ЈЕДАН ПРЕДЛОГ СЕГМЕНТАЦИЈЕ ИЗГОВОРЕНОГ АУДИО МАТЕРИЈАЛА КОРИШЋЕЊЕМ *TEAGER* ЕНЕРГИЈЕ

Петар Узуновић, Мирко Вермезовић,
Институт за примењену математику и електронику, Београд

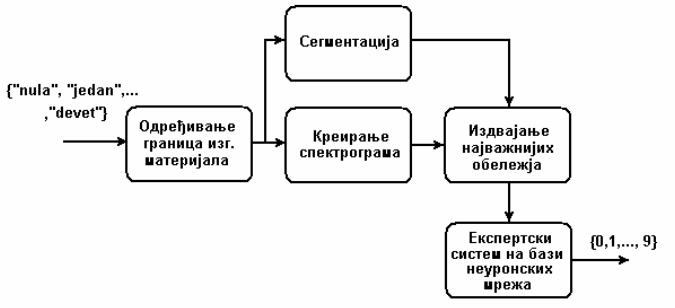
Садржај - У раду је представљена једна метода за сегментацију аудио материјала изговореног на српском језику. Сегментација се изводи на нивоу фонема коришћењем Teager енергије. Као аудио материјал усвојена је база изговорених цифара декадног бројног система на српском језику.

1. УВОД

Проблем препознавања говора је јако сложене природе и још увек не постоји генерално јединствена метода за препознавање говора. Методе се углавном везују за поједине уско-специфичне проблеме препознавања одређеног типа текста изговореног на одређеном језику. Поступак препознавања говора се може поделити на два подједнако важна подпроцеса: процес сегментације изговореног материјала и процес препознавања добијених сегмената. Са гледишта препознавања фонема, односно основних градивних делова говора, поступак сегментације је од кључног значаја.

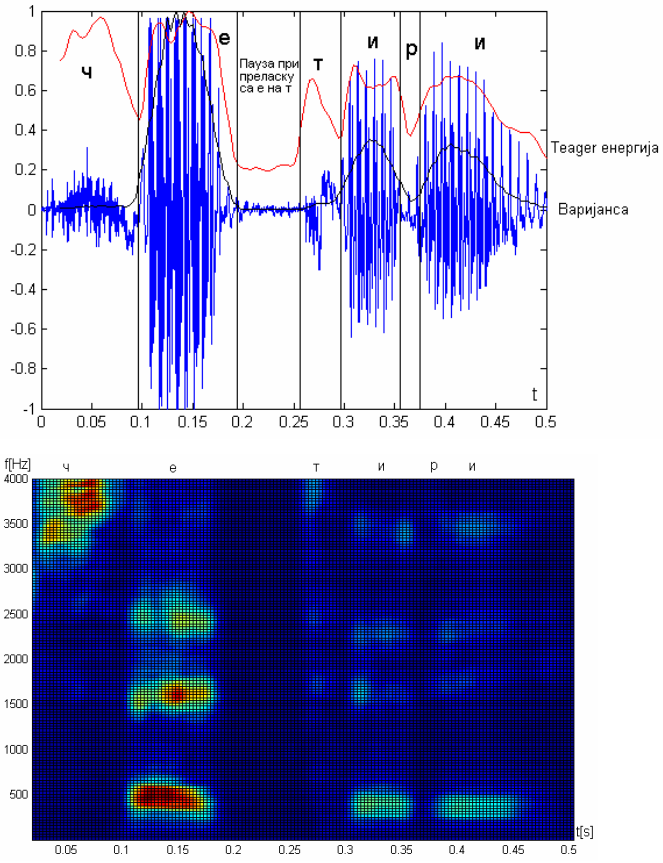
2. ДЕФИНИЦИЈА ПРОБЛЕМА

Разматрани систем представља систем за препознавање изговорених цифара на српском језику. Овакав систем би по свему судећи представљао један важан и комерцијално исплатив производ. Могао би се применити у случајевима говорних аутомата, или издавања говорних команди рачунару. Блок шема оваквог ситема приказана је на слици 1.



