

Adaptivni sistem za detekciju osnovne učestanosti govornog signala zasnovan na AMDF metodi

Knežević D., Pekar D., Milošević I., and Vujnović Sedlar N.

Sadržaj — U ovom radu je opisan način adaptivne procene osnovne učestanosti govornog signala uz primenu AMDF algoritma. Metod opisan u ovom radu značajno unapređuje rezultate originalnog algoritma, a zasnovan je na poznavanju vrsta grešaka koje se javljaju pri upotrebi AMDF metode za detekciju osnovne učestanosti govornog signala.

Ključne reči — Adaptivno, AMDF, govor, osnovna učestanost.

I. UVOD

ALGORITMI za detekciju osnovne učestanosti pseudoperiodičnog zvučnog signala imaju dosta različitih primena (u fonetici, u analizi muzičkih signala itd.). Ni jedan od postojećih algoritama nije savršen, a mogu značajno da se razlikuju po efikasnosti i kompleksnosti. Svaka od primena ima različite zahteve po pitanju kvaliteta i performansi sistema tako da je razvijeno više algoritama za detekciju osnovne učestanosti. Ovi algoritmi se mogu grupisati u dve osnovne kategorije: algoritmi sa analizom u vremenskom domenu i algoritmi sa analizom u frekvenциjskom domenu [1], [3], [4].

- a) Algoritmi sa analizom u vremenskom domenu
 1. Autokorelaciona metoda
 - mala osetljivost na šum
 - računski zahtevna
 2. Metoda preseka sa nulama
 - jednostavna i računski jeftina
 - neprecizna, osetljiva na šum
 3. Maximum likelihood
 - precizan
 - računski zahtevan
 4. AMDF (Average Magnitude Difference Function)

Ovaj rad je realizovan u okviru istraživačkog projekta koga finansira Ministarstvo za nauku države Srbije pod nazivom "Razvoj govornih tehnologija za srpski jezik i primena u "Telekomu Srbija" (TR-6144A).

Dragan D. Knežević, Alfanum d.o.o. u Novom Sadu, Trg Dositeja Obradovića 6, 21000 Novi Sad, Srbija (telefon: 381-21-4750080; faks: 381-21-4752997; e-mail: dragan.knezevic@alfanum.co.yu).

Darko J. Pekar, Alfanum d.o.o. u Novom Sadu, Trg Dositeja Obradovića 6, 21000 Novi Sad, Srbija (telefon: 381-21-4750080; faks: 381-21-4750080; e-mail: darko.pekar@alfanum.co.yu).

Ivan I. Milošević, Alfanum d.o.o. u Novom Sadu, Trg Dositeja Obradovića 6, 21000 Novi Sad, Srbija (telefon: 381-21-4750080; faks: 381-21-4750080; e-mail: ivan.milosevic@alfanum.co.yu).

Nataša M. Vujnović Sedlar, Fakultet tehničkih nauka u Novom Sadu, Trg Dositeja Obradovića 6, 21000 Novi Sad, Srbija (telefon: 381-21-4852521; faks: 381-21-4752997; e-mail: natasav@uns.ns.ac.yu).

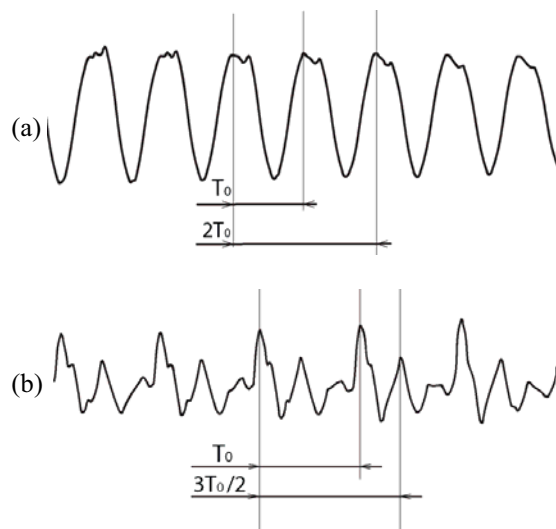
- precizan
- relativno brz metod

- b) Algoritmi sa analizom u frekvenциjskom domenu
 1. Proizvod harmonijskih spektara (HPS, Harmonic Product Spectrum)
 - računski nezahtevna, relativno otporna na šum
 - efikasnost zavisna od frekventnog opsega
 2. Kepstralna metoda
 - brza (zasnovana na FFT)

Ipak, nijedan od algoritama ne daje idealne rezultate. Određene vrste pseudoperiodičnih signala predstavljaju skoro nepremostivu prepreku za sve postojeće algoritme za detekciju osnovne učestanosti.

