УНИВЕРЗИТЕТ У БЕОГРАДУ

ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

Б Р А Н К О Р. М А Р К О В И Ћ

АНАЛИЗА ОБЕЛЕЖЈА У ГОВОРНОМ СИГНАЛУ ЗА ПОТРЕБЕ ПРЕПОЗНАВАЊА МУЛТИМОДАЛНОГ ГОВОРА

ДОКТОРСКА ДИСЕРТАЦИЈА

Београд, 2018.

UNIVERSITY OF BELGRADE

SCHOOL OF ELECTRICAL ENGINEERING

B R A N K O R. M A R K O V I Ć

ANALYSIS OF THE SPEECH SIGNALS FEATURES FOR MULTIMODAL SPEECH RECOGNITION

DOCTORAL DISSERTATION

Belgrade, 2018.

МЕНТОР:

_____

др Миомир Мијић, редовни професор, Електротехнички факултет, Београд

ЧЛАНОВИ КОМИСИЈЕ:

_____

др Миомир Мијић, редовни професор, Електротехнички факултет, Београд

_____

др Драгана Шумарац-Павловић, ванредни професор, Електротехнички факултет, Београд

_____

др Мишко Суботић, научни сарадник, Центар за унапређење животних активности, Београд

датум одбране:

_____

ЗАХВАЛНИЦА

*Захваљујем се **професору др Слободану Т. Јовичићу** на посебној бризи, усмеравању и помоћи током дугогодишњег научно-истраживачког рада.*

*Захваљујем се **професору др Миомиру Мијићу**, ментору овог рада, на срдачном односу и сарадњи, а такође и **професорки др Драгани Шумарац-Павловић** на корисним саветима и подршци.*

*Захваљујем се **колегама са ЕТФ-а** и **са ВШТСС Чачак** на подршци и сарадњи као и **студентима волонтерима** који су учествовали у снимању говорне базе Whi-Spe.*

*Захваљујем се **члановима своје породице** који се ми били узданица и подршка у сваком, па и у овом подухвату.*

АНАЛИЗА ОБЕЛЕЖЈА У ГОВОРНОМ СИГНАЛУ ЗА ПОТРЕБЕ ПРЕПОЗНАВАЊА МУЛТИМОДАЛНОГ ГОВОРА

Резиме

У системима аутоматског препознавања говора све значајнију улогу игра и мултимодални говор. Различити модови говора носе са собом и различита акустичка обележја. Од посебног интереса су модалитети нормалног (природног) говора и говора генерисаног шапатом као и њихове комбинације (када је систем трениран нормалним говором, а препознаје се шапат као и остале опције). Стога постоје различити сценарији и могу се поделити на усаглашене („нормалан/нормалан“ и „шапат/шапат“) и неусаглашене („нормалан/шапат“ и „шапат/нормалан“).

За експериментални део овог рада коришћено је једанаест векторских обележја (LPCC, LFCC, TELFCC, MFCC, TEMFCC, GFCC, TEGFCC, PLPCC, TEPLPCC, RASTACC и TERASTACC) која на различите начине описују акустичке карактеристике мултимодалног сигнала. Урађена је статистичка анализа где је ниво поузданости постављен на 95% и на бази тога су рачунати интервали поузданости. Посебно је истакнута потреба за употребом нормализације (Cepstral Mean Subtraction и RelAtive SpecTrA) и показано је да се коришћењем исте успешност препознавања може знатно поправити. То је посебно изражено када су у питању неусаглашени сценарији (нпр. код сценарија „шапат/нормалан“ побољшање може ићи и до 47%). Применом нелинеарног Teager Energy оператора који је добар за опис турбулентног кретања ваздуха, а које се јавља приликом шапата, додатно се може повећати успешност препознавања (нпр. и до 21% код сценарија „шапат/нормалан“ за векторско обележје TERASTACC). Показано је да и у усаглашеним сценаријима примена Teager Energy оператора даје побољшање успешности препознавања.

Са аспекта различитих векторских обележја испоставља се да TERASTACC скоро у свим случајевима даје најбоље резултате. Затим по успеху следе TEPLPCC, TEMFCC, MFCC и тако даље.

Што се тиче дужине вектора, резултати са вектором од 24 коефицијента од којих су 12 кепстрални и 12 делта кепстрални (са применом нормализације) су се показали најбољим. Осим овог типа вектора коришћена су још три (један од 12 коефицијената без нормализације и два са нормализацијом - од 12 и 36 коефицијената).

Метода за препознавање, базирана на динамичком програмирању и DTW алгоритму показала се као ефикасана и поуздана за брзо налажење оптималне стазе по којој се рачуна дистанца између вектора кепстралних коефицијената.

Добијени резултати указују на неопходност коришћења нормализације и Teager Energy оператора при препознавању мултимодалног говора са акцентом на препознавање шапата.

Кључне речи: Мултимодални говор, акустичка обележја, кепстрални коефицијенти, Teager Energy оператор, нормализација, DTW.

Научна област: Електротехника

Ужа научна област: Акустика

УДК: 621.3

ANALYSIS OF THE SPEECH SIGNALS FEATURES FOR MULTIMODAL SPEECH RECOGNITION

Summary

In automatic speech recognition systems, multimodal speech is playing an increasingly important role. Different modes of speech involve different acoustic features. Particular importance is given to normal (natural) speech, whisper and their combinations (when the system is trained by normal speech and then used to recognize whisper, and vice versa). The match scenarios: „normal/normal“ and „whisper/whisper“, and mismatch scenarios: „normal/whisper“ and „whisper/normal“ are considered here.

The experiment was performed using eleven vector features (LPCC, LFCC, TELFCC, MFCC, TEMFCC, GFCC, TEGFCC, PLPCC, TEPLPCC, RASTACC and TERASTACC) which provide different descriptions of acoustic properties of multimodal speech. A statistical analysis was conducted at a 95% confidence level, based on which confidence intervals were calculated. Particular emphasis was placed on the use of normalization (Cepstral Mean Subtraction and RelAtive SpecTrA) which considerably increased the word recognition rate. This was especially evident in mismatch scenarios (e.g. in „whisper/normal“ scenarios, the increase was up to 47%). The use of the nonlinear Teager Energy operator in describing turbulent airflow occurring during whisper production additionally increased the word recognition rate (i.e. up to 21% in a „whisper/normal“ scenario for the TERASTACC vector feature). The word recognition rate in match scenarios was also improved by the Teager Energy operator.

As regards vector features, the best results were achieved by TERASTACC in almost all cases, followed by TEPLPCC, TEMFCC, MFCC etc.

As for vector length, very good performance was obtained by the vector consisting of 24 coefficients i.e. 12 cepstral and 12 delta cepstral coefficients (with normalization included). Three additional types of vectors were used (one non-

normalized vector containing 12 coefficients, and two normalized vectors consisting of 12 and 36 coefficients, respectively).

Recognition was performed by the DTW method based on dynamic programming. The method proved effective and reliable in finding the optimal path to be used in computing the distance between feature vectors.

The final results showed that normalization and the Teager Energy operator were beneficial for multimodal speech recognition with a specific focus on whisper.

Keywords: Multimodal speech, Acoustical features, Cepstral coefficients, Teager Energy operator, Normalization, DTW.

Scientific area: Electrical Engineering

Scientific subarea: Acoustics

UDC number: 621.3

С А Д Р Ж А Ј

СПИСАК СЛИКА

СПИСАК ТАБЕЛА

„У почетку беше Реч, и Реч беше у Бога, и Бог беше Реч.“

(св. Јеванђеље по Јовану)

1. УВОД

Данашње време је створило нове потребе и могућности за комуникацију између човека и рачунара, односно човека и неког аутоматизованог система. Ова област се посебно актуелизовала са напретком савремених технологија, а пре свега повећањем брзине рада микропроцесора, повећањем капацитета радних меморија и развојем нових алгоритама. Нано технологије су унеле додатни импулс у овај свет те се данас, при изради савремених рачунарских као и других компоненти, прорачуни врше „у нанометрима“. Све ово је утицало и на нове научне помераје у телекомуникацијама. Омогућено је да се неки ранији инжењерски проблеми у процесирању сигнала превазиђу и извршавање веома сложених, спорих алгоритама, знатно убрза.

Комуникација човека и аутоматизоване машине постала је посебно популарна када је „машина“ могла да препознаје човеков глас и да извршава његове команде. Дакле, без физичког додира, „машина“ бива руковођена говором у циљу обављања одређених задатака. Тако је област аутоматског препознавања говора (АПГ) постала врло актуелна у разним применама роботике. У телекомуникационим системима она је такође нашла широку примену, а један од примера је иницирање и завршетак комуникације са саговорником без употребе руку (у фиксној и мобилној телефонији).

Уопште узев, када су у питању људски глас и аутоматизовани систем, постоје два аспекта којима се наука деценијама интензивно бави, а то су:

1) препознавање говора од стране аутоматизованог система и извршавање одговарајућих команди на бази препознатог (аутоматско препознавање говора);
2) генерисање говора на бази писаних улазних информација (обично текста) од стране аутоматизованог система (синтеза говора).

Говор, као акустички производ, може се појавити у разним варијацијама. Најчешће се под говором подразумева продукција одговарајућег, разумљивог сигнала при чему је особа која говори здрава и у нормалним, релаксираним условима. Међутим, услед различитих узрочника говор се може манифестовати и у другим облицима-модовима. Једна од опште прихваћених подела је да говор може имати облик: шапата, полутихог, нормалног (природног), гласног говора и вике. Сви ови облици имају одређене заједничке особине, а међу њима, у поређењу са нормалним говором, најразличитији је шапат. Стога је и овај рад фокусиран на модалитете шапата и нормалног говора као и њихове комбинације.

Са аспекта система за АПГ најчешће се разматрају следеће комбинације нормалног говора и шапата и то:

- „нормалан/нормалан“ - систем трениран нормалним говором, а узорци за препознавање су нормалан говор,
- „шапат/шапат“ - систем трениран шапатом, а узорци за препознавање су такође шапат,
- „нормалан/шапат“ - систем трениран нормалним говором, а узорци за препознавање су шапат и
- „шапат/нормалан“ - систем трениран шапатом, а узорци за препознавање су нормалан говор.

Прве две комбинације се називају „усаглашени сценарији“ док друге две „неусаглашени сценарији“. Овај рад се управо бави разматрањем ових сценарија и тиме како акустичка обележја (која су репрезентована кроз скуп одабраних вектора) утичу на успешност препознавања. Коришћен је систем за препознавање који је зависан од говорника (speaker’s dependent system).

Пошто је говорни сигнал нестационаран процес он се на кратким временским интервалима може апроксимирати квазистационарним. Говорни сигнал се представља у одговарајућем математичком облику да би могао да се анализира и обрађује користећи познате алгоритме и софтверске алате. За обраду се користе различити приступи тј. методе. Неки од њих боље описују једне карактеристике говорног сигнала, док други боље описују неке друге. Данас највише коришћене методе за опис говора и његову обраду су: метода процесирања сигнала, метода теорије информација, метода препознавања одмерака, метода вештачке интелигенције, статистичка метода, метода неуронских мрежа, метода усклађивања узорака и њихове комбинације.

Због своје практичне примене и ефикасности посебно су популарна три приступа и то: метода на бази усклађивања узорака (Dynamic Time Warping – DTW), статистичка метода заснована на скривеним Марковљевим моделима (Hidden Markov Models - HMM) и метода неуронских мрежа (Artificial Neural Networks – ANN). Могу се користити и хибриди ових метода.

Динамичко усклађивање узорака (или развлачење) у времену (DTW) даје могућности поређења две секвенце и проналажење њихове мере сличности. Ова метода је посебно интересантна и ефикасна за употребу код система аутоматског препознавања говора када је у питању исти говорник.

Са друге стране, скривени Марковљеви модели (HMM) су врло успешни приликом препознавања говора код система који су независни од говорника. Ту се једна реч може представити са једним или више модела, а сваки модел са више стања. Модели који се најчешће користе су типа „с-лева-на-десно“ („left-to-right“) мада се могу користити и други типови (нпр. ергодични и сл.). Код ових модела свако од стања у принципу може генерисати

одређени скуп вектора. Скривени Марковљеви модели су двоструко стохастични процеси. Ова двострука стохастичност се огледа кроз случајан догађај прелаза система из једног у друго стање као и случајно генерисање симбола (вектора) по стањима.

Метода неуронских мрежа користи принцип перцепције говора у људском мозгу. Креира се од више слојева неурона који су међусобно повезани. Постоје улазни, средњи (више њих) и излазни слој. Систем је у стању да учи и на бази одређених улазних параметара, пролазећи кроз „унутрашње“ слојеве, даје одговарајуће излазне резултате.

Мотивација за овај рад лежи у потреби да се испита како различита акустичка обележја утичу на успешност препознавања мултимодалног говора са посебним нагласком на шапат и нормалан говор и њихове усаглашене/неусаглашене сценарије. Циљ је био да се на великом скупу најчешће коришћених векторских обележја испита њихов утицај, изврши компаративна анализа, а такође испита и значај нормализације и нелинеарног Teager Energy оператора. Посебно је истакнут значај нормализације приликом препознавања у неусаглашеним сценаријима, а такође је показано како примена Teager Energy оператора утиче на резултате и како у појединим ситуацијама доводи до побољшања. За препознавање је коришћена DTW метода.

Друга глава овог рада посвећена је говору као акустичком сигналу и врстама говора. Овде се детаљно описују процеси генерисања и пријема говора као и органи који у томе учествују. Описан је говор као акустички сигнал уопште и објашњена особина његове нестационарности. Анализа говорног сигнала врши се у временском и спектралном домену те су и основне каратеристике говора у овим доменима овде истакуте. Посебно је наглашен значај спектралног домена који заузима централно место у системима за препознавање говора. Један од проблема који се често јавља при препознавању говора је и утицај маскирања и шума. Он се такође овде у кратким цртама анализира и због тога што је енергија шапата често на нивоу шума те је присуство шума отежавајући фактор у препознавању шапата. Што се тиче врсте говора основна подела је дата на бази услова у којима се говор остварује (пре свега мислећи на врсту и ниво шума) као и на бази здравственог и емотивног стања у коме се говорник налази (здрав, прехлађен, после операције, уплашен и слично). Стање здравља може да утиче и на стање вокалног тракта, али се вокални тракт може и прилагођавати одређеним ситуацијама. Дато је и детаљно објашњење како настаје шапат (са аспекта функције вокалног тракта) као и акустичка и прецептивна обележја шапата, са посебним акцентом на померање првих форманата при артикулацији вокала. Појмови „доброг“ и „лошег“ шапата, као и турбулентно кретање ваздуха током креирања шапата су такође размотрени. У том смислу овде је дат и преглед шта се по питању мултимодалног говора до сада истраживало.

Акустичка обележја за препознавање нормалног говора и шапата презентована су у трећој глави. При томе су се користила следећа обележја: LPCC (Linear Prediction Cepstral Coefficients), LFCC (Linear Frequency Cepstral Coefficients), MFCC (Mel Frequency Cepstral

Coefficients), GFCC (Gammatone Filterbank Cepstral Coefficients), PLPCC (Perceptual Linear Prediction Cepstral Coefficients) и RASTACC (RelAtive SpecTrA Cepstral Coefficients). На векторска обележја је примењена нормализација (типа CMS или RASTA), а на она, где је то могуће, примењен је и нелинеарни Teager Energy оператор. Применом TE оператора добила су се додатна обележја типа TELFCC, TEMFCC, TEGFCC, TEPLPCC и TERASTACC. На овај начин формирано је једанаест различитих обележја. Анализа утицаја нормализације и TE оператора су од кључне важности у овим истраживањима па се њима посветила и посебна пажња. За сва наведена обележја изложено је теоретски и приказано детаљно (блок дијаграмима) како се добијају одговарајући вектори који их репрезентују.

Четврта глава даје опис говорне базе „Whi-Spe“ која је креирана за потребе ових истраживања коришћењем српског језика. Најпре је дат дизајн базе у коме је објашњено како је база конципирана, колико говорника је учествовало у снимању и како су одговарајући узорци дигитализовани. Потом је објашњен поступак снимања, карактеристике употребљене опреме (микрофон, рачунар и софтверски пакет) и сл. Изложени су и проблеми који су уочени током снимања, различити ефекти и грешке приликом аквизиције узорака. На крају су дати елементи говорне базе који представљају три подскупа, а то су: речи основних боја, речи изабраних бројева и акустички балансиране речи тј. дат је речник који је коришћен. Приказана је и одговарајућа IPA (International Phonetic Alphabet) нотација овог речника.

У петој глави описана је метода за поређење говорних узорака. Ова метода се базира на динамичком програмирању, а да би се спровела најпре је потребно креирати векторе кепстралних коефицијената који репрезентују говорне узорке. Описан је начин добијања мере за поређење вектора као и векторска дистанца. Потом је детаљно објашњена метода који омогућава поређење скупа вектора тражењем оптималне стазе у 2D равни и на бази тога проналажење најбољег решења при препознавању. Метода под називом DTW (Dynamic Time Warping) омогућава врло ефикасно поређење говорних узорака и примењива је када је број узорака ограничен (мали) и када је систем за АПГ „зависан од говорника“. Овде је представљена детаљна анализа одговарајућег алгоритма динамичког програмирања као и начин имплементације DTW методе. Објашњена су и ограничења која се приликом рачунања DTW-а могу користити.

Резултати истраживања и финални закључци рада изложени су у шестој глави која уједно представља и најобимнији део. Најпре су објашњени типови вектора који се састоје од одговарајућих коефицијената. Постоје четири типа ових вектора и то:

- вектор од 12 кепстралних коефицијената без нормализације,
- вектор од 12 кепстралних коефицијената са нормализацијом,
- вектор од 24 коефицијента (12 кепстралних плус 12 делта) са нормализацијом и

- вектор од 36 коефицијената (12 кепстралних, 12 делта и 12 делта-делта) са нормализацијом.

Потом су детаљно приказани резултати препознавања за следеће врсте векторских обележја: LPCC, LFCC, TELFCC, MFCC, TEMFCC, GFCC, TEGFCC, PLPCC, TEPLPCC, RASTACC и TERASTACC. Резултати су дати за сваки од четири типа вектора и за сваки од четири мултимодална сценарија и то: „нормалан/нормалан“, „шапат/шапат“, „нормалан/шапат“ и „шапат/нормалан“. Примена нормализације и побољшање препознавања за неусаглашене сценарије посебно су истакнути. Такође и утицај Teager Energy оператора је наглашен и показано је како се примена овог оператора одражава на успешност препознавања. Сви резултати су дати табеларно и у облику дијаграма. На крају ове главе извршена је сумарна компарација свих векторских обележја, извучени одговарајући закључци и предложена решења.

У седмој глави дат је закључак у коме су сумирани резултати, истакнути доприноси ове дисертације и препоручени даљи правци истраживања које би било могуће предузети користећи ове резултате као полазиште. Истакнута је оригиналност овог рада која се огледа у препорукама за одабир најповољнијих акустичких обележја за препознавање мултимодалног говора, тестираних на бази Whi-Spe говорног корпуса српског језика.

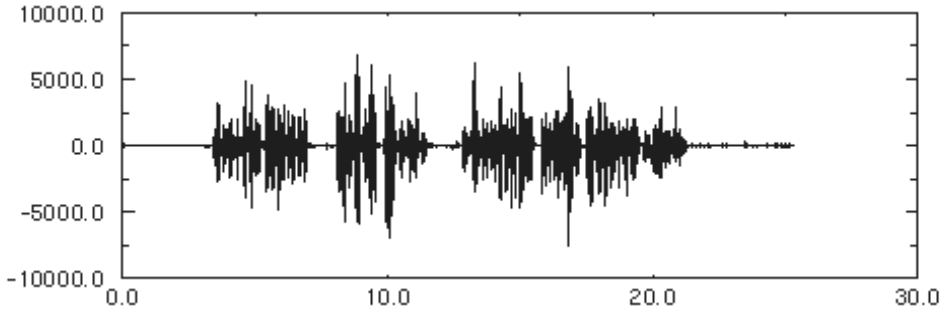
На самом крају, у Прилогу А, дати су HMM резултати за пет векторских обележја и то: LFCC, TELFCC, MFCC, TEMFCC и GFCC. Они су резултат тимског рада и користили су идентичне улазне векторе као што су употребљени за DTW анализу.

2. ГОВОР И ВРСТЕ ГОВОРА

Говор је акустички производ који се добија покретањем делова вокалног тракта. Говорник има могућност да прими, анализира и коригује свој говор. Дакле, постоји повратна спрега која се формира пријемом звука преко слушног механизма, а затим прослеђивањем истог путем нервних влакана до мозга, где се примљена информација процењује и на основу те процене врши се по потреби корекција (нпр. гласности). Положај појединих органа у вокалном тракту током генерисања говора даје различите параметре говорног сигнала од којих су најинтересантнији: интезитет звука, висина и боја тона, разумљивост и тд. Са друге стране уво које се састоји од спољашњег, средњег и унутрашњег дела служи као пријемник и трансформише аналогне сигнале говора у нервне импулсе.

Као сигнал, говор поседује низ специфичности. Тако, на пример, исти говорник није у стању да два пута изговори истоветно једну исту реч. Са друге стране говор поседује и особину редундантности што значи да без обзира који говорник изговори једну одређену реч из заједничког речника, у формираном говорном сигналу садржана је информација која се може издвојити и анализирати као заједничка за све говорнике. Та информација је и циљ многих истраживања и кључ за обучавање система аутоматског препознавања говора.

Код говорног сигнала записаног у некој електронској форми често можемо уочити паузе. Оне представљају стања мировања вокалног тракта. На слици 2.1 приказан је један такав сигнал који представља изговор реченице састављене од пет речи.



Слика 2.1 *Временски облик сигнала говора.*

Почетак и крај говорног сигнала се могу сматрати стања када нема активности у одређеном временском интервалу. Говорни сигнали се обично анализирају у временском и/или спектралном домену. Спектрални домен је атрактивнији са гледишта математичког апарата. Што се временског домена тиче, у зависности од интервала на коме се говорни сигнал посматра, могу се уочити микро и макро карактеристике. Тако је, на пример,

квазистационарност особина која се везује за посматрање говора на малим временским интервалима.

Са становишта пријема и анализе говорног сигнала интересантно је колико уво може да разликују ефекте који настају када истовремено имамо више звучних побуда. То зависи да ли ове побуде припадају истој или различитим фреквенцијским групама. Базиларна мембрана, која је осетљива на различите фреквенције, може бити тако побуђена да се неки звуци не могу чути и тада настаје ефекат маскирања, а то је посебно изражено у условима шума.

За анализу говорног сигнала користе се различите методе. Једна од најинтересантнијих је статистичка код које се свака изговорена реч апроксимира неким математичким моделом, а затим се рачунају одговарајуће вероватноће подударности. Приликом формирања модела од посебног је значаја да се изврши одговарајућа предобрада сигнала и да се на основу ње формирају вектори који ће бити репрезенти тог сигнала. У току предобраде аналогни говорни сигнал се из временског домена пребацује у скуп дискретних параметара (вектора) који одсликавају особине оригиналног сигнала. Енергија говорног сигнала, односно њена промена, је врло битна па је као параметер често једна од компонената ових вектора.

Када су у питању изоловано изговорене речи показало се да су слогови један од најприроднијих начина за њихову анализу. Повезивањем слогова могу се анализирати и веће целине, укључујући реченице. Унутар слога јавља се тзв. коартикулациони ефекат, док је између слогова исти знатно мањи.

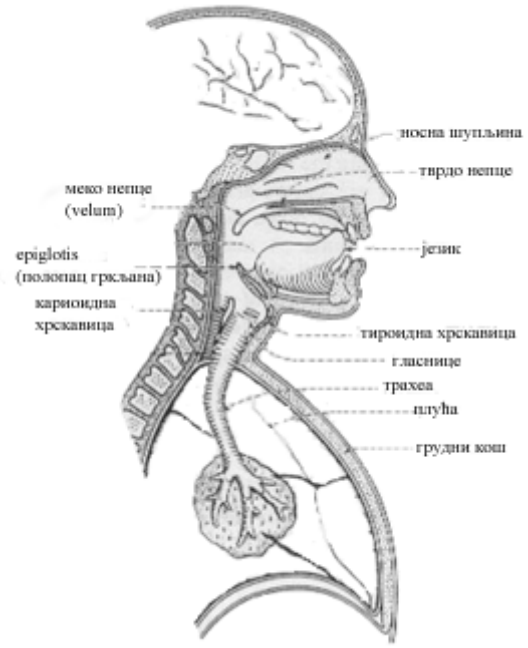
2.1. ФОРМИРАЊЕ ГОВОРНОГ СИГНАЛА

Приликом формирања говорног сигнала долази до струјања ваздуха кроз делове вокалног механизма, а такође и до контракција (скупљања и опуштања) појединих мускулаторних делова овог механизма. Основи делови вокалног механизма који учествују у формирању говора приказани су на слици 2.2.

Делови вокалног механизма који су од посебног значаја за формирање говорног сигнала су: грудни кош, плућа, трахеа, гласнице (гласне жице), носна и усна шупљина, непца, језик, велум (отвор између носне и усне шупљине), усне и ноздрве. Читав овај механизам такође има свог удела и у два веома важна механизма за човека: у механизмима једења и дисања.

Осим трахее кроз коју ваздух циркулише из и у плућа, постоје још две шупљине које играју доминантну улогу у формирању говора: вокални и назални тракт.

Вокални тракт започиње са завршетком трахее (где се налазе гласнице), а завршава се са уснама. Ова акустична цев има велики утицај на формирање говора, а померањем појединих делова унутар ње (усана, језика, вилица и велума) долази до модификације гласа. Дужина ове акустичке цеви је око 17 cm код одраслих особа.



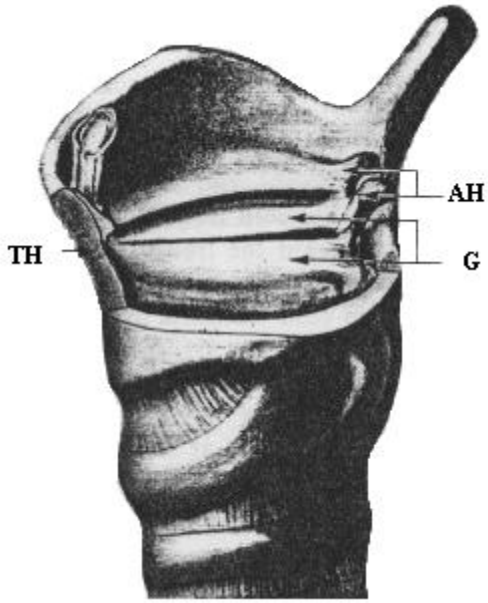
Слика 2.2 *Органи који учествују у формирању говорног сигнала [Jovičić, 1999].*

Назални тракт је образован шупљином која започиње велумом, а завршава се ноздрвама. Он је мање подложан промени унутрашњег стања у поређењу са вокалним трактом. Његова дужина се креће око 12 cm код одраслих особа.

Сама интеракција између вокалног и назалног тракта регулише се отвором велума. За звуке који нису носни велум се подиже и затвара назални тракт тако да ваздух струји само кроз уста.

Што се тиче трахее, на самом њеном врху налази се структура која по потреби врши прекидање ваздушног тока који иде из плућа. Ова структура се назива ларинкс (larynx) и приказана је на слици 2.3.

Постоји хрскавичави омотач који садржи две покретна дела формирана од лигамената и мишића. Они су на слици означени са “G” и представљају гласнице. Простор између њих који подсећа на процеп најчешће у облику троугла назива се глотис (glottis). Затварањем и отварањем овог процепа, односно смањивањем и повећавањем његовог отвора долази до вибрирања гласница и до иницирања читавог спектра говорних сигнала. Вибрирање гласница производи звучне таласе тако да се побуђује вибрирање ваздуха. Као резултат тога формирају се квази-периодични таласи ваздуха који даље интерферирају са вокалним и назалним трактом. Основна фреквенција вибрирања гласница је субјективна карактеристика и може се мерити за сваког говорника понаособ. Често се ова фреквенција користи у системима за препознавање говора где је од интереса верификација говорника.



Слика 2.3 *Ларинкс.*

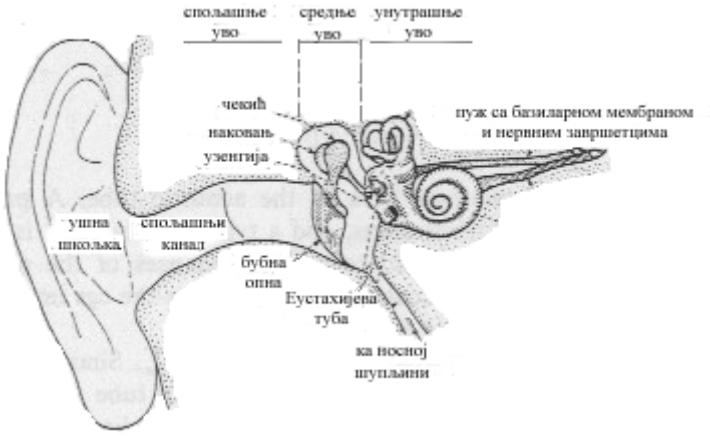
Осим формирања говорног сигнала на напред наведени начин (проласком ваздуха из трахее у вокални тракт уз треперење гласница), постоје још два начина. Први се састоји у стварању звука на основу турбулентног кретање ваздуха унутар самог вокалног тракта. Ово кретање се изазива покретањем неког од делова вокалног тракта. Други начин формирања звука је наглим отварањем односно затварањем крајева вокалног тракта тако да се ствара одређени притисак ваздуха односно његова промена.

Када говор разматрамо са аспекта једног језика онда је јасно да тај језик мора садржати одређен, коначан број, за све кориснике тог језика разумљивих, посебних гласова. Језик се конструише на бази лингвистичких јединица које имају особину да уколико се једна замени другом - онда се и значење мења. При томе акустичке манифестације основне јединице могу бити раличите (када их изговарају различите особе) али значење за све кориснике тог језика мора остати исто. Овај основни лингвистички елеменат се назива фонем [Симић, Остојић, 1996]. Стога када се приступа неком новом непознатом језику, први корак који треба учинити је идентификација скупа фонема који тај језик користи, односно треба изврши транскрипцију у којој се сваком могуће разумљивом гласу додељује одређени симбол. Када се ради о српском језику проблем је знатно упрошћен јер он има особине алфабетског језика. Код ових језика, на основу слушања, може се одмах вршити транскрипција, односно писање текста и то тако што су сугласници и самогласници јасно уочљиви. Код неких других језика ово није могуће већ је потребно анализирати групу симбола и онда вршити одлучивање. Ова предност српског језика требала би да омогући и једноставнији механизам за систем АПГ-а, а такође и повећање вероватноће успешно препознатих говорних сигнала.

2.2. ПРИЈЕМ ГОВОРНОГ СИГНАЛА

У ланцу активности које се односе на регистровање и препознавање говорног сигнала водећу улогу игра пријемни механизам код слушаоца. Да би могао да се довољно добро опише процес препознавање, као и да би се математички моделовао, потребно је најпре упознати како овај процес функционише код човека.

Пријем код слушаоца започиње регистровањем акустичког сигнала и тај процес се одвија у уву. Од ува, где се започиње прикупљање акустичких информација, па до мозга, где се оне обрађују и на основу њих се доноси одређени закључак, постоји велики број компонената које директно или индиректно учествују у трансформацијама акустичког сигнала. У том смислу треба идентификовати три дела ува (као што се може и видети са слике 2.4) и то су: спољашње, средње и унутрашње уво.



Слика 2.4 *Функционални делови ува [Jovičić, 1999].*

Спољашње уво се састоји из следећих елемената: ушна шкољка, спољашњи канал и бубна опна. Пошто бубна опна дели спољашње од средњег ува то се може виртуелно посматрати као заједнички део. Ушна шкољка има за циљ да прикупи акустичке информације и да их усмери ка унутрашњости ува, а такође и да заштити спољашњи канал. У одређеној мери ушна шкољка служи и да се одреди правац одакле долази звук. Спољашњи канал усмерава акустичке вибрације ка бубној опни (мембрани), а сама бубна опна има улогу да ове вибрације проследи даље према средњем уву и да их трансформише у механичке покрете. Такође бубна опна има улогу да заштити унутрашњост ува од спољашњих, механичких утицаја.

Средње уво прима преко бубне опне акустичке вибрације и претвара их у механичке. Главни делови срењег ува који обављају овај прецес су слушне кошчице: чекић, наковањ и узенгија. Чекић први добија вибрације од бубне опне те их преко наковања преноси ка узенгији. Узенгија даље вибрације преноси на унутрашње уво. Од делова средњег ува потребно је поменути још и Еустахијеву тубу која повезује средње уво са носном шупљином.

