

Fakultet elektrotehnike i računarstva
Zagreb, Unska 3

Zavod za elektroničke sustave
i obradu informacija

Alati za fonetsku analizu i sintezu

Igor Cigrovski
0036378101

Uvod

U slijedećem tekstu biti će dat kratak pregled razvoja i problema sa kojima je suočena tehnologija govora, objašnjen termin govornog koda koji bi, jednom kada bude nađen, trebao omogućiti istu razinu komunikacije sa strojevima kakvu danas imamo međusobno sa ljudima. Također je dat kratak pregled i opis tri programa za fonetsku sintezu i analizu govora koji predstavljaju trenutnu razinu napretka na ovom području.

Govorni kod i trenutno stanje razvoja tehnologije govora

Tehnologija govora je omogućila mnoge važne alate za aplikacije koje ostvaruju komunikaciju između ljudi i strojeva, te se očekuje i njezin daljnji napredak. Međutim počinju se pojavljivati i prvi problemi: simbioza između tehnologije i nauke koja je do sada gurala naprijed počinje se polako cijepati. Tehnologija govora je jako ovisna o statističkim alatima i velikim bazama podataka, dok se fonetska istraživanja usredotočuju na usko definirane probleme i abstraktna pitanja nepovezana sa problemom nalaženja „koda ljudskog govora“.

Termin „govorni kod“ ili „fonetski kod“ odnosi se na znanje kako se lingvistički definirane jedinice realiziraju u procesu govora. Ta znanja bi se zatim mogla upotrijebiti u programima za sintezu teksta u govor i obrnuto, za strategije prepoznavanja govora preko fonetske analize obilježja govornog vala.

Glavno pitanje ovog pristupa je na koji način su jedinice izgovorene poruke u govornom kanalu kodirane u akustični signal i zatim u slušnom kanalu dekodirane. Govorni signal predstaviti ćemo govornim valom koji možemo zabilježiti pomoću oscilografa i spektrografa. Na taj način bit će maksimalno sačuvan fizički aspekt govora, uz gubitak podataka o facijalnoj mimici i govoru tijela. Uzorci govornog vala su izrazito složeni i nemoguće ih je analizirati bez povezivanja sa artikulacijskom interpretacijom. Artikulacijska interpretacija je ključ razumjevanja govornog koda.

Prilikom analize govornog koda suočeni smo s dvije različite komunikacijske situacije: jedna sa čovjekom kao slušateljem i druga sa čovjekom ili računalom koje čita

spektografski prikaz nepoznatog teksta. Prilikom njihove analize moramo naročito obratiti pažnju na slušatelja, preciznije na ono što on očekuje da će biti izgovoreno. Vrlo često čujemo samo ono što očekujemo da ćemo čuti i ako na to ne obratimo pažnju ne možemo ostvariti napredak.

Trenutno su računala u stanju pretvoriti bilo kakav tekst u običnom zapisu u razmjerno razumljiv sintetički govor. Takve aplikacije primjenjuju se u komunikaciji slijepih osoba sa računalima, kao pomoć djeci sa poteškoćama u čitanju i pisanju i slično. Međutim ako želimo da računala nas razumiju nailazimo na probleme. Dosadašnji napredak ograničio je prepoznavanje govora na razinu pojedinih riječi i jednostavnijih rečenica s tim da se prepoznavanje vrši uspoređivanjem s prethodno pohranjenim uzorcima tih istih riječi i rečenica. To omogućuje vrlo ograničenu komunikaciju sa računalom (jednostavne naredbe) ili se može koristiti u sigurnosti u sklopu provjere identiteta.