Clika 1.*Блок шема система за препознавање изговорених цифара на српском језику*

Централни део рада представља блок који врши сегментацију речи. Издвајање обележја се врши на основу спектрограма снимљеног говорног сигнала, по следећем поступку: спектрограм се дели по времену и по фреквенцији на седам делова чинећи тако вектор обележја од 49 елемента. Затим се примењује неки од поступака редукције димензије смањујући број елемената на минимални неопходни број. Поступак сегментације има важну улогу у одређивању најинформативнијих места за сечење говорног сигнала дуж временске осе. У пракси се показало да се најбољи резултати добијају пресецима на местима између фонема, јер се баш у тим тачкама фреквенцијске карактеристике мењају. На слици 2. приказан је таласни облик говорног сигнала и одговарајући спектрограм за цифру „четири”.



Слика 2. *Таласни облик и спектрограм изговорене цифре „четири”*

Последњи блок у систему представља експертски систем на бази неуронске мреже, који се обучава на основу базе изговорених цифара.

Циљ подсистема за сегментацију је да изолује фреквенцијски најкарактеристичније зоне изговореног материјала. Основне карактеристике говора зависе како од самог говорника тако и од говорног подручја. Претпоставка је да динамика прелаза између фонема и фреквенцијска слика фонема највише зависе од говорног подручја, док брзина говора и позиције форманата доминантно зависе од говорника. *Teager* енергија представља један од скаларних показатеља богатсва спектра говорног сигнала, како ниским тако и високим учестаностима које су од кључног значаја за препознавање фонема. Како показатељ, *Teager* енергија,

не зависи много од тачне позиције форманата у говору, што шредставља карактеристику говора, већ само од постојања одређених фреквенцијских компоненти, представља добар показатељ могућих места за сегментацију.

3. *TEAGER* ЕНЕРГИЈА

У комерцијалним системима за детекцију изговорених цифара од првенственог значаја је да се поступак детекције обави што је могуће брже. Пре свега потребно је да процес детекције буде рачунарски једноставан. *Teager* енергија преставља скаларну величину која се рачуна веома једноставно. По карактеристикама *Teager* енергија се разликује од обичне енергетске контуре јер узима у обзир и више учестаности. На слици 2. приказани су облици *Teager* енергије и варијансе за изговорену реч „четири" и лако се може приметити да фонеме „ч" и „т" поседују доминанте високо фреквентне компоненте, па их зато *Teager* енергија примећује, док их варијанса не примећује. Вокали носе енергију, па се јасно могу уочити и на контури варијансе и на контури *Teager* енергије. Шум снимања најчешће има поприлично равну карактеристику без доминантних компоненти, па према томе није приметан ни у једном ни у другом случају.

Дефинишимо *Teager* енергију једног одбирка. *Teager* енергија одбирка не зависи само од амлитуде сигнала већ и од фреквенције. Ако је дат дискретан сигнал $x_i = A\cos(\Omega \cdot i + \phi)$, тада се *Teager* енергија одбирака x_i рачуна на следећи начин

$$E_{i=} = x_i^2 - x_{i+1}x_{i-1} = A^2 \sin^2 \Omega \cong A^2\Omega^2$$

Приликом рачунања најчешће се рачуна модификована *Teager* енергија која се рачуна методом клизећег прозора по следећем алгоритму:

- Сигнал се прво дели на преклапајуће прозоре фиксне дужине.
- За сваки прозор рачуна се Брза Фуријеова трансформација (FFT) $X(\omega_k)$.
- Спектралне компоненте се тежински множе са квадратом одговарајуће фреквенције.
- На крају се *Teager* енергија рачуна као

$$T_i = \left(\sum_{k=d}^{g} \left(X_i(\omega_k)\omega_k^2 \right) \right)^{1/2}$$

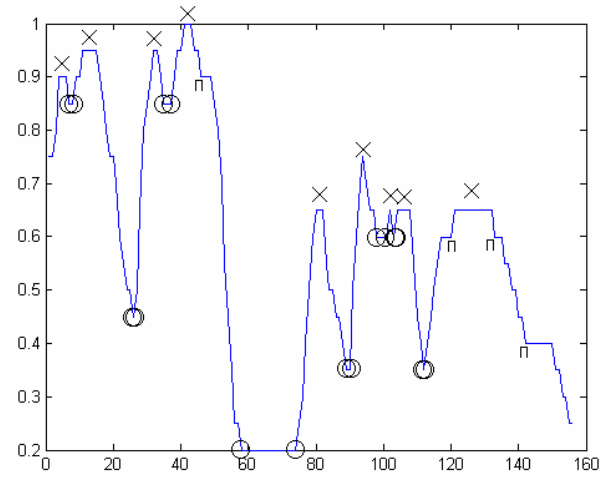
где i предтавља индекс по времену, док k представља индекс по фреквенцији. Бројеви g и d представљају индекс горње и доње граничне компоненте спектра говорног сигнала који одговарају фреквенцијама од 250Hz и 3750Hz.