Средње уво је испуњено ваздухом и важно је напоменути да оно такође има заштитину улогу као и спољашње јер најосетљивији део ува - унутрашње, штити од претерано гласних звучних ефеката. Дакле, постоји одређена редукција осцилација које долазе из спољашње средине, те се овај део ува често описује као пропусник нижих фреквенција.

Унутрашњи део ува је насложенији, што по саставу то и по функцијама. Ипак могу се идентификовати најважнији делови и то су: пуж са базиларном мембраном, нервни завршеци од акустичког живца и апаратура за оријентацију-равнотежу. Са аспекта препознавања звука од интереса је пуж са базиларном мембраном и нервним завршецима. Наиме пуж је испуњен течношћу која има два пута већу вискозност од воде и карактеристичну специфичну тежину. Унутар пужа налази се базиларна мембрана и нервни завршеци који су повезани на унутрашња ћелијска влакна. Када се механичким вибрацијама изазове таласање унутар пужа долази и до побуђивања на појединим деловима базиларне мембране сагласно са фреквенцијом осциловања. Тада се иницирају унутрашња ћелијска влакна, а како су за њих везани завршеци нервних влакана - то и они бивају побуђени. Око 10 нервних завршетака је везано за свако унутрашње ћелијско влакно (IHC – inner hair cell) тако да је укупно око 30.000 нервних влакана повезано на акустички живац. На овај начин се фреквенцијске промене трансформишу у нервне импулсе који преко акустичког живца одлазе до мозга. У мозгу се порука дешифрује и тиме је процес препознавања говора завршен. Треба још напоменути да се, због особина базиларне мембране, унутрашње уво може моделовати са банком филтара пропусника опсега учестаности.

2.3. ВРЕМЕНСКЕ И ФРЕКВЕНЦИЈСКЕ КАРАКТЕРИСТИКЕ

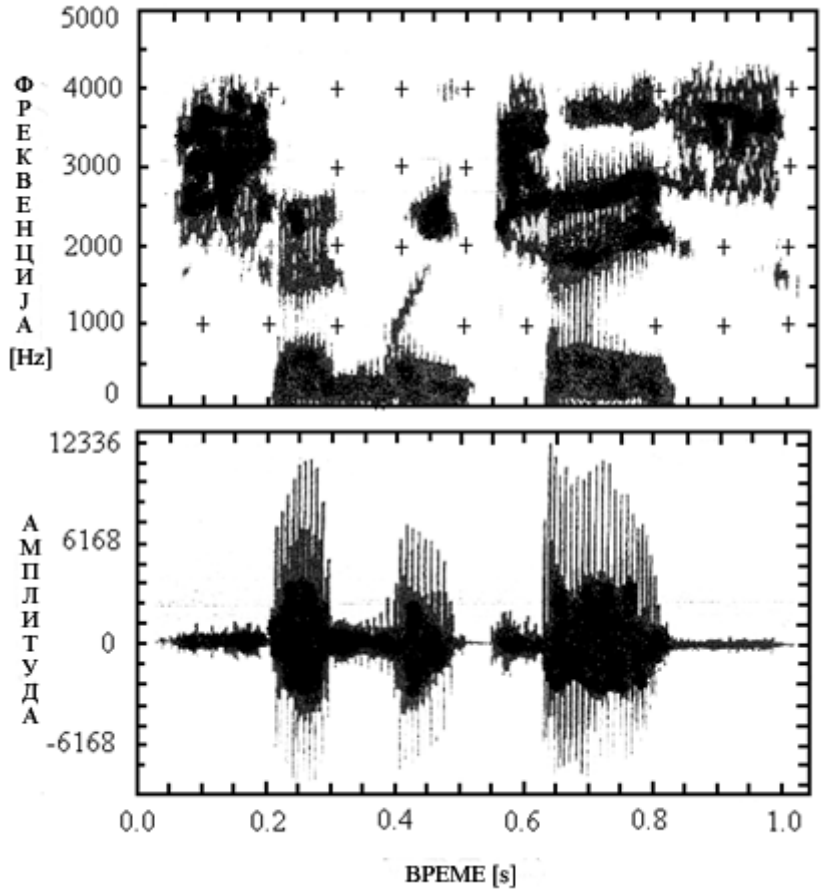
Говорни сигнал је споро променљив уколико се посматра на кратким временским интервалима реда до 100ms. Зато се и користи та његова особина квазистационарности приликом описа, а приликом моделовања користи се осим временске карактеристике и фреквенцијска и то најчешће кратковременски спектрални делови сигнала. Слика 2.5 приказује како за једну изговорену реченицу изгледају спектрални и временски облици сигнала.

Са слике се може уочити да постоје три основна стања кроз која сигнал пролази и то су:

а) стање ћутања (када нема сигнала),

б) стање безвучних гласова (када је таласни облик апериодичан и нема вибрирања гласница) и

в) стање звучних гласова који садрже периодичност (јер приликом њиховог формирања гласнице вибрирају).



Слика 2.5 *Спектрални и временски облици једног говорног сигнала.*

Уопштено гледано може се појавити и комбиновано стање (од задања два), када се на пример појаве звучни фрикативи (/з/, /ж/). Фрикативи се јављају као последица турбулентног кретања ваздуха у делу вокалног тракта и детаљан опис њихових карактеристика је описан у [Subotić et al., 2013].

Ова подела на стања се може користити приликом моделовања говорног сигнала, одређивање крајева речи као и других релевантних параметара који су неопходни током процеса препознавања говора.

2.3.1. ВРЕМЕНСКЕ КАРАКТЕРИСТИКЕ ГОВОРНОГ СИГНАЛА

Са аспекта временске анализе говорног сигнала потребно је исти посматрати на одговарајућим еквидистантним интервалима. Уколико се сигнал посматра на интервалу реда неколико десетина милисекунди (нпр. до 50 ms) онда се све уочене особине сигнала називају микродинамичке карактеристике. Пошто се положај вокалног тракта (који генерише овај сигнал) скоро и не мења у тако кратким временским интервалима онда се сигнал сматра скоро стационаран. Међутим, када се говорни сигнал посматра на интервалима од 0,2s до неколико секунди онда особине које се могу уочити су тзв. макродинамичке карактеристике. Оне носе информације о разумљивости говора, интонацији, нагласку и слично и неопходне

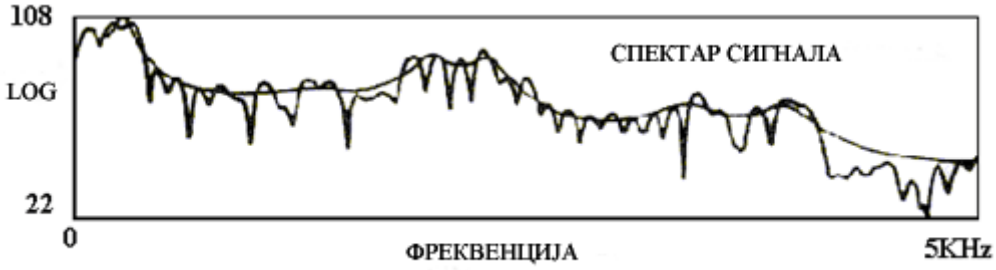
су да би перцептивни механизам вршио препознавање говора, односно разумевање онога што је изговорено. На слици 2.5 где временски облик сигнала репрезентује три речи, јасно се уочавају велике промене на интервалима од нпр. 200ms.

2.3.2. ФРЕКВЕНЦИЈСКЕ КАРАКТЕРИСТИКЕ ГОВОРНОГ СИГНАЛА

За разматрање сигнала говора посебно је интересантан фреквенцијски домен. Спектрограмом се могу одредити вредности спектралних хармоника (као нпр. слика 2.5) за задати улазни сигнал. Спектрална анализа се сматра као основа за процесирање сигнала у системима аутоматског препознавања говора. Користе се два основна модела за спектралну анализу и то:

- модел базиран на банци филтера;
- модел базиран на линеарном предиктивном кодирању.

За спектралну анализу највише се користи кратковременска спектрална густина снаге која се зове и кратовременски спектар (short-time spectrum). У одређивању истог може се користити нпр. брза Фуријеова трансформација (FFT) или модел LPC-а који подразумева преносну карактеристику са свим половима унутар јединичног круга. Слика 2.6 приказује један такав спектар добијен коришћењем поменута два начина: крива са израженим врховима представља спектар добијен помоћу FFT анализе, а заравњенија крива припада LPC анализи.



Слика 2.6 *Спектар сигнала добијен помоћу FFT и LPC метода [Rabiner, Juang, 1993].*

Са слике се може уочити одређено поклапање спектара добијених на ова два начина што је и природно јер описују исту појаву, само су математички приступи различити.

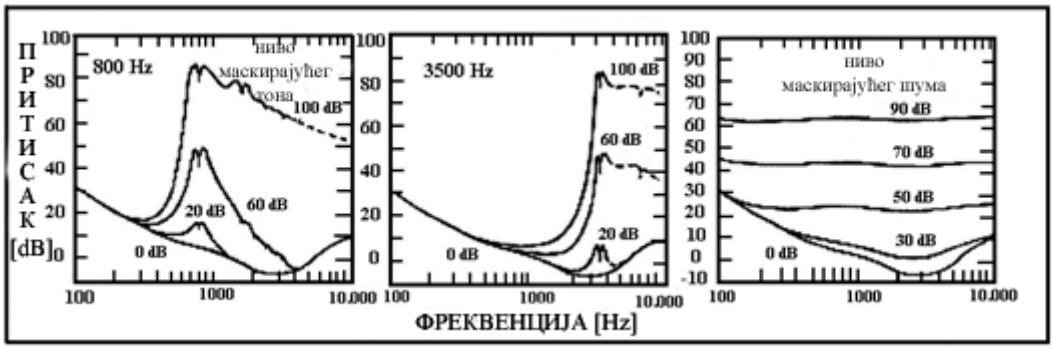
Што се тиче опсега чујности код анализе говорног сигнала обично се разматрају фреквенцијски опсези који покривају телефонски канал (200 – 3.200Hz) као и опсег чујности људског ува (око 20Hz до 20.000Hz). Закључено је да је уво мање осетљиво на појаве изобличења у области ниских него у области високих учестаности и да уво чује скоро логаритамски на фреквенцији преко 1kHz, а до ове фреквенције да је чујност линеарна. На бази различите осетљивости ува по фреквенцијској скали утврђени су фреквенцијски интервали на којима уво не осећа промену висине тона. Та скала је названа мелодијска скала или скраћено – “mel”.

Такође постоје и друге поделе и скале, а међу њима је од значаја и “Bark” скала код које се разматра да ли и како фазни став компонената сложеног звука долази до изражаја дуж базиларне мембране. На основу овог утицаја читав опсег фреквенција се дели на фреквенцијске групе. Унутар једне фреквенцијске групе људско уво није у стању да разликује два одвојена тона ако оба припадају истој групи.

2.4. ЕФЕКТИ КОАРТИКУЛАЦИЈЕ И МАСКИРАЊА

Свака изговорена реч може се посматрати као скуп јединица које обрађује уво на пријему. Те јединице се могу идентификовати као слогови и приликом говора долази до узајамног утицаја између њих и то унутар речи и између суседних речи (краја једне и почетка друге речи). Ова појава се назива коартикулација и од значаја је приликом моделовања система за препознавање говора. Показује се да је највећи коартикулациони ефекат унутар слога између гласова који чине дати слог.

Маскирајући ефекат је појава када на уво делује више различитих звукова тако да се дејство неког од њих уопште не примећује. Дешава се да маскирајући звук активира одређени број слушних ћелија на базиларној мембрани док други, маскирани звук, није у стању да активира довољан број нових ћелија. Зато се његово присуство и не запажа. Ниво маскирања се обично дефинише у децибелима као “праг маскирања” који је потребно прећи да би се изашло из стања маскирања. Да би се избегло маскирање потребно је да нови звучни догађаји буду на одређеном и довољном фреквенцијском растојању. Постоји низ мерења на основу којих су одређене фреквенцијске групе за које се ефекат маскирања своди на минимум. Посебно је интерсантна појава маскирања са позадинским шумом јер је то и најчешћи случај у пракси.



Слика 2.7 *Маскирање сигналима а) 800Hz, б) 3.500Hz и в) широкопојасним шумом.*

На слици 2.7 приказана је промена нивоа чујности кад је маскирање вршено: тоновима од 800Hz, 3.500Hz и широкопојасним белим шумом.

Маскирање може бити понекад и пријатна појава. Тако се на неким раскрсницама великих градова постављају фонтане чији звук ублажује буку саобраћаја и даје пријатнији и лепши осећај пролазницима.

2.5 МУЛТИМОДАЛНИ ГОВОР

Говор као сигнал који у себи носи одговарајућу информацију може бити генерисан у различитим условима и при различитим стањима говорника. Што се услова тиче ту се пре свега мисли на одговарајући позадински шум, а што се стања говорника тиче ту може бити различито здравствено стање говорика, различита емотивна стања и различита потреба за наглашавањем/ненаглашавањем говора.

Уколико се говор разматра са аспекта спољашњих услова постоји:

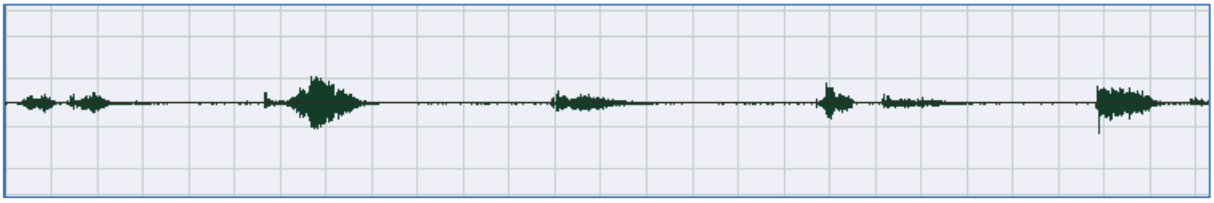
- „чист“ говор – то је говор код кога је максимално потиснут позадински шум;
- говор са „белим“ адитивним шумом;
- говор са шумом „у боји“.

У емотивном говору говорник може исказати различита осећања [Jovičić et al., 2004]. Неке од најчешће анализираних емоција су:

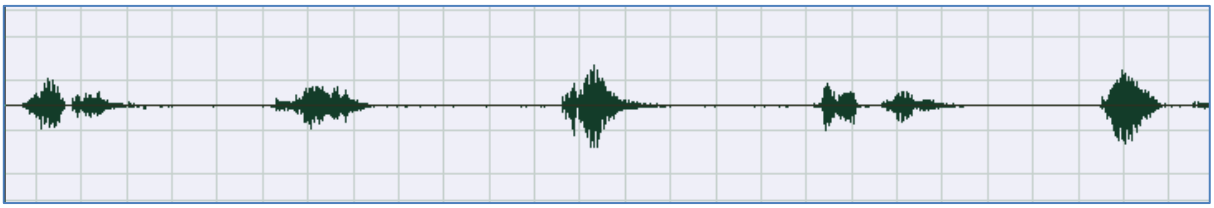
- љутња,
- срећа,
- страх,
- туга и
- неутрално расположење.

Ако се разматра стање вокалног „механизма“ говорника приликом генерисања говора, говор се може поделити на:

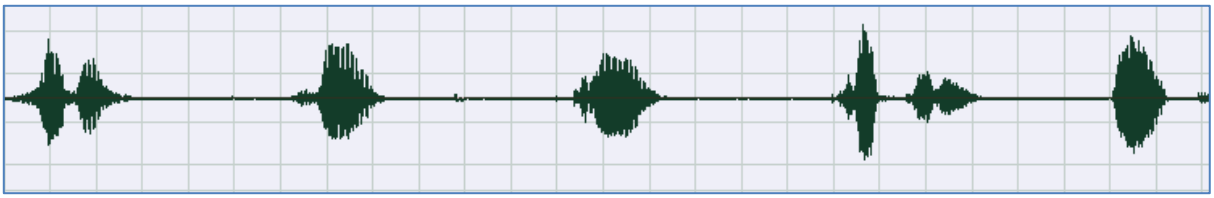
- шапат
- полутих,
- нормалан,
- гласан и
- вику.



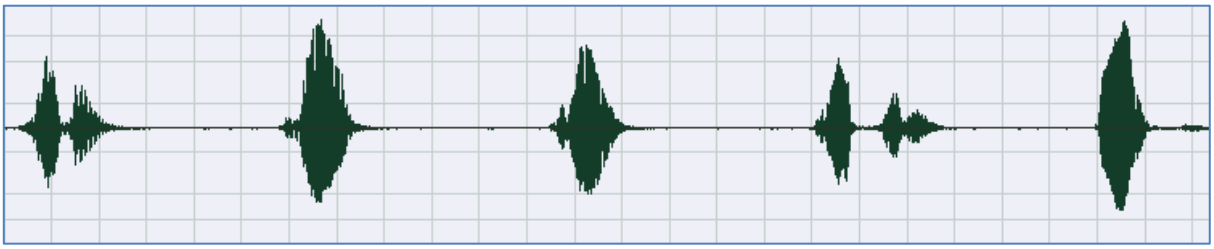
а) Шапат



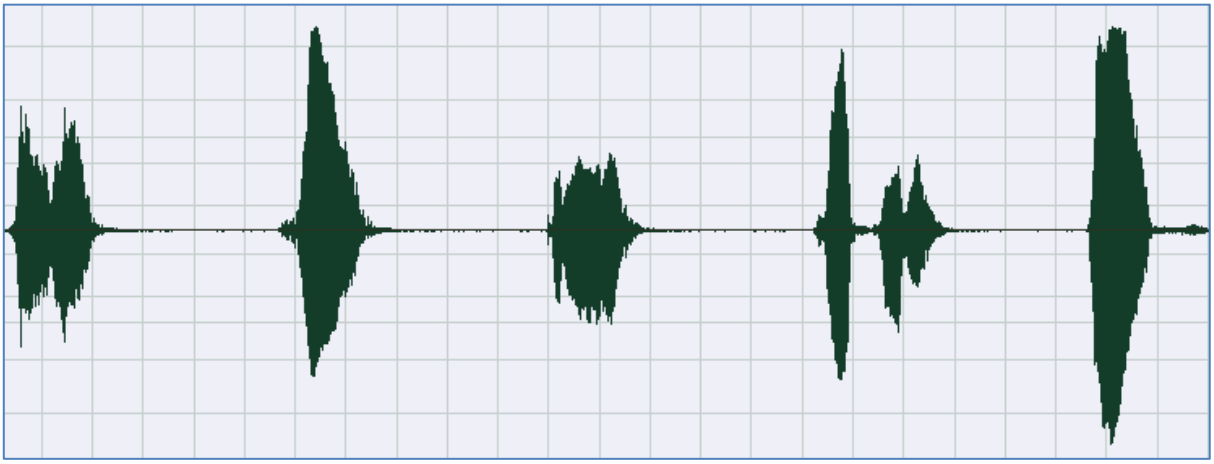
б) Полутих



в) Нормалан



г) Гласан



д) Вика

Слика 2.8 *Различити модалитети говор: од шапата до вике.*

Добар начин класификације говора са аспекта стања вокалног механизма дат је у истраживању Зенга и Хансена [Zhang, Hansen, 2007]. Разматрани су одређени параметри за класификацију говора од шапата до вике и то:

- интензитет говора,
- проценат и дужина паузе,
- дистрибуција енергије по рамовима и
- спектрални нагиб.

На основу ових параметара показано је како извршити одређивање ком моду говор припада.

Нормалан говор је најчешће присутан у свакодневној комуникацији и он је највише истраживан и најдетаљније објашњен са аспекта аутоматског препознавања говора укључујући горе поменуте спољашње услове. Основне особине су му релаксирано стање механизма за продукцију говора, слободно кретање ваздуха кроз вокални тракт и треперење гласница приликом продукције самогласника.

Полутих говор је нешто између шапата и нормалног говора. О њему је дато мало истраживања јер се може приближити нормалном говору [Hansen, 1988].

Шапат је врло актуелан за истраживање у данашње време јер по својим особинама има много специфичности које га раздвајају од остала четири мода [Zhang, Hansen, 2007], а при томе је после нормалног говора највише заступљен у свакодневној комуникацији. Посебно је интересантно поређење нормалног говора и шапата [Jovičić, 1998], [Wilson, 1998] као и могућности њиховог поређења и препознавања (нормалног на бази шапата и шапата на бази нормалног). Због тога је овај рад највећим делом посвећен овој врсти анализе.

Вика је говор који има највиши интензитет и драматичну промену вокалног тракта током екситације. Трајање реченице и пауза између речу су најкраће [Zhang, Hansen, 2007]. Често је повезана са стресним ситуацијама које могу бити узрок вике. Акустичке особине вике су детаљно описане од више аутора [Rostolland, 1982 a], [Rostolland, 1982 b], [Bou-Ghazle, Hansen, 2000]. Говор под стресом се подразумева као наглашавање одређених слогова (лингвистичка дефиниција). Говорник који производи говор под стресом је под психолошким притиском који условљава такву врсту „деформације“ говора. Стрес је патолошко стање и то је реакција особе на коју је извршен неки притисак или постављен неки задатак. Обично је говор под стресом обојен и емоцијама као што су страх, љутња, дезоријентисаност и слично [Hansen, Patil, 2007].

Гласан говор је нешто између нормалног и вике, а по својим особинама је приближан нормалном говору. Зато се и посебно истраживање о њему није детаљно спроводило.

Емотиван говор даје додатне информације о говорнику и “боји“ говора [Douglas-Cowie et al., 2003]. Он се може анализирати помоћу параметара као што су: основна фреквенција, дужина говорних сегмената/пауза и енергије на нивоу рамова података [Jovičić et al., 2004]. У неким од радова који су се бавили емоцијама коришћен је Teager Energy оператор [Georgogiannis, Digalakis, 2012]. За ову врсту говора постоји све већи број заинтересованих истраживача.

Развијени системи за аутоматско препознавање говора према врсти речи могу се поделити на системе за препознавање:

- изоловано изговорених речи,
- везано изговорених речи и

- континуалног говора.

На основу препознавања оног који генерише говор ови системи се даље деле на:

- зависне од говорика (speaker's dependent systems) и
- независне од говорника (speaker's independent systems).

На бази горњих подела приступа се и различитим техникама за предобраду и поређење говорних узорака. Предобрадом се добијају одговарајућа акустичка обележја обично у облику скупа вектора који репрезентују говорне узорке, а техником одлучивања врши се поређење улазних (тест) и референтних узорака и на бази одређених критеријума врши одлучивање која је реч (или речи) дошла на улаз система за одлучивање.

Најчешће коришћене технике за поређење говорних узорака су:

- DTW (Dynamic Time Warping) – техника динамичког усклађивања у времену,
- HMM (Hidden Markov Models.) – техника скривених Марковљевих модела,
- NN (Neural Networks) – техника неуронских мрежа и
- различита хибридна решења.

Свака од поменутих техника има своје предности и мане, а све оне налазе своје место у системима за аутоматско препознавање говора.

Предности DTW методе су што не захтева велики број узорака за поређење. То је посебно значајно код мањих речника и када „брзо" могу да се реализују алгоритми за одлучивање. Интересантна је највише за системе који су зависни од говорника [Марковић, 2004].

HMM метода ради на принципу прелаза из стања у стање са одговарајућом вероватноћом генерисања одређених вектора по стању. То је двоструко стохастички процес и често се за препознавање говора користи модел „с-лева-на-десно" [Rabiner, Juang, 1993].

NN метода је базирана на постулатима неуронских мрежа. Формира се скуп чворова и скуп нивоа/слојева (од улазног, преко средњих до излазних слојева). Метода паралелног програмирања омогућава добијање одличних резултата у разним областима па и у препознавању говора [Kostek, 1999].

Суштина напора многих истраживача се своди на одређивању одговарајућих акустичких обележја која ће послужити да се остваре што бољи резултати препознавања. Технике за поређење које се иначе користе за препознавање нормалног говора на сличан начин се могу применити и на шапат узимајући у обзир одговарајуће специфичности шапата и могуће корекције усклађивања са шапатом.

2.5.1 НОРМАЛАН ГОВОР

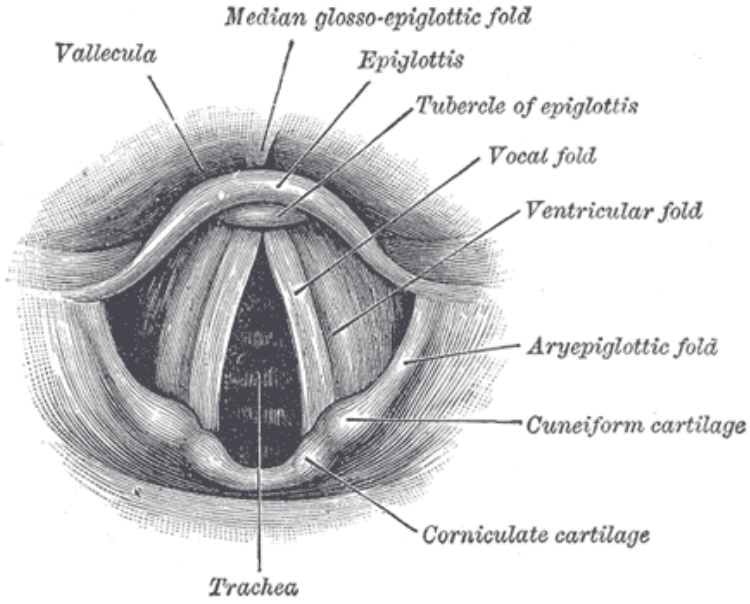
Природан говор (нормалан) подразумева креирање звучног таласа који садржи одређену информацију која се жели пренети. Нормалан говор претпоставља да је говорник у опуштеном стању, да су делови вокалног механизма исправни (да су здрави) и да нема посебних препрека на путу генерисања звучног таласа.

Ова врста говора се проучава већ дуже време и највише је истражена област. Различити су аспекти проучавања: од акустичких, преко артикулационих, физиолошких до прецептивних. Развијен је велики број система за аутоматско препознавање говора и сви се они углавном базирају на напред поменутим методама (динамичко усклађивање у времену, скривени Марковљеви модели са Гаусовим микстурама, неуронске мреже итд.). Данас је посебно актуелно коришћење дубоких неуронских мрежа (DNN–Deep Neural Networks) у проблемима овог типа [Ghaffarzadegan et al., 2016].

2.5.2 ШАПАТ

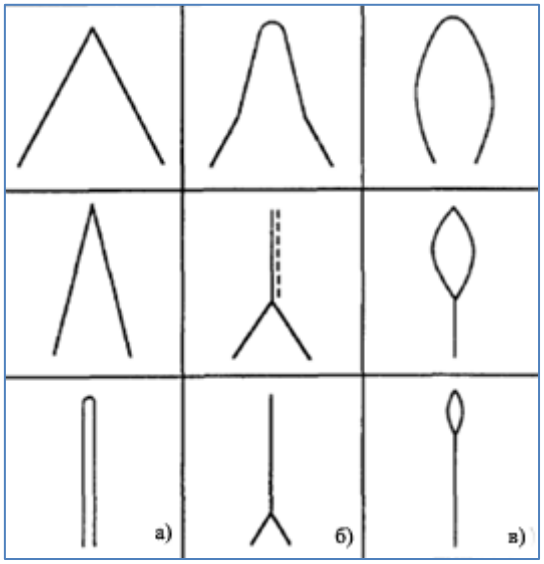
Шапат представља специфичну врсту говора која може настати из основна два разлога: особа жели да шапуће и због тога подешава вокални тракт тако да се гернерише тих, специфичан говор или особа је имала проблема са ларинксом (нпр. услед болести) и због тога није у могућности да генерише нормалан говор већ тај говор подсећа највише на шапат.

Ова врста говора има низ својих специфичности које га разликују у односу на горе поменуте врсте. Пре свега гласнице (Vocal folds – слика 2.9) не трепере за време изговора звучних гласова [Catford, 1977]. Ваздух који иде из плућа са циљем да произведе звук долази до pharynix-a, а овај тако подешава облик да гласнице не трепере [Gavidia-Ceballos, 1995], [Gavidia-Ceballos, Hansen, 1996]. Услед тога јавља се турбулентно кретање ваздуха у делу испод глотиса и на тај начин ствара се сигнал који је комплексан и налик шуму [Mathur et al., 2012]. Слика 2.9 даје детаљан изглед отвора глотиса и елемената који га окружују (са горње стране). Поред глотиса и гласница значајно место у генерисању шапата заузимају и области „лажних“ гласница (на слици 2.9 приказане као Ventricular folds). Кроз отвор глотиса види се трахеа.



Слика 2.9 *Глотис-детаљно [Wikipedia].*

Облици глотиса могу бити различити, а истраживања су показала да су следећи доминанти: облик инверзног “V” (слика 2.10а), облик инверзног “Y” (слика 2.10б) и облик савијеног (напрегнутог, извијеног) предњег дела глотиса (слика 2.10в) [Solomon et al., 1989].



Слика 2.10 *Облици глотиса при шапату [Solomon et al., 1989].*

Такође, временски облик таласа који репрезентује шапат је много мање амплитуде него што је то случај са нормалним говором што говори о односима енергија ових таласа (слика 2.8а и в). Спектрални нагиб се такође значајно разликује у односу на нормалан говор.

Шапат се може и класификовати на више начина. Једна од њих је на „меки“ шапат и „позоришни“ шапат [Sharifzadeh et al., 2009]. „Меки“ је онај који се обично говори некоме на

уво, док је „позоришни“ онај којим „шаптач“ шапуће глумцу на сцени текст (ако га је глумац заборавио).

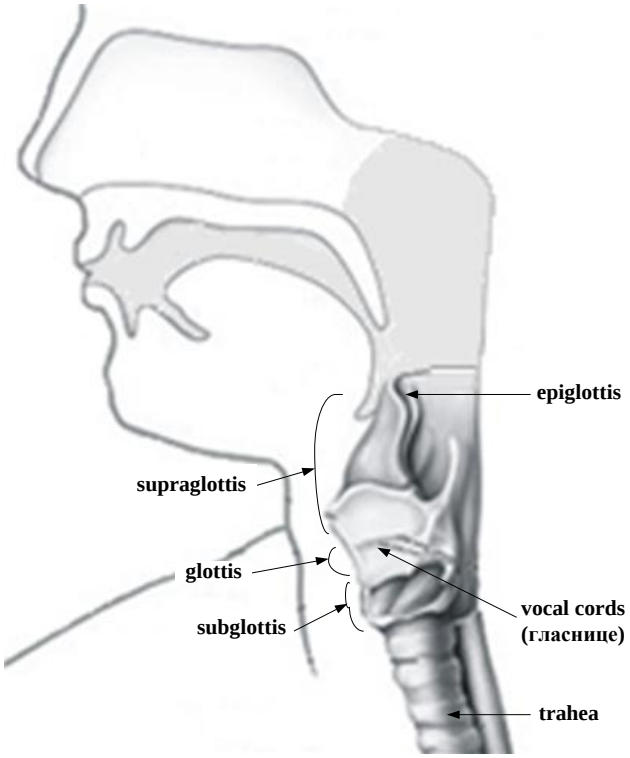
Сандберг и остали [Sundberg et al., 2010] класификују шапата у четири врсте: хиперфункционални (hyperfunctional), неутрални (neutral), хипофункционални (hypofunctional) и шапат после нормалног говора (postphonation). Неки од истраживача су разматрали хипер функционисање ларинкса током шапата на бази оптичких сензора [Rubin et al., 2004].

Посебно је интересантна класификација на „добар“ и „лош“ шапат [Hansen, 1988], [Fan, Hansen, 2010], [Fan, Hansen, 2011] која може бити врло важна за одређивање узорака приликом тернирања и приликом одлучивања. Предложена решења подразумевају коришћење спектралног нагиба и односа енергија које се налазе у областима од 1000-2000Hz и од 1000-8000Hz. На бази тога добија се SR (Spectral Representative) параметар који се може користити за процену квалитета шапата. Ако се SR параметар комбинује са SNR (Signal-to-Noise Ratio) параметром добија се дводимензиони простор поверења за „добар“ односно „лош“ шапат. Постоје различите технике за побољшање SNR параметра [Boll, 1979]. Анализом „доброг“ и „лошег“ шапата закључило се да је шапат „добар“ ако има више спектралне енергије у горњем делу скале (на вишим фреквенцијама од 2000Hz). Такође, прилико шапата фереквенције форманата F3 и F4 се скоро и не померају ка вишим фреквенцијама (за разлику од F1 и F2) па је та чињеница добра за идентификацију говорника.