II. GREŠKE PRI UPOTREBI AMDF ALGORITMA

Najčešće greške koje se javljaju pri određivanju maksimuma signala kod AMDF metode [1],[2] su greška dvostruke periode koja se dešava kada se analizira signal koji ima značajno višu f_0 od one za koju je sistem prilagođen i greška kod signala sa izraženim drugim ili trećim harmonikom (Slika 1).



Sl. 1. Najčešće greške kod AMDF metode:

- a) Dvostruka vrednost T_0
- b) Izražen treći harmonik

U ovom radu korišćen je AMDF metod koji je modifikovan tako da daje što bolje rezultate pri analizi

govornog signala.

Predstavićemo metodu za automatizovanu ocenu pouzdanosti dobijenih markera i način na koji se ta ocena može iskoristiti za automatsko prilagođavanje sistema na konkretan signal.

III. ZAHTEVI SPECIFIČNI ZA SINTEZU GOVORA

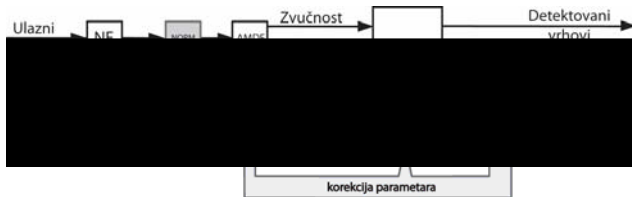
Za potrebe sinteze govora procena osnovne učestanosti ne mora da bude brza (pošto može da se radi samo jednom pri analizi govorne baze), ali je od izuzetnog značaja da bude precizna. Loše procenjena osnovna učestanost govora koji se koristi za izgradnju govorne baze izaziva izuzetno loše rezultate sinteze.

Zbog toga smo pažnju posvetili kvalitetu rezultata, a zanemarili smo računsku kompleksnost.

IV. STRUKTURA ADAPTIVNOG SISTEMA ZA DETEKCIJU

Originalni sistem za detekciju osnovne učestanosti zasnovan na AMDF metodi unapređen je (Slika 2) tako što je:

1. Analizirani signal dodatno obrađen pre ulaska u AMDF blok.
2. Uvedena dodatna provera izlaznog signala koja je moguća zahvaljujući znanju o vrstama grešaka koje se javljaju kod ove metode. Na osnovu podataka o greški sistem se modifikuje tako da se ta greška umanjí.



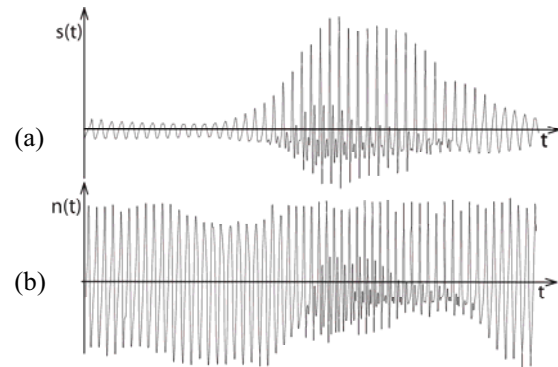
Sl. 2. Blok dijagram predloženog unapređenog sistema za detekciju osnovne učestanosti govornog signala

A. Obrada signala na ulazu u AMDF blok

Uticaj izraženog višeg harmonika može se jednostavno ukloniti filtriranjem ulaznog signala. Međutim, uticaj izraženog drugog ili trećeg harmonika se na ovaj način neprimetno ublažuje jer filter mora biti projektovan tako da dozvoli oscilacije f_0 koje su prirodne u govoru pa se ne sme dati suviše niska propusna učestanost filtra. Greške izazvane drugim ili trećim harmonikom redukovane su dodatnom automatskom proverom rezultata.