Посматрањем контуре *Teager* енергије могу се уочити следеће правилности:

- На променама између сугласника прве групе (т,к,ф,з,с,ш,ж,х,ц,ћ,ч,ц,д,г,ђ,б,п) и самогласника долази до наглог пада *Teager* енергије. Сугласници прве групе имају доминантне компоненте на високим учестаностима, за разлику од самогласника.
- На променама између сугласника друге групе (в,м,н,л,љ,њ,ј,р) и самогласника долази до благе промене у *Teager* енергији, па се прелази теже уочавају. Сугласници друге групе немају тако доминантне компоненте на вишим учестаностима, јер су јако блиски вокалима.
- На променама између сугласника прве групе и сугласника друге групе, поново се уочава нагли прелаз у *Teager* енергији.
- Прелаз исмеђу два самогласника, вокала, се врло тешко уочава помоћу *Teager* енергије.

4. ПОСТУПАК СЕГМЕНТАЦИЈЕ

Ради пројектовања што једноставнијег алгоритма, извршена је дискретизација добијене *Teager* енергије на 20 дискретних нивоа као што је приказано на слици 3. На тај начин ће се локални екстремуми лакше проналазити.



Слика 3. Дискретизована *Teager* енергија на 20 нивоа

На основу претходних запажања о *Teager* енергији креиран је низ правила на основу којих ће се обављати сегментација. Неопходно је уочити са слике 3. карактеристичне тачке које су означене следећим ознакама: централне позиције локалних максимума „X", границе локалних минимума „O" и значајне превојне тачке „П". Правила по којима се врши сегментација су следећа:

1П. Резове треба потражити у областима локалних минимума и превојних тачака, у даљем тексту су то **погодни минимуми**. Треба напоменути да локални превој задовољава особину да довољно дуго задржава исту дискретну вредност, односно да је довољно широк. Више узастопних превојних тачака истог типа еквивалентира се једном превојном тачком и то оном која има највећу ширину. У случају да више њих има исту ширину, усваја се она најнижа превојна тачка.

2П. За сваки локални минимум (леви и десни), утврдити два суседна локална максимума, а за сваку превојну тачку утврдити један суседни локални максимум.

3П. Срачунати разлике између погодних минимума и одговарајућих максимума (у наставку **висине**).

Сваком погодном минимуму придружити **потенцијални рез** (који се налази на 10-15% висине почевши од погодног мимимума), одговарајући максимум и висину.

4П. Применити операцију лимитирања (*treshhold*) и избацити све оне погодне минимуме који имају висину мању од унапред задатог прага (1.5 дискретни ниво) слика 4.

5П. Само за локалне мимимуме, код којих су остала оба потенцијална реза (леви и десни), проверити да ли су леви и десни потенцијални рез довољно размакнути, (размак већи од 20ms). У случају да нису, такав минимум називамо **гранични,** односно он одређује прелаз између две фонеме, и у том случају усвајамо нови потенцијални рез који се налази на аритметичкој средини левог и десног потенцијалног реза. У супротном случају остају нам оба потенцијална реза, а такав минимум називамо **фонемски** (слика 5.).

6П. Сви добијени потенцијални резови се могу сврстати у следећих пет категорија: гранични минимум, леви минимум, десни минимум, лева превојна тачка и десна превојна тачка .

7П. Све преостале потенцијалне резове треба рангирати према следећој формули

$$K_i = \lambda \cdot h_i(1 - y_i)$$

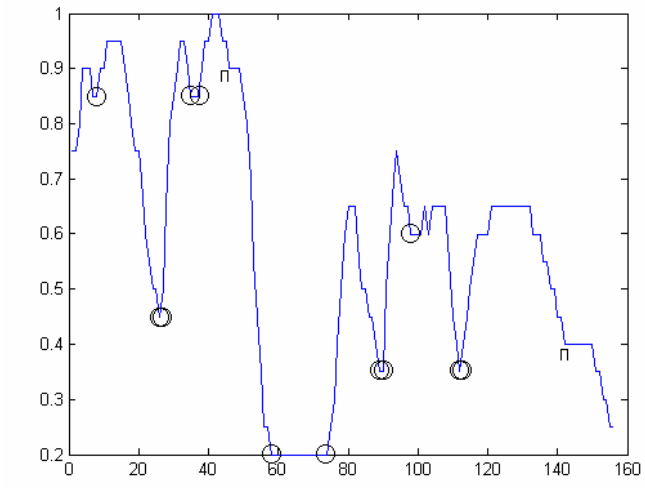
$$\lambda = \begin{cases} 1, \textit{ за локалне минимуме} \\ 0.8, \quad \textit{за превојне тачке} \end{cases}$$