Постоји већи број и других истраживања и анализа везаних за шапат и на основу њих могу се изнети неки општи закључци. Посебно су занимљива она која су подразумевала анализу форманата [Tartter, 1986], [Jovičić, 1998], [Benesty et al., 2008]. У својим акустичким и артикулационим анализама везаним за шапат, а примењеним на дуге вокале српског језик, (анализирани су: централна фреквенција форманта и ширина спектра (bandwith) форманта) Јовичић показује да:

- фреквенције форманата (првог и другог форманта) за самогласнике /и/, /е/, /а/ и /о/ су померене на више док су фреквенције свих форманата за /у/ померене на ниже (у односу на нормалан говор);
- ширина спектра за форманте шапата је већа него код нормалног говора;
- спектар прва три форманта за шапат је прилично раван;
- логаритамска вредност спектралне снаге у области од 200Hz до 2000Hz је равна;
- мања фарингеална шупљина за прва четири самогласника: условљава фреквенцију првог форманта па следи да је и фреквенција виша; обрнуто када је у питању вокал /у/ (јер има већу фарингеалну шупљину);
- највеће повећање ширине спектра форманта је за први формант, а најмање за четврти.

У својој акустичкој анализи вокалног тракта при генерисања шапата Матсуда и Касуја [Matsuda, Kasuya, 1999] су вршили 3D мерења коришћењем MRI (Magnetic Resonance Image) технике. Током шапата супраглотална структура (supraglottis - слика 2.11) није била само укочена већ и померане на доле - тако да је спречавала вибрирање гласница (vocal cords) [Tsunoda et al., 1997].



Слика 2.11 *Глотални делови који утичу на шапат.*

Закључено је да се због тога фреквенције нижих форманата померају ка вишим вредностима. Такође, при шапату се јавља и турбулентно кретање ваздуха које подсећа на шум. Ово кретање ваздуха током шапата се ствара на око 0,85 cm ниже у глотису. Коришћен је ларинарни ендоскоп за посматрање (провучен кроз назални тракт) ларинкса. Мерењем је утврђено да за шапат важи:

- супраглотална структура (supraglottis) је сужена у областима „лажних“ гласница (Ventricular folds) и гласнице су „покривене“ са „лажним“ гласницама.
- глотис је отворен са малим додатком.

Редукује се простор између глотиса и врха epiglottis-а (при шапутању) посебно у областима „лажних“ гласница. Промене вокалног тракта услед промена величине глотиса такође су овде разматране.

Закључак овог истраживања је да слаба акустичка веза са подглоталним системом (subglottis) и сужење у области „лажних“ гласница су главни узроци раста нижих фреквенција форманата код шапата.

Тартер [Tartter, 1986] даје одређену анализу шапата и то перцептивну и акустичку. Ова анализа је базирана на 18 фонема који су се састојали из комбинације консонант-вокал (за вокал је коришћен /a/), а који су изговарани шапатом. Показало се да је основна фреквенција, као база за идентификацију говорника, одсутна у шапату. Такође је уочено да жене производе више фреквенције за самогласнике него мушкарци. И други истраживачи су разматрали разлике шапата код женског и мушког пола и могућност идентификације полова [Lass et al., 1976], [Eklund, Traunmuller, 1996], [Smith, 2016]. Мерени су: трајање фрикције, фреквенција за прва три форманта, време транзиције, „burst“ и аспирација, трајање назалне резонанције. За анализу грешака коришћене су матрице конфузије (према Милеру и Никелију [Miller, Nicely, 1955]). Закључено је да, када су у питању фрикативи, безвучни консонанти имају дуже трајање артикулације него звучни консонанти.

Безвучни консонанти код шапата и нормалног говора су слични. Када су у питању звучни консонанти и вокали – велика је разлика за шапат и нормални говор. Звучни сугласници, када је шапат у питању, мигрирају ка безвучним. Облик, структуру и утицај вокала у комбинацији са консонаната разматрало је више аутора [Tartter, 1991], [Matsuda et al., 2000], [Jovičić, Šarić, 2008]. Показало се да је спектрална енергија безвучних гласова изражена на вишим фреквенцијама, док је код звучних, а посебно вокала, изражена на нижим фреквенцијама.

Пошто је већина спектралних информација о безвучним гласовима смештена у горњем делу спектра (вишим фреквенцијама) то векторско обележје типа MFCC које фаворизује ниже фреквенције не може најбоље описати ову врсту говора (изражену у шапату). Због тога се препоручују друге врсте скала (као што су експоненцијална, линеарна и/или мешовита скала) па на основу тога и друга векторска обележја. Резултати на бази кепстралних коефицијената рачунатих на линеарној фреквенцијској скали (LFCC), као и резултати на бази кепстралних коефицијената рачунатих на експоненцијалној фреквенцијској скали (EFCC) дали су боље резултате за безвучне консонанте у шапату у односу на оне који су добијени са обележјем рачунатом на „mel“ скали (MFCC).

Поједини аутори су вршили и анализе активности мозга приликом афоније (која је била узрокована бронхитисом и ларингитисом), као и приликом престанка проблема [Tsunoda et al., 2012]. Снимање активности мозга је вршено помоћу f-MRI (functional Magnetic Resonance Imaging) метода. Уочене су одговарајуће разлике можданих активности приликом шапутања и афоније код пацијента.

2.5.3 НОРМАЛАН ГОВОР И ШАПАТ – ПРЕГЛЕД ИСТРАЖИВАЊА

Основне разлике између нормалног говора и шапата које се огледају у артикулационим и акустичим особинама манифестују се и у процесу препознавања говора. У том смислу и успешност препознавања нормалног говора у односу на шапат је знатно боља. Многи од истраживача се труде да ову разлику у успешности препознавања минимизирају.

Осим усаглашених сценарија који подразумевају тренинг/тест сценарије типа „нормалан/нормалан“ и „шапат/шапат“ посебно је интересантно како ће систем који је трениран нормалним говором препознавати шапат (сценарио „нормалан/шапат“) као и како ће систем који је трениран шапатом препознавати нормалан говор (сценарио „шапат/нормалан“).

У литератури су приказани различити експерименти са различитим говорним корпусима, различитим векторским обележјима и различитим методама препознавања. Неки од најинтересантнијих су овде наведени.

Јапански истраживачи [Ito et al., 2005] су осим нормалног говора и шапата снимали и израз лица приликом формирања говорног корпуса. Добијени резултати на бази монофонског HMM система за препознавање и великог речника су били за „нормалан/шапат“ око 40% док за „шапат/нормалан“ око 56% са одговарајућом MLLR (Maximum Likelihood Linear Regression) адаптацијом.

На сличан начин Фан и Хансен [Fan, Hansen, 2011], на бази акустичке разлике између нормалног говора и шапата, предлажу коришћење линеарних и експоненцијалних фреквенцијских скала да би што боље „испратили“ спектре форманата који се померају ка вишим фреквенцијама (за идентификацију говорника). У експериментима су коришћена векторска обележја типа MFCC, LFCC и EFCC као и одговарајуће комбинације, а резултати добијени за идентификацију (зависно од говорника и врсте текста) су били за сценарио „нормалан/шапат“ 79,29% (MFCC), 88,35% (MFCC+LFCC) и 88,14% (MFCC+EFCC). Резултати за сценарио „шапат/нормалан“ су били врло слаби и то око 10%. У овим истраживањима потенциран је развој система који би на бази тренирања са нормалним говором препознавао говорника који шапуће (али не и обрнуто).

Радови на томе да се добије такозвани „псеудо-шапат“ рађени су од стране Гафарзадегана и осталих [Ghaffarzadegan et al., 2015], [Ghaffarzadegan et al., 2016]. Направљен је такозвани „псеудо модел“ који се тренирао и са нормалним говором и са шапатом. Користио се VTS (Vector Taylor Series) алгоритам. Овај модел је комбинован са VTLN (Vocal Tract Length Normalization) моделом [Lee, Rose, 1996] и SFN (Shift Frequency Normalization) моделом [Boril, Hansen, 2010] и показао је знатан добитак у препознавању говора. Векторска обележја коришћена у овим експериментима су била типа MFCC и PLP.

Разматрана су препознавања „зависно од говорника“ и „независно од говорника“. Треба нагласити да је овај модел препоручен када се поседује „мала количина“ шапата. Показано је да се успешност препознавања у овом случају повећава за око 10% (када је у питању препознавање шапата) и 0,5% (када је у питању препознавање нормалног говора).

Методу за проналажење делова шапата унутар нормалног говора предложили су Матур и остали [Mathur et al., 2012]. Ова метода је базирана на рачунању спектралних односа добијених методом линеарне предикције (LP-Linear Prediction) и методом минималне варијансе немодификованог одзива (MVDR–Minimum Variance Distortion-less Response). Дошло се до закључка да делови где је спектрална разлика глатка представљају део шапата унутар нормалног говора. Међутим, ова метода не даје и тачне прелазе са нормалног говора на шапат и обрнуто. Стога се за такве потребе користила BIC (Bayesian Information Criterion) метода [Acquah, 2010], [Tritschler, Gopinath, 1999]. Предложена LP-MVDR метода се показала и отпорна на шум, тј. дала је добре резултате за различите односе SNR. За препознавање говора је коришћен HMM модел трениран са нормалним говором уз одговарајућу адаптацију према деловима који су шапат. Векторско обележје је било MFCC типа. Адаптација је урађена коришћењем MLLR (Maximum Likelihood Linear Regression) методе [Gales, 1996], [Gales, Woodland, 1996] при чему су средње вредности и варијансе „померана ка“ адаптационим подацима. Разматрана је и примена овог поступка за препознавање шапата при коришћењу мобилних телефона.

И други аутори су се бавили проналажењем „острва шапата“ унутар нормалног говора, а пример је и коришћење ентропије [Zhang, Hansen, 2011].

У раду Шарифзадеха и осталих [Sharifzadeh et al., 2009] разматрана је могућност како да се пацијентима после операције, којом је проузрокована промена вокалног тракта (тако да могу само да шапућу), омогући облик говора који је приближан нормалном (дакле трансформација шапата у нормалан говор). Да би се произвео што разумљивији нормалан говор предложено је коришћење модификованог CELP (Code Excited Linear Prediction) кодека тако што се анализирају, модификују и реконструишу делови шапата [Ahmadi et al., 2008], [Sharifzadeh et al., 2008]. CELP омогућава да се шапат декомпонује према вокалном тракту (на екситације и „pitch“ компоненте), а потом извршивши прилагођавање ових параметара убацивањем значајних „pitch“ сигнала и на тај начин да се шапат приближи нормалном (померањем локације форманата и убацивањем поменутих сигнала). За разлику од стандардног CELP кодека [Kondoz, 1994], модификовани користи шаблон за „pitch“ који одговара процени нивоа сигнала. Параметри прилагођења се користе са циљем генерисања „pitch“ фактора и примењује се неопходна LSP (Line Spectral Pairs) модификација [McLoughlin, 2007]. И други аутори у својим истраживањима везаним за шапат одређивалу су „pitch“ [Thomas, 1969].

Интересантна су и истраживања везана за одређивање равни форманата F1xF2 за самогласнике шапата у енглеском језику. То је рађено са циљем да се шапат код пацијената (после проблема са фариниксом) преведе у нормалан говор [Sharifzadeh et al., 2012].

Поједини аутори су разматрали и употребу других врста микрофона (микрофоне који бележе вибрацију коже и костију и који су прикачени на грло - throat microphone), као и одговарајуће врсте адаптација артикулационих могућности како би повећали успешност препознавања немуштог говора [Jou et al., 2005].

Истраживања везана за препознавање нормалног говора и шапата са аспекта усаглашених и неусаглашених сценарија вршена су за српски језик коришћењем скривених Марковљевих модела и дата су у радовима Галића ([Galić et al., 2011], [Galić et al., 2013 a], [Galić et al., 2013 b] и [Galić et al., 2014 a]). Разматрано је коришћење модела за монофоне, трифоне и целе речи [Galić et al., 2014 b], а коришћена су различита векторска обележја (LFCC, MFCC, PLPCC, итд). Софтверски алат за реализацију HMM-а је био HTK [HTK]. Део ових резултата је приказан у Прилогу А.

На истом говорном корпусу, коришћењем неуронских мрежа, разматрано је препознавање за поменуте врсте сценарија у радовима Гроздића ([Grozdić et al., 2013 a], [Grozdić et al., 2013 b] и [Grozdić et al., 2013 c]). Коришћена су векторска обележја типа MFCC, TEMFCC, TECC и тд. Интересантни су резултати добијени коришћењем инверзног филтрирања у процесу препознавања шапата [Grozdić et al., 2014], а такође и анализа одређених техника за нормализацију [Grozdić et al., 2015], [Grozdić et al., 2017].

И други софтверски алати су коришћени у препознавању шапата, као што је нпр. Kaldi [KALDI]. Пример су радови пољских истраживача [Kozierski et al., 2016].

Многи истраживачи који су се фокусирали на разматрање нормалног говора и шапата своје експерименте су базирали на налажењу начина како да се на основу постојећих знања из области препознавања нормалног говора иста примене и на шапат. У том смисли постоји стална потреба за развијањем нових, напреднијих алгоритама и техника. Стога и ово истраживање даје предлог нових, бољих решења за препознавање мултимодалног говора са различитим векторским обележјима, при различитим сценаријима и типовима вектора.

3. ВЕКТОРСКА ОБЕЛЕЖЈА

У циљу обраде и поређења говорних узорака у системима за препознавање говора исти се трансформишу у одговарајући скуп вектора. Процес превођења говорних узорака из облика говорних фајлова у скуп вектора назива се предобрада (претпроцесирање) говорног сигнала. На овај начин речи се претварају у одговарајуће векторске репрезенте па је њихово поређење могуће коришћењем различитих техника од којих су најпознатије: динамичко усклађивање у времену (DTW), примена скривених Марковљевих модела (HMM), коришћење неуронских мрежа (NN) као и разна хибридна решења.

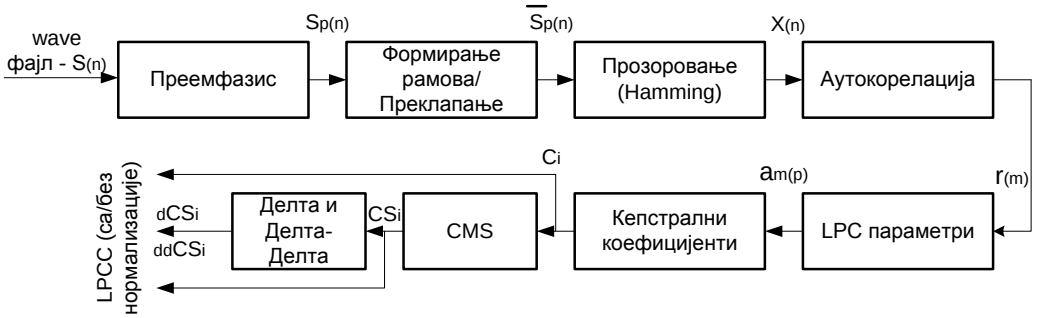
У најпознатија и најчешће коришћена векторска обележја убрајају се LPCC (Linear Prediction Cepstral Coefficients), LFCC (Linear Frequency Cepstral Coefficients), MFCC (Mel Frequency Cepstral Coefficients), GFCC (Gammatone Filterbank Cepstral Coefficients), PLPCC (Perceptual Linear Prediction Cepstral Coefficients), RASTACC (RelAtive SpecTrA Cepstral Coefficients) и сл.

Применом Teager Energy оператора на неке од њих могу се добити обележја типа TELFCC, TEMFCC, TEGFCC (у литератури познато и као TECC [Dimitriadis et al., 2005]), као и нова обележја типа TEPLPCC и TERASTACC.

Уколико се користе и одговарајуће методе за нормализацију (као што су CMS, RASTA и сл.) могуће је добити додатне скупове вектора и са њима вршити одговарајућа тестирања и анализе.

3.1 ВЕКТОРСКО ОБЕЛЕЖЈЕ ТИПА LPCC

Добијање векторског обележја на бази линеарног предикционог кодирања (LPC) је врло распрострањено и дуго времена било доминантно код система за аутоматско препознавање говора јер поседује погодан математички алат. Шематски приказ блок дијаграма за добијање кепстралних коефицијената и њихових извода базираних на LPC-у, а који укључује и могућност коришћења CMS нормализације, приказан је на слици 3.1.



Слика 3.1 *Блок дијаграм за добијање векторских обележја LPCC типа.*

На улазу у систем долази сигнал $S(n)$ који је у облику wave фајла (дигитализован говорни сигнал), а на излазу се добија скуп вектора састављених од кепстралних коефицијената. Ови коефицијенти су обележени са c_i (ако не пролазе кроз блок за нормализацију), односно са cs_i (ако пролазе кроз блок за нормализацију). Такође су размотрени и изводи (деривати) кесптралних коефицијената и то за случај када кепстрални коефицијенти подлежу нормализацији. Први извод (делта) кепстралних коефицијената је обележен са dcs_i, а други извод (делта-делта) са $ddcs_i$.

На бази ове предобраде креирају се четири врсте вектора за поређење и то:

- $V1=\{c_1,c_2,..c_R\}$ - вектор који је састављен од кепстралних коефицијената који нису нормализовани,
- $V2=\{cs_1,cs_2,..cs_R\}$ - вектор који је састављен од кепстралних коефицијената који су нормализовани,
- $V3=\{cs_1,cs_2,..cs_R,dcs_1,dcs_2,...,dcs_R\}$ - вектор који је састављен од кепстралних и делта кепстралних коефицијената који су нормализовани,
- $V4=\{cs_1,cs_2,..cs_R,dcs_1,dcs_2,...,dcs_R,ddcs_1,ddcs_2,...,ddcs_R\}$ - вектор који је састављен од кепстралних, делта и делта-делта кепстралних коефицијената који су нормализовани.

Број кепстралних коефицијената по вектору је R и за случају овог истраживања користи се $R=12.$

Систем предобраде за добијања LPCC векторских обележја је састављен од релативно независних блокова који су назначени на претходној слици. Функционалност сваког од њих ће бити појединачно размотрена.

3.1.1 ПРЕЕМФАЗИС

На почетку предобраде користи се преемфазис за обликовање спектра. Дигитални сигнал се тиме пропушта кроз FIR (Finite Impulse Response) филтер који има задатак да „испегла“ сигнал у фреквенцијском домену и тако га учини мање осетљивим на ефекте ограничене прецизности у каснијој обради. Филтер за преемфазис може бити или фиксан или споро адаптиван [Rabiner, Juang, 1993]. Најчешће се користи фиксни филтер првог реда са једним коефицијентом (ознака *a*) чија преносна карактеристика (у *z* домену) је дата са:

$$H(z)=1+\frac{a}{z} \quad \text{где је} \quad a\in[-1,-0.4] \tag{3.1}$$

Стога је излазни сигнал из блока за преемфазис дат у облику:

$$S_p(n)=S(n)-a*S(n-1) \tag{3.2}$$

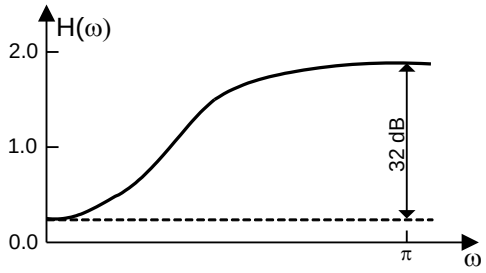
Филтер за преемфазис има за циљ да подигне спектар сигнала за око 20 dB/dec, а разлози су што:

- делови говора имају слабљење 20dB/dec па овај филтер ради инверзну операцију,
- уво је више осетљиво на опсег фреквенција изнад 1kHz, а овај филтер управо појачава тај опсег.

Пример често коришћеног филтра за преемфазис је:

$$H(z) = 1 - \frac{0.95}{z} \tag{3.3}$$

У пракси је вредност за коефицијенат *a* око 0,95, а једна од честих имплементација са фиксном тачком је и *a*=15/16=0,9375.



Слика 3.2 *Преносна карактеристика преемфазиса за* a *=0,95.*

Преносна карактеристика са слике 3.2 приказује како преемфазис фаворизује више учестаности и подиже их и преко 20dB/dec.

3.1.2 ПРЕКЛАПАЊЕ И ПРОЗОРОВАЊЕ

После блока за преемфазис сигнал $S_p(n)$ долази на блок у коме се врши формирање рамова, преклапање и прозоровање. Најпре се цео сигнал подели на рамове дужине *N* одмерака. Потом се формирају помоћни рамови дужине *M* при чему важи да је *M<N*, а затим се врши преклапање на дужини *N-M*. На овај начин се добија нови скуп рамова који су дужине *N* и који у себи садрже информације и од других рамова (због преклапања) и представљају нову репрезентацију посматраног говорног узорка. Преклапање се обично врши на пола или на трећину, а може и на друге вредности.

Прозоровање сигнала се своди на множење вредности сваког индивидуалног рама (који је настао преклапањем) са функцијом прозоровања - $w(n)$**.** Ова функција има одређену вредност у границама где је $0 \le n \le N-1$ док је ван тих граница њена вредност једнака нули. Циљ прозоровања је да се минимизује дисконтинуитет сигнала на почетку и на крају сваког од рамова. Постоје различите врсте прозора (Hamming, Hanning, Triangular [Marković**,** Luković, 2012]) али за аутокорелациону методу линеарне предикције најчешће се користи

Hamming-ово прозоровање. Сигнал подвргнут овој врсти прозоровања приказан је на слици 3.3 при чему је број одмерака у раму *N*=200, а дужина преклапања *M*=14.

Слика 3.3 *Рам од N=200 одмерака отежан Hamming-овим прозором[Rabiner, Juang, 1993].*

Математички облик Hamming-овог прозора дат је формулом (3.4):

$$w(n) = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}) \text{ за } 0 \le n \le N-1 \tag{3.4}$$

Ако се са $\overline{S_p(n)}$ обележи сигнал који је настао после формирања рамова података и њиховог преклапања, онда одговарајући сигнал после прозоровања може се представити у следећем облику:

$$X_l(n) = \overline{S_p(n)}w(n) \tag{3.5}$$

Овако добијени сигнал се даље прослеђује на блок за аутокорелациону анализу.

3.1.3 АУТОКОРЕЛАЦИОНА АНАЛИЗА

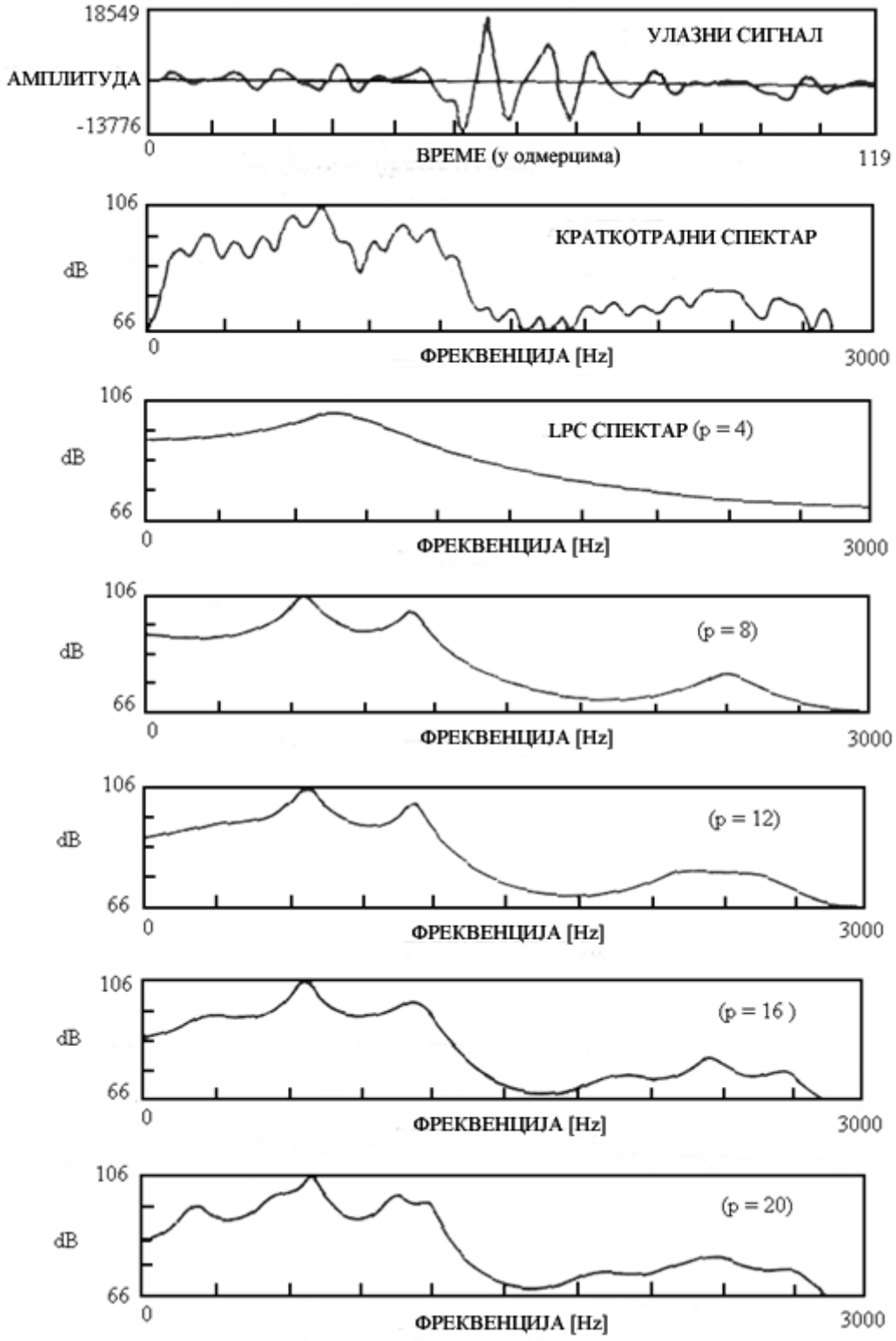
Аутокорелациона анализа је следећи блок у процесу предобраде (према слици 3.1). На сваки од рамова који је прошао отежавање Hamming-овим прозором примењује се аутокорелација. При томе значајну улогу игра ред аутокорелације који се најчешће означава са *p*. Вредност за *p* се обично узима од 8 до 16.

На слици 3.4 је дат приказ како краткотрајни LPC спектар сигнала зависи од реда аутокорелације (где *p* узима вредност од 4 до 20).

Аутокорелациони коефицијенти се одређују помоћу следеће формуле:

$$r_l(m) = \sum_{n=0}^{N-1-m} X_l(n)X_l(n+m) \text{ за } m = 0,1,2,...,p \tag{3.6}$$

Треба напоменути да нулта аутокорелација, $r_l(0)$ представља енергију l -тог рама.



Слика 3.4 *Зависност LPC спектра од реда аутокорелације p [Rabiner, Juang, 1993].*

3.1.4 LPC ПАРАМЕТРИ

За добијање коефицијената линеарне предикције користе се преходно добијене вредности аутокорелационих коефицијената тако што се сваки рам од *p+1*-не аутокорелације трансформише у скуп LPC параметара. Метода за ову конверзију познат је као Дарбинова (Durbin's method). Конверзија се врши рекурзивним поступком коришћењем следећег скупа једначина:

$$E^{(0)} = r(0) \tag{3.7}$$

$$k_i = \{r(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} r(|i-j|)\} / E^{(i-1)} \text{ где је } 1 \le i \le p \tag{3.8}$$

$$\alpha_i^{(i)} = k_i \tag{3.9}$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \tag{3.10}$$

$$E^{(i)} = (1-k_i^2) E^{(i-1)} \tag{3.11}$$

Ове једначине се решавају рекурзивно за $i = 1,2,...,p$. На бази тога добијају се тзв. **директни LPC коефицијенти** који се обележавају са a_m, а рачунају на следећи начин:

$$a_m = \alpha_m^{(p)} \text{ , } m = 1,2,...,p \tag{3.12}$$

Такође се могу користити и **рефлексиони (PARCOR - PARtial CORrelation) коефицијенти** – k_m, а у одређеним случајевима и **логаритамски коефицијенти односа** - g_m који се рачунају по формули (3.13):

$$g_m = \log(\frac{1-k_m}{1+k_m}) \tag{3.13}$$

У овом истраживању коришћени су директни LPC коефицијенти (a_m) при чему је за ред аутокорелације узето *p*=8.

3.1.5 LPC КЕПСТРАЛНИ КОЕФИЦИЈЕНТИ

Приликом поређења скупа вектора који репрезентују говорне сигнале кепстрални коефицијенти су се показали много поузданији него сами LPC коефицијенти. Због тога се у пракси спроводи додатни корак који омогућава добијање кепстралних коефицијената, а често и њихових извода: делта и делта-делта коефицијената. Ако се користе директни LPC коефицијенти (a_m) онда се одговарајући кепстрални коефицијенти могу добити помоћу формула (3.14-3.16):

$$c_0 = \ln(\sigma^2) \tag{3.14}$$

$$c_m = a_m + \sum_{k=1}^{m-1} (\frac{k}{m}) c_k a_{m-k} \text{ за } 1 \le m \le p \tag{3.15}$$

$$c_m = \sum_{k=m-p}^{m-1} (\frac{k}{m}) c_k a_{m-k} \quad \text{за } m > p \tag{3.16}$$

Овде σ^2 означава фактор добитка за LPC модел. Обично се узима R кепстралних LPC коефицијената при чему се бира да је $R \approx (3/2)p$.

Због осетљивости нижег реда кепстралних коефицијената на укупан спектрални нагиб, а такође и вишег реда на шум, ови коефицијенти се могу и додатно отежати специфичном функцијом. Поступак отежавања подразумева множење кепстралних коефицијента са функциом w_m која има облик:

$$w_m = [1 + \frac{R}{2}\sin(\frac{\pi m}{R})] \quad \text{за } 1 \le m \le R \tag{3.17}$$

3.1.6 ДЕЛТА И ДЕЛТА-ДЕЛТА КЕПСТРАЛНИ КОЕФИЦИЈЕНТИ

Изводи кепстралних коефицијента (први и други) се такође могу укључити при предобади и они, иако усложњавају, додатно побољшавају алгоритме за препознавање говора [Marković**,** et al., 2013 c].

Први извод кепстралних коефицијената (тзв. делта коефицијент) се може апроксимативно рачунати по формули:

$$\frac{\delta c_m(t)}{\delta t} = dc_m(t) \approx \mu \sum_{k=-K}^{K} k * c_m(t+k) \tag{3.18}$$

где је μ нормализациона константа, а рачунање се врши на 2*K*+1 суседних рамова. За *K* се најчешће узима вредност 3 или 4.

Други извод кепстралних коефицијената (тзв. делта-делта коефицијент) се рачуна тако што се одреди извод од делта коефицијената аналогно као што је дато у (3.18). Математички се то може представити у облику:

$$\frac{\delta dc_m(t)}{\delta t} = ddc_m(t) \approx \mu \sum_{k=-K}^{K} k * dc_m(t+k) \tag{3.19}$$

На овај начин може се добити вектор који садржи 3**R* елемената (*R* - кепстралних, *R* - делта кепстралних и *R* - делта-делта кепстралних коефицијената), а чији је облик:

$$V = \{c_1, c_2, ..., c_R, dc_1, dc_2, ..., dc_R, ddc_1, ddc_2, ..., ddc_R\} \tag{3.20}$$

Наравно, уколико се изоставе делта и/или делта-делта коефицијенати добијају се варијације овог вектора различитих дужина. У овом истраживању разматрана су сва три типа тј. варијације вектора (3.20).

3.1.7 CMS НОРМАЛИЗАЦИЈА

Нормализација се често користи у процесирању говора јер омогућава да се сачувају важне информације и пониште нежељене дисторзије. Разне методе нормализације (смањења утицаја) телекомуникационог канала су предлагане тако што се вршила компензација одговарајућих карактеристика ([Atal, 1974], [Furui, 1981]). Једна од најпопуларнијих метода је одузимање средње вредности тј. CMS (Cepstral Mean Subtraction) [De Veth, Boves, 1998]. Ова метода се показала и врло ефикасна у „уобличавању“ спектра шапата и нормалног говора посебно у случајевима када треба вршити тестирање неусаглашених сценарија [Grozdić et al., 2015]. Циљ је да кепстрални коефицијенти оба модалитета (а посебно c_0 који је одговоран за енергију и c_1 који је одговоран за спектрални нагиб) буду што више усклађени, а то се управо постиже CMS нормализацијом. Такође, ова нормализација утиче на смањење варијација које настају током изговора у истом моду.