Pokazalo se, međutim, da još jedna pojava kod govornih signala izaziva lošu procenu osnovne učestanosti osnovnom AMDF metodom. Reč je o pojavi nagle promene snage govornog signala. Na mestima gde se dešava takva promena osnovni AMDF sistem, teže je detektovao sličnost između susednih perioda, i zbog toga se dešavalo da se pogrešno proceni f_0 . Uticaj prelaznih oblasti između zvučnih delova signala sa značajno različitim snagama potisnut je primenom normalizacije signala na ulazu u AMDF blok.

Normalizacija je realizovana tako što je prvo detektovana energetska envelope signala, a zatim je svaki odmerak filtriranog signala podeljen sa interpoliranom vrednošću energetske envelope za dati odmerak. Energetska envelope je računata upotrebom prozorske funkcije čija je širina dvostruko veća od maksimalne periode posmatranog signala, a ne na nivou čitavog signala.



Sl. 3. Filtriran (a) i normalizovan signal (b).
Pojedinačno izgovoreno „v“

Na ovaj način dobijen je signal ujednačene amplitude, a bez značajnih izobličenja koja bi uticala na rezultat detekcije.

B. Analiza izlaza bloka za detekciju vrhova na osnovu poznatih najčešćih vrsta greške.

Zahvaljujući poznatim najčešćim vrstama greške, u mogućnosti smo da testiramo rezultat i utvrdimo da li se neka od grešaka javila u toku detekcije.

1. Prva vrsta greške je greška $2T_0$. Za ovu grešku je karakteristično da se između susednih detektovanih vrhova nalazi i jedan koji nije detektovan. Prisustvo ove greške može se automatski utvrditi pronalaženjem najveće vrednosti normalizovanog signala na delu između dva detektovana vrha.

Ukoliko se utvrdi da je maksimalna vrednost signala na sredini intervala između detektovanih vrhova približne vrednosti kao i vrednost detektovanih vrhova vrlo je verovatno da je posredi nedetektovan vrh.

$$e_1 = 1 - \frac{1}{N} \sum_{i=1}^N \left| \left(\frac{s_i + s_{i+1}}{2} \right) - S_{\max_i} \right|, \quad (1)$$

gde je i redni broj periode unutar posmatranog segmenta signala sačinjenog od N perioda,

S_{\max_i} označava maksimalni vrednost normalizovanog signala na posmatranom intervalu između detektovanih vrhova,

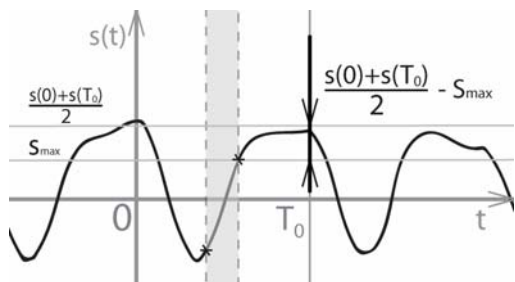
s_i i s_{i+1} su susedne detektovane vršne amplitude normalizovanog signala,

e_1 predstavlja srednju grešku prvog tipa na posmatranom segmentu normalizovanog signala.

Ukoliko je e_1 veće od unapred definisanog praga ε_1 , proglašavamo da je detektovana greška. Prag ε_1 se određuje empirijski.

Mora se obratiti pažnja na to da vrh signala u nekim

slučajevima može da bude proširen pa se mora posmatrati samo uzak deo signala na središtu intervala između dva detektovana vrha da se ne bi javljale lažno prepoznate greške.

