

WiseEdit 1.0 – СОФТВЕРСКИ ПАКЕТ ЗА АКВИЗИЦИЈУ ГОВОРНИХ УЗОРАКА У ОБЛИКУ “wave” ФАЈЛОВА

Бранко Р. Марковић, Виша техничка школа -Чачак
e-mail: brankomarko@yahoo.com

Садржај – Овај рад детаљно објашњава софтверски пакет *WiseEdit 1.0* који је намењен за снимање и анализу говорних узорака. Формат фајлова којим се представљају снимљени узорци је типа “wave”. Софтверски пакет омогућава различите опције које се могу користити приликом снимања као што су: подешавање параметара за снимани узорак, графички приказ, репродукцију снимљеног узорака, чување, ресетовање и друге основне опције које се користе приликом рада са фајловима. Овај софтверски пакет је компатибилан са *Microsoft*-овим стандардом за аудио фајлове. Такође компатибилан је и са свим другим софтверским пакетима из “Wise” фамилије (*WiseHMM*, *WiseWave*, *WiseDigit*).

1. УВОД

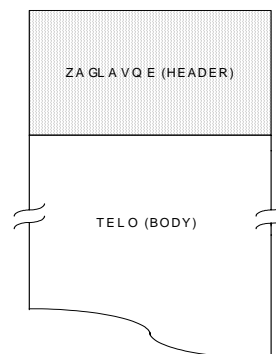
Софтверски пакет *WiseEdit 1.0* је развијен у *Microsoft*-овом *Windows* окружењу. Оперативни системи на којима се може инсталирати и успешно извршавати су *Windows 98/2000/ME/XP*. Хардверски захтеви су релативно скромни: процесор *Pentium III*, меморија од 128МВ, звучна картица и графичка картица од минимално 32МВ.

Овај графички алат омогућава кориснику да снима говорне (звучне) узорке у облику *wave* фајлова, да их прикаже графички у две димензије, да врши њихову анализу, чување и друге манипулације које су стандардне за фајлове. Софтвер такође омогућава да корисник може подешавати различите параметре сниманих узорака као што су: фреквенција одмеравања, број канала, резолуција и тд.

Основне карактеристике софтверског пакета *WiseEdit* су да је једноставан за коришћење, лак за инсталацију на оперативним системима типа *Windows*, компатибилан са *Microsoft*-овим производима из сличног домена, релативно јефтин и са скромних хардверских захтева. То омогућава да се он употребљава код широког круга корисника који се баве проучавањем аудио сигнала типа “wave”, а користе веома распрострањене *Microsoft*-ове оперативне системе.

2. WAVE ФАЈЛ – ФОРМАТ И КАРАКТЕРИСТИКЕ

Wave фајлови представљају бинарни запис снимљених говорних узорака. Формат ове врсте фајлова приказан је на слици 2.1.



Слика 2. 1: Формат *wave* фајла

Са слике се може уочити да се сваки фајл састоји из две целине: заглавља и тела. Заглавље *wave* фајла садржи више врста података о фајлу, а најзначајнији су следећи: укупна дужина фајла, број канала, фреквенција одмеравања, број бајтова у секунди, број бита по одмерку итд. Тело *wave* фајла садржи на почетку ознаку “data” која је карактеристична реч и означава почетак података који описују сам говорни узорак. Затим следи број који означава дужину података говорног узорака, а потом следе подаци који репрезентују сам говорни узорак.

Оваква организација формата *wave* фајла омогућава да се програмски може лако приступити сваком од делова фајла унутар заглавља и тела. На овај начин могуће је једноставно уписивање/ишчитавање вредности елемената *wave* фајла на основу задатих/тражених параметара сниманог говорног узорака (нпр. фреквенција одмеравања, број канала итд.).