Када говорни сигнал прође кроз временски непроменљив канал ($C(\omega)$) конволуција дисторзија постаје мултипликативна у спектралном домену, а адитивна у логаритамском спектралном домену. Пошто је кепструм линеарна трансформација логаритамског спектра, оба могу да се посматрају на исти начин. Ако се на говорни сигнал примени краткотрајна спектрална анализа – добија се $S_t(\omega)$. Такав сигнал пролази кроз канал $K(\omega)$, а резултујући спектар је $Y_t(\omega)$. Са индексом *t* обележена је временска зависност. Резултујући сигнал се може представити формулом:

$$Y_t(\omega) = K(\omega) * S_t(\omega) \tag{3.21}$$

а одговарајући кепструм (логаритам спектра) са:

$$y_t = k + s_t \tag{3.22}$$

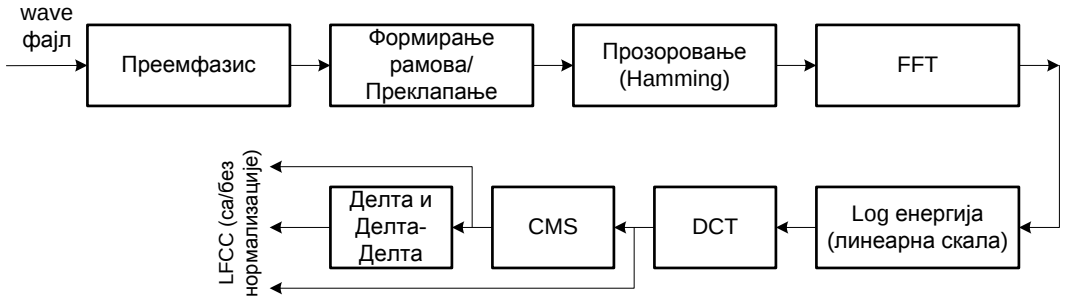
Пошто је канал константан ($K(\omega) = const$), то се може извршити компензација одузимањем средње вредности и на тај начин добити нови CMS параметар са ознаком cs_t :

$$cs_t = y_t - \overline{y_t} = k + s_t - (k + \overline{s_t}) = s_t - \overline{s_t} \tag{3.23}$$

Тако се добијају кепстрални коефицијенти на бази CMS нормализације. Потом се од њих могу добити и одговарајући делта и делта-делта коефицијенти према ознакама које су дате на слици 3.1.

3.2 ВЕКТОРСКА ОБЕЛЕЖЈА ТИПА LFCC И TELFCC

Добијање векторских обележја типа LFCC је базирано на коришћењу линеарне фреквенцијске скале. Поступак њиховог добијања дат је на основу блок дијаграма са слике 3.5.



Слика 3.5 *Блок дијаграм за добијање векторских обележја LFCC типа.*

Поступци преемфазиса, формирања рамова, преклапања и прозоровања су идентични као што је напред описано за кепстралне коефицијенте типа LPCC. После тога над добијеним рамовима се примењује брза Фуријеова трансформација (FFT-Fast Fourier Transformation), а затим се рачуна логаритам енергије према линеарној фреквенцијској скали. Након тога користи се дискретна косинусна трансформација (DCT – Discrete Cosine Transformation) и на бази ње добијају се кепстрални коефицијенти. Ако се примени нормализација (CMS) онда се могу добити и нормализовани кепстрални коефицијенти, а на основу њих и први и други изводи (делта и делта-делта коефицијенти). На овај начин добија се скуп векторских обележја који може имати различит број коефицијената по вектору (12, 24 или 36).

3.2.1 БРЗА ФУРИЈЕОВА ТРАНСФОРМАЦИЈА

Брза Фуријеова трансформација је поступак којим се рачунски врло ефикасно добија спектар сигнала у кратким временским интервалима. Много је ефикаснија од дискретне Фуријеове трансформације (DFT –Discrete Fourier Transformation). Дефинише се за ниску симбола $s(n)$ коначне дужине N на следећи начин:

$$S(f) = \sum_{n=0}^{N-1} s(n) * e^{-j\frac{2\pi * f}{f_s} n} \tag{3.24}$$

где је N дужина прозора (број одмерака), а f_s фреквенција одмеравања.

Брзом Фуријеовом трансформацијом добијају се одговарајуће вредности у еквидистантним тачкама које су дефинисане са $k * f_s / N$ па претходна једначина добија облик:

$$S(k) = \sum_{n=0}^{N-1} s(n) * e^{-j\frac{2\pi * k}{N} n} \tag{3.25}$$

где k узима вредности $k = 0,1,...,N-1$.

Често се бира да вредност за N буде облика 2^p. Тада се за рачунање брзе Фуријеове трансформације користи „radix 2“ алгоритам. Реализација брзе Фуријеове трансформације се у том случају може спровести на два начина:

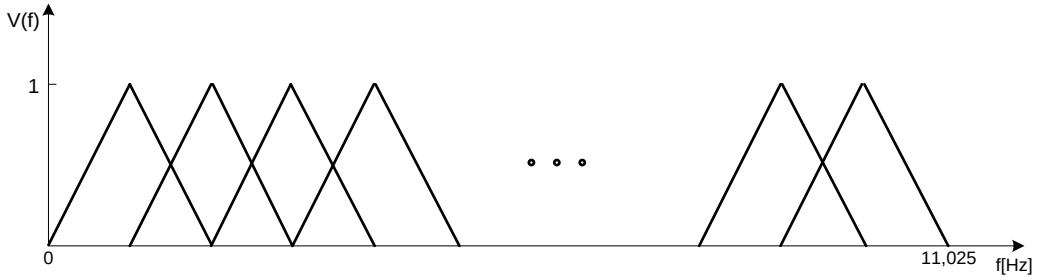
- развијањем по времену (DIT – Decimation In Time)
- развијањем по фреквенцијама (DIF – Decimation In Frequency)

Избор алгоритма програмирања и коришћење једног или другог алгоритма зависи од корисника, а оба, путем рекурзивних корака, брзо доводе до јединственог скупа решења.

Рачунање дискретне Фуријеове трансформације (DFT) подразумева број операција реда величине N^2, док за брзу Фуријеову трансформацију тај број је $N*Log_2N$. То је и главни разлог што се у пракси углавном користи брза Фуријеова трансформација.

3.2.2 ЛИНЕАРНА ФРЕКВЕНЦИЈСКА СКАЛА

Коришћење линеарне фреквенцијске скале подразумева формирање скупа филтера који су еквидистантно распоређени на фреквенцијама од 0 (најниже) до 11.025Hz (највише). Филтери су троугаони и имају преклапање на централним фреквенцијама. Подела скала на 30 подопсега, еквивалентних филтера, је приказана на слици 3.6.



Слика 3.6 *Распоред филтера на линеарној скали.*

На сваком од подопсега рачуна се логаритам енергије и та вредност се прослеђује у блок за дискретну косинусну трансформацију. Енергија сваког од подопсега се рачуна као:

$$E_i = \sum_{j=D_i}^{G_i} |X_f(j)|^2 \tag{3.26}$$

где је $i = 1,2,...N_f$, а D_i и G_i су доња и горња граница i-тог подопсега на линеарној скали. N_f представља укупан број подопсега на линеарној скали што је у овом случају 30.

3.2.3 ДИСКРЕТНА КОСИНУСНА ТРАНСФОРМАЦИЈА

После блока у коме се рачуна енергија по подопсезима, тј. њен логаритамски еквивалент, примењује се дискретна косинусна трансформација са циљем да се добију кепстрални коефицијенти. То је у ствари множење логаритма енергије са одговарајућом косинусном функцијом. Тако се кепстрални коефицијенти одређују по следећој формули:

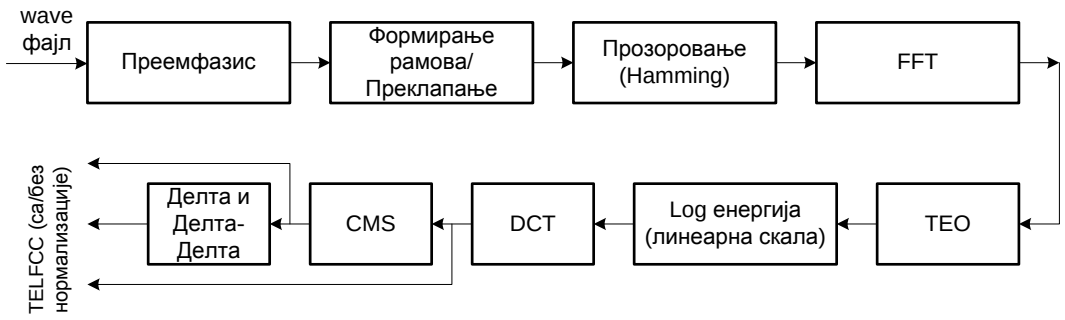
$$c_i = \sum_{j=1}^{Nf} \log(E_j) * \cos(\frac{i*(j-0.5)*\pi}{N_f}) \tag{3.27}$$

где је N_f број подопсега на скали, а $i = 1,2,...N_c \leq N_f - 1$ редни број кепстралног коефицијената. У овом раду разматрано је коришћење 12 кепстралних коефицијената.

После одређивања кепстралних коефицијената могуће је применити нормализацију (CMS) као што је представљено на слици 3.6, а такође добити изводе кепстралних коефицијената првог (делта) и другог (делта-делта) реда.

Добијање извода кепстралних коефицијената је идентично као што је описано у 3.1.6.

Векторска обележја типа TELFCC добијају се коришћењем блок шеме која је приказна на слици 3.7.



Слика 3.7 *Блок дијаграм за добијање векторских обележја TELFCC типа.*

Кораци као што су преемфазис, формирање рамова, преклапање, прозоровање Hamming-овим прозором и брза Фуријеова трансформација су идентични као што је већ назначено при добијању LFCC кепстралних коефицијената. Ново је то да се после брзе Фуријеове трансформације примењује нелинарни Teager Energy оператор (TEO) на тако добијени сигнал.

3.2.4 TEAGER ENERGY ОПЕРАТОР

Овај оператор се у литератури назива и Teager-Kaiser Energy Operator [Kaiser, 1990 a], [Kaiser, 1990 b] и омогућава опис наглих промена енергије унутар глоталног дела говорног апарата. За реалне дискретне временске сигнале овај оператор се дефинише као:

$$\Psi(x[n]) = x^2[n] - x[n-1] * x[n+1] \tag{3.28}$$

где су $x[n]$ одмерци тог реалног сигнала.

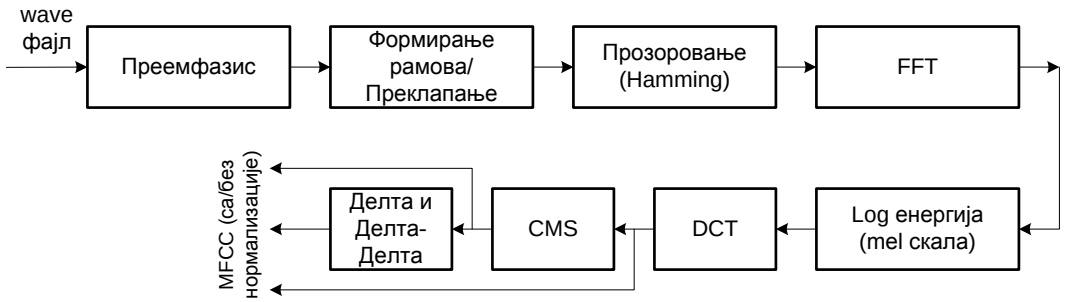
За комплексне дискретне сигнале TEO се рачуна као сума енергија реалног и имагинарног дела тог комплексног сигнала:

$$\Phi(x[n]) = \Psi(\mathrm{Re}\{x[n]\}) + \Psi(\mathrm{Im}\{x[n]\}) \tag{3.29}$$

Израчуната вредност $\Phi[]$ се отежава према филтрима линеарне фреквенцијске скале (слика 3.7). Потом следе поступци израчунавања логаритма енергије, дискретне косинусне трансформације, и тако се добијају кепстрални коефицијенти са и без нормализације. Дакле, остатак процеса је потпуно идентичан са оним који је описан за добијање LFCC коефицијената.

3.3 ВЕКТОРСКА ОБЕЛЕЖЈА ТИПА MFCC И TEMFCC

За добијање векторских обележја типа MFCC користи се блок шема приказана на слици 3.8. У питању су кепстрални коефицијенти базирани на мелодијској ("mel") фреквенцијској скали при чему се такође разматрају опције са и без нормализације.



Слика 3.8 *Блок дијаграм за добијање векторских обележја MFCC типа.*

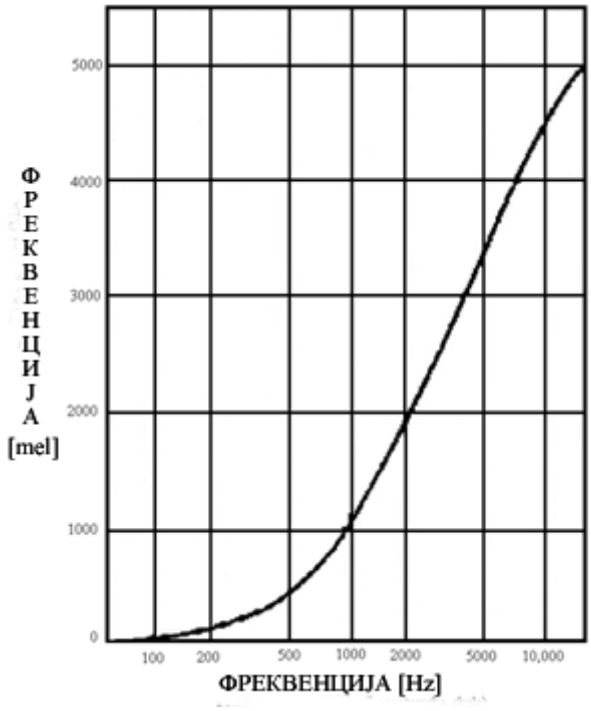
Поступак добијања MFCC коефицијената је идентичан са напред описаним поступком LFCC коефицијената. Једина разлика је да се уместо линарне скале користи мелодијска скала.

3.3.1 МЕЛОДИЈСКА ФРЕКВЕНЦИЈСКА СКАЛА

Мелодијска скала се користи да опише особину чујности ува која није линеарна. Наиме, експериментално се дошло до закључка да се чујност може апроксимирати линеарном функцијом до фреквенције од око 1000Hz, а преко те фреквенције са логаритамском функцијом [Rabiner, Juang, 1993]. Слика 3.9 даје приказ мелодијске скале у интервалу од 0 до 11.025Hz.

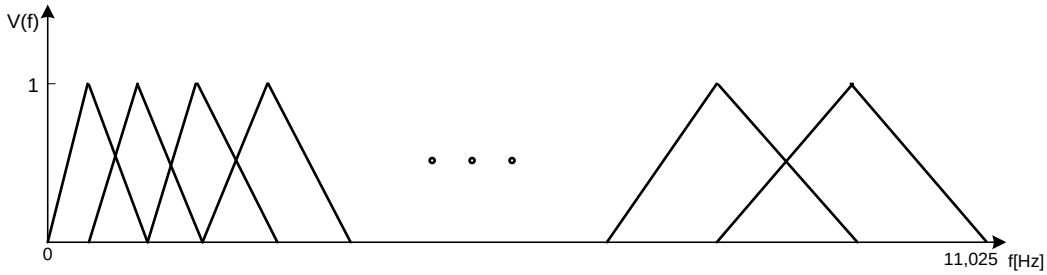
„Фреквенција" у "mel"-има се рачуна на основу следеће формуле:

$$f_{mel} = 2595 * \log_{10}(1 + f/700) \tag{3.30}$$



Слика 3.9 *Мелодијска ("mel") скала.*

После брзе Фуријеве трансформације (слика 3.8) сигнал долази на блок за отежање према "mel" скали. Овај блок се састоји од скупа троугаоних филтера који су распоређени према "mel" скали на растојању од 0 до 11.025Hz (слика 3.10). Укупан број филтера је 30.

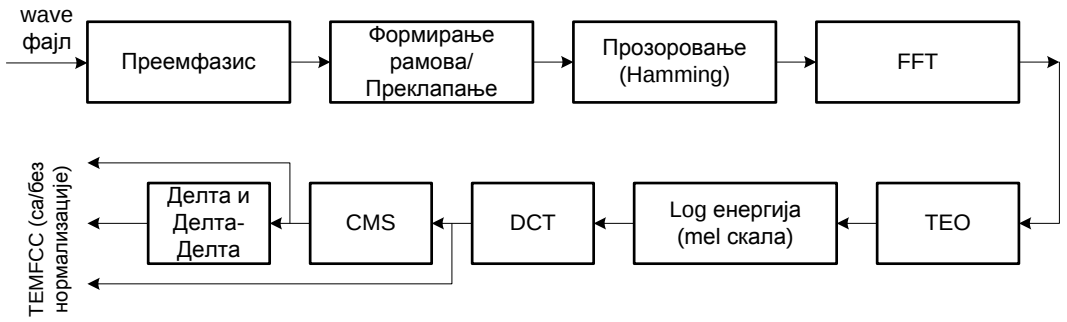


Слика 3.10 *Распоред филтера на мелодијској скали.*

После рачунања логаритма енергије на појединим подопсезима врши се рачунање кепстралних коефицијената према (3.27). Потом следи примена нормализације, па одређивање делта и делта-делта кепстралних коефицијената. Тиме се добија заокружени сет MFCC параметара.

Добијање TEMFCC векторских обележја омогућава систем приказан на слици 3.11.

На идентичан начин као што је већ описано добијање TELFCC коефицијената, добијају се и TEMFCC коефицијенти с том разликом што се уместо линеарне скале сада користи "mel" скала. Остали кораци обраде су истоветни са напред назначеним.

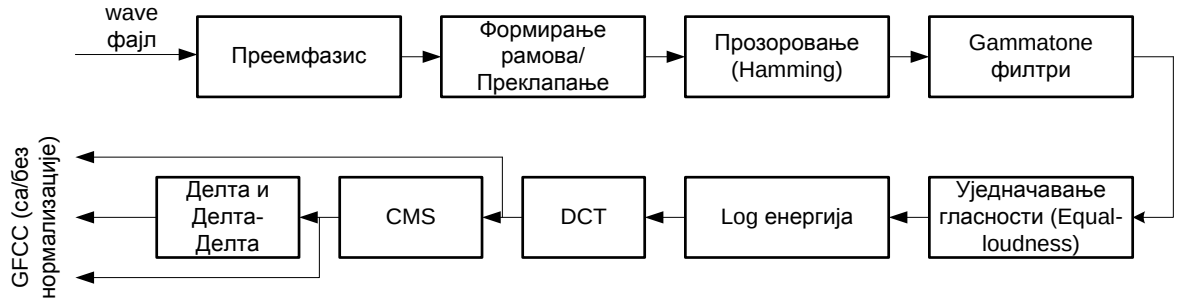


Слика 3.11 *Блок дијаграм за добијање векторских обележја TEMFCC типа.*

3.4 ВЕКТОРСКА ОБЕЛЕЖЈА ТИПА GFCC И TEGFCC

Коришћењем скупа Gammatone филтера као основе за селекцију одређеног дела у говорном спектру могу се добити Gammatone кепстрални коефицијенти (GFCC) и њихове варијације (познати у литератури и као GTCC [Cheng et al., 2005]). Популарност ове врсте филтера огледа се у доброј симулацији аудиторних процеса у људском уву [Petterson, 1992].

Систем приказан на слици 3.12 приказује начин добијања Gammatone кепстралних коефицијената са и без нормализације, као и добијање њихових извода.



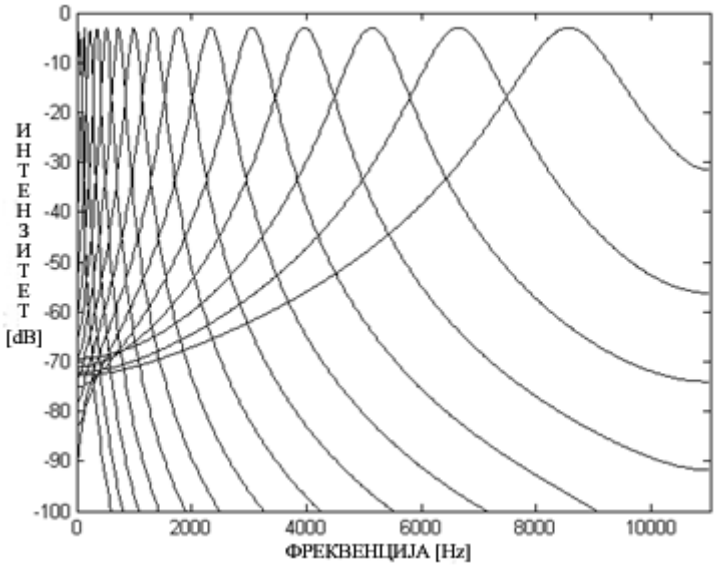
Слика 3.12 *Блок дијаграм за добијање векторских обележја GFCC типа.*

Процес предобраде започиње блоковима за преемфазис, формирање рамова и преклапање, па отежавањем Hamming-овим прозором. Затим сигнал пролази кроз скуп Gammatone филтера, па се врши уједначавање гласности (Equal-loudness), израчунавање логаритма енергије на појединим опсезима, дискретна косинусна трансформација и тако добијају Gammatone кепстрални коефицијенати (GFCC) са или без нормализације. На основу креираних кепстралних коефицијента може се одредити њихов први (делта) и други (делта-делта) извод.

3.4.1 GAMMATONE СКУП ФИЛТЕРА

Скуп Gammatone филтера се у анализи често користи зато што добро апроксимира понашање осетљивости пужа (cochlea) у људском уву. У овом раду коришћено је 30 филтера

који су распоређени на фреквенцијама од 0 до 11.025Hz као што је приказано на слици 3.13 (при чему је сваки други нацртан ради боље прегледности слике).

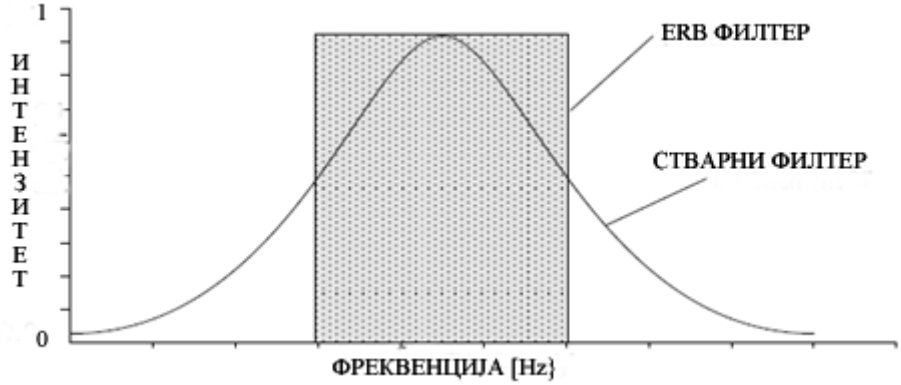


Слика 3.13 *Скуп Gammatone филтера.*

Импулсни одзив сваког од филтера се може рачунати према формули [Petterson, 1992]:

$$g(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \phi) \tag{3.31}$$

где је a константа и обично има вредност 1, n је ред филтра, ϕ фазни померај, f_c је централна фреквенција, а b (bandwidth) ширина спектра у Херцима. Централне фреквенције се распоређују према Bark скали. У пракси скуп Gammatone филтера се моделује низом правоугаоних аудио филтера који се називају ERB (Equivalent Rectangular Bandwidth) као што је приказано на слици 3.14.



Слика 3.14 *Однос ERB и стварног филтра.*

Glasberg и Moore [Glasberg, Moore, 1990] су предложили следећу везу између ERB и централне фреквенције f_c сваког од филтера:

$$ERB(f_c) = 24.7(\frac{4.37 f_c}{1000} + 1) \tag{3.32}$$

Petterson је сугерисао да ширина спектра/опсега сваког од Gammatone филтра буде:

$$b = 1.019ERB \tag{3.33}$$

Што се тиче реда n за Gammatone филтре предлаже се да буде n=4 .

После проласка кроз поменуте филтре сигнал долази на блок за уједначавање гласности.

3.4.2 УЈЕДНАЧАВАЊЕ ГЛАСНОСТИ

Крива уједначавања гласности (Equal-loudness curve) апроксимира осетљивост људског ува на звук различитих фреквенција [Hermansky, 1986]. У овом процесу предобраде сигнал који се добија као излаз из Gammatone филтера бива отежан са одговарајућом функцијом. Функција отежања за Никвистову фреквенцију која је изнад 5kHz дата је у облику:

$$E = \frac{(\omega^2 + 56.8*10^6)\omega^4}{(\omega^2 + 6.3*10^6)^2(\omega^2 + 0.38*10^9)(\omega^6 + 9.58*10^{26})} \tag{3.34}$$

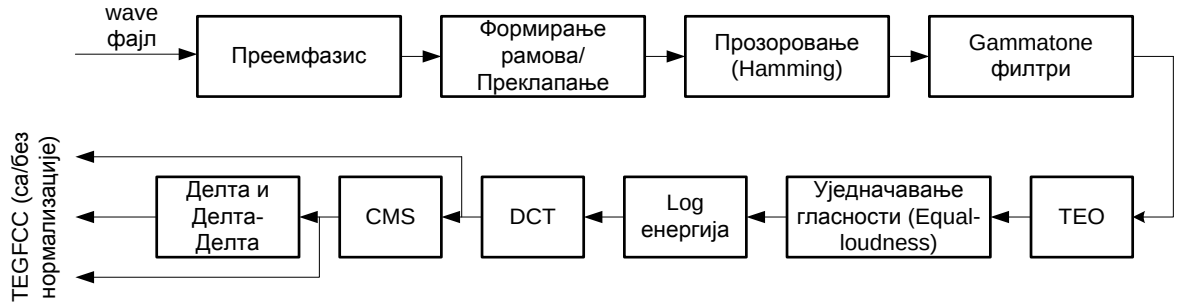
где је ω угаона фреквенција ($\omega = 2\pi f$).

У случају да је Никвистова фреквенција испод 5 kHz користи се следећа формула:

$$E = \frac{(\omega^2 + 56.8*10^6)\omega^4}{(\omega^2 + 6.3*10^6)^2(\omega^2 + 0.38*10^9)} \tag{3.35}$$

Пошто је фреквенција одмеравања за сигнал у овом раду 22.050Hz то се користи отежање дато формулом (3.34).

Добијање векторских обележја типа TEGFCC врши се према шеми која је дата на слици 3.15.

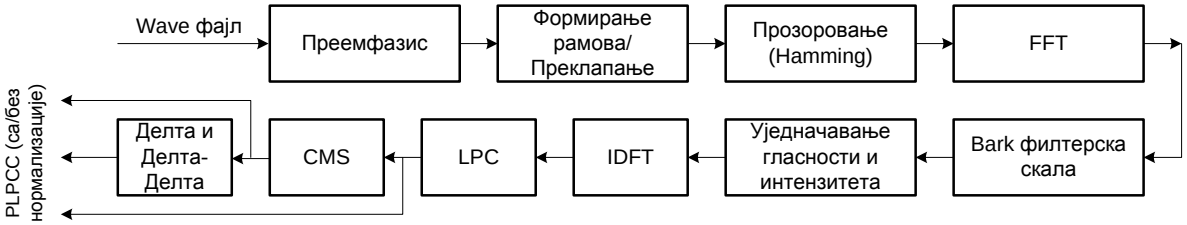


Слика 3.15 *Блок дијаграм за добијање векторских обележја TEGFCC типа.*

Блок са Teager Energy оператором се поставља иза скупа Gammatone филтера и примењује на тако филтриран сигнал. Остали елементи предобраде су идентични као што је описано за добијање GFCC коефицијената.

3.5 ВЕКТОРСКА ОБЕЛЕЖЈА ТИПА PLPCC И TEPLPCC

Скуп обележја који је у литератури познат као PLP (Perceptual Linear Predictive) представља перцептивну линеарну предикцију уведену од стране Херманског [Hermansky, 1986]. Добијање PLPCC векторског обележја са одговарајућим опцијама (нормализација, делта и делта-делта коефицијенти) приказано је на слици 3.16.



Слика 3.16 *Блок дијаграм за добијање векторских обележја PLPCC типа.*

Први део предобраде (преемфазис, формирање рамова/преклапање, прозоровање и брза Фуријеова трансформација) је идентичан као код раније описаних обележја. Затим следи скуп филтера према Bark скали. Bark скала дефинише фреквенције на основу једначине:

$$f_{Bark} = 6 * \ln(\frac{f}{600} + ((\frac{f}{600})^2 + 1)^{0.5}) \tag{3.36}$$

Централне фреквенције филтера су једнако распоређене на основу Bark скале, а према препоруци Херманског између њих је размак од 1 Bark. Облик филтера је идентичан и задовољава следеће релације:

$$\psi = \begin{cases} 0 & f_{Bark} - f_{c(Bark)} < -2.5 \\ 10^{(f_{Bark} - f_{c(Bark)} + 0.5)} & -2.5 \le f_{Bark} - f_{c(Bark)} \le -0.5 \\ 1 & -0.5 < f_{Bark} - f_{c(Bark)} < 0.5 \\ 10^{-2.5(f_{Bark} - f_{c(Bark)} - 0.5)} & 0.5 \le f_{Bark} - f_{c(Bark)} \le 1.3 \\ 0 & f_{Bark} - f_{c(Bark)} > 1.3 \end{cases} \tag{3.37}$$

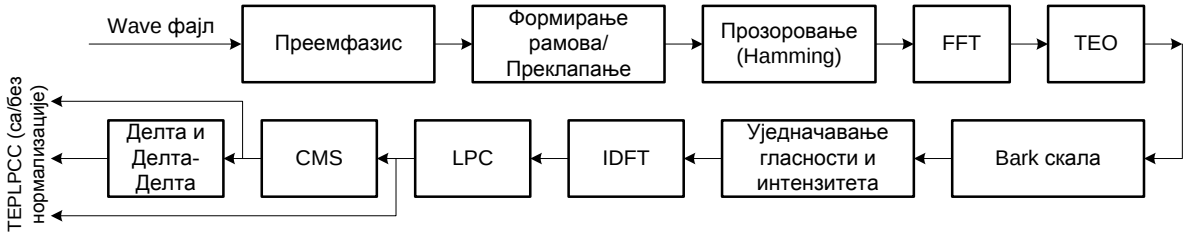
Уједначавање гласности се рачуна на идентичан начин као што је наведено у једначини (3.34). Потом се користи компресија интензитета гласности. Она апроксимира нелинеарни однос између интензитета звука и перцепције гласности. Одговарајућа релација за компресију интензитет гласности је дата са:

$$\Phi_m = (\mathrm{X}_m)^{0.33} \qquad 1 \le m \le M \tag{3.38}$$

где је m ред филтра, а M=30 укупан број примењених филтера.

Потом се примењује инверзна дискретна Фуријеова трансформација (IDFT – Inverse Discrete Fourier Transformation) на сигнал Φ_m. Следи добијање LPC коефицијената (поступком који је објашњен у делу 3.1), а од њих кепстралних са одговарајућим варијацијама (са и без нормализације, делта и делта-делта).

Добијање TEPLPCC векторског обележја са одговарајућим кепстралним коефицијентима приказано је шематски на слици 3.17.

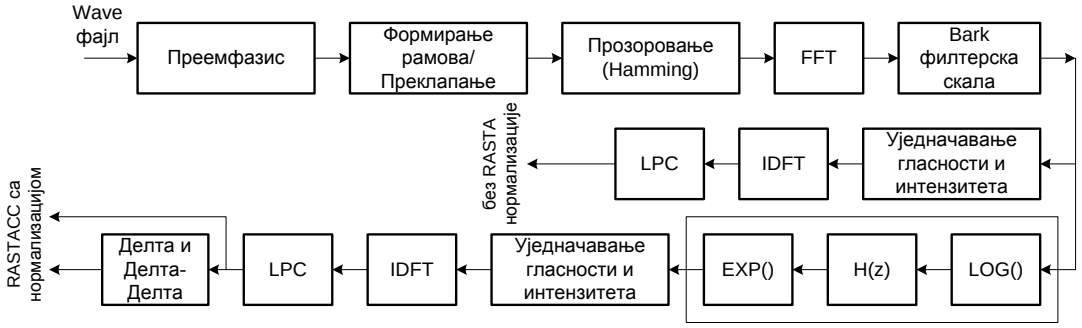


Слика 3.17 *Блок дијаграм за добијање векторских обележја TEPLPCC типа.*

Са слике се уочава да је TE оператор примењен после блока за брзу Фуријеову трансформацију. Остали делови и њихова улога су идентични као што је описано за векторско обележје PLPCC типа.

3.6 ВЕКТОРСКА ОБЕЛЕЖЈА ТИПА RASTACC И TERASTACC

Добијање векторског обележја типа RASTACC (RelAtive SpecTrA Cepstral Coefficients) заснива се на PLP анализи на коју је примењена одговарајућа модификација спектралних компонената коју су предложили Хермански и Морган [Hermansky, Morgan, 1994]. Циљ RASTA-е је да потисне спектралне компоненте које су „спорије“ или „брже“ од уобичајених промена у говору. На овај начин се RASTA третира и као један од облика нормализације [De Veth, Boves, 1998]. Да би се то остварило уводе се додатни кораци у односу на PLP, а шема за добијање ових обележја је дата на слици 3.18.