Fonetski alati

ACTOR

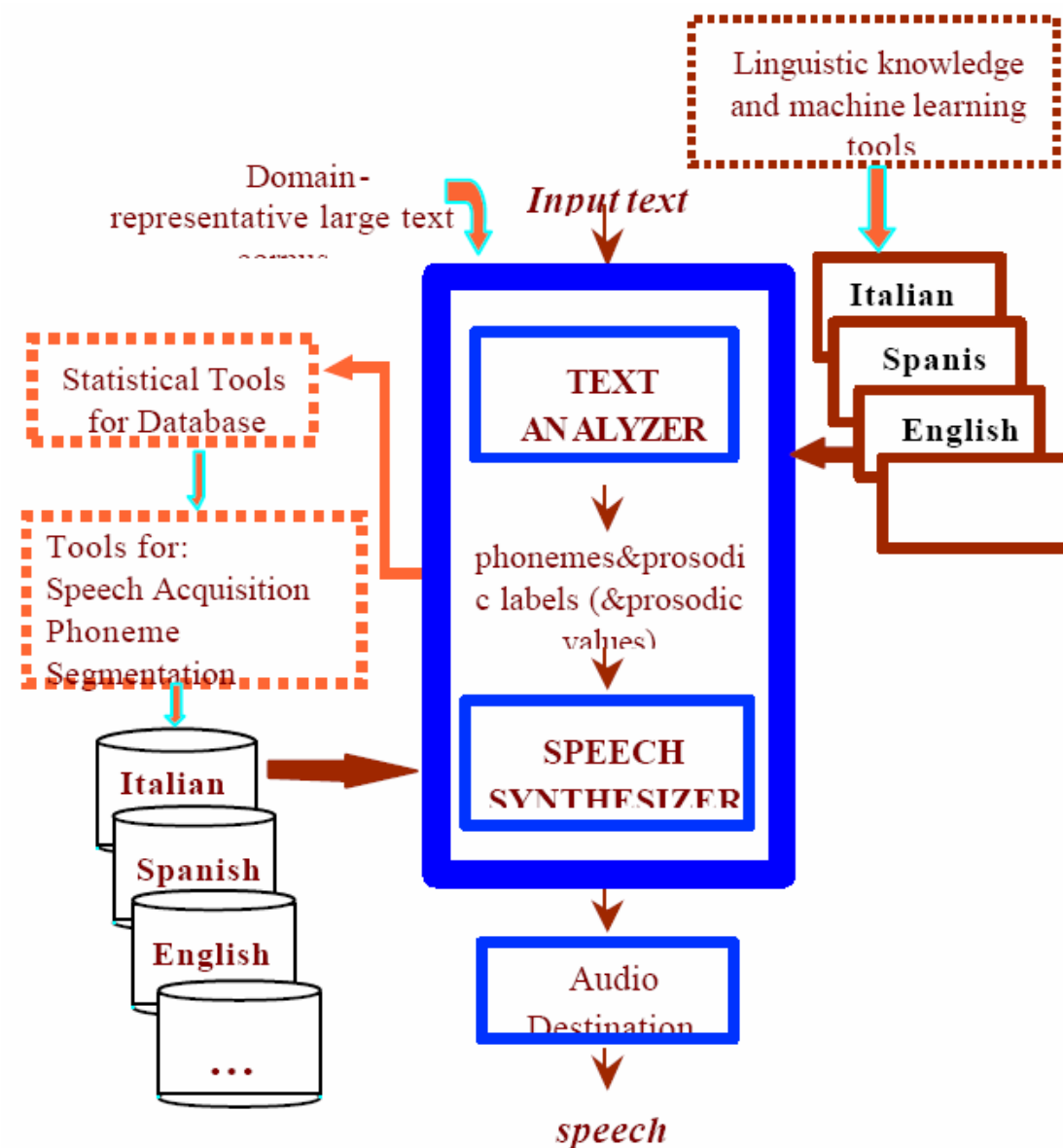
ACTOR je komercijalni multijezični/multiglasovni sintetizator teksta u govor. Radi na principu povezivanja govornih segmenata izdvojenih direktno iz glasa govornika, pohranjenim u eksternim multijezičnim bazama podataka. Performanse prilagođuje mogućnostima sistema na koji je instaliran: veće baze podataka zahtjevaju jača računala ali i omogućuju bolju kvalitetu govora i mogućnost odabira različitih glasova (muški, ženski, boja glasa i slično). Dosta je fleksibilan i omogućuje nadogradnju tako da se mogu implementirati i novi jezici. Trenutno su podržani slijedeći jezici: talijanski, španjolski, meksički španjolski, britanski i američki engleski, francuski, njemački, portugalski i brazilski portugalski, dok su argentinski španjolski i grčki još u razvoju.