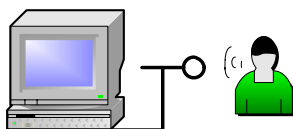
3. УСЛОВИ СНИМАЊА ГОВОРНИХ УЗОРАКА

Снимање говорних узорака може се извести у различитим условима. Пре свега ту треба обратити пажњу на ниво буке (шума). Шум који је присутан потиче од различитих спољњих извора, али такође један део настаје и радом рачунара на коме се врши снимање. Уколико се жели што бољи квалитет говорних узорака пожељно је да услови снимања буду што приближнији “глувој соби”.

Поступак снимања коришћењем софтверског пакета “*WiseEdit 1.0*” захтева персонални компјутер (PC) са напред наведеним хардверским карактеристикама, стандардну звучну картицу и микрофон. На овом PC-ју треба да је инсталиран неки од раније наведених оперативних система. Микрофон (MC) се бира са одговарајућим карактеристикама тако да пропушта онај опсег учестаности који је од интереса за анализу, а такође и да ради као добар

поништавач ефекат еха и позадинског шума. У неким случајевима ради што бољег поништавања еха користи се микрофонски ред.

Слика 3.2 приказује корисника и одговарајући хардвер неопходан за аквизицију говорних узорака.

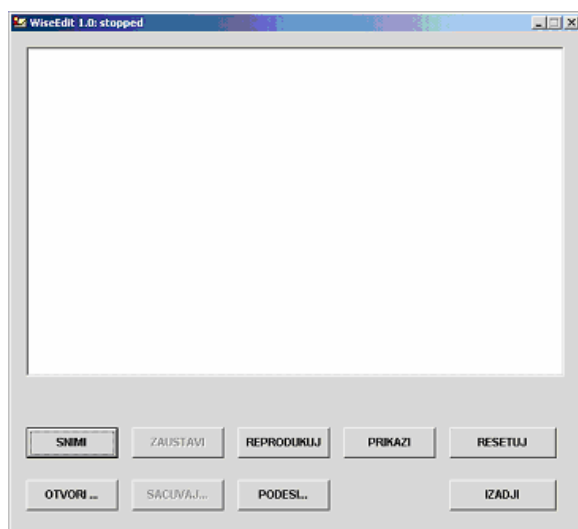


Слика 3.2: Поступак аквизиције говорних узорака

После активирања програма WiseEdit корисник треба да подеси одговарајуће параметре као што су фреквенција одмеравања, број канала, и резолуција о чему ће бити више речи у наставку па да затим приступи поступку снимања.

4. ПОСТУПАК СНИМАЊА

Када су сви хардверски и софтверски услови који су напред наведени задовољени, онда се приступа поступку снимања говорних узорака. Овај поступак састоји се у активирању апликације "WiseEdit_v1_0". После уводног (splash) прозора корисник добија екран за аквизицију говорних узорака који има изглед као на слици 4.1:

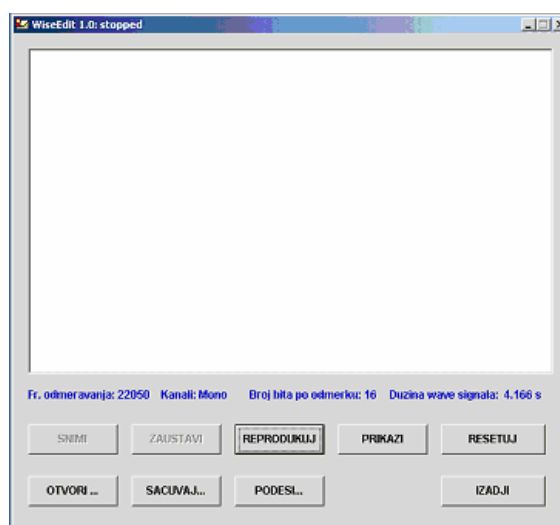


Слика 4.1: Пре снимања говорних узорака

Овај део апликације омогућава да се започне снимање говорних узорака. У доњем делу екрана налазе се два скупа команди за аквизицију. Први скуп команди је: "SNIMI" (Record), "ZAUSTAVI" (Stop), "REPRODUKUJ" (Play), "PRIKAZI" (Show) и "RESETUJ" (Reset). Он се односи на могућности рада са wave сигналом. Другу скуп команди, у доњем делу, односи се на рад са wave фајлом и то су:

"OTVORI..." (Open...), "SACUVAJ..." (Save...), "PODESI ..." (Set up...) и "IZADJI" (Exit).