Слика 3.18 *Блок дијаграм за добијање векторских обележја RASTACC типа.*

Додатни кораци подразумевају:

1) трансформацију спектралних амплитуда помоћу нелинеарне компресионе функције (изражене помоћу *LOG()* блока на слици 3.18),
2) филтрирање временске путање за сваку претходно трансформисану спектралну компоненету (изражено помоћу *H(z)* блока),
3) трансформацију филтрираних спектралних компонената помоћу нелинеарне експанзионе функције (представљене помоћу *EXP()* блока на претходној слици).

Циљ овог поступка је да се свака константна или споропроменљива компонената потисне пре него што се примени LPC анализа.

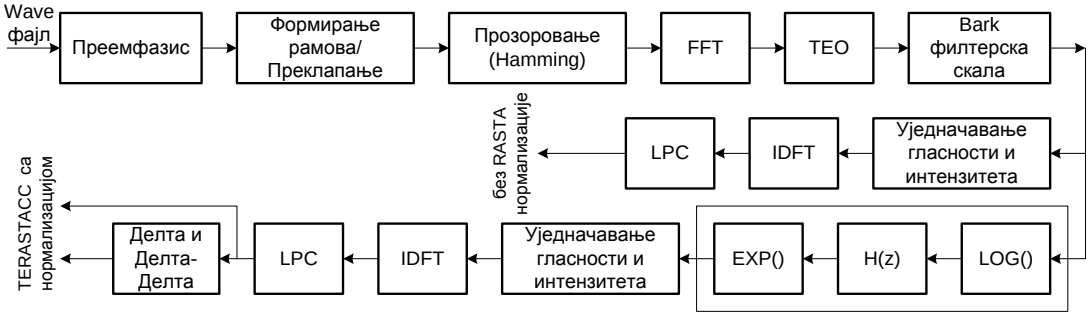
Филтер (у *z* нотацији) који је предложио Хермански је IIR (Infinite Impulse Response) облика дефинисан са:

$$H(z) = 0{,}1 * z^{4} * \frac{2 + z^{-1} - z^{-3} - 2 * z^{-4}}{1 - 0{,}98 * z^{-1}} \tag{3.39}$$

и он омогућава да се потисну одговарајуће споропроменљиве компоненте.

Применом блок дијаграма са слике 3.18 могу се добити одговарајући RASTA кепстрални коефицијенти (RASTACC) без нормализације, са нормализацијом као и одговарајући изводи (делта и делта-делта кепстрални коефицијенти).

Добијање TERASTACC векторских обележја остварује се тако што се примени Teager Energy оператор после блока за брзу Фуријеову трансформацију, као што је приказано на слици 3.19.



Слика 3.19 *Блок дијаграм за добијање векторских обележја TERASTACC типа.*

На сличан начин као и за RASTACC добија се скуп векторских обележја који се могу састојати од кепстралних коефицијената без нормализације, са нормализацијом, делта и делта-делта коефицијената. Овим се заокружује скуп разматраних векторских обележја.

4. WHI-SPE ГОВОРНА БАЗА

Да би се успешно извршила анализа и тестирање мултимодалног говора најпре се приступило креирању базе говорних узорака. Ова база је унапред осмишљена и представља скуп од 10.000 датотека. Коришћено је 50 различитих речи које су тако одабране да формирају три корпуса: скуп боја, скуп бројева и скуп акустички балансираних речи. Ове речи су изговаране у два мода: нормалним говором и шапатом. Назив базе говорних узорака је Whi-Spe што асоцира на шапат (**Whi**spered **Spe**ech).

4.1 ДИЗАЈН ГОВОРНЕ БАЗЕ

База је дизајнирана тако да садржи два основна дела: први део су говорни узорци који се односе на нормалан говор, а други део су узорци који се односе на шапат. За снимање узорака ангажовано је 10 говорника и то пет женских и пет мушких који су били старости од 20 до 30 година. Женски и мушки говорници су обележени различитим индексима тако да је током анализе могуће пратити и како пол говорника утиче на препознавање говора.

Сваки од говорника је изговарао скуп од 50 речи у оба мода и то 10 пута при чему је постојао временски размак између снимања од неколико дана. На тај начин је формирана база од укупно 10.000 говорних узорака [Marković et al., 2013 a]. Коришћене су три основне категорије речи и то: називи основних боја (укупно 6 боја), бројева (укупно 14 бројева) и акустички балансиране речи (укупно 30 речи). Акустички балансиране речи су раније дефинисане на бази GEES-a [Jovičić et al., 2004] и оне су пажљиво изабране да покрију основне лингвистичке критеријуме српског језика као што су распоред фонема, акцентна структура, креирање слогова, груписање сугласника и слично.

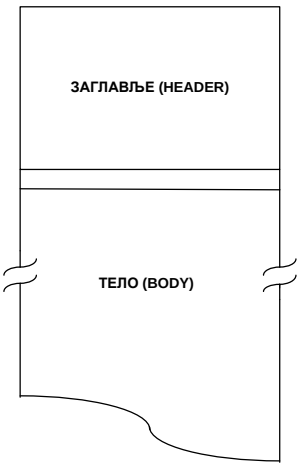
Сви подаци у бази су једнозначно обележени тако да их је лако користити. Следеће правило је коришћено при обележавању говорних узорака:

- ако је говорни узорак везан за нормалан говор његов облик је: *rec**x_y_z**n.wav***
- а ако је везан за шапат онда је облика: *rec**x_y_z**s.wav***

Слова ‘*n*’ и ‘*s*’ испред екстензије ‘*wav*’ означавају да се фајл односи на нормалан говор, односно шапат, респективно. Ознаке “***x***”, “***y***” и “***z***” су природни бројеви такви да “***x***” означава индекс (редни број) речи из датог речника (може узимати вредности од 1 до 50), “***y***” означава редни број говорника који реч изговара (и може бити од 1 до 10 при чему су женски говорници обележени бројевима од 1 до 5, а мушки бројевима од 6 до 10) и “***z***” означава редни број изговора појединачне речи коју говорник “***y***” изговара (може бити од 1 до 10 за оба мода). На овај начин успостављен је јединствен начин обележавања свих говорних узорака за све говорнике и све њихове изговоре што омогућава једноставан приступ и обраду узорака током креирања одговарајућих програма. Тиме се даје могућност

одабира одређене категорије речи (нпр. боје или бројеви или акустички балансиране речи) и примене одговарајућег софтверског решења.

Сви говорни узорци у овој бази су смештени у „wave“ формату [Марковић, 2005]. Формат ових датотека дат је на слици 4.1.



Слика 4.1 *Формат “wave” датотеке.*

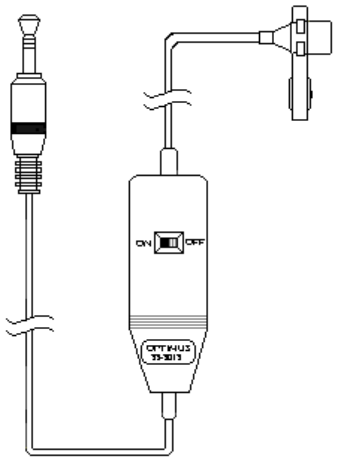
Са слике се може уочити да се датотека састоји из два дела: заглавље (header) и тело (body). Заглавље садржи више врста података, а од интереса су: укупна дужина датотеке, број канала, фреквенција одмеравања, број бајтова у секунди, број бита по одмерку и тд. Што се тиче тела оно садржи ознаку да почињу корисни подаци (коришћењем карактеристичне речи „data“), затим број који означава дужину корисних података, а потом корисне податке који репрезентују снимљени сигнал. Ово је данас широко распрострањени формат кога подржава много оперативних система, а посебно они који су Microsoft-ови производи. На бази ових датотека током процеса предобраде добијају се одговарајућа векторска обележја у облику кепстралних, делта и делта-делта кепстралних коефицијената која се потом користе у процесу обучавања и препознавања.

4.2 СНИМАЊЕ И ОБРАДА УЗОРАКА

Да би се ова говорна база реализовала коришћен је специфични амбијент тихе собе на Високој школи техничких струковних студија Чачак. Ова соба је омогућила да се пре свега, шапат фаворизује у односу на позадински шум.

Учесници овог снимања су били студенти Високе школе у Чачку који су као волонтери изговарали скуп од 50 речи у два мода: нормалним говором и шапатом. Овај процес је понављан више од 10 пута (да би се успешно добило бар 10 комплета узорака) при чему између сваког понављања је постојала пауза од неколико дана.

За процес снимања је коришћен лап-топ рачунар са софтвером Adobe Audition 1.5, и одговарајућим омни-дирекционим микрофоном типа “Optimus” (слика 4.2).



Слика 4.2 *Омнидирекциони микрофон типа Optimus.*

Овај микрофон има добре фреквенцијске карактеристике у опсегу до 20 kHz (слика 4.3).

Слика 4.3 *Фреквенцијска карактеристика омнидирекционог микрофона.*

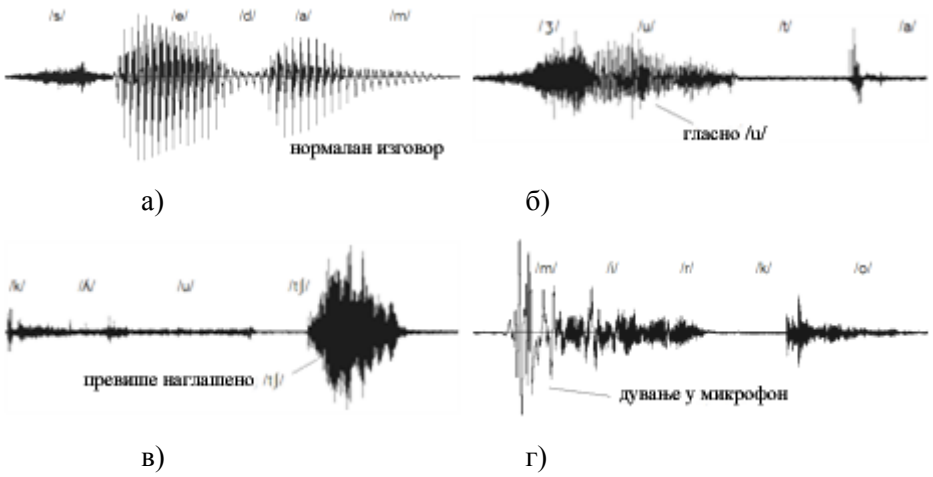
Фреквенција одмеравања је подешена на 22050 Hz, а сви узорци су снимани у облику стандарних Windows PCM wav датотека.

Приликом снимања нормалног говора микрофон се налазио на око 20cm од уста говорника, а приликом снимања шапата био је око 5cm са стране уста говорника да би се избегли одговарајући нежељени ефекти (као што је нпр. дување у микрофон и слично). Снимање шапата захтевало је посебну пажњу и већи број понављања јер су се у неким говорним сетовима појављивали нежељени ефекти као што су: низак ниво говорног сигнала у односу на амбијентални шум, погрешан изговор слова или слогова, дување у микрофон, изостављање одређених фонема из речи која се изговара и слично.

После снимања приступило се обради говорних узорака и провери њиховог квалитета. У овом процесу су учествовала два експерта и један фонетичар. Они су најпре све снимљене комплете узорака преслушали и извршили мануелну сегментацију и обележавање према напред наведеном правилу. Сваки од узорака добио је своју лабелу и смештен у Whi-Spe говорну базу. Међутим, током овог процеса јавили су се и одређени проблеми јер неки од узорака нису били довољно доброг квалитета. Основни проблеми су били везани за:

неправилну артикулацију, погрешан изговор, низак ниво говорног сигнала код шапата што се посебно често манифестовало код женских говорника.

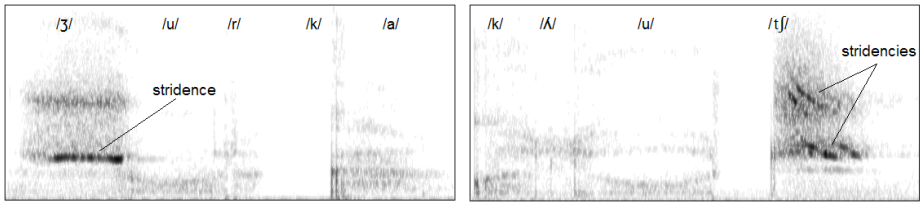
При креирању дела базе са узорцима шапата појавиле су се две врсте грешака које се могу класификовати као: контролисане и неконтролисане. У контролисане грешке спадају појаве као што су: неправилна артикулација, „пенетрација сонарности“, „пренаглашен“ изговор африката, дување у микрофон, неправилан изговор, изостављање неког фонема током изговора и слично (слика 4.4). Ове грешке се могу елиминисати поновним снимањем и применом контроле квалитета на снимљене узорке.



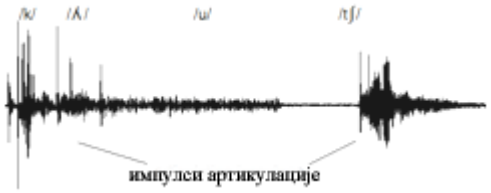
Слика 4.4 *Грешке при изговору: а) нормалан говор б) „гласан“ сегмент у шапату в) наглашено изговарање африката г) дување у микрофон.*

У неконтролисане грешке спадају артикулационе акције одређеног говорника. Оне могу бити случајне и систематске. Уколико су везане и за патологију говорника онда се тај говорник елиминише из даљег процеса снимања. Једна од случајних грешака је и „stridence“ [Jovičić et al., 2008] која је карактеристична при шапату и настаје трењем језика о одређени део непца приликом изговора одговарајућих фонема (слика 4.5.).

У зависности од врсте грешака, експерти и фонетичар су предузимали одговарајуће мере тако да се формирала база од 10.000 говорних узорака који су задовољили унапред постављене критеријуме.



а) б)



в)

Слика 4.5 *Специфичне манифестације при изговору: а) наглашени "stridence" у гласном фрикативу б) вишеструки "stridence" у африкативу в) ефекат контакта језика са непцима.*

4.3 ЕЛЕМЕНТИ ГОВОРНЕ БАЗЕ

Речник говорне базе је подељен у три корпуса. Први део представљају називи неких основних боја, други су бројеви који осим основних цифара садрже и друге најчешће коришћене вишецифрене бројеве, а трећи део су акустички балансиране речи које су раније коришћене у бази GEES. Укупан број ових речи је 50 и оне су детаљно представљене у табели 4.1 са одговарајућом IPA нотацијом.

Табела 4.1 *Речник Whi-Spe говорне базе*

	IPA нотација	***Српски***		***IPA нотација***	***Српски***
Боје	/bela/	бела	**Акустички балансиране речи**	/Mirko/	Мирко
	/ʒuta/	жута		/ʒurka/	журка
	/tsrna/	црна		/Petar/	Петар
	/tsrvena/	црвена		/demonstratsije/	демонстрације
	/plava/	плава		/standard/	стандард
	/zelena/	зелена		/pijatsa/	пијаца
Бројеви	/nula/	нула		/padavine/	падавине
	/jedan/	један		/ponedeʌak/	понедељак
	/dva/	два		/godina/	година
	/tri/	три		/predstava/	представа

	/tʃetiri/	четири		/kompjuteri/	компјутери
	/pet/	пет		/inostranstvo/	иностранство
	/ʃest/	шест		/drvo/	дрво
	/sedam/	седам		/Mirjana/	Мирјана
	/osam/	осам		/more/	море
	/devet/	девет		/kiʃa/	киша
	/deset/	десет		/zgrade/	зграде
	/sto/	сто		/klintsi/	клинци
	/hiʌadu/	хиљаду		/Milan/	Милан
	/milion/	милион		/rezultati/	резултати
				/telefon/	телефон
				/svetlo/	светло
				/prozor/	прозор
				/ruke/	руке
				/lokal/	локал
				/kʌutʃ/	кључ
				/suntse/	сунце
				/pare/	паре
				/sef/	сеф
				/blok/	блок

Српска и IPA нотација су исте за сугласнике и самогласнике изузев следећих сугласника: ʃ(ш), h(x), ʒ(ж), ts(ц), tɕ(ћ), tʃ(ч), dʑ(ђ), dʒ(џ), ɲ(њ), ʌ(љ).

5. ПОРЕЂЕЊЕ ГОВОРНИХ УЗОРАКА

На бази добијених векторских обележја потребно је извршити њихову евалуацију помоћу одговарајућег "back-end" система, тј. одговарајуће методе за препознавање говора. На овај начин се испитује која векторска обележја дају најбоље резултате. Постоје различите мере за поређење вектора (који репрезентују говорне узорке) и на основу њих се дефинишу одговарајућа растојања која ће у наредном делу бити детаљније изложена. Могу се користити различите методе за поређење узорака [Марковић, 2004]. Овде изложена метода се врло често користи када је у питању препознавање говора зависног од говорника и када је број изговора ограничен. То је техника динамичког усклађивања у времену (DTW – Dynamic Time Warping). Друге методе, базиране на скривеним Марковљевим моделима [Марковић, 2002], [Galić et al., 2014 a] и неуронским мрежама [Grozdić et al., 2012], су такође врло популарне.

5.1 МЕРА ЗА ПОРЕЂЕЊЕ ВЕКТОРА

Основни изазови које треба решити у системима за препознавање, а који су базирани на поређењу говорних узорака, јесте како вршити поређење између вектора који репрезентују говорне узорке и како експлицитно одредити меру њихове сличности.

Ако постоје два вектора *x* и *y* из векторског простора *Z* тада се може дефинисати дистанца *d* која у овом векторском простор представља различитост између посматраних вектора *x* и *y*. То је ненегативна фукција која се може записати у облику:

$$d(x, y) \geq 0 \tag{5.1}$$

У пракси се показало да, са аспекта математичке тачности, а такође и са становишта психиоакустичког осећаја чујности, модел који је заснован на спектралној разлици говорних узорака даје добре резултате. Наиме, како се говорни сигнал може сматрати квазистационаран на кратком временском интервалу реда десетак милисекунди онда се баш на таквим интервалима и рачуна кратковремански спектар сигнала и особине спектра и то се користи при поређењу говорних узорака. На овим кратким интервалима се примењују и брза Фуријеова трансформација, особине аутокорелације, модел „сви-полови" (all-poles) за представљање спектра (облика $\sigma / A(z)$ - често коришћена код LPC анализе), као и спектрална густина снаге $S(\omega)$.

Постоје различите врсте дистанци које се користе при поређењу говорних узорака, а оне које се најчешће користе су [Rabiner, Juang, 1993]:

- логаритамско спектрално растојање;

- кепстрално растојање;
- отежано кепстрално растојање;
- растојање на бази максималне веродостојности;
- спектрално растојање на специфичним фреквенцијским скалама (Linear, Mel, Bark) итд.

Логаритамско спектрално растојање се често користи при поређењу говорних узорка. Ако је спектрална густина снаге једног узорка дата са $S(\omega)$, а другог са $S'(\omega)$, тада се дефинише разлика логаритамских вредности ова два спектра као $V(\omega)$:

$$V(\omega) = \log S(\omega) - \log S'(\omega) \tag{5.2}$$

Мера различитости између вектора може да се прикаже као дисторзија између $S(\omega)$ и $S'(\omega)$, а коришћењем параметра *p* (*p* је обично или 1 или 2) дефинише се и момент те дисторзије (средњи, квадратни и сл.). У том случају ова дистанца добија облик:

$$d(S,S')^p = (d_p)^p = \int_{-\pi}^{\pi} |V(\omega)|^p \frac{d\omega}{2\pi} \tag{5.3}$$

На слици 5.1 приказане су спектралне густине снага $S(\omega)$ и $S'(\omega)$ као и фукција $V(\omega)$ добијени коришћењем брзе Фуријеове трансформације, док су на слици 5.2 приказане спектралне густине снага добијене применом LPC анализе и апроксимацијом $S(\omega)$ са „сви-полови“ функцијом ($\sigma / A(z)$).

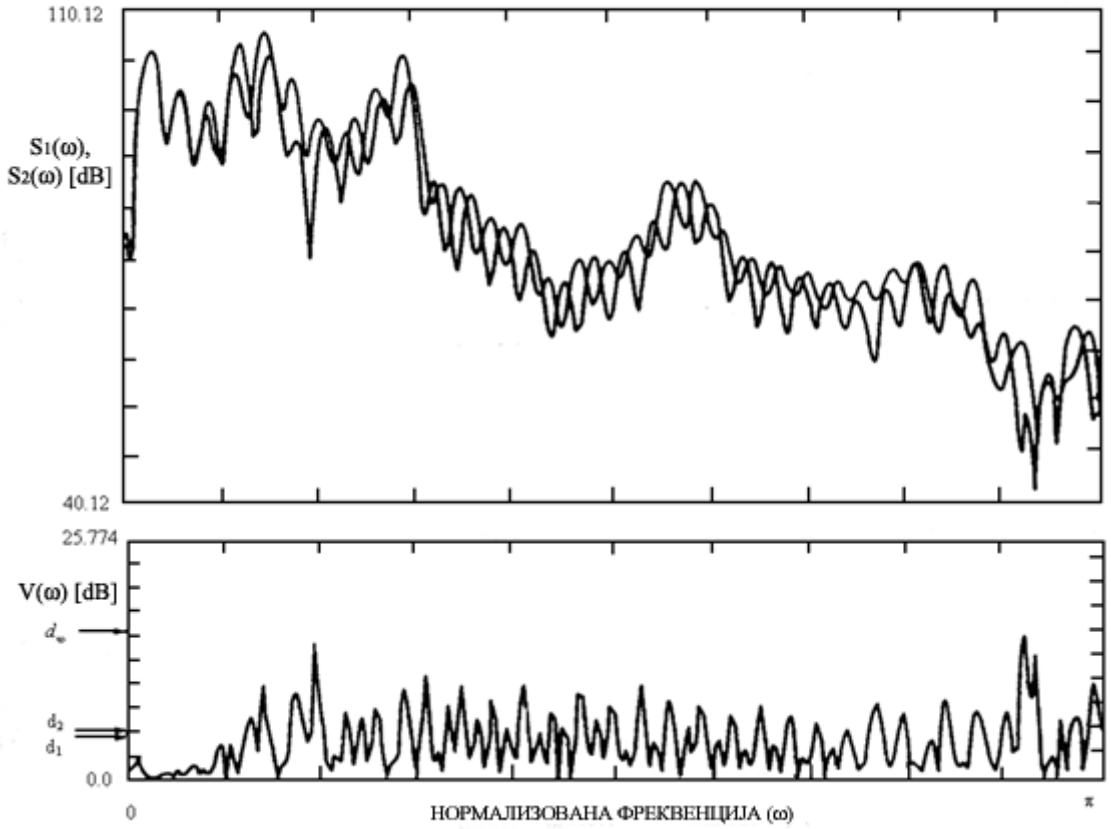
Кепстрално растојање се добија на бази поређења кепстралних коефицијената који су репрезенти одговарајућих спектралних снага. Пошто се комплексни кепструм (cepstrum) дефинише као Фуријеова трансформација од логаритма спектра сигнала, онда се $\log S(\omega)$ може представити у облику:

$$\log S(\omega) = \sum_{n=-\infty}^{\infty} C_n e^{-jn\omega} \tag{5.4}$$

Како важи да је $C_n = C_{-n}$ то су ове вредности реални бројеви и називају се кепстралним коефицијентима. Користећи такође Парсевалову (Paseval) теорему кепстрално растојање се може даље изразити у следећем облику:

$$d_{cp}^2 = \int_{-\pi}^{\pi} |\log S(\omega) - \log S'(\omega)|^2 \frac{d\omega}{2\pi} = \sum_{n=-\infty}^{\infty} (C_n - C'_n)^2 \tag{5.5}$$

где су C_n, C'_n кепстрални коефицијенти од спектралних снага $S(\omega)$ и $S'(\omega)$.



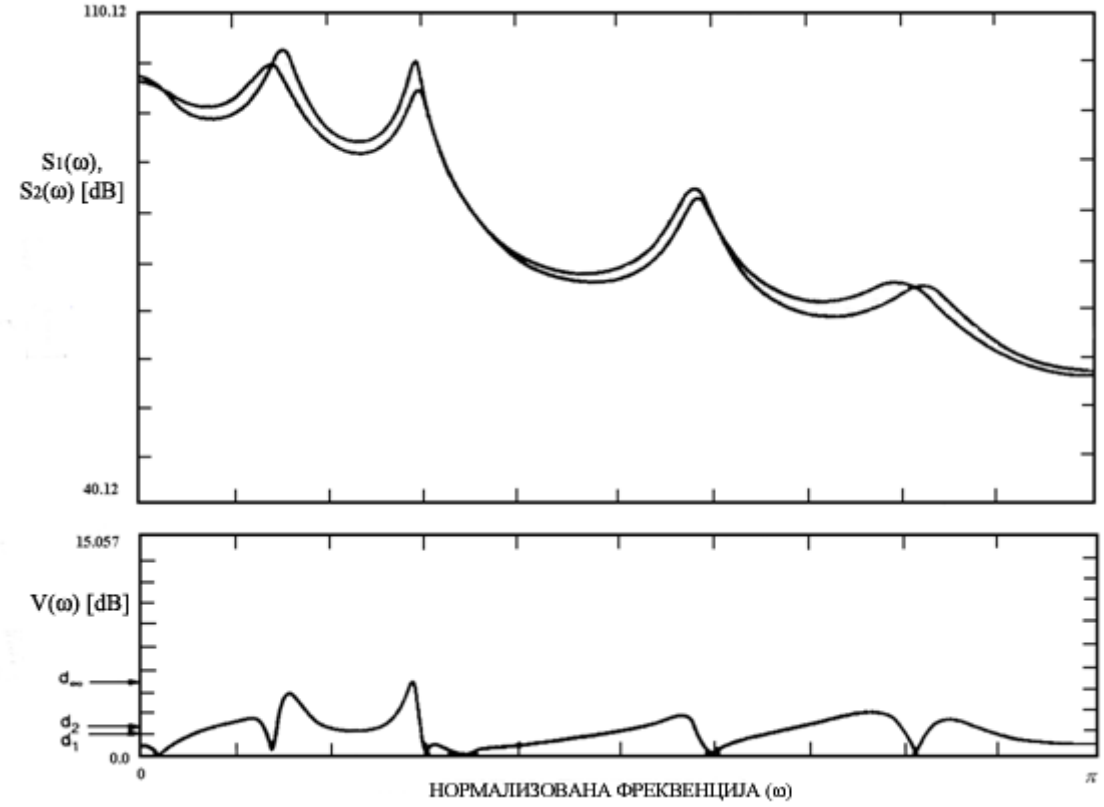
Слика 5.1 *Спектралне густине снага ($S(\omega)$ и $S'(\omega)$) и лог разлика $V(\omega)$ добијени помоћу FFT-а [Rabiner, Juang, 1993].*

У реалности број кепстралних коефицијената није бесконачан већ се ограничава на неки број *K*. Тиме се тражено кепстрално растојање апроксимира и добија се формула облика:

$$d^2_{cpK}(K) = \sum_{n=1}^{K} (C_n - C'_n)^2 \tag{5.6}$$

Отежано кепстрално растојање се често користи при поређењу вектора због интересантних особина кепстралних коефицијената [Tohkura, 1986]. Наиме, показује се да кепстрални коефицијенти (изузев нултог) имају средњу вредност једнаку нули, а варијанса им је инверзно пропорционална са квадратом реда посматраног кепстралног коефицијената, односно:

$$E\{C_n^2\} \sim \frac{1}{n^2} \tag{5.7}$$



Слика 5.2 *Спектралне густине снага ($S(\omega)$ и $S'(\omega)$) и лог разлика ($V(\omega)$) добијени помоћу LPC анализе [Rabiner, Juang, 1993].*

На овај начин могуће је увести отежавање разлике кепстралних коефицијената па формула (5.5) добија нови облик:

$$d_{cpW}^2 = \sum_{n=-\infty}^{\infty} n^2 (C_n - C'_n)^2 \tag{5.8}$$

Такође, показује се да кепстрални коефицијенти нижег и вишег реда имају различиту осетљивост на врсту отежавања па се стога предлажу и различите фукције $w(n)$ којима би се они отежавали. Ако се узме одређен број *K* кепстралних коефицијената, онда формула (5.8) добија нови облик:

$$d_{cpW}^2(K) = \sum_{n=1}^{K} (w(n)C_n - w(n)C'_n)^2 \tag{5.9}$$

Растојање максималне веродостојности има више облика, а најпознатији су следећа три: Itakura-Satio-ова мера максималне веродостојности, Itakura-ина мера максималне веродостојности и мера односа веродостојности. На основу раније поменутих дефиниција за $S(\omega)$, $S'(\omega)$ и $V(\omega)$ Itakura-Satio-ова мера максималне веродостојности може се дефинисати следећом формулом:

$$d_{IS}(S,S') = \int_{-\pi}^{\pi} [e^{V(\omega)} - V(\omega) - 1] \frac{d\omega}{2\pi} = \int_{-\pi}^{\pi} \frac{S(\omega)}{S'(\omega)} \frac{d\omega}{2\pi} - \log \frac{\sigma_\infty^2}{\sigma_\infty'^2} - 1 \qquad (5.10)$$

где су σ_∞^2 и $\sigma_\infty'^2$ грешке предикције једног корака за $S(\omega)$ и $S'(\omega)$ респективно. Остале поменуте дистанце веродостојности могу се извести на основу горње формуле.

За анализу спектралног растојања могу се користити различите фреквенцијске скале, као и различити облици подела подопсега на овим скалама. Тако се у овом раду користе линеарна, “mel”, “bark” скале као и други распореди филтера све са циљем да се пронађе најбољи резултат за препознавање мултимодалног говора са акцентом на шапат и његове сценарије.

У овој дисертацији осим кепстралних коефицијената базираних на различитим векторским обележјима коришћени су и њихови изводи (делта коефицијенти) и изводи извода (делта-делта коефицијенти). Такође, примењен је и одговарајући облик нормализације.

5.2 DTW МЕТОДА ЗА ПОРЕЂЕЊЕ УЗОРАКА

Да би се поредили говорни узорци потребно је да се они на неки начин најпре уједначе (нпр. да им се поклопе крајеви тј. почеци и завршеци), а такође и друге флуктуације (нпр. енергије) да се доведу у одређене границе. На тај начин процес поређења добија смисло. Са тим циљем врши се нормализација и усклађивање посматраних сигнала. Као једна од најефикаснијих метода за сам процес поређења, посебно када је у питању мањи број узорака, показао се DTW алгоритам (алгоритам динамичког усклађивања-уједначавања у времену). Он је базиран на технициди динамичког програмирања [Davis, Mermelstein, 1980].

5.2.1. УСКЛАЂИВАЊЕ И НОРМАЛИЗАЦИЈА

Приликом поређења два говорна узорка појављују се различити проблеми. На пример брзина изговора једне исте речи је различита, а то се дешава чак и кад је у питању један исти говорник (зависи да ли је јутро или вече, да ли је уморан или не и сл.). Стога је потребно, и поред различите брзине изговора, различите наглашености појединих слогова или посебног истицања неког вокала, наћи начин да се и такви узорци говорног сигнала што боље упореде и одреди мера њихове сличности. Због ових разлога јавља се потреба да се изврши нека врста уједначавања узорака и да се флуктуације нормализују, односно сведу у неке дозвољене границе.

Нека су два говорна узорка X и Y представљени са својим спектралним секвенцама (векторима) $(x_1, x_2, ..., x_{T_x})$ и $(y_1, y_2, ..., y_{T_y})$. Онда се тражи дистанца између ова два вектора као нека функција кратовременске спектралне дисторзије компонената ових вектора. Ако се

са $d(x_i, y_j)$ обележи кратковременска спектрална дисторзија између два елемента i и j од два различита вектора X и Y (или упрошћено $d(i, j)$), тада у случају коришћења најједноставније нормализације, а то је линеарна временска нормализација, мера растојања између вектора X и Y је облика:

$$d(X,Y) = \sum_{i=1}^{T_x} d(i, j) \tag{5.11}$$

При чему као услов линеарне нормализације мора важити:

$$j = \frac{T_y}{T_x} i \tag{5.12}$$

У овом поступку линеарне временске нормализације подразумева се да је варијација брзине говора пропорционална дужини узорка, а независна од врсте узорка који се користи. Мерење одступања врши се строго по правој линији која је дијагонала правоугаоника кога образују дужине узорака X и Y (T_x и T_y).

