

SINTEZA HRVATSKOG GOVORA UPORABOM SUSTAVA FESTIVAL

Renato Šoić

Fakultet elektrotehnike i računarstva

renato.soic@fer.hr

prof. dr. sc. Šandor Dembitz

Fakultet elektrotehnike i računarstva

sandor.dembitz@fer.hr

Sažetak

Tema ovog rada je izgradnja sustava za sintezu govora na hrvatskom jeziku korištenjem sustava Festival i njegovih pomoćnih sustava. Cilj je postizanje sinteze razumljivoga govora na hrvatskom jeziku. U radu su opisane značajke odabrane metode sinteze govora i koraci u postupku stvaranja novoga glasa. Rezultat rada je analiziran i definirana su moguća poboljšanja kvalitete sintetiziranoga govora.

1. Uvod

Razvoj sustava za sintezu govora je vrlo složen i skup posao. Za izgradnju potpunog, funkcionalnog i kvalitetnog sustava potreban je prilično velik broj stručnjaka iz različitih domena, financijska potpora i mnogo vremena. Zbog toga su takvi sustavi razvijeni uglavnom za centralne jezike, kod kojih je to ekonomski isplativo.

2. Festival i pomoćni alati

Festival je slobodan sustav otvorenog koda koji omogućuje razvoj strojne tvorbe govora za bilo koji jezik. Dobro je dokumentiran i vrlo popularan, zato se nametnuo kao prvi izbor kod razvoja sustava za sintezu hrvatskoga govora. Za razvoj novoga glasa treba koristiti pomoćne alate sadržane u sustavu Festvox koji je razvijen kao dodatak Festivalu.

U okviru ovog rada korištena je metoda sinteze govora Statistical Parametric Synthesis. Ona pripada skupini statističkih metoda sinteze govora. Temelji se na izgradnji skrivenih Markovljevih modela iz parametara dobivenih analizom snimljenoga govora i pripadnih transkripata [2].

3. Rad sa sustavom Festival

Stvaranje glasa u sustavu Festival sastoji se od nekoliko koraka. Potrebno je definirati skup fonema u jeziku, pravila obrade tokena (brojeva, kratica, itd.), stvoriti leksikon, definirati pravila LTS (Letter-To-Sound), stvoriti modele fraziranja, intonacije i trajanja te sintetizirati

valni oblik. U okviru ovog rada nisu izrađeni svi koraci. Modul za normalizaciju teksta nije razvijen zato što je složen i nije ga praktično razvijati u sklopu Festivala. Leksikon i pravila LTS za hrvatski jezik nije potrebno definirati zato što se fonemski slijed poklapa s grafemskim.

3.1. Stvaranje korpusa

Potrebno je stvoriti korpus koji će se koristiti prilikom izgradnje novoga glasa. Prikupljeni tekst ne bi smio sadržavati znakove koji nisu u hrvatskom jeziku, a rečenice bi trebale biti razumne duljine. Tekstualni korpus sastojao se od otprilike 4500 rečenica, 110000 riječi.

Festival sadrži alate za postupak analize tekstualnog korpusa: iz cijelog korpusa odabere se skup rečenica s najboljom fonetskom strukturom. Te rečenice je potrebno snimiti i one predstavljaju govorni korpus. Snimljeno je 270 rečenica, 17 minuta govora.

3.2. Definiranje skupa fonema

Osnova postupka stvaranja novoga glasa jest definiranje skupa fonema koji se pojavljuju u jeziku. Za svaki fonem potrebno je definirati značajke poput tipa fonema, duljine, mjesta tvorbe, naglašavanja, itd.

3.3. Automatski postupci

Označavanje podataka odnosi se na označavanje snimljenih zvučnih datoteka i ekstrakciju potrebnih parametara. Postupak izgradnje strukture izjava za svaku snimljenu rečenicu stvara strukturu kao da je rečenica pravilno sintetizirana. Rečenice se analiziraju i kako bi se izvukli parametri fundamentalne frekvencije i mel-cepstruma. Zatim se gradi parametarski model koji se koristi kod postupka sinteze govora [1].

4. Zaključak

Rezultat ovog rada je sustav za strojnu tvorbu govora na hrvatskom jeziku. Sintetizirani govor nije prirodan, ali zadovoljava kriterij razumljivosti. Kvalitetu govora moguće je poboljšati izradom bolje baze snimljenoga govora i ručnim označavanjem snimaka. Ipak, za spomenuta poboljšanja potrebno je utrošiti mnogo vremena zbog složenosti problema.

Literatura

[1] Black, A. W.; Lenzo, K.: Building Synthetic Voices, <http://festvox.org/bsv/>

[2] Black, A. W.; Zen, H.; Tokuda, K.: Statistical Parametric Speech Synthesis, *Speech Communication*, 51(11), studeni 2009., 1039-1064