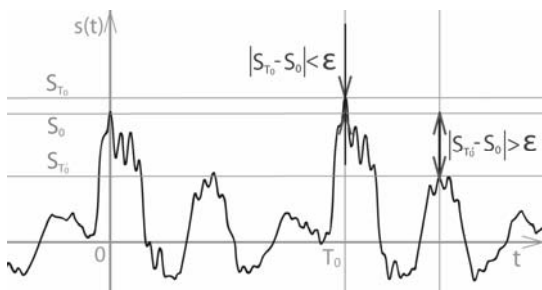


Sl. 3. Provera greške dvostruke periode; Siva oblast na vremeskoj osi predstavlja interval na kom se traži maksimalna vrednost signala

2. Ukoliko je pogrešno prepoznat vrh koji je posledica nekog od harmonika postojaće značajna razlika u amplitudi između susednih vrhova. Ova vrsta greške može se detektovati proverom srednje vrednosti modula razlike u amplitudi susednih detektovanih vrhova:

$$e_2 = \frac{1}{N-1} \sum_{i=1}^{N-1} |S_{T_i} - S_{T_{i+1}}|, \quad (2)$$

Ako je ta srednja vrednost veća od unapred definisanog praga ϵ_2 , znamo da je vrlo verovatno došlo do greške pri proceni osnovne učestanosti.



Sl. 4. Provera razlike detektovanih vrhova

Ova saznanja o greškama nam omogućuju da korigujemo podrazumevane parametre sistema za detekciju osnovne učestanosti i ponovimo proces sa smanjenom verovatnoćom pojave greške.

Zaključak koji nam daju ova dva testa je sledeći: Ukoliko je zadovoljen samo prvi kriterijum zaključujemo da se radi o greški dvostruko veće procenjene periode T_0 od stvarne, a ukoliko je ispunjen drugi kriterijum onda je u pitanju greška izazvana izraženim drugim ili višim harmonikom (drugi tip greške). Dodatno, ukoliko su ispunjena oba kriterijuma za grešku znamo da je takođe u pitanju drugi tip greške s dodatnom informacijom da je procenjena prevelika perioda.

Ukupna procena greške pri detekciji za konkretan segment signala se dobija kombinovanjem dva faktora:

$$e = (ae_1 + b) / ce_2 \quad (3)$$

gde su a , b i c koeficijenti koji se određuju empirijski.

Na osnovu ovih saznanja menjamo parametre sistema i to:

- Za slučaj pojave prve greške, pokrećemo čitav sistem sa povećanom očekivanom osnovnom učestanošću (parametar AMDF bloka).
- Za slučaj kad je uočena pojava drugog tipa greške, ukoliko se radi o procenjenoj prevelikoj periodi, takođe povećavamo očekivanu osnovnu učestanost, a u suprotnom povećavamo širinu posmatranog prozora u bloku za detekciju vrhova signala.

V. ZAKLJUČAK

Od govornih baza koje se obrađuju u sistemima za sintezu govora umnogome zavisi kvalitet dobijenog govora. Za kvalitetniju sintezu je potrebna obimnija baza, a to sa sobom donosi i veliku verovatnoću pojave loše detektovane osnovne učestanosti, a time i loš rezultat sinteze. Neophodno je automatski prepoznati loše detektovane segmente koji na kraju mogu biti izbačeni iz baze. Predložena metoda značajno popravljja rezultate detekcije, a u slučajevima gde se ne može precizno odrediti osnovna učestanost ni nakon korekcije parametara (vocal fry, loš snimak itd.), omogućuje automatsko odbacivanje loših segmenata ili ubrzan pregled ručni segmenata sortiranih opadajući po kvalitetu.

LITERATURA

- [1] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, H. J. Manley, "Average magnitude difference function pitch extractor". IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-22, pp. 353-362, Oct. 1974.
- [2] M. Sečujski, R. Obradović, D. Pekar, L.J. Jovanov, V. Delić, "AlfaNum System for Speech Synthesis in Serbian Language", Text, Speech and Dialogue : 5th International Conference, TSD 2002, Brno, Czech Republic, September 9-12, 2002. Proceedings., pp. 199-218, 2002.
- [3] G. Middleton, "Connexion project", Rice University, Houston, <http://cnx.org/content/m11714>
- [4] D. Gerhard, "Pitch Extraction and Fundamental Frequency: History and Current Techniques", Technical Report TR-CS, Department of Computer Science, University of Regina, 2003-06, November, 2003.

ABSTRACT

This paper describes a method of adaptive pitch detection based on AMDF algorithm. Shown method improves original algorithm performance thanks to the known characteristics of two most frequent error types.

ADAPTIVE SYSTEM FOR PITCH DETECTION BASED ON AMDF

Knežević D., Pekar D., Milošević I., Vujnović N.