Овај начин временског усклађивања и нормализације није реалан па се уводе тзв. функције усклађивања у времену (warping functions). Могу се дефинисати две функције развлачења ϕ_x и ϕ_y које ће бити кореспондентне са i и j и служиће да се ове вредности нормирају на заједничку, k осу. Математички се то може представити једначинама:

$$i = \phi_x(k),\ k = 1,2,...,T \tag{5.13}$$

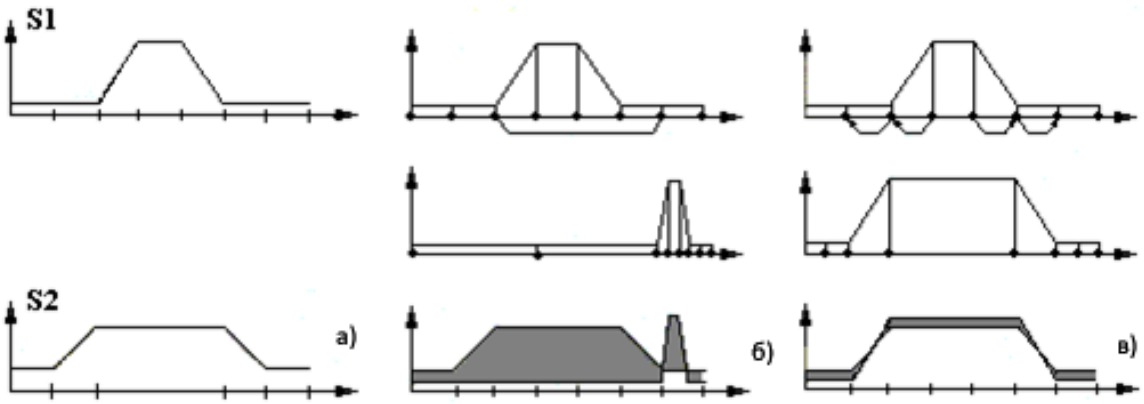
$$j = \phi_y(k),\ k = 1,2,...,T \tag{5.14}$$

Проблем усклађивања (развлачења) у времену може се илустровати сликом 5.3. Овде се пореде две секвенце $S1$ и $S2$. Ако се замисли да $S1$ има скуп тачака које се могу „ухватити“ и као еластични ластиш развући и тако померити по апсциси - онда се може видети да померењем у одређеном смеру разлика између $S1$ и $S2$ графикона може бити већа (под б) или мања (под в) (шрафирани део). Сврха развлачења у времену код говорних сигнала је управо да се секвенце параметара (кепстралних коефицијената) тако подесе да ова разлика буде што мања.

Полазећи од уведених функција усклађивања ϕ_x и ϕ_y може се дефинисати и функционални пар $\phi = (\phi_x, \phi_y)$. У том случају, у зависности од ове дефиниције, општа мера различитости између два говорна узорка X и Y може да се изрази у следећем облику:

$$d_\phi(X,Y) = \sum_{k=1}^{T} w(k) d(\phi_x(k), \phi_y(k)) \frac{1}{M_\phi} \tag{5.15}$$

где $w(k)$ представља коефицијенат за отежање пута, а M_ϕ нормализациони фактор стазе.



Слика 5.3 *Усклађивање (warping) у времену у циљу поређења секвенци* $S1$ *и* $S2$.

Даље се проблем поређења вектора своди на проналажење најбоље стазе којом би се вршило поређење односно проналажење одговарајућег функционалног пара $\phi = (\phi_x, \phi_y)$ за који би се минимизовало растојање вектора X и Y. У том смислу, минимизација се представља као:

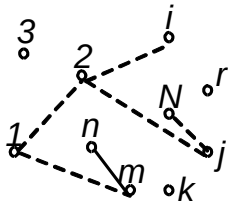
$$d(X,Y) \equiv \min d_{\phi}(X,Y) \tag{5.16}$$

Може се користити велики број парова функција за усклађивање $\phi = (\phi_x, \phi_y)$, али се тражи да се одреди такав пар који ће дати најбоље преклапање и такав да се мерење врши са конзистентношћу. Функције усклађивања ϕ_x и ϕ_y морају бити монотоно неопадајуће, а то значи да нема промене редоследа спектралних елемената вектора X и Y, односно мора се поређење вршити секвенцијално, без прескакања или кретања уназад.

Употреба динамичког програмирања даје одличне резултате за ову врсту проблематике, тј. када је у питању секвенцијално одлучивање.

5.2.2. ПРИМЕНА ДИНАМИЧКОГ ПРОГРАМИРАЊА

Динамичко програмирање има широку примену код разних проблема, а може се користити где је потребно пронаћи оптимални пут између, на пример, две тачке у векторском простору. Нека су те две тачке i и j, као што је представљено на слици 5.4 и нека у простору има укупно N тачака. Нека је цена прелаза између било које две тачке m и n у том простору дефинисана као одређена вредност - $c(m,n)$. Поставља се питање проналажења пута између i и j који би при томе имао минималну цену.



Слика 5.4 *Векторски простор од* N *елемената.*

У општем случају постоји више путева да се од полазишта i дође до дестинације j па је циљ пронаћи оптимални пут (онај са минималном ценом). Основна два приступа у решавању овог проблема су:

- асинхрона метода секвенцијалног одлучивања;
- синхрона метода секвенцијалног одлучивања.

Асинхрона метода подразумева да се од полазишта до дестинације дође уз минималну цену без ограничења броја корака [Mitrović et al., 2012]. За разлику од ње, синхрона унапред дефинише колики је тај број корака уз претпоставку, наравно, да такав пут постоји.

Кључну улогу у динамичком програмирању игра "правило одлучивања". Оно одређује како ће се алгоритам кретати од полазне тачке (рецимо i) до дестинације (рецимо j) уз минималну цену коштања. Ту се поставља питање које "правило одлучивања" користити да се добије минимала цена за прелаз од i до j. Нека је та цена обележена са $\lambda(i, j)$. Белманова [Bellman, 1957] теорема оптималности даје смерницу за избор правила „оптималног одлучивања" и она гласи:

"Оптимално правило одлучивања има такву особину да што год да су почетно стање и почетна одлука, преостале одлуке морају формирати оптимално правило одлучивања у односу на стање у које се дошло после прве одлуке."

Ово правило се у математичком облику може исказати следећим примером: нека се тражи оптимално правило одлучивања за прелаз из стања i у j и нека се прелаз може остварити преко једног или више међустања. Тада, нека је први корак прелаз од стања i у неко „међустање" k и нека је минимална цена прелаза $\lambda(i, k)$. Следећи корак би био наћи минималну цену коштања од k до j кроз било колико стања, односно наћи $\lambda(k, j)$. Отимално "правило одлучивања" треба да задовољи следећу једнакост:

$$\lambda(i, j) = \min_{k}[\lambda(i, k) + \lambda(k, j)] \tag{5.17}$$

Може се закључити да оптимално правило одлучивања захтева и проналажење међустања која су оптималне тачке прелаза и њихова укупна цена коштања би била минимална за задату почетну тачку и дестинацију.

Код синхроне методе секвенцијалног одлучивања задатак је да се стаза од полазне до крајње тачке пређе у тачно одређеном броју корака – на пример у M корака. У процесу

проналажења оптималног пута за овај случај користи се решеткасти дијаграм. Алгоритам посматра како из прве тачке i (полазишта) може да се дође до следеће тачке, рецимо n (где је $n=1,2,...,N$ - међукорак) на N могућих начина. Тражи се минимална цена коштања између стања i и n и нека је она изражена са $\lambda_1(i,n)$. После m-тог корака где је $m<M$ нека је систем у стању l (где је $l=1,2,...,N$) са одговарајућом минималном ценом $\lambda_m(i,l)$. Нека алгоритам у наредном кораку ($m+1$-ом) долази до следеће тачке, нпр. k. За ту тачку треба да важи:

$$\lambda_{m+1}(i,k)=\min_l[\lambda_m(i,l)+c(l,k)] \tag{5.18}$$

Ова једначина указује да се одређивање оптималне стазе тражи рекурзивном методом. Стога, алгоритам динамичког програмирања за синхрони начин одлучивања, где се решење тражи у M корака, а при чему је укупан број стања у систему N, може да се прикаже у следећем облику:

а) **Иницијализација:**

$\lambda_1(i,n)=c(i,n)$ - рачунање цене од полазишта до прве тачке (5.19)

$s_1(n)=i$ - функција која показује стазу односно претходно стање из кога се дошло,

а рачуна се за $n=1,2,...,N$ (5.20)

б) **Рекурзија:**

$\lambda_{m+1}(i,n)=\min_{1\le l\le N}[\lambda_m(i,l)+c(l,n)]$ - рачунање мин. цене следећег корака (5.21)

$s_{m+1}(n)=\arg\min_{1\le l\le N}[\lambda_m(i,l)+c(l,n)]$ - чување информ. о претходном кораку, (5.22)

а рачуна се за $n=1,2,...,N$ и $m=1,2,...,M-2$

в) **Завршетак:**

$\lambda_M(i,j)=\min_{1\le l\le N}[\lambda_{M-1}(i,l)+c(l,j)]$ - укупна минимална цена између *i* и *j* (5.23)

$s_M(j)=\arg\min_{1\le l\le N}[\lambda_{M-1}(i,l)+c(l,j)]$ - чување информ. о претходном стању (5.24)

г) **Одређивање стазе уназад:**

Оптимална стаза је секвенца облика $(i,i_1,i_2,..i_{M-1},j)$ и налази се помоћу кретања уназад узимајући у обзир да је $i_M=j$ и користећи $i_m=s_{m+1}(i_{m+1})$ где се за m узима $m=M-1,M-2,...,1$

Алгоритми динамичког програмирања су од посебног значаја за област препознавања говора што ће се и практично показати у наредном делу овога рада кроз објашњење DTW алгоритма.

5.2.3. ОГРАНИЧЕЊА ВРЕМЕНСКЕ НОРМАЛИЗАЦИЈЕ

Поређење вектора који репрезентују говорне узорке коришћењем функција усклађивања $\phi = (\phi_x, \phi_y)$ подразумева увођење одговарајућих ограничења у облику услова под којима ово поређење треба вршити. Ови услови се називају и ограничења функције усклађивања (warping constraints). Могу се класификовати у следеће групе:

- ограничења крајева речи;
- ограничења монотоности;
- ограничења локалног континуитета;
- ограничења општег пута;
- ограничења отежаног нагиба.

Овде ће бити поменуте само њихове основне особине и одговарајући математички облици.

Ограничење крајева речи подразумева да су почеци и завршеци речи које се пореде тачно дефинисани. То се математички може изразити (за почетно стање) са:

$$\phi_x(1) = 1 \text{ и } \phi_y(1) = 1 \tag{5.25}$$

За завршно стање треба да важи:

$$\phi_x(T) = T_x \text{ и } \phi_y(T) = T_y \tag{5.26}$$

На овај начин се два говорна узорка тако усклађују да се њихови почеци и завршеци поклапају, па се потом врши поређење.

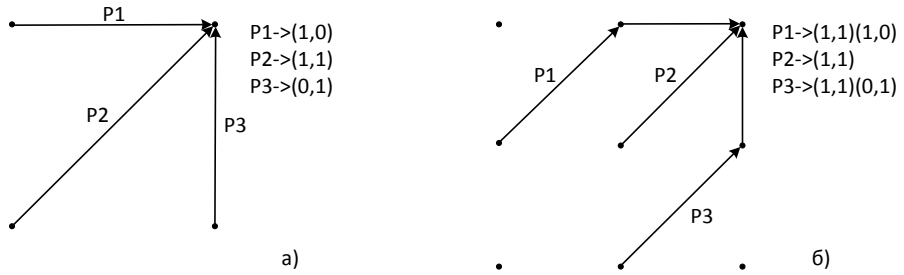
Ограничења монотоности омогућавају очување редослед елемената вектора приликом поређења, односно да редослед елемента буде растући (тачније неопадајући). То значи да стаза по којој се врши прорачун минималног растојања вектора $d_\phi(X, Y)$ неће бити негативног нагиба. Овај услов монотоности математички се може приказати у облику:

$$\phi_x(k+1) \geq \phi_x(k) \text{ и } \phi_y(k+1) \geq \phi_y(k) \tag{5.27}$$

Ограничења локалног континуитета се могу поставити у више различитих облика. Она прописују које су путање (за прелазе из стања у стање) дозвољене, а које нису. Тако нпр. ако целу стазу дефинишемо као скуп координата p и q онда се путања може представити као:

$$P \rightarrow (p_1, q_1)(p_2, q_2)...(p_T, q_T) \tag{5.28}$$

па се могу дефинисати дозвољени прелази (неки од њих, као пример, дати су на слици 5.5).



Слика 5.5 *Примери ограничења локалног континуитета.*

Поред ова два наведена постоји и читав низ других ограничења локалног континуитета и за све њих је карактеристично да имају неопадајући карактер. Увек је исти циљ: да се крене из почетне тачке (нпр. $(1,1)$) и да се стигне до крајње тачке (нпр. (T_x, T_y)).

Ограничењима општег пута (глобалне стазе) дефинише се област у којој треба да се нађе оптимална стаза. У том смислу дефинишу се два параметра $Q_{\max}$ и $Q_{\min}$ на следећи начин:

$$Q_{\max} = \max_{l}\left[\frac{\sum_{i=1}^{T_l} p_i^{(l)}}{\sum_{i=1}^{T_l} q_i^{(l)}}\right] \qquad Q_{\min} = \min_{l}\left[\frac{\sum_{i=1}^{T_l} p_i^{(l)}}{\sum_{i=1}^{T_l} q_i^{(l)}}\right] \tag{5.29}$$

где l означава индекс могуће стазе P_l, а T_l је укупан број корака на стази P_l. Параметри $Q_{\max}$ и $Q_{\min}$ означавају максималну и минималну могућу експанзију приликом растезања. Обично је $Q_{\max}$ =1/ $Q_{\min}$, а ове вредности зависе од раније поменутих ограничења локалног нагиба. Ако се усвоје горе наведени параметри и њихова релација даље се могу дефинисати ограничења општег пута у следећем облику:

$$1 + \frac{[\phi_x(k) - 1]}{Q_{\max}} \le \phi_y(k) \le 1 + Q_{\max}[\phi_x(k) - 1] \quad \text{и} \tag{5.30}$$

$$T_y + Q_{\max}[\phi_x(k) - T_x] \le \phi_y(k) \le T_y + \frac{[\phi_x(k) - T_x]}{Q_{\max}} \tag{5.31}$$

Додатно ограничење за глобалну стазу може се применити по препоруци Сакоеа и Чајба [Sakoe, Chiba, 1978] које гласи: $|\phi_x(k) - \phi_y(k)| \le T_o$ где је T_o максимално дозвољена апсолутна временска разлика између два узорка на било ком раму. Овај услов додатно смањује површину у координатном систему у којој је дозвољено да се нађе оптимална стаза.

Ограничења отежаног нагиба омогућава да сваки од прелаза (из једног у друго стање) има одређену вредност. На тај начин неки прелази имају већу вредност, а други мању. Отежавање има и своје физичко оправдање, а на основу формуле (5.9) се види да се функцијом $w(k)$ отежава допринос сваког кратковременског растојања $d(\phi_x(k), \phi_y(k))$. Има више облика ових ограничења сагласно са више типова локалних ограничења, а један од

њих, примера ради, је и тип ограничења отежаног нагиба предложен од стране Сакоеа и Чајба:

$$w(k) = \phi_x(k) - \phi_x(k-1) \tag{5.32}$$

Фактор глобалне нормализације M_ϕ има за циљ да средњу дисторзију стазе учини независном од дужина узорака који се пореде. Његова вредност се рачуна као:

$$M_\phi = \sum_{k=1}^{T} w(k) \tag{5.33}$$

Циљ је да овај фактор буде независан од функције усклађивања те да се на тај начин избегну претеране компликације приликом рачунања. У појединим случајевима, избором типа локалних ограничења (континуитета и отежавања) може се постићи да је $M_\phi = T_x$, а тиме је M_ϕ независно од функције усклађивања.

5.2.4. УПОТРЕБА АЛГОРИТМА

Улога алгоритма динамичког усклађивања у времену (DTW) је да, користећи напред објашњену технику динамичког програмирања, реши проблем како да се одреди минимална дистанца између вектора којима су репрезентовани говорни узорци. Под претпоставком да су задовољена ограничења крајева речи, да је глобални фактор стазе M_ϕ независан од функције ϕ и да постоји поклапања крајева речи (T_x, T_y) онда се једначина (5.15) може модификовати у:

$$M_\phi * d(X,Y) = D(T_x, T_y) = \min_{\phi_x, \phi_y} \sum_{k=1}^{T} d(\phi_x(k), \phi_y(k)) * w(k) \tag{5.34}$$

На основу ове формуле може писати да је минимална парцијална акумулирана дисторзија (растојање) дуж стазе која спаја тачке (стања) $(1,1)$ и (i,j) једнака:

$$D(i,j) = \min_{\phi_x, \phi_y, T'} \sum_{k=1}^{T'} d(\phi_x(k), \phi_y(k)) * w(k) \tag{5.35}$$

где су $i = \phi_x(T')$ и $j = \phi_y(T')$.

Примењујући технику динамичког програмирања, поступак рекурзије би био следећи:

$$D(i,j) = \min_{(i',j')}[D(i',j') + c((i',j'),(i,j))] \tag{5.36}$$

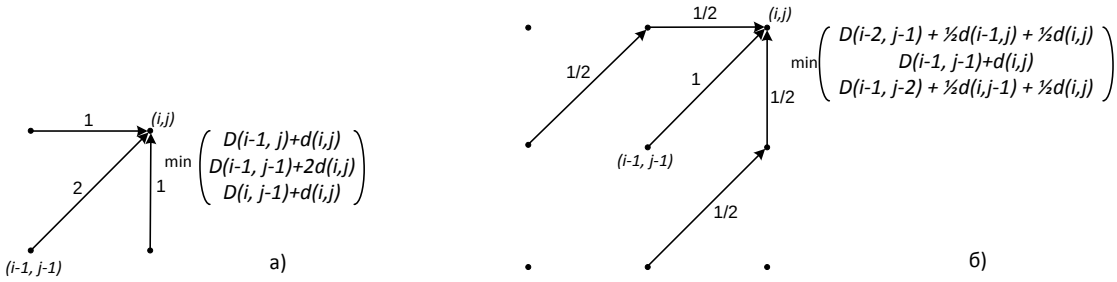
где је са $c((i',j'),(i,j))$ означена отежана акумулирана локална дистанца између тачака (i',j') и (i,j). Она се може представити у скраћеном облику:

$$c((i',j'),(i,j)) = \sum_{l=0}^{L_s} d(\phi_x(T'-l), \phi_y(T'-l)) w(T'-l) \tag{5.37}$$

где L_s представља број корака на стази којом се иде од тачке (i', j') до тачке (i, j). Такође, и за функције усклађивања важи: $i' = \phi_x(T' - L_s)$ и $j' = \phi_y(T' - L_s)$.

Функција дисторзије $c(,)$ је растућа и рачуна се само дуж специфичних путева који су одређени ограничењима локалног континуитета и отежаног нагиба. Стога се алгоритам динамичког програмирања примењује само на специфичан, лимитиран број тачака, а не на цео правоугаоник облика (T_x, T_y).

Ради стицања бољег увида у ову проблематику на слици 5.6 дата су два примера ограничења локалног континуитета са одговарајућим отежањима нагиба. За сваки од примера дата је и рекурзивна формула динамичког програмирања.



Слика 5.6 *Примери облика ограничења са формулама за рекурзију.*

Под претпоставком да је простор на коме се започиње алгоритам динамичког програмирања ограничен са почетном тачком $(1, 1)$ и са завршном (T_x, T_y), онда DTW алгоритам подразумева следеће кораке [Marković et al., 2013b]:

1. **Иницијализација:**

$$D(1,1) = d(1,1) * w(1) \tag{5.38}$$

2. **Рекурзија:**

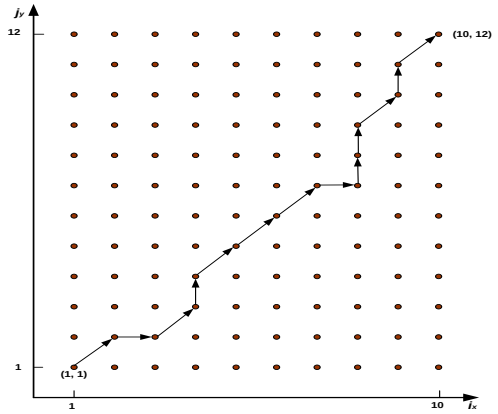
$$D(i, j) = \min_{(i', j')}[D(i', j') + c((i', j'), (i, j))] \tag{5.39}$$

где су i и j унутар дозвољеног простора тако да је: $1 \le i \le T_x$, $1 \le j \le T_y$

3. **Завршетак:**

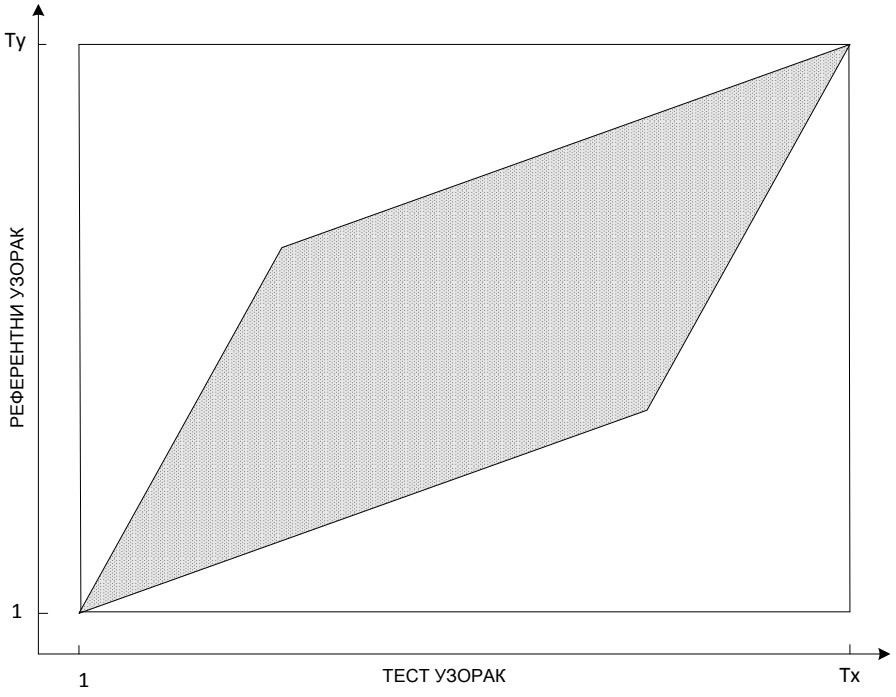
$$d(X, Y) = \frac{D(T_x, T_y)}{M_\phi} \tag{5.40}$$

Основна идеја овог алгоритма је да се рекурзивни кораци рачунају за све локалне стазе којима се може доћи до тачке (i, j) у тачно једном прелазу од (i', j') и да се при томе поштују постављена локална ограничења. На слици 5.7 дат је пример једне такве DTW стазе.



Слика 5.7 *Пример DTW стазе.*

Слика 5.8 приказује један пример скупа могућих тачака које се налазе у шрафираном делу ограниченим почетном тачком $(1,1)$ и завршном тачком (T_x, T_y). Само унутар назначеног простора врши се израчунавање локалних стаза у оквиру DTW алгоритма.

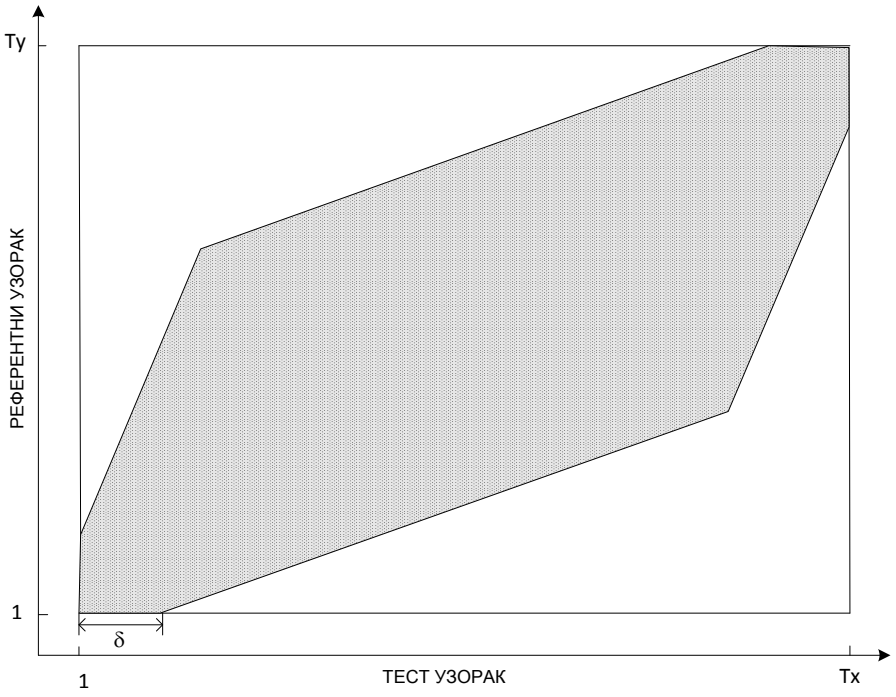


Слика 5.8 *Скуп тачака по којима се примењује DTW алгоритам.*

Увођењем ограничења број могућих тачака (стања) се смањује што знатно убрзава процес рачунања оптималне стазе.

Поред овог случаја када се користе "стриктна" ограничења крајева речи постоје и ситуације када се она могу направити флексибилнијим, односно тада почетак поређења не мора бити тачка $(1,1)$ и крај не мора бити (T_x, T_y). У том случају уводи се неки померај δ - којим се померају ове почетне/крајње тачке. На тај начин даје се више флексибилности

приликом поређења јер у неким реалним ситуацијама се и не може прецизно одредити почетак и крај речи (нпр. утицај позадинског шума, пуцкетања усана, тешког дисања и слично). Негативна страна се у овом случају огледа у томе што је отежано израчунавање оптималне стазе, односно повећан је број корака и стаза по којима се одређује минимална дистанца.



Слика 5.9 *Проширење услова за ограничење крајева речи.*

На слици 5.9 приказана је област по којој се врши рачунање минималне дистанце за два узорка (тест и референтни). У односу на преходну слику (слика 5.8) број тачака за израчунавање се повећава.

У пракси је доста истраживано коришћење глобалног ограничења типа CE2-1 [Marković et al., 2013d] које даје интересантне резултате, а постоји и могућност графичке представе стазе којом се врши поређење [Marković, Marković, 2008].

Поред поступака којима се налази једна оптимална стаза често се примењују и поступци где се налази већи број, рецимо *к*, најоптималнијих стаза. Тај проблем се решава помоћу паралелног, серијског или решеткастог алгоритма.

6. ЕКСПЕРИМЕНТАЛНИ РЕЗУЛТАТИ

На основу раније дефинисаних полазних хипотеза и метода истраживања извршено је креирање одговарајућих векторских обележја. Са аспекта параметара од којих су састављени, вектори се могу класификовати у четири типа:

1) вектори састављени од 12 кепстралних коефицијената без нормализације;
2) вектори састављени од 12 кепстралних коефицијената са нормализацијом;
3) вектори састављени од 24 коефицијента (12 кепстралних плус 12 делта кепстралних) при чему је примењена нормализација и
4) вектори састављени од 36 коефицијената (12 кепстралних плус 12 делта кепстралних плус 12 делта-делта кепстралних) при чему је примењена нормализација.

Векторска обележја коришћена за ове експерименте су:

1) LPCC – кепстрални коефицијенти базирани на линеарном предикционом кодирању;
2) LFCC – кепстрални коефицијенти базирани на линеарној фреквенцијској скали;
3) TELFCC – кепстрални коефицијенти базирани на линеарној фреквенцијској скали уз примену Teager Energy оператора;
4) MFCC – кепстрални коефицијенти базирани на “mel” фреквенцијској скали;
5) TEMFCC – кепстрални коефицијенти базирани на “mel” фреквенцијској скали уз примену Teager Energy оператора;
6) GFCC– кепстрални коефицијенти базирани на Gammatone филтер скали;
7) TEGFCC – кепстрални коефицијенти базирани на Gammatone филтер скали уз примену Teager Energy оператора;
8) PLPCC – кепстрални коефицијенти базирани на перцептивној линеарној предикцији;
9) TEPLPCC – кепстрални коефицијенти базирани на перцептивној линеарној предикцији уз примену Teager Energy оператора;
10) RASTACC – кепстрални коефицијенти базирани на перцептивној линеарној предикцију на коју је примењена RASTA нормализација и
11) TERASTACC – кепстрални коефицијенти базирани на перцептивној линеарној предикцију на коју је примењен Teager Energy оператор и RASTA нормализација.

Добијање вектора омогућено је реализацијом одговарајућих функција у софтверском пакету MATLAB [MATLAB].

На основу добијених векторских обележја извршено је њихово поређење са акцентом на примену нормализације и Teager Energy оператора где год је то изводљиво. Размотрена се четири основна сценарија (два усаглашена и два неусаглашена) за мултимодални говор и то:

- **„нормалан/нормалан“** (референтни узорци су снимљени у нормалном моду, а тест узорци такође у нормалном моду);
- **„шапат/шапат“** (референтни узорци су снимљени у моду шапата, а тест узорци такође у моду шапата);
- **„нормалан/шапат“** (референтни узорци су снимљени у нормалном моду, а тест узорци у моду шапата) и
- **„шапат/нормалан“** (референтни узорци су снимљени у моду шапата, а тест узорци у нормалном моду).

За поређење говорних узорака користила се класична метода динамичког усклађивања у времену (DTW) базирана на Еуклидској дистанци која се рачунала између одговарајућих вектора (као што је детаљно објашњено у поглављу 5).

6.1 КРЕИРАЊЕ И ПОРЕЂЕЊЕ ВЕКТОРА

Да би се креирала одговарајућа векторска обележја и да би се извршило њихово поређење развијен је одређен број софтверских модула. Алат за развој је био MATLAB верзија R2006a [Ljubić et al., 2014] и Visual Basic 6.0 [Deitel et al., 1999].

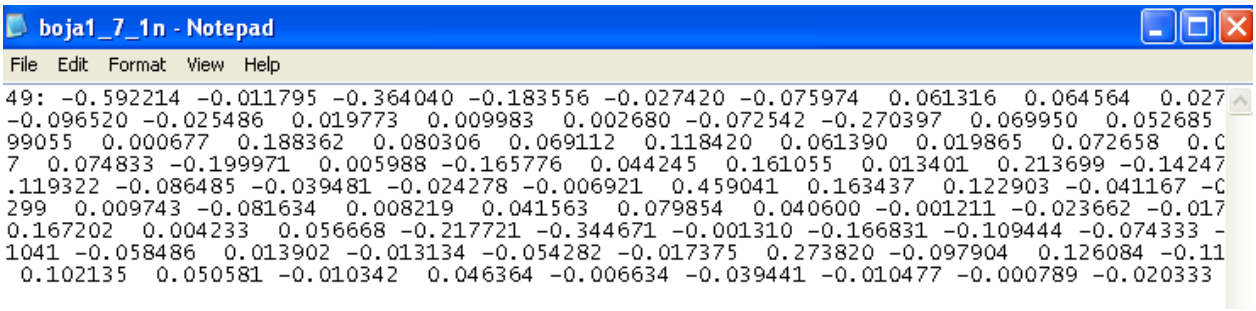
6.1.1 КРЕИРАЊЕ ВЕКТОРА

За сваку појединачну врсту обележја као и за сваку категорију вектора развијени су одговарајуће софтверске функције тј. модули. Њихов задатак је био да све говорне узорке забележене у облику ***wav*** датотека претворе у скуп вектора према раније наведеном поступку предобраде (поглавље 3). Стога је број ових софтверских модула пропорционалан броју векторских обележја.

Креиране MATLAB функције као аргументе садрже одређене параметре којима се може додатно вршити подешавање жељене предобраде. Тако нпр. функција: ***MFCC_text_cms12('Ulazni_fajl.txt', 22050, 1, 30)*** означава софтверски модул који учитава све ***wav*** датотеке који се налазе у наведеној датотеци ***'Ulazni_fajl.txt'***, чија фреквенција одмеравања је ***22050Hz***, који су у ***wav*** формату (трећи аргумент чија је вредност „***1***“), и који користи ***30*** филтера распоређених према “mel” скали. Ова функција креира векторе од по 12 кепстралних коефицијената на које је примењена нормализација CMS („***cms12***“). На слици 6.1 дат је део садржаја улазне ***txt*** датотеке за једног говорника (Govornik7). Он садржи списак ***wav*** датотека за обраду.

Слика 6.1 *Пример листе улазних датотека.*

Поменута MATLAB функција обрађује једну по једну ***wav*** датотеку и креира нове датотеке типа ***txt*** који садрже векторе од по 12 (или 24 или 36) коефицијената, већ према раније дефинисаном типу. На почетку сваке од ових излазних датотека најпре је један број који означава колико укупно вектора садржи одговарајући излазни репрезент ***wav*** датотеке, затим двотачка („:“) па следе бројчане вредности коефицијената. Коефицијенти су раздвојени празнином. Део излазне датотеке **boja1_7_1n.txt** овог типа приказан је на слици 6.2.