У делу пре почетка снимања команде "ZAUSTAVI" и "SACUVAJ..." су неактивне. Ако корисник хоће да сними нови говорни узорак приступа активирању команде "SNIMI" притиском на командно дугме. Од тог тренутка снимање говорног узорка започиње, команда "SNIMI" бива неактивна, а команда "ZAUSTAVI" постаје активна. Када хоћемо да завршимо снимање једног говорног узорка притискамо командно дугме "ZAUSTAVI" и добија се нови облик екрана као на слици 4.2:



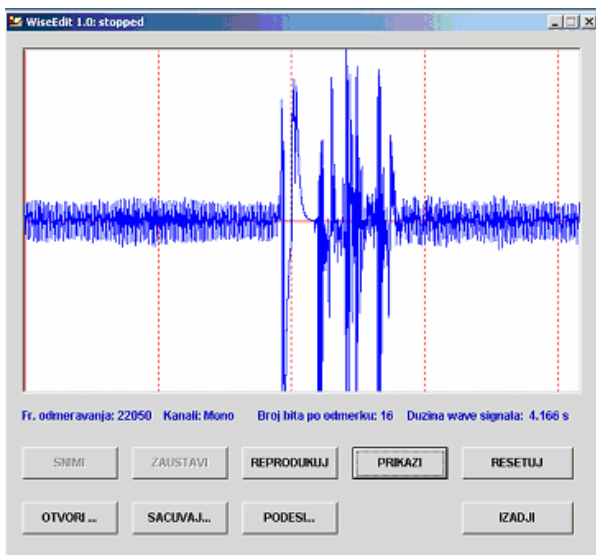
Слика 4.2: По завршетку снимања

У тренутку завршетка снимања на екрану команде "SNIMI" и "ZAUSTAVI" постају неактивне док се команда "REPRODUKUJ" активира. Такође, испод горњег дела екрана који служи као активна површина за приказивање изгледа говорног сигнала појављују се основни подаци о снимљеном сигналу и то: фреквенција одмеравања (Fr. odmeravanja), број канала (Kanali), резолуција (Broj bita po odmerku) и дужина снимљеног сигнала (Duzina wave signala). Сви ови подаци су по дифолту, а о њиховом подешавању ће бити више речи у наставку.

Уколико у старту није постојала потреба за снимањем новог сигнала већ анализом постојећег онда треба активирати команду "OTVORI..." којом се отвара већ постојећи wave фајл.

Ако се после завршеног снимања активира команда "REPRODUKUJ" онда се може чути звучни запис одговарајућег говорног узорка.

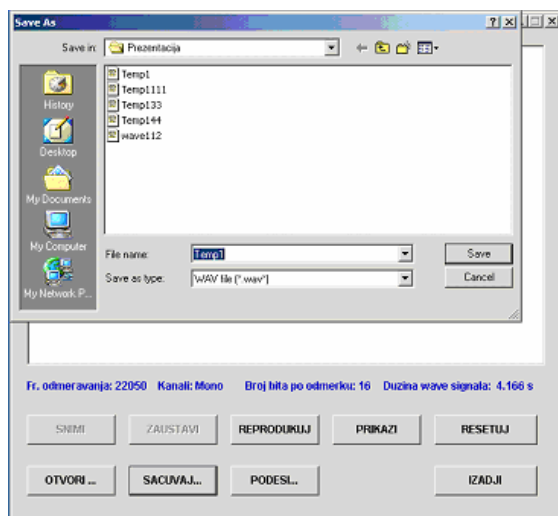
Следећа, и вероватно најзанимљивија опција је да се добије визуелна представа снимљеног говорног узорка. То се остварује притиском на команду "PRIKAZI". За један конкретан случај изглед екрана би био као на слици 4.3:



Слика 4.3: Приказ снимљеног говорног узорка

Уколико снимљени сигнал не одговара кориснику, а он то процењује на основу слушања и визуелне представе, онда се приступа поништавање овог сигнала притиском на команду **“RESETUJ”**. Потом се напред описани поступак снимања понавља.