Građa:

Značajno obilježje ACTOR-a sastoji se u tome da se jezično ovisni podaci drže što dalje od centralnih algoritama. Time se postiže nezavisnost programa o jeziku: prilikom promjene jezika potrebno je samo nadodati bazu podataka i program nastavlja svoju funkciju na drugom jeziku bez potrebe za dodatnim preinakama.

Građa programa prikazana je slikom 1. Crveno uokvireni dio predstavlja centralni algoritam sa bazama znanja i tokom podataka od ulaznog teksta do izlaznog govora. Oko centralnog dijela nalazi se razvojno okruženje i moduli sa podacima o pojedinim jezicima

Centralni algoritam sastoji se od dva modula: Analizatora teksta i Sintetizatora govora. Analizator teksta pretvara ulazni tekst u detaljnu fonetsku i prozadičnu reprezentaciju. Oslanja se na jezično-ovisno znanje, leksikone i pravila definirana od strane stručnjaka. Sintetizator govora pretvara apstraktni fonetski i prozadični zapis u uzorke signala koji se zatim pretvaraju u govor. Uzorke dobija od akustičnih rječnika, od kojih svaki predstavlja različiti glas. Akustični rječnici su razvijeni u laboratorijskim uvjetima tako da predstavljaju visoku fonetsku i prozadičnu pokrivenost određenog jezika.



Slika 1.

ACTOR zaključak

ACTOR pripada novoj generaciji konvertera teksta u govor sa kvalitetom govora koje se približava ljudskom. Najznačajnije mu je svojstvo njegova fleksibilnost, odnosno mogućnost prebacivanja između različitih glasova, jezika, leksikona i akustičnih baza podataka.

CSLU alat

CSLU je skraćenica od centra za razumjevanje govornih jezika (Center for Spoken Language Understanding). Centar se bavi integriranjem najnovijih govornih tehnologija u prijenosno lako shvatljivo i upotrebljivo software-sko okruženje. Najnovije razvijeni software nazvan je CSLU alat. CSLU alat objedinjuje materijale za učenje, autorske alate i osnovne tehnologije: prepoznavanje govora, sinteza teksta u govor, facijalna animacija i čitanje govora. Alatt je dizajniran da podržava razvojne i edukacijske aktivnosti vezane uz govorne jezike i sučelju između ljudi i računala.

Raspoznavanje govora:

Alat podržava nekoliko pristupa problemu raspoznavanja govora: umjetna neuronska mreža (ANN) i skriveni Markov model (HMM) zajedno sa njihovim tutorialima i alatima za treniranje novih ANN i HMM raspoznavatelja. Uključuje mogućnost samostalnog prikupljanja uzoraka a točnost mu se kreće oko 70%.

Sinteza govora:

Alat koristi sistem sinteze teksta u govor nazvanog Festival i razvijenog na Edinburškom sveučilištu. CSLU je dodatno razvio „plug-in“ komponentu za waveform sintezu koja uključuje 6 različitih muških i ženskih glasova na američkom engleskom i meksičkom španjolskom. Festival omogućuje potpuno okruženje za učenje istraživanje i razvoj sintetičkog govora. Uključuje module za normalizaciju teksta (prepoznavanje skraćenica) i pretvaranje teksta u fonetske dijelove prikladnog trajanja i intonacije.

Facijalna animacija:

Alat koristi Baldi, animiranu trodimenzionalnu glavu razvijenu na kalifornijskom sveučilištu u Santa Cruzu. Baldi je sposoban, unutar predviđenog jezika, za automatsku sinhronizaciju prirodnog ili sintetskog govorasa realističnim usnicama, jezikom, ustima i

facijalnim pokretima što ga čini moćnim alatom za učenje pravilnog govora i izgovora. Lice se može po volji okretati i promatrati iz različitih kuteva, a i postati prozirno (slika 2.) tako da možemo promatrati pokrete čeljusti i jezika pri govoru. Zanimljivo je da može prikazivati i emocije poput veselja, srdžbe, iznenađenja, tuge, gađenja i straha te se na taj način dodatno približiti ravnopravnoj komunikaciji sa ljudima.