Слика 6.2 *Пример излазне датотеке која садржи кепстралне коефицијенте.*

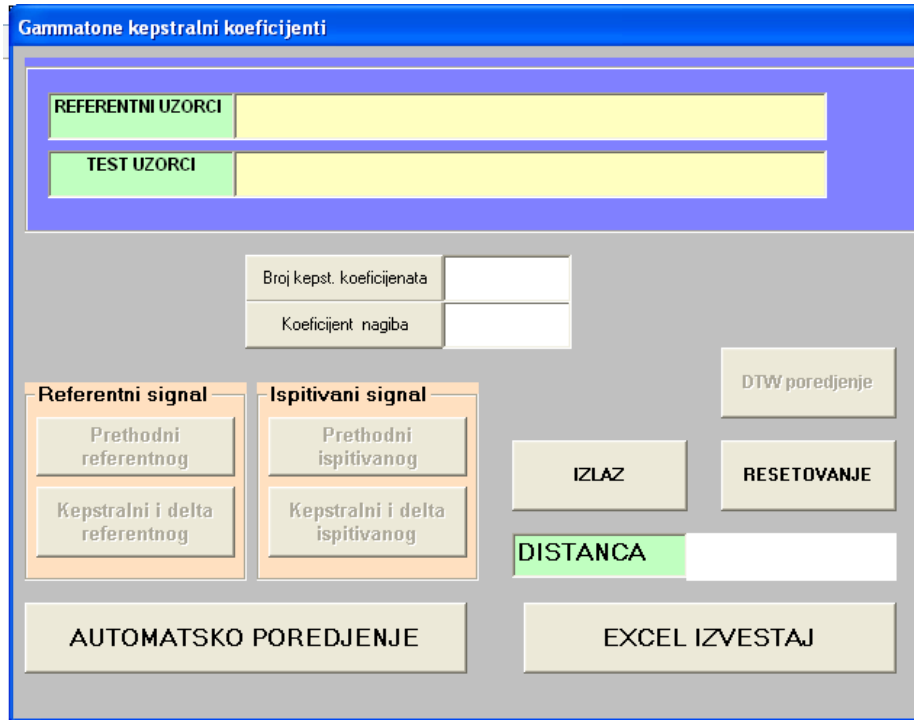
На основу списка свих улазних ***wav*** датотека из Whi-Spe базе (њих 10.000) на овај начин креира се 10.000 излазних датотека по свакој појединачној категорији (4 категорије) и за

свако појединачно векторско обележје (11 векторских обележја). Стога, за овај експеримент је креирано и коришћено око 440.000 датотека облика као што је она приказана на слици 6.2.

6.1.2 ПОРЕЂЕЊЕ ВЕКТОРА

Поређење говорних узорака, односно одговарајућих вектора који их репрезентују, вршено је помоћу софтверског модула WiseWave [Marković, 2002 a], [Marković**,** Luković, 2012]. Овај модул представља графички кориснички интерфејс за реализацију DTW алгоритма и развијен је коришћењем софтверског пакета Visual Basic 6.0.

После одговарајућег уводног екрана добија се апликација као што је приказана на слици 6.3.

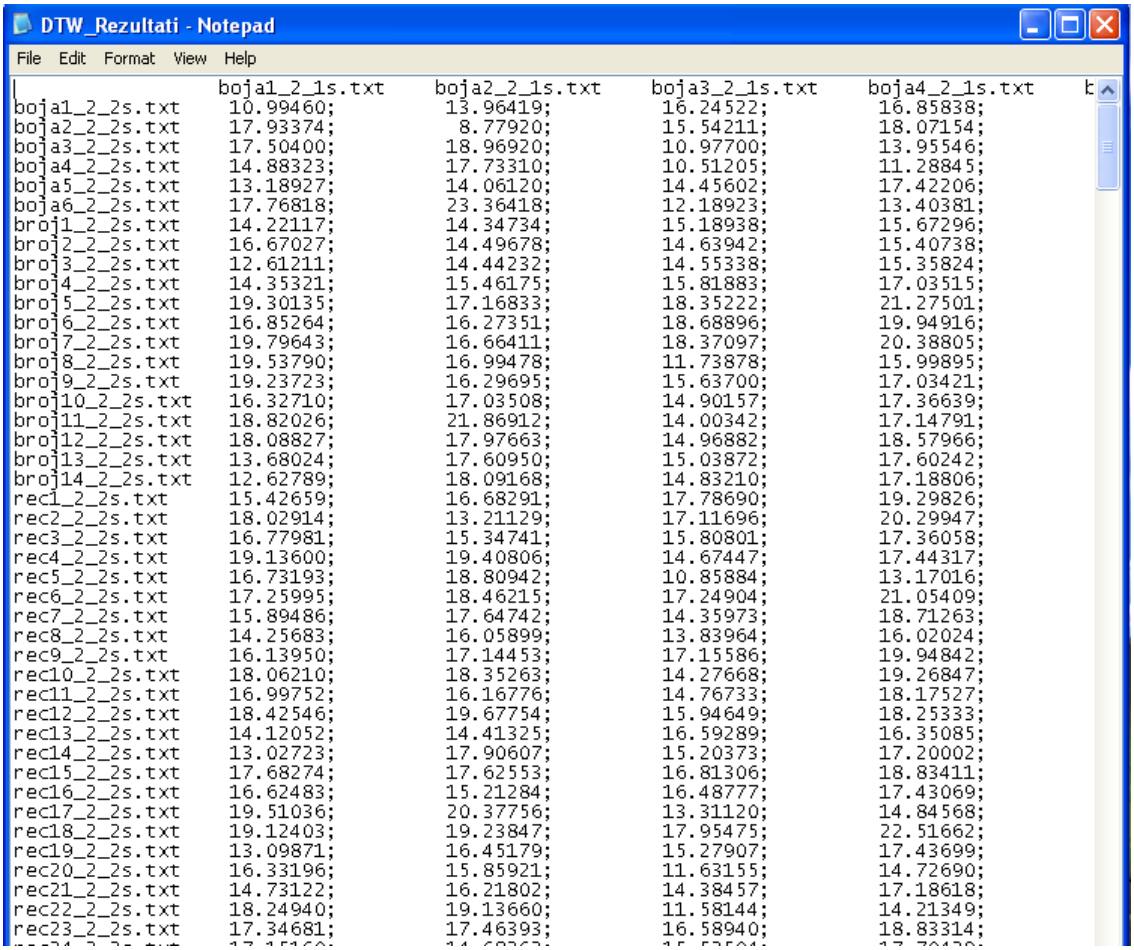


Слика 6.3 *WiseWave апликација за поређење говорних узорака.*

У поље означено са „**REFERENTNI UZORCI**“ уноси се назив ***txt*** датотеке која садржи називе свих ***txt*** датотека (добијених помоћу MATLAB функције на раније описани начин), а које се користе као референтне речи, а у поље означено са „**TEST UZORCI**“ уноси се такође назив ***txt*** датотеке који садржи називе датотека који представљају тест речи. У овим експериментима за референтне узорке су се користили скупови од по 50 речи тј. одговарајући векторски репрезенти (нормалног говора и/или шапата). Као тест узорци користили су се преостали сетови, сваки по 50 речи (репрезенти нормалног говора и/или шапата). У поље **„Broj kepst. koeficijenata“** уноси се колико параметара садржи сваки вектор понаособ, а то је или 12 или 24 или 36. **„Koeficijent nagiba“** је у ствари коефицијент локалног континуитета (нагиба) и за

њега се користи природан број. Обично се узима 1 или 2. У овим експериментима за коефицијент локалног нагиба коришћена је вредност 1.

Притиском на командно дугме „**AUTOMATSKO POREDJENJE**“ започиње поступак DTW поређења свих референтних узорака са свим тест узорцима који су специфицирани у поменутим датотекама. Резултат овог поређења је матрица која даје дистанце између референтних и тест узорака (као што је приказано на слици 6.4).



	boja1_2_1s.txt	boja2_2_1s.txt	boja3_2_1s.txt	boja4_2_1s.txt
boja1_2_2s.txt	10.99460;	13.96419;	16.24522;	16.85838;
boja2_2_2s.txt	17.93374;	8.77920;	15.54211;	18.07154;
boja3_2_2s.txt	17.50400;	18.96920;	10.97700;	13.95546;
boja4_2_2s.txt	14.88323;	17.73310;	10.51205;	11.28845;
boja5_2_2s.txt	13.18927;	14.06120;	14.45602;	17.42206;
boja6_2_2s.txt	17.76818;	23.36418;	12.18923;	13.40381;
broj1_2_2s.txt	14.22117;	14.34734;	15.18938;	15.67296;
broj2_2_2s.txt	16.67027;	14.49678;	14.63942;	15.40738;
broj3_2_2s.txt	12.61211;	14.44232;	14.55338;	15.35824;
broj4_2_2s.txt	14.35321;	15.46175;	15.81883;	17.03515;
broj5_2_2s.txt	19.30135;	17.16833;	18.35222;	21.27501;
broj6_2_2s.txt	16.85264;	16.27351;	18.68896;	19.94916;
broj7_2_2s.txt	19.79643;	16.66411;	18.37097;	20.38805;
broj8_2_2s.txt	19.53790;	16.99478;	11.73878;	15.99895;
broj9_2_2s.txt	19.23723;	16.29695;	15.63700;	17.03421;
broj10_2_2s.txt	16.32710;	17.03508;	14.90157;	17.36639;
broj11_2_2s.txt	18.82026;	21.86912;	14.00342;	17.14791;
broj12_2_2s.txt	18.08827;	17.97663;	14.96882;	18.57966;
broj13_2_2s.txt	13.68024;	17.60950;	15.03872;	17.60242;
broj14_2_2s.txt	12.62789;	18.09168;	14.83210;	17.18806;
rec1_2_2s.txt	15.42659;	16.68291;	17.78690;	19.29826;
rec2_2_2s.txt	18.02914;	13.21129;	17.11696;	20.29947;
rec3_2_2s.txt	16.77981;	15.34741;	15.80801;	17.36058;
rec4_2_2s.txt	19.13600;	19.40806;	14.67447;	17.44317;
rec5_2_2s.txt	16.73193;	18.80942;	10.85884;	13.17016;
rec6_2_2s.txt	17.25995;	18.46215;	17.24904;	21.05409;
rec7_2_2s.txt	15.89486;	17.64742;	14.35973;	18.71263;
rec8_2_2s.txt	14.25683;	16.05899;	13.83964;	16.02024;
rec9_2_2s.txt	16.13950;	17.14453;	17.15586;	19.94842;
rec10_2_2s.txt	18.06210;	18.35263;	14.27668;	19.26847;
rec11_2_2s.txt	16.99752;	16.16776;	14.76733;	18.17527;
rec12_2_2s.txt	18.42546;	19.67754;	15.94649;	18.25333;
rec13_2_2s.txt	14.12052;	14.41325;	16.59289;	16.35085;
rec14_2_2s.txt	13.02723;	17.90607;	15.20373;	17.20002;
rec15_2_2s.txt	17.68274;	17.62553;	16.81306;	18.83411;
rec16_2_2s.txt	16.62483;	15.21284;	16.48777;	17.43069;
rec17_2_2s.txt	19.51036;	20.37756;	13.31120;	14.84568;
rec18_2_2s.txt	19.12403;	19.23847;	17.95475;	22.51662;
rec19_2_2s.txt	13.09871;	16.45179;	15.27907;	17.43699;
rec20_2_2s.txt	16.33196;	15.85921;	11.63155;	14.72690;
rec21_2_2s.txt	14.73122;	16.21802;	14.38457;	17.18618;
rec22_2_2s.txt	18.24940;	19.13660;	11.58144;	14.21349;
rec23_2_2s.txt	17.34681;	17.46393;	16.58940;	18.83314;

Слика 6.4 *Приказ дела резултата поређења.*

Са слике се може уочити да у горњем делу је листа назива референтних датотека (листа садржи 50 назива који су кореспондентни са 50 референтних узорака). У левом делу слике (по вертикали) налази се листа назива тест сетова од по 50 датотека. У пресеку овако дефинисаних колона (референтни узорци) и врста (тест узорци) налази се дистанца израчуната помоћу DTW методе између одговарајућег референтног и тест узорка. Уколико је по дијагонали најмања вредност онда је реч успешно препозната, у противном дошло је до погрешне одлуке.

Када се заврши поступак „**AUTOMATSKO POREDJENJE**“ добија се одговарајућа порука и резултујућа датотека (као што је пример приказан на слици 6.4) је комплетирана. Потом се притисне командно дугме „**EXCEL IZVESTAJ**“ и аутоматски се генерише одговарајућа Excel

датотека на бази ***txt*** датотеке (која је приказана на слици 6.4). Excel датотека даје приказ које су речи успешно препознате (обележено зеленом бојом), а које су речи погрешно препознате (обележено црвеном бојом). Део приказа те датотеке је дат на слици 6.5. Такође добија се матрица конфузије (у Excel варијанти) која показује које су речи „замењене“ са другима (део приказа је на слици 6.6) као и статистички извештај са појединачним процентима успешно препознатих речи (WRR – Word Recognition Rate) за сваку од речи и укупним процентом успешног препознавања за одговарајућег говорника тј. за све речи сумарно (део приказа на слици 6.7).

DTW_Rezultati

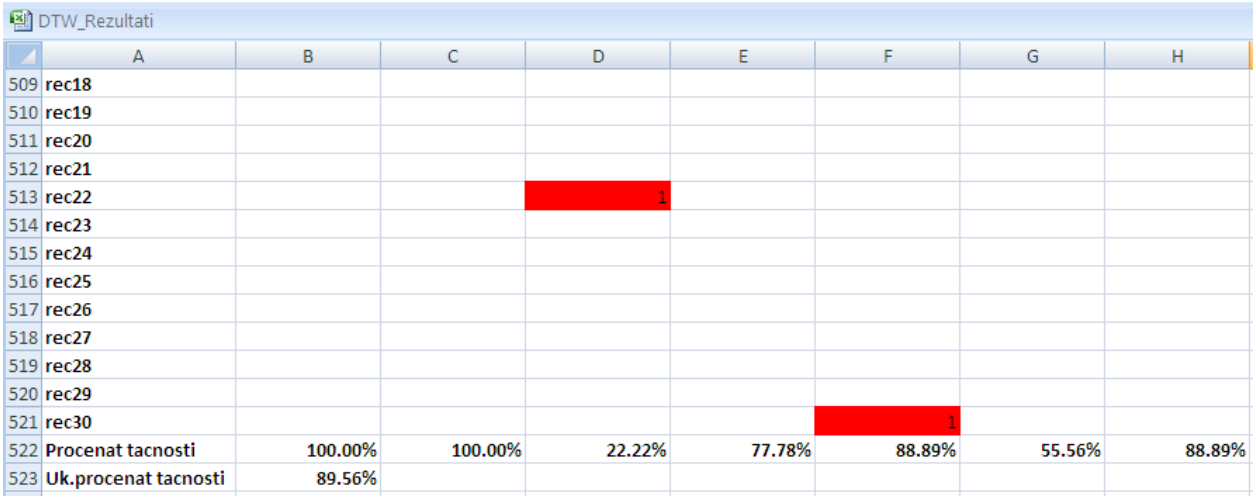
	A	B	C	D	E	F	G	H
1								
2		boja1_2_1s.txt	boja2_2_1s.txt	boja3_2_1s.txt	boja4_2_1s.txt	boja5_2_1s.txt	boja6_2_1s.txt	broj1_2_1s.txt
3	boja1_2_2s.txt	10.99460	13.96419	16.24522	16.85838	11.69386	16.36755	14.86616
4	boja2_2_2s.txt	17.93374	8.77920	15.54211	18.07154	14.42646	18.15330	16.50041
5	boja3_2_2s.txt	17.50400	18.96920	10.97700	13.95546	16.89604	11.31543	16.58184
6	boja4_2_2s.txt	14.88323	17.73310	10.51205	11.28845	14.32993	13.95900	12.85708
7	boja5_2_2s.txt	13.18927	14.06120	14.45602	17.42206	9.59669	17.53820	14.49493
8	boja6_2_2s.txt	17.76818	23.36418	12.18923	13.40381	18.41545	12.86824	16.84147
9	broj1_2_2s.txt	14.22117	14.34734	15.18938	15.67296	11.52213	15.18627	10.45843
10	broj2_2_2s.txt	16.67027	14.49678	14.63942	15.40738	13.34214	15.66401	16.01442
11	broj3_2_2s.txt	12.61211	14.44232	14.55338	15.35824	9.95958	14.68897	12.01573
12	broj4_2_2s.txt	14.35321	15.46175	15.81883	17.03515	12.33700	15.83061	16.33272
13	broj5_2_2s.txt	19.30135	17.16833	18.35222	21.27501	16.60021	18.18186	20.44815
14	broj6_2_2s.txt	16.85264	16.27351	18.68896	19.94916	13.78001	19.35305	18.90007
15	broj7_2_2s.txt	19.79643	16.66411	18.37097	20.38805	16.85137	22.59327	19.13824
16	broj8_2_2s.txt	19.53790	16.99478	11.73878	15.99895	16.63083	13.16094	17.55534
17	broj9_2_2s.txt	19.23723	16.29695	15.63700	17.03421	15.08976	13.58639	19.00827
18	broj10_2_2s.txt	16.32710	17.03508	14.90157	17.36639	13.66295	17.43146	15.47576
19	broj11_2_2s.txt	18.82026	21.86912	14.00342	17.14791	16.96665	16.38467	19.85911
20	broj12_2_2s.txt	18.08827	17.97663	14.96882	18.57966	18.21155	13.95521	17.37282
21	broj13_2_2s.txt	13.68024	17.60950	15.03872	17.60242	14.87419	17.81412	15.45111
22	broj14_2_2s.txt	12.62789	18.09168	14.83210	17.18806	13.85524	18.45289	14.59730

Слика 6.5 *Приказ дела резултата поређења у облику Excel извештаја.*

DTW_Rezultati

	A	B	C	D	E	F	G	H
470	Statisticki izvestaj							
471		boja1_2_1s.txt	boja2_2_1s.txt	boja3_2_1s.txt	boja4_2_1s.txt	boja5_2_1s.txt	boja6_2_1s.txt	broj1_2_1s.txt
472	boja1	9						
473	boja2		9					
474	boja3			2	1		3	
475	boja4			1	7		1	
476	boja5					8		
477	boja6			4	1		5	
478	broj1							8
479	broj2							
480	broj3							1
481	broj4							
482	broj5							
483	broj6							
484	broj7							

Слика 6.6 *Приказ дела матрице конфузије.*



DTW_Rezultati

	A	B	C	D	E	F	G	H
509	rec18							
510	rec19							
511	rec20							
512	rec21							
513	rec22			1				
514	rec23							
515	rec24							
516	rec25							
517	rec26							
518	rec27							
519	rec28							
520	rec29							
521	rec30					1		
522	Procenat tacnosti	100.00%	100.00%	22.22%	77.78%	88.89%	55.56%	88.89%
523	Uk.procenat tacnosti	89.56%						

Слика 6.7 *Приказ процента успешно препознатих речи појединачно и сумарно.*

Вредност означена на извештају са „**Uk. procenat tacnosti**“ (слика 6.7) се уписује у одговарајућу табелу резултата па се поступак поређења понавља за све преостале говорнике.

Осим поменутих команди на форми са слике 6.3 постоје и команде „**RESETOVANJE**“ и „**IZLAZ**“. Њихове улоге су да се ресетују вредности свих променљивих које се тренутно користе у програму, односно да се изађе из програма.

6.2 РЕЗУЛТАТИ ПОРЕЂЕЊА

Резултати препознавања у виду процента тачно препознатих речи (WRR–Word Recognition Rate) за векторска обележја LPCC, LFCC, TELFCC, MFCC, TEMFCC, GFCC, TEGFCC, PLPCC, TEPLPCC, RASTACC и TERASTACC дати су у облику табела и дијаграма. На бази сваког од поменутих обележја извршено је поређење резултата са аспекта коришћења/некоришћења нормализације, а такође и са аспекта примене статичких (кепстралних) и динамичких (делта и делта-делта) коефицијената.

Свако поље у табели представља резултат од 22.500 поређења па је за комплетан експериментални део овог рада број DTW поређења износио преко 40 милиона.

Урађена је статистичка анализа где је ниво поузданости (confidence level) постављен на 95% и на бази тога су рачунати интервали поузданости (confidence intervals). Добијене вредности су представљене у табелама 6.1-6.44. Интервали поузданости су се рачунали као „**Ср. вред. ± Грешка**“ и приказани су на одговарајућим дијаграмима.

За све случајева где је било могуће коришћење Teager Energy оператора дато је и поређење резултата препознавања са сродним обележјима (тако нпр. поређени су TELFCC и LFCC, TEMFCC и MFCC итд.).

6.2.1 РЕЗУЛТАТИ НА БАЗИ LPCC ВЕКТОРСКОГ ОБЕЛЕЖЈА

На бази линеарне предикционе кодне анализе (LPC), (која је детаљно описана у делу 3.1) добијен је скуп вектора састављених од кепстралних коефицијената, делта кепстралних коефицијената и делта-делта кепстралних коефицијената [Marković**,** Grozdić, 2014] као што је раније наведено.

Разматрана су следећа четири сценарија: усаглашени („нормалан/нормалан“ и „шапат/шапат“) и неусаглашени („нормалан/шапат“ и „шапат/нормалан“). Резултати препознавања у облику процента тачно препознатих речи дати су у табелама 6.1 и 6.2 за усаглашене, а у табелама 6.3 и 6.4 за неусаглашене сценарије.

Табела 6.1 *LPCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	LPCC (без CMS-а)	LPCC (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	96.00	96.67	97.78	97.56
Говорник 2	94.00	94.44	95.33	95.33
Говорник 3	97.11	99.33	98.89	98.67
Говорник 4	96.00	97.11	97.33	97.33
Говорник 5	95.78	97.33	97.56	97.78
Говорник 6	92.67	94.00	93.33	93.11
Говорник 7	92.44	94.89	95.33	94.89
Говорник 8	98.22	98.44	98.44	98.00
Говорник 9	95.78	96.89	96.89	96.67
Говорник 10	92.89	95.33	94.89	94.89
Ср. вред. ± Грешка	**95.09±1.23**	**96.44±1.08**	**96.58±1.10**	**96.42±1.10**

Табела 6.2 *LPCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	LPCC (без CMS-а)	LPCC (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	93.33	94.89	94.67	94.67
Говорник 2	94.00	96.00	96.44	95.56
Говорник 3	97.33	98.22	98.89	98.89
Говорник 4	97.11	97.33	97.33	97.11
Говорник 5	90.44	92.67	92.22	92.00
Говорник 6	76.22	87.56	86.44	85.56
Говорник 7	86.00	91.56	91.11	90.44
Говорник 8	85.11	94.44	94.22	93.33
Говорник 9	89.11	94.22	94.67	94.67
Говорник 10	79.11	83.33	83.78	83.11
Ср. вред. ± Грешка	**88.78±4.46**	**93.02±2.83**	**92.98±2.95**	**92.53±3.08**

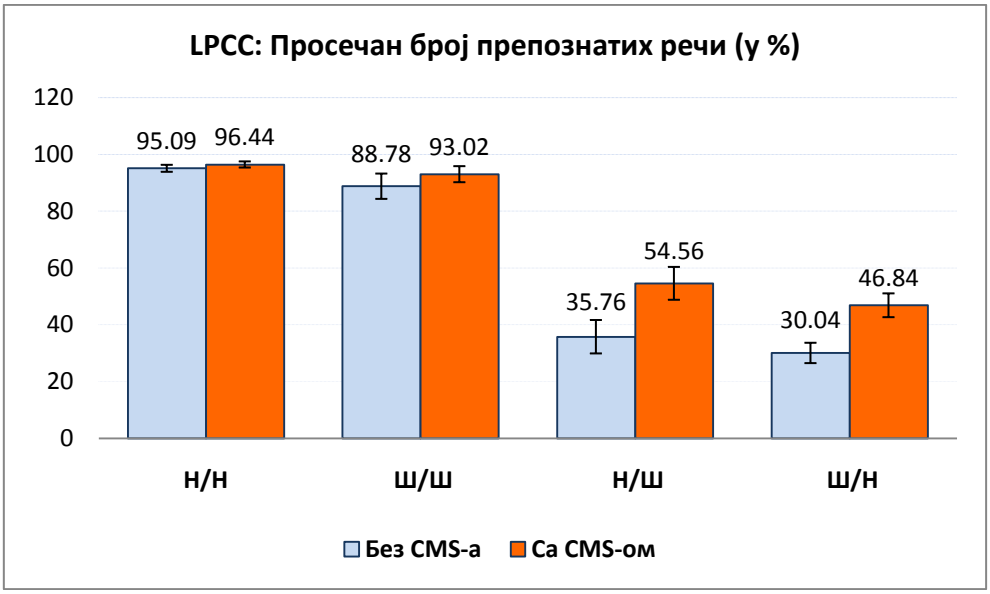
Табела 6.3 *LPCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	LPCC (без CMS-а)	LPCC (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	53.56	67.78	67.78	68.00
Говорник 2	16.44	38.00	39.56	38.89
Говорник 3	38.22	60.89	62.89	62.22
Говорник 4	40.67	65.11	66.89	66.00
Говорник 5	29.11	47.56	49.11	47.11
Говорник 6	35.33	49.78	48.00	47.33
Говорник 7	30.22	50.22	48.89	47.56
Говорник 8	36.67	62.22	65.11	62.67
Говорник 9	39.11	55.33	55.78	55.56
Говорник 10	38.22	48.67	48.67	47.11
Ср. вред. ± Грешка	**35.76±5.89**	**54.56±5.79**	**55.27±6.09**	**54.24±6.18**

Табела 6.4 *LPCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	LPCC (без CMS-а)	LPCC (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	28.22	49.56	49.11	47.11
Говорник 2	21.33	33.33	34.44	33.56
Говорник 3	33.33	50.67	53.56	53.78
Говорник 4	26.44	48.67	52.67	52.00
Говорник 5	20.89	57.33	39.33	39.56
Говорник 6	31.11	39.33	38.44	39.56
Говорник 7	36.89	43.56	44.00	44.00
Говорник 8	33.56	55.11	56.89	56.89
Говорник 9	31.56	48.00	48.44	47.78
Говорник 10	37.11	42.89	43.78	43.33
Ср. вред. ± Грешка	**30.04±3.58**	**46.84±4.19**	**46.07±4.53**	**45.76±4.47**

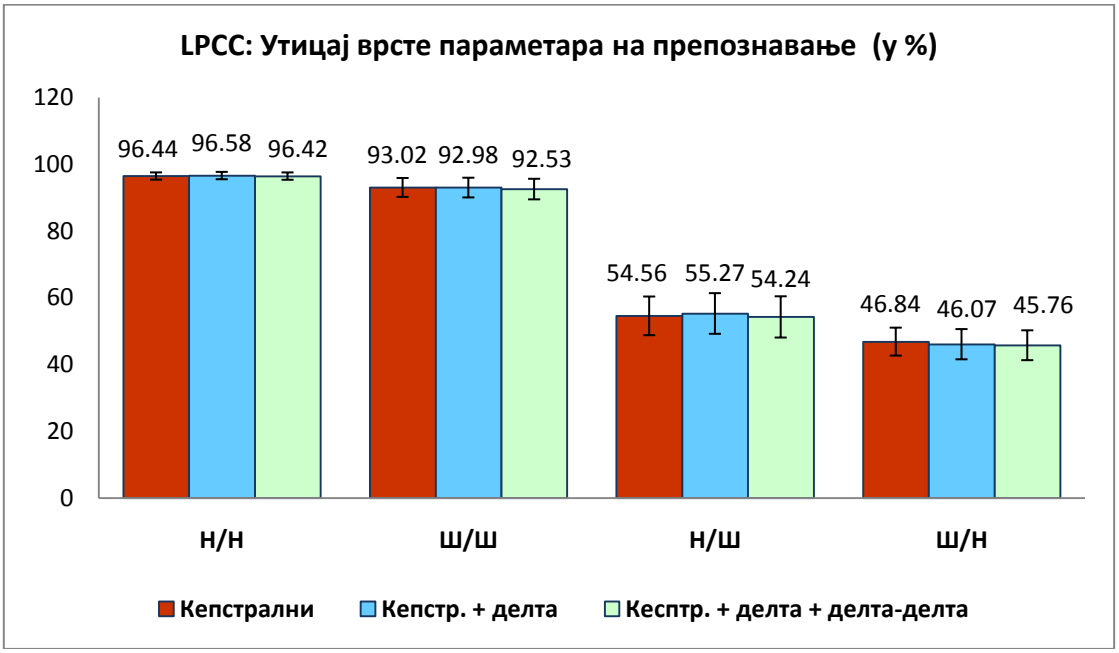
На основу добијених резултата из претходне четири табеле може се формирати дијаграм који показује однос препознавања за случајеве када је и када није примењена CMS нормализација, а дужина вектора је 12 кепстралних коефицијената (слика 6.8).



Слика 6.8 *Резултати препознавања за LPCC обележје без и са CMS-ом.*

На основу добијеног дијаграма може се закључити да CMS нормализација даје побољшање за сваки од размотрених сценарија. Оно се креће од око 1,3% (за сценарио „нормалан/нормалан“) до скоро 19% (за сценарио „нормалан/шапат“).

Утицај врсте параметара (кепстрални, делта и делта-делта), а тиме и одговарајуће дужине вектора, на препознавање дати су на слици 6.9.



Слика 6.9 *Утицај врсте параметара на препознавање за LPCC обележје.*

Анализом дијаграма са слике 6.9 може се уочити да је врло мала разлика између ове три врсте параметара и да се за сценарио „шапат/нормалан“ код примене делта-делта кепстралних коефицијената јавља смањење препознавања за око 1%. Оно се може објаснити као адитивни утицај грешке рачунања јер су делта-делта доприноси врло мали. Њиховим

акумулирањем ова грешака може добити на значају али је и даље у области малих вредности (статистичке грешке).

6.2.2 РЕЗУЛТАТИ НА БАЗИ LFCC И TELFCC ВЕКТОРСКИХ ОБЕЛЕЖЈА

Коришћењем линеарне фреквенцијске скале и одговарајућег TE оператора добијају се векторска обележја одговарајућег типа (као што је описано у потпоглављу 3.2) [Marković et al., 2018]. На бази векторског обележје типа LFCC, а за напред наведене сценарије и варијације параметара, добијају се резултати препознавања који су дати у табелама 6.5-6.8.