h_i - *висина локалног екстремума*

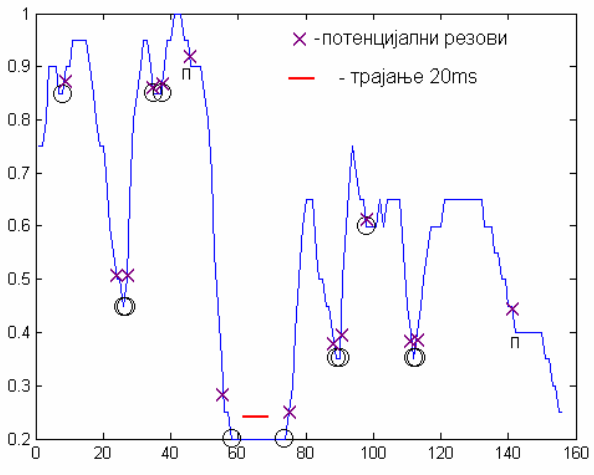
y_i - *у координата лок. екстремума*

где i представља индекс одговарајућег потенцијалног реза. За граничне минимуме као висину усвојити средњу вредност обе висине.

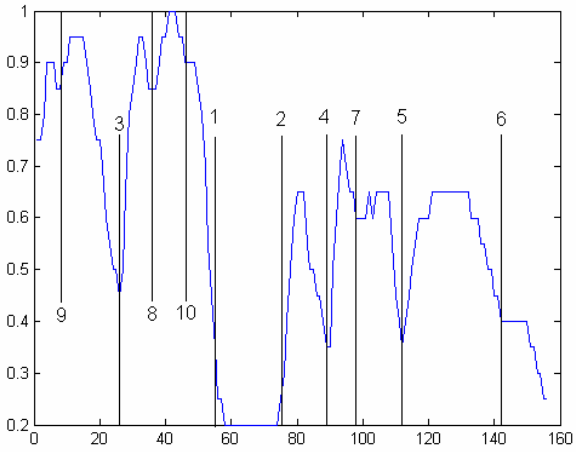
8П. Пресечна места бирати према индексу K од највећег ка најмањем, притом водити рачуна да резови не буду преблизу (растојање не мање од 20ms), слика 6.



Слика 4. Преостали погодни минимуми након операције лимитирања



Слика 5. Позиције потенцијалних резова и поређење са трајањем од 20ms.



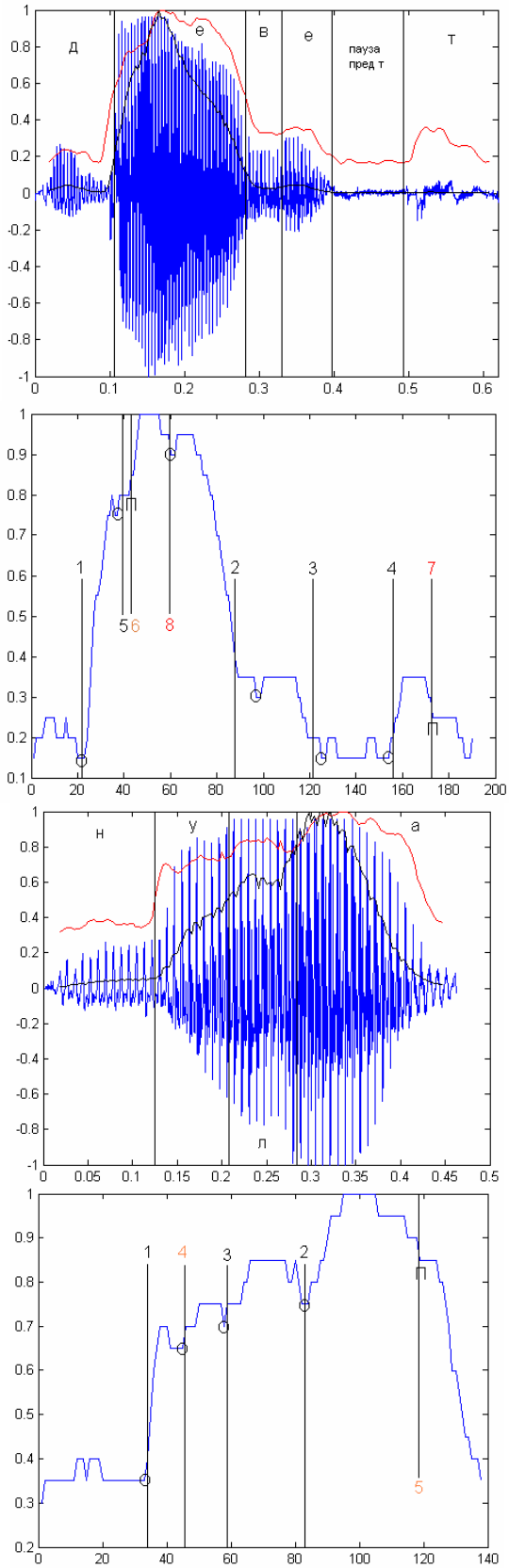
Слика 6. Добијена рангирана пресечна места након целокупног поступка.