Уколико је корисник задовољан са снимљеним говорним узорком онда може да га сачува под одређеним именом. То се остварује притиском на команду **“SACUVAJ...”**. После ове команде појављује се прозор као на слици 4.4:

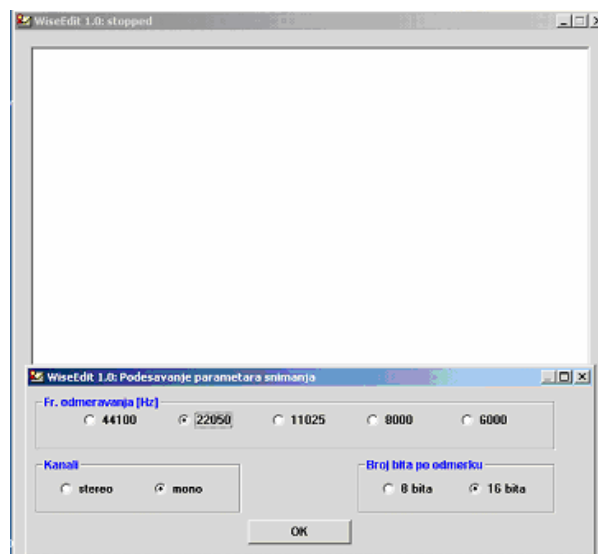


Слика 4.4: Чување снимљеног *wave* фајла

У овом случају отвара се одговарајући директоријум, а *wave* фајл за чување по дефолту носи назив **“Temp1.wav”**. Корисник треба да одабере одговарајући директоријум, упише или ново име фајла (ако жели да фајл носи другачији назив) или да га сачува под програмски задатим именом (**“Temp1.wav”**). Овим би био заокружен процес снимања, анализе и чувања *wave* фајла. Даље се овако добијени фајлови могу организovati и употребљавати у оне сврхе у које их је корисник планирао.

Завршетак и излазак из програма се обавља притиском на командно дугме **“IZADJI”**.

У овом програму постоји и опција за подешавање параметара сниманог говорног узорка и она се може активирати пре почетка снимања притиском на командно дугме **“PODESL...”**. После активирања добија се екран као на слици 4.5:



Слика 4.5: Подешавање параметара за снимање

Са слике се може уочити нови прозор који је по димензијама мањи од главног. Оба прозора су активна, а на новоотвореном имамо следеће опције за подешавање:

- Фреквенција одмеравања (Fr. odmeravanja [Hz]) са могућим изборима: 44100, 22050, 11025, 800 и 600;
- Број канала (Kanali) са могућим изборима: stereo и mono;
- Резолуција (Broj bita po odmerku) са следећим изборима: 8 bita и 16 bita.

Избором одговарајућих опција и притиском на команду **“OK”**, долази се поново до главног екрана сада са новим параметрима за снимање говорних узорака. Ови параметри ће се рефлектовати током прве следеће аквизиције (снимања) на делу прозора као што је раније објашњено (слика 4.2).

5. АНАЛИЗА СНИМЉЕНИХ ГОВОРНИХ УЗОРАКА

У зависности од потребе корисника снимљени говорни узорци могу се даље анализирати по више критеријума, а такође и користити употребом различитих софтверских алата.

Прва део анализе снимљених говорних узорака може се вршити током снимања. Он подразумева пре свега две врсте анализе:

- звучну и
- визуелну.

