Infants & Young Children*, 16(4):296–316, 2003.

Laura Schreibman i Brooke Ingersoll. Behavioral interventions to promote learning in individuals with autism. U *Handbook of autism and pervasive developmental disorders (3rd. Edition)*, stranice 882–896, Hoboken, NJ, 2005. John Wiley & Sons, Inc.

Sanja Šimleša, Cepanec Maja, Damjan Miklić, Frano Petric, i Zdenko Kovačić. Can humanoid robots be used in the assessement of autism spectrum disorder? U *XI Autism-Europe International Congress*, 2016.

Tristram Smith, Annette D Groen, i Jacqueline W Wynn. Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, 105(4):269–285, 2000.

Steve J Young i Sj Young. *The HTK hidden Markov model toolkit: Design and philosophy*. University of Cambridge, Department of Engineering, 1993.

Dong Yu i Li Deng. *Automatic speech recognition: A deep learning approach*. Springer, 2014.

Automatski sustav za poboljšavanje izgovora

Sažetak

U sklopu ovog diplomskog rada napravljen je automatski sustav za raspoznavanje i poticanje govora. Cilj sustava je potaknuti razvijanje govora kod dječaka s poremećajima iz autističnog spektra u dobi od pet do sedam godina. Sustav je ostvaren u obliku računalne igre napravljene prema bihevioralnim principima učenja, a njena glavna komponenta je sustav za automatsko prepoznavanje govora na hrvatskom jeziku. Razvijeni sustav prepoznaje govor djece iz ciljane skupine uz točnost od 87.47%.

Ključne riječi: Raspoznavanje govora, poremećaji autističnog spektra, Mel-frekvencijski kepstar, skriveni Markovljevi modeli, računalna igra.

Automatic System for Pronunciation Improvement

Abstract

The goal of this Master's thesis is to build an automatic system for speech recognition and speech encouragement. It is meant to help boys with autism spectrum disorder between the age of five and seven years old to improve their speech abilities. The system is in the form of a computer game which follows behavioral learning principles. The main component of the game is the speech recognition system for Croatian language. The accuracy of the speech recognition model within the target user group is 87.47%.

Keywords: Speech recognition, Autism Spectrum Disorder, Mel-frequency cepstrum, Hidden Markov Models, computer game.