Slika 2. Baldi

PRAAT

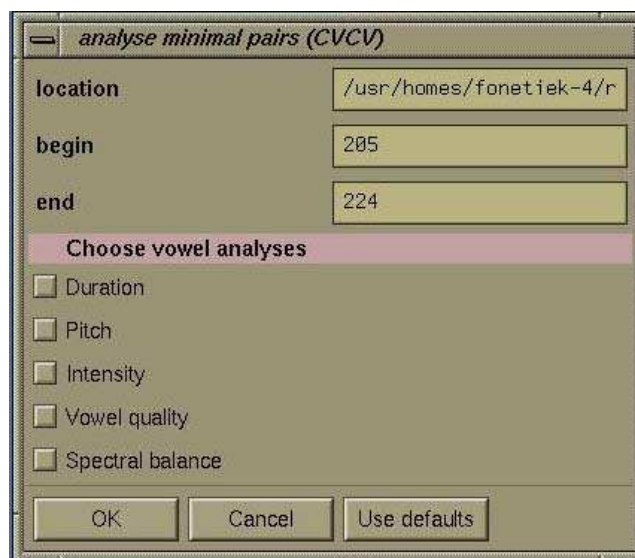
Općenito

Praat je alat za obavljanje fonetske analize i sinteze putem računala. Dizajnirali su ga Paul Boersma i David Weenink na odjelu za fonetiku Amsterdamskog sveučilišta.). Program je shareware tipa i može se naći na stranici: <http://fonsg3.let.uva.nl/praat>.

Konstantno se usavršava, nova verzija se objavljuje skoro svaki tjedan. Dizajniran je za rad na različitim platformama i operativnim sistemima: Unix, Macintosh, Windows 95/NT. Trenutno se radi na verziji 4.2.34. Koristan je za širok raspon manipulacija i analiza.

Akustična analiza

Praat je program koji podržava velik broj različitih pristupa analizi govornih signala. Podržane su slijedeći načini analize: pitch, intensity, formants i spectrograms i spectral balance. Kada radimo sa uniformnim skupinama podataka korisnik može isprogramirati ponavljanje mjernih procedura upotrebom skripti. Skripte je lako naučiti koristiti zbog vrlo detaljnog user manuala. Na slici 3. je prikazano korisničko sučelje za analizu akustičnih korelacija između izgovorenog i pohranjenog govora



Slika 3.

Označavanje teksta upotrebom Praata

Osnovu dobrog dijela lingvističkog posla čine transkripti govora različite razine detaljnosti. Praat nudi dva načina obilježavanja načina izgovora riječi i rečenica: IntervalTier i PointTier. IntervalTier se sastoji od oznaka koje imaju svoje trajanje sa definiranom početnom i završnom točkom. PointTier povezuje oznake sa pojedinim vremenskim točkama, IPA simboli se mogu upotrebljavati u oba slučaja.

Zaključak

U ovom seminaru prikazana su tri različita govorna alata različitih namjena: ACTOR, CSLU i Praat. ACTOR je vrlo fleksibilan alat za sintezu teksta u govor kojemu građa omogućuje laku nadogradnju novih jezika. Podržava izgovor na različitim jezicima, a podaci o njima dobivaju se iz akustičnih rječnika nastalih u laboratorijskim uvjetima. CSLU osim izgovora ima i prepoznavanje govora, podržava znatno manje jezika, ali uključuje animiranu trodimenzionalnu glavu idealnu za vježbanje izgovora. Praat je besplatni program namijenjen komunikaciji računala i čovjeka i kao pomoć u lingvističkoj znanosti gdje može svojim raznim opcijama znatno olakšati i skratiti vrijeme rada. Prepoznavanje govora mu je ograničeno na riječi i fraze.

Literatura:

1. Gunnar Fant: On the speech code
2. Silvia Quazza, Laura Donetti, Loreta Moisa, Pier Luigi Salza: ACTOR®: A MULTILINGUAL UNIT-SELECTION SPEECH SYNTHESIS SYSTEM
3. Stephen Sutton, Ronald Cole, Jacques de Villiers, Johan Schalkwyk, Pieter Vermeulen¹, Mike Macon, Yonghong Yan, Ed Kaiser, Brian Rundle, Khaldoun Shobaki, Paul Hosom, Alex Kain, Johan Wouters, Dominic Massaro, Michael Cohen: UNIVERSAL SPEECH TOOLS:THE CSLU TOOLKIT
4. Paul Boersma and David Weenink: Praat: doing phonetics by computer