Звучна анализа се врши тако што се коришћењем команде **“REPRODUKUJ”** може чути како је сигнал снимљен и проверити да ли задовољава унапред постављене захтеве корисника. Да ли је то баш та информација која се желела снимити? Ту се пре свега

мисли на почетак и крај сигнала, јасноћу, гласност, затим на присуство шума и тд.

Под визуелном анализом се подразумева коришћење команде “**PRIKAZI**” и на бази ње посматрање и анализа сигнала. Визуелном методом могу се уочити амплитуде сигнала, континуалност, изненадни прекиди или пикови и слично и на основу тога такође се може оценити да ли је снимљени узорак погодан за даље коришћење.

Осим ових врста анализа које директно пружа софтверски пакет **WiseEdit 1.0** постоје и многе друге базиране на различитим моделима и алгоритмима. Неке од њих захтевају већи број узорака као што су нпр. DTW, HMM итд. Софтверски алати који су раније помињани (**WiseDigit**[7], **WiseHMM**[6], **WiseWave**[8] итд.) су потпуно компатибилни са овим софтвером тако да заједно заокружују једну целину за обраду говорних узорака.

6. ЗАКЉУЧАК

Софтверски пакет **WiseEdit 1.0** пружа корисницима лак за коришћење и једноставан за инсталацију алат којим могу прикупљати и анализирати говорне и друге аудио узорке у облику **wave** фајлова. Хардверски и софтверски захтеви за овај пакет су релативно скромни па овај софтвер може имати широк круг потенцијалних корисника.

Анализа говорних узорака може се вршити у облику преслушавања снимљених узорака, затим коришћењем визуелног приказа као и неким од метода које се користе у областима препознавања говора (DTW, HMM итд.). Иако у свету постоји одређен број сличних програмских пакета јединственост овог је и та што је написан за српско говорно подручје.

Даљи правци усавршавања овог софтверског пакета би се односили на повећање броја функција које би могле да се пруже корисницима (нпр: филтрирање, конверзија из временског у фреквенцијски домен, повећање броја опција као што су активни екран који интерактивно приказује изглед снимљеног сигнала током аквизиције, интеграција са другим софтверским пакетима из **Wise** фамилије итд.). Ове и друге функције побољшања софтверског

пакета би требало да се нађу у некој од следећих верзија **WiseEdit-a**.

ЛИТЕРАТУРА

- [1] L. Rabiner, B.-H. Juang, “Fundamentals of Speech Recognition”, Prentice Hall, 1993.
- [2] Deitel & Deitel, T.R. Nieto, “Visual Basic 6, How to program”, Prentice Hall, 1999, New Jersey.
- [3] www.goldwave.com
- [4] www.world-voices.com
- [5] www.syntrillium.com
- [6] Бранко Марковић “*WiseHMM – Графички интерфејс за обучавање скривених Марковљевих модела (ХММ) континуалног типа*”, ЕТРАН, XLVI Конференција, Теслић, 2002;
- [7] Branko Marković “*WiseDigit – Softver za prepoznavawe vezano izgovorenih cifara u uslovima šuma*”, DOGS, Sombor, 2004.
- [8] Branko Marković “*Grafički interfejs za poređewe govornih signala korišćewem DTW metoda*”, DOGS, Већеј, 2002.
- [9] Deitel & Deitel, “C++, How to program”, Prentice Hall, 1998, New Jersey.

Abstract - This paper explains in details a software package **WiseEdit 1.0** which can be used for recording and for analysis of spoken patterns. The format of files used in this case has a type “wave”. The software package allows different options and they are: setting up the parameters, visualization of signal, playing the recorded patterns, saving, resetting and other options which are used for manipulation of the files. This package is compatible with the Microsoft’s standard for the audio files. Also, it’s compatible with all other software packages from the “Wise” family (**WiseHMM**, **WiseWave**, **WiseDigit** etc.).

WiseEdit 1.0 – THE SOFTWARE PACKAGE FOR ACQUISITION OF SPOKEN PATTERNS IN FORMS OF “WAVE” FILES

Branko Markovic