Критеријум K_i, на основу кога је извршено рангирање пресечних места, фаворизује оне погодне минимуме који припадају класи минимума који имају ниску вредност по Y координати и што већу висину до суседног максимума.

Табела 1. Добијени резултати критеријума

Позиција	Вредн. критеријума	тип погодног мин.
55	0.6400	лев. фонемски мин.
76	0.3680	дес. фонемски мин.
26	0.2646	гранични мин.
89	0.2112	гранични мин.
112	0.1890	гранични мин.
142	0.1160	лев. превојна тачка
97	0.0380	лев фонемски мин.
36	0.0208	гранични мин.
9	0.0130	дес. фонемски мин
46	0.0051	лев. превојна тачка

Што се тиче физичког процеса, најбољи кандидати за пресечне тачке су баш она места на којима *Teager* енергија има најизраженије промене.



Слика 7. Добијена пресечна места за цифре „девет” и „нула”

5. ЕКПЕРИМЕНТАЛНИ РЕЗУЛТАТИ

На основу снимљене базе изговорених цифара на српском језику, извршено је тестирање претходно наведеног алгоритма и у наставку су приказани неки од резултата.

6. ЗАКЉУЧАК

Изложена метода показује доста добре резултате приликом сегментеације изговорених цифара на српском језику. Најбоље резултате показује код цифара које имају пуно сугласника прве групе, као што су 4,6,7,8,3,5,9, нешто лошије резултате показује код цифара 1,2, а најлошије резултате показује код цифре 0.

Правилном сегментацијом речи издвајају се најпогоднија места за дискретизацију спектрограма. Оваквим поступком изводимо велику редукцију димензија улазног обележја, притом губећи минимум информација.

Овакав алгоритам могао би пронаћи интересантну примену у брзом препознавању фонема, односно говора. Његова првенствена вредност је у томе што захтева мале рачунарске ресурсе.

7. ЛИТЕРАТУРА

[1] Lingyun Gu, Stephen A. Zahorian, “A New Robust Algorithm For Isolated Word Endpoint Detection”, ICASSP2002, Orlando, USA

[2] Jia-lin Shen, Jeih-weih Hung, Lin-shan Lee, “Robust Entropy-based Endpoint Detection For Speech Recognition in Noisy Environment”, Proc. Int. Conf. on Spoken Lang. Processing, Sydney ICSLP, 1998

[3] M. Bilginer Gülmezoğlu, Vakif Dzafarov, Mustafa Keskin, Atalay Barkana, “A Novel Approach to Isolated Word Recognition”, IEEE Transaction of Speech and Audio Processing, VOL 7. NO 6, November 1999

[4] Gin-Der Wu, Chin-Teng Lee, “Words Boundary Detection With Mel-Scale Frequency Bank in Noisy Environment”, IEEE Transaction of Speech and Audio Processing, VOL 8. NO 5, September 2000

Abstract- In this paper a method for segmentation of spoken digits in Serbian language using Teager energy is proposed. This simple method can be used for foneme recognition as well.

A PROPOSAL OF METHOD FOR SEGMENTATION OF SPOKEN AUDIO MATERIAL USING TEAGER ENERGY

Petar Uzunović, Mirko Vermezović