Табела 6.5 *LFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	97.11	98.89	98.67	98.67
Говорник 2	96.89	97.33	97.11	97.33
Говорник 3	98.67	99.56	99.56	99.56
Говорник 4	98.67	99.56	99.56	99.56
Говорник 5	97.56	99.11	98.89	98.89
Говорник 6	96.47	96.44	96.22	96.67
Говорник 7	95.78	97.33	97.78	97.56
Говорник 8	98.89	99.11	99.33	99.11
Говорник 9	98.22	99.11	99.11	98.67
Говорник 10	94.22	96.44	97.11	97.33
Ср. вред. ± Грешка	**97.25±0.92**	**98.29±0.78**	**98.33±0.74**	**98.34±0.64**

Табела 6.6 *LFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	94.89	95.56	96.00	96.22
Говорник 2	95.56	97.11	97.33	96.89
Говорник 3	98.44	99.33	99.33	99.33
Говорник 4	97.78	98.22	97.78	97.78
Говорник 5	90.89	93.11	93.56	93.11
Говорник 6	82.22	89.33	88.67	87.78
Говорник 7	88.89	94.44	94.67	94.22
Говорник 8	91.11	95.33	96.67	96.22
Говорник 9	92.44	95.56	96.44	96.67
Говорник 10	82.67	85.33	86.22	86.00
Ср. вред. ± Грешка	**91.45±3.51**	**94.33±2.61**	**94.67±2.58**	**94.42±2.69**

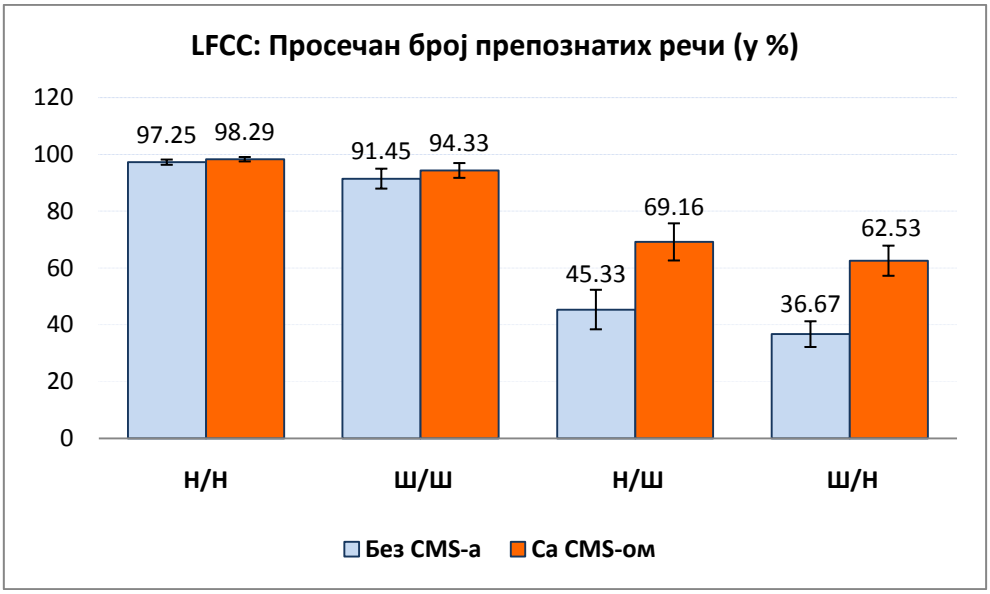
Табела 6.7 *LFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	66.44	81.56	82.00	80.67
Говорник 2	24.67	52.00	50.67	50.00
Говорник 3	48.44	81.11	81.56	81.33
Говорник 4	50.89	77.78	78.22	78.00
Говорник 5	33.56	58.89	58.89	58.44
Говорник 6	45.33	62.00	63.11	63.33
Говорник 7	39.11	65.33	65.56	64.22
Говорник 8	51.56	78.00	77.78	76.44
Говорник 9	46.00	73.78	72.44	69.56
Говорник 10	47.33	61.11	61.78	60.67
Ср. вред. ± Грешка	**45.33±6.96**	**69.16±6.54**	**69.20±6.65**	**68.27±6.59**

Табела 6.8 *LFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	39.78	67.78	66.44	66.00
Говорник 2	23.33	58.22	60.89	60.22
Говорник 3	37.56	77.56	77.56	78.00
Говорник 4	30.89	66.00	68.67	67.56
Говорник 5	27.56	52.00	54.67	54.44
Говорник 6	37.56	51.33	51.33	51.11
Говорник 7	45.56	60.44	60.89	60.00
Говорник 8	43.33	71.78	73.78	72.67
Говорник 9	37.11	64.44	65.11	64.67
Говорник 10	43.78	55.78	57.33	56.67
Ср. вред. ± Грешка	**36.67±4.53**	**62.53±5.31**	**63.67±5.14**	**63.13±5.15**

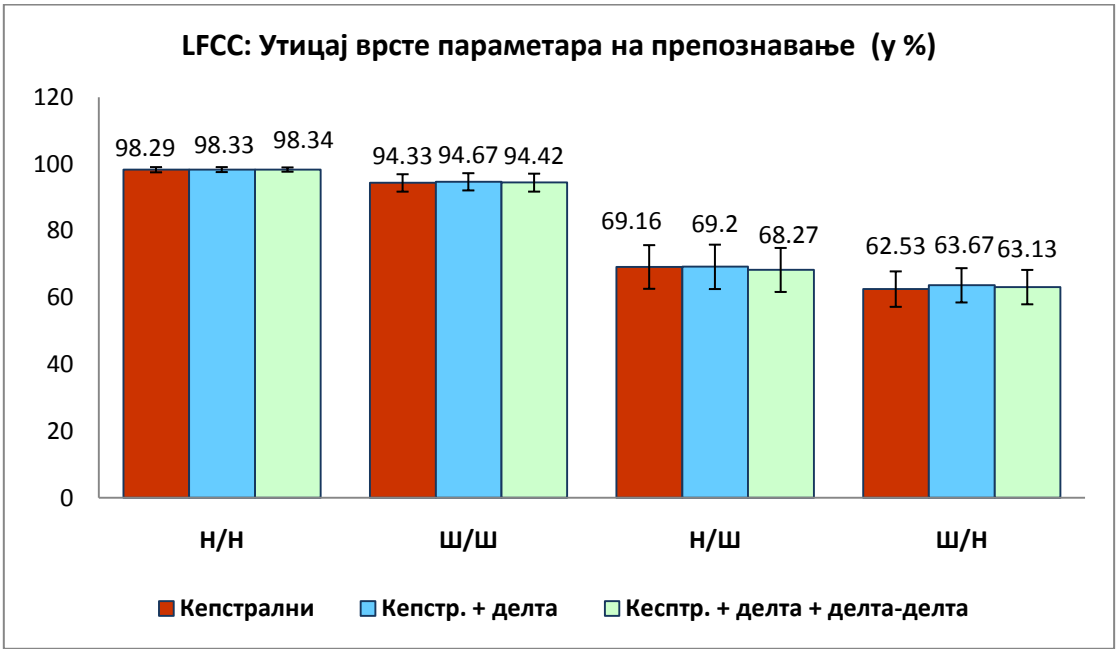
На основу резултата приказаних у претходним табелама дат је дијаграм препознавања за сва четири сценарија и векторе састављене од 12 кепстралних коефицијената за случајеве када није и када је коришћена нормализација (слика 6.10).



Слика 6.10 *Резултати препознавања за LFCC обележје без и са CMS-ом.*

Може се уочити са дијаграма да примена CMS нормализације повећава успешност препознавања за све наведене сценарије и то од око 1% за сценарио „нормалан/нормалан“ па до око 26% за сценарио „шапат/нормалан“.

Поређење успешности препознавања када се користе различити параметри за векторе (кепстрални, кепстрални са делта кепстралним и кепстрални са делта и делта-делта кепстралним коефицијентима, а при томе је примењена нормализација), за поменуте сценарије дато је на слици 6.11.



Слика 6.11 *Утицај врсте параметара на препознавање за LFCC обележје.*

На бази овог дијаграма може се уочити да је врло мала разлика између наведене три врсте параметара и да је добитак највише за око 1% у случају сценарија „шапат/нормалан“ и при коришћењу вектора састављеног од кепстралних и делта кепстралних коефицијената.

Применом Teager Energy оператора добија се векторско обележје типа TELFCC. За напред наведене сценарије и поменуте варијације параметара резултати препознавања су дати у табелама 6.9-6.12.

Табела 6.9 *TELFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	97.11	98.89	98.67	98.44
Говорник 2	96.44	97.78	97.56	97.56
Говорник 3	98.67	99.56	99.56	99.56
Говорник 4	98.67	99.56	99.56	99.56
Говорник 5	97.33	99.11	98.44	98.44
Говорник 6	95.78	97.11	96.44	96.67
Говорник 7	96.22	96.89	97.33	97.56
Говорник 8	99.33	99.33	99.11	98.89
Говорник 9	97.78	98.67	98.89	98.67
Говорник 10	94.89	97.11	98.00	97.56
Ср. вред. ± Грешка	**97.22±0.88**	**98.40±0.67**	**98.36±0.63**	**98.29±0.58**

Табела 6.10 *TELFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	94.44	96.22	96.22	96.22
Говорник 2	95.78	97.11	96.89	96.22
Говорник 3	98.44	99.11	99.33	99.33
Говорник 4	97.33	98.22	97.78	98.00
Говорник 5	90.89	93.33	93.33	93.33
Говорник 6	81.33	89.78	89.11	88.22
Говорник 7	89.56	94.89	94.89	94.67
Говорник 8	90.22	96.00	96.44	96.44
Говорник 9	93.33	96.00	96.44	96.67
Говорник 10	82.89	85.78	85.78	85.78
Ср. вред. ± Грешка	**91.42±3.55**	**94.64±2.52**	**94.62±2.57**	**94.49±2.67**

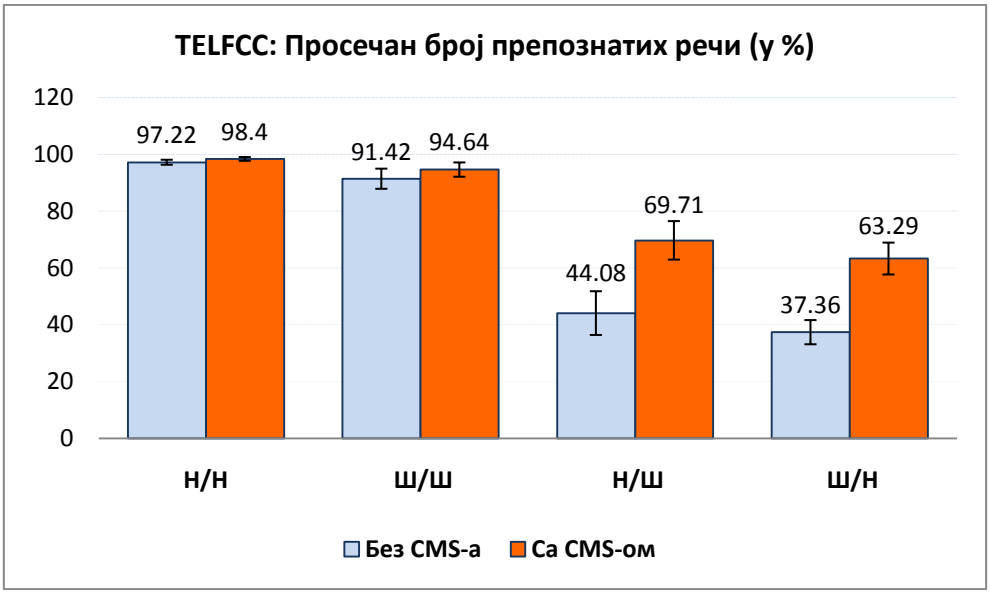
Табела 6.11 *TELFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	67.56	84.00	82.89	82.44
Говорник 2	25.78	52.44	51.56	52.22
Говорник 3	48.56	82.89	83.11	82.22
Говорник 4	52.00	78.89	79.11	77.78
Говорник 5	32.89	61.11	61.11	59.33
Говорник 6	29.11	61.33	62.00	61.56
Говорник 7	40.44	62.89	64.67	64.00
Говорник 8	50.22	77.78	77.33	76.89
Говорник 9	46.67	73.11	71.11	70.22
Говорник 10	47.56	62.67	62.44	61.11
Ср. вред. ± Грешка	**44.08±7.70**	**69.71±6.78**	**69.53±6.67**	**68.78±6.58**

Табела 6.12 *TELFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	41.56	66.67	67.56	66.00
Говорник 2	25.11	60.44	61.11	61.11
Говорник 3	38.00	79.33	80.89	80.00
Говорник 4	32.22	70.00	71.78	71.11
Говорник 5	28.67	52.89	56.00	56.00
Говорник 6	38.44	51.11	50.67	50.67
Говорник 7	45.56	59.33	59.56	60.00
Говорник 8	43.33	72.67	73.56	72.00
Говорник 9	36.22	64.67	66.22	64.67
Говорник 10	44.44	55.78	57.33	56.00
Ср. вред. ± Грешка	**37.36±4.26**	**63.29±5.62**	**64.47±5.72**	**63.76±5.48**

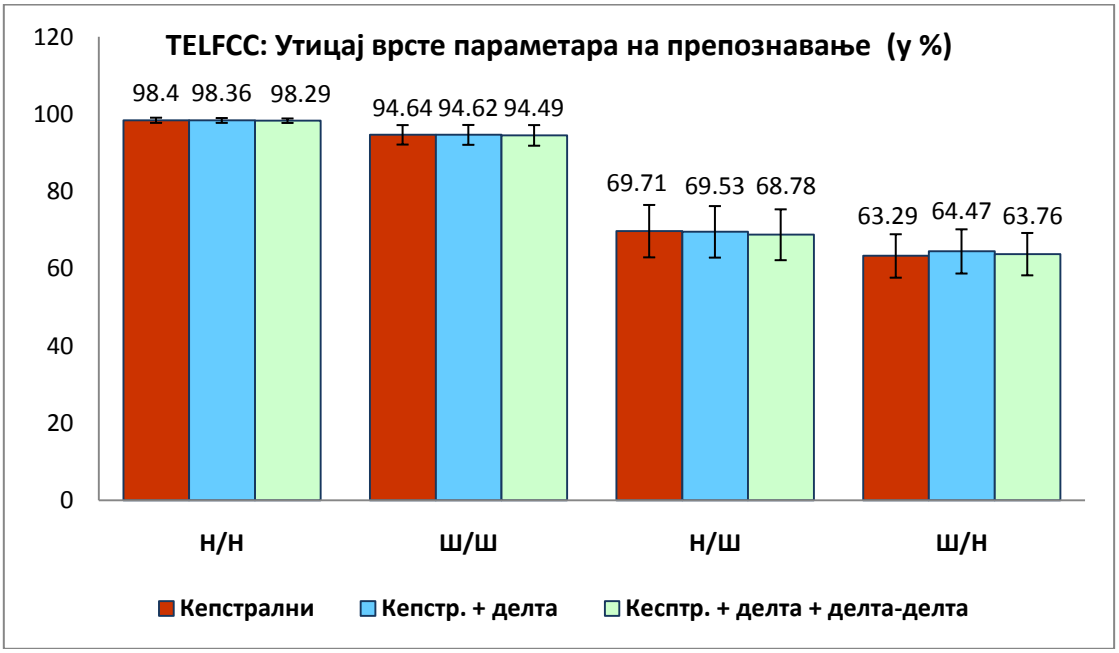
На бази резултата приказаних у табелама 6.9-6.12 дат је дијаграм препознавања (у процентима) за сва четири сценарија и кепстралне коефицијенте и то за случајеве без и са CMS нормализацијом (слика 6.12).



Слика 6.12 *Резултати препознавања за TELFCC обележје без и са CMS-ом.*

Анализом дијаграма са слике 6.12 може се уочити да примена нормализације типа CMS у свим случајевима даје побољшања. Она се крећу од око 1% за сценарио „нормалан/нормалан“ до чак 26% за сценарио „шапат/нормалан“.

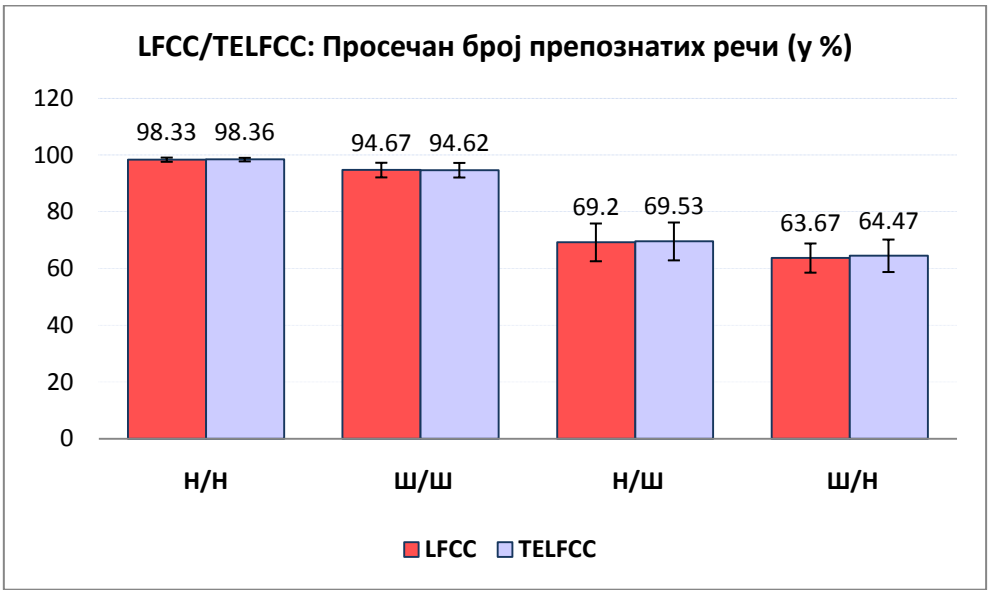
Поређење успешности препознавања када се користе различите врсте параметра и различите дужине вектора (само кепстрални, кепстрални са делта кепстралним и кепстрални са делта и делта-делта кепстралним коефицијентима) за поменуте сценарије дато је на слици 6.13.



Слика 6.13 *Утицај врсте параметара на препознавање за TELFCC обележје.*

Анализом вредности са дијаграма може се уочити да неме велике разлике између добијених резултата без обзира на врсту параметара и дужину вектора који се користе. Највећа разлика је нешто изнад 1% код сценарија „шапат/нормалан“.

На бази напред приказаних резултата могуће је поредити LFCC и TELFCC векторска обележја. Ово поређење извршено је коришћењем вектора састављених од 24 елемента који се састоје од кепстралних и делта кепстралних коефицијената (јер дају у просеку најбоље резултате). Слика 6.14 даје приказ овог поређења.



Слика 6.14 *Упоредна анализа препознавања за LFCC и TELFCC обележја.*

Анализом претходног дијаграма може се уочити да применом TE оператора скоро у свим случајевима се добија одређено побољшање препознавања, а највеће је за сценарије „нормалан/шапат“ и „шапат/нормалан“. Међутим и то повећање је испод 1%.

6.2.3 РЕЗУЛТАТИ НА БАЗИ MFCC И TEMFCC ВЕКТОРСКИХ ОБЕЛЕЖЈА

Употребом “mel” фреквенцијске скале и нелинеарног Teager Energy оператора добијају се векторска обележја одговарајућег типа (као што је описано у потпоглављу 3.3) и [Marković et al., 2018].

Приказ резултата препознавања за MFCC векторско обележје у форми процента успешно препознатих речи, за раније објашњене сценарије и варијације параметара вектора, дат је у табелама 6.13-6.16.

Табела 6.13 *MFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.11	99.78	99.56	99.56
Говорник 2	99.11	99.78	99.78	99.56
Говорник 3	98.67	99.11	99.11	99.11
Говорник 4	99.78	99.56	99.33	99.11
Говорник 5	99.33	99.56	99.56	99.56
Говорник 6	98.89	99.33	99.11	98.00
Говорник 7	98.44	98.67	98.44	98.44
Говорник 8	98.89	98.67	98.67	98.67
Говорник 9	99.11	99.56	99.56	99.56
Говорник 10	98.67	98.89	98.89	98.89
Ср. вред. ± Грешка	**99.00±0.24**	**99.29±0.27**	**99.20±0.27**	**99.05±0.34**

Табела 6.14 *MFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	96.44	97.78	97.78	97.78
Говорник 2	98.44	99.11	99.56	99.56
Говорник 3	99.56	99.78	99.78	99.78
Говорник 4	98.89	99.33	99.56	98.89
Говорник 5	95.11	97.33	97.11	97.11
Говорник 6	90.44	95.11	94.89	94.44
Говорник 7	96.22	97.11	97.33	96.89
Говорник 8	94.67	97.78	97.78	97.78
Говорник 9	95.56	97.56	97.78	97.78
Говорник 10	86.89	90.89	90.89	91.11
Ср. вред. ± Грешка	**95.22±2.43**	**97.18±1.60**	**97.25±1.65**	**97.11±1.61**

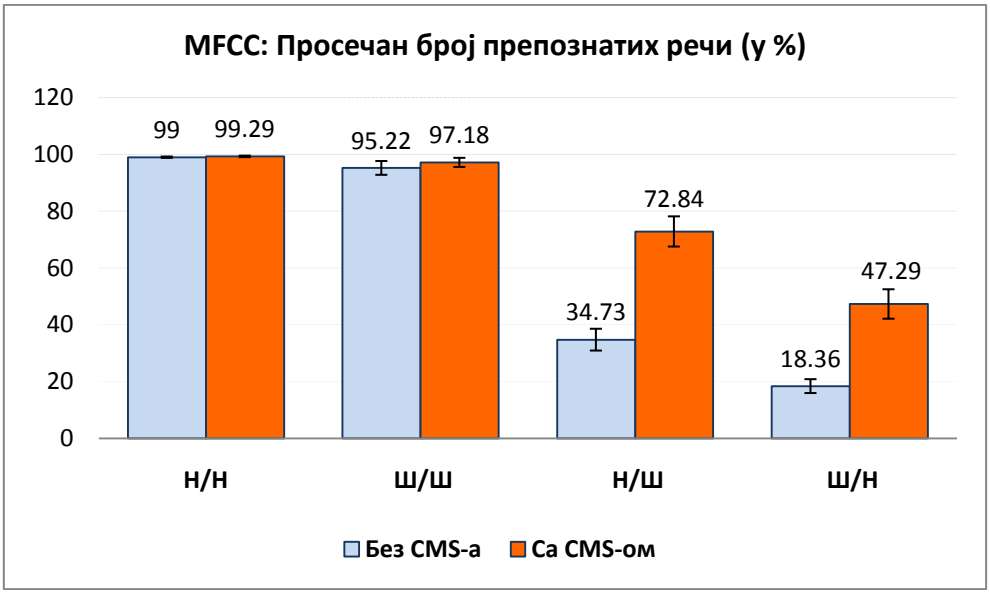
Табела 6.15 *MFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	42.00	78.22	78.22	77.78
Говорник 2	22.67	63.11	63.33	62.44
Говорник 3	35.56	74.67	75.78	75.33
Говорник 4	40.00	82.22	83.11	81.78
Говорник 5	31.78	72.00	71.11	69.78
Говорник 6	34.00	68.00	67.56	65.11
Говорник 7	38.44	73.33	73.56	72.67
Говорник 8	34.00	82.00	81.11	78.00
Говорник 9	41.33	79.33	79.33	78.22
Говорник 10	27.56	55.56	53.78	52.89
Ср. вред. ± Грешка	**34.73±3.84**	**72.84±5.32**	**72.69±5.61**	**71.40±5.55**

Табела 6.16 *MFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	17.33	46.44	45.78	45.78
Говорник 2	12.89	41.78	43.33	43.78
Говорник 3	19.78	49.78	51.56	52.44
Говорник 4	16.67	54.89	57.56	56.67
Говорник 5	13.33	38.44	41.11	42.22
Говорник 6	21.11	36.44	37.56	37.56
Говорник 7	25.78	52.00	53.78	53.78
Говорник 8	15.78	58.89	58.44	57.33
Говорник 9	20.22	56.22	59.33	59.56
Говорник 10	20.67	38.00	40.00	39.78
Ср. вред. ± Грешка	**18.36±2.44**	**47.29±5.19**	**49.05±5.12**	**48.89±4.95**

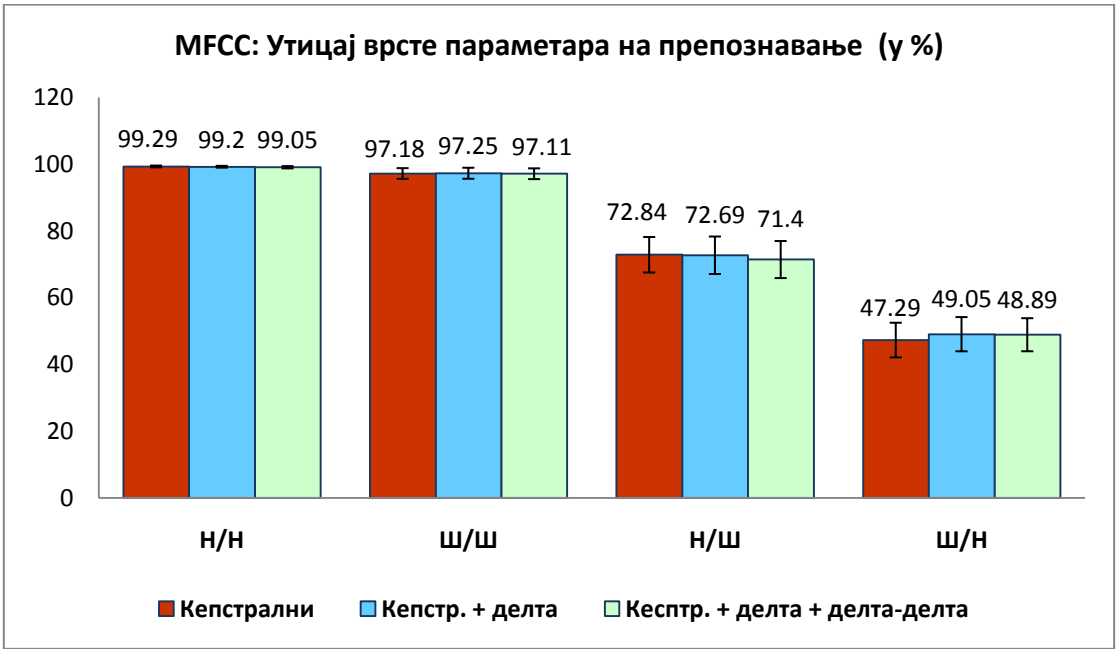
Коришћењем резултата који су приказани у претходне четири табеле дат је дијаграм резултата препознавања за четири сценарија при чему су вектори састављени само од кепстралних коефицијената и то за случајеве када није и када јесте коришћена нормализација CMS типа (слика 6.15).



Слика 6.15 *Резултати препознавања за MFCC обележје без и са CMS-ом.*

Примена CMS-а за ово векторско обележје даје у свим случајевима побољшање препознавања. Оно се креће од трећине процента (за сценарио „нормалан/нормалан“) па чак до 38 % (за сценарио „нормалан/шапат“).

Ако се пореде вектори различитих параметара, па стога и различитих дужина, (састављени од кепстралних, делта и делта-делта коефицијената) на којима је примењена нормализација, добија се следећи дијаграм (слика 6.16).



Слика 6.16 *Утицај врсте параметара на препознавање за MFCC обележје.*

На бази овог дијаграма уочава се да су вредности препознавања скоро идентичне за све три врсте параметара. Изузетак је сценарио „шапат/нормалан“ где се увођењем делта коефицијената добија побољшање које је изнад 1%.

Коришћењем векторског обележја типа TEMFCC добијени су резултати који су приказани у табелама 6.17-6.20.

Табела 6.17 *TEMFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.11	99.78	99.78	99.56
Говорник 2	99.11	100.00	99.56	99.78
Говорник 3	98.89	99.11	99.11	99.11
Говорник 4	99.56	99.33	99.11	99.11
Говорник 5	99.33	99.78	99.78	99.78
Говорник 6	98.67	99.11	98.89	98.44
Говорник 7	98.00	98.44	98.44	98.44
Говорник 8	98.89	98.67	98.67	98.67
Говорник 9	98.89	99.33	99.33	99.11
Говорник 10	98.22	98.89	99.11	98.89
Ср. вред. ± Грешка	**98.87±0.29**	**99.24±0.31**	**99.18±0.28**	**99.09±0.31**

Табела 6.18 *TEMFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	96.44	97.78	97.33	97.33
Говорник 2	98.00	99.11	99.56	99.11
Говорник 3	99.56	99.78	99.78	99.56
Говорник 4	99.11	99.33	99.11	99.33
Говорник 5	95.33	97.33	96.89	96.22
Говорник 6	90.44	95.33	94.67	94.44
Говорник 7	96.22	97.33	97.33	96.67
Говорник 8	96.22	97.78	98.00	98.44
Говорник 9	96.00	97.78	98.22	98.22
Говорник 10	86.00	90.22	90.22	90.22
Ср. вред. ± Грешка	**95.33±2.56**	**97.18±1.70**	**97.11±1.76**	**96.95±1.77**

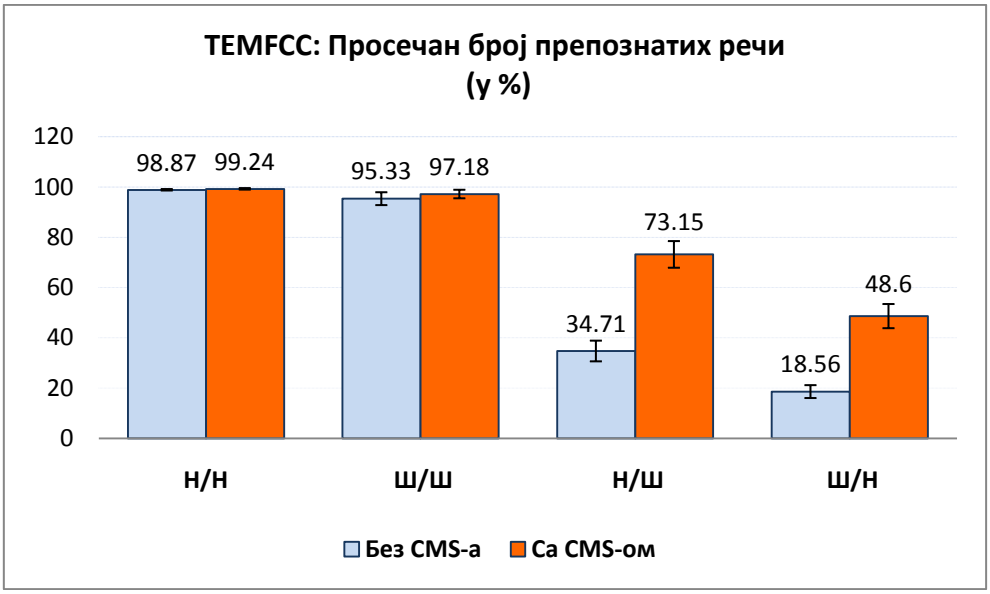
Табела 6.19 *TEMFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	42.22	78.89	78.67	78.22
Говорник 2	22.00	64.00	64.44	63.78
Говорник 3	37.11	76.22	76.44	76.00
Говорник 4	40.00	83.33	82.00	81.78
Говорник 5	30.44	72.22	72.00	70.44
Говорник 6	34.89	66.22	68.89	68.00
Говорник 7	38.44	74.00	73.78	72.44
Говорник 8	35.11	81.33	80.22	78.00
Говорник 9	40.89	78.89	79.11	78.89
Говорник 10	26.00	56.44	56.67	54.00
Ср. вред. ± Грешка	**34.71±4.12**	**73.15±5.30**	**73.22±4.94**	**72.16±5.24**

Табела 6.20 *TEMFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	17.11	46.22	46.22	45.56
Говорник 2	13.56	42.22	44.67	44.44
Говорник 3	19.33	49.78	52.22	54.00
Говорник 4	16.00	57.11	58.89	58.44
Говорник 5	13.56	42.00	43.33	44.00
Говорник 6	22.00	38.67	38.00	38.44
Говорник 7	26.67	52.67	54.67	53.11
Говорник 8	16.00	60.67	58.89	57.11
Говорник 9	19.78	56.00	58.89	59.33
Говорник 10	21.56	40.67	39.78	40.67
Ср. вред. ± Грешка	**18.56±2.56**	**48.60±4.81**	**49.56±5.05**	**49.51±4.80**

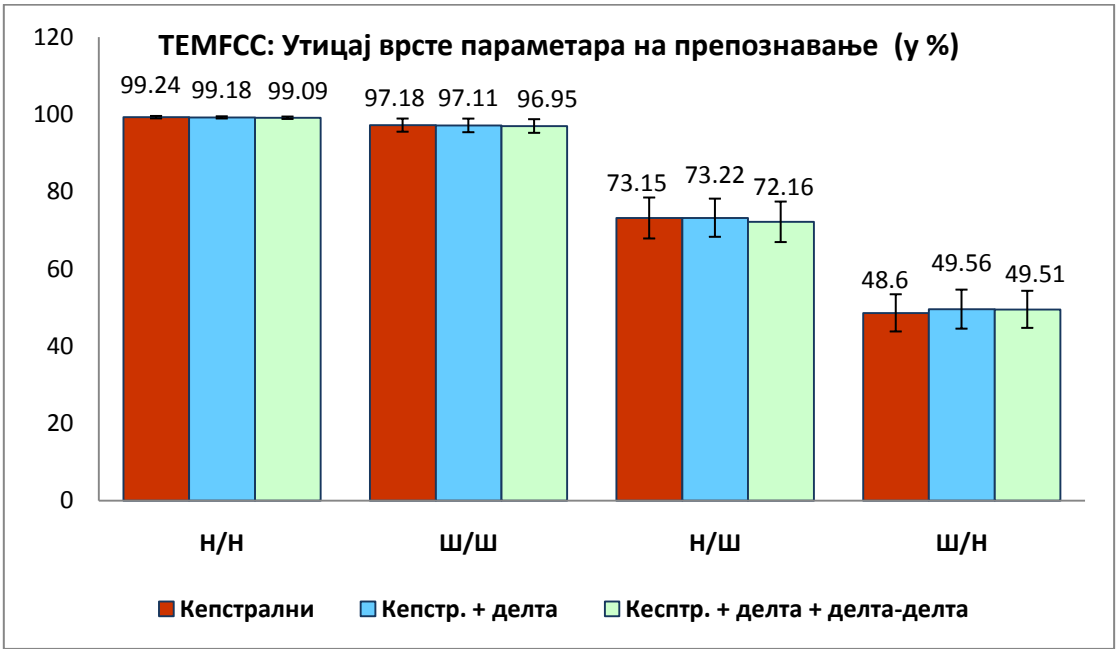
На основу резултата који су приказани у табелама 6.17-6.20 дат је дијаграм за резултате препознавања за сва четири сценарија при чему су вектори састављени само од кепстралних коефицијената и то за случајеве када није и када јесте коришћена нормализација (слика 6.17).



Слика 6.17 *Резултати препознавања за TEMFCC обележје без и са CMS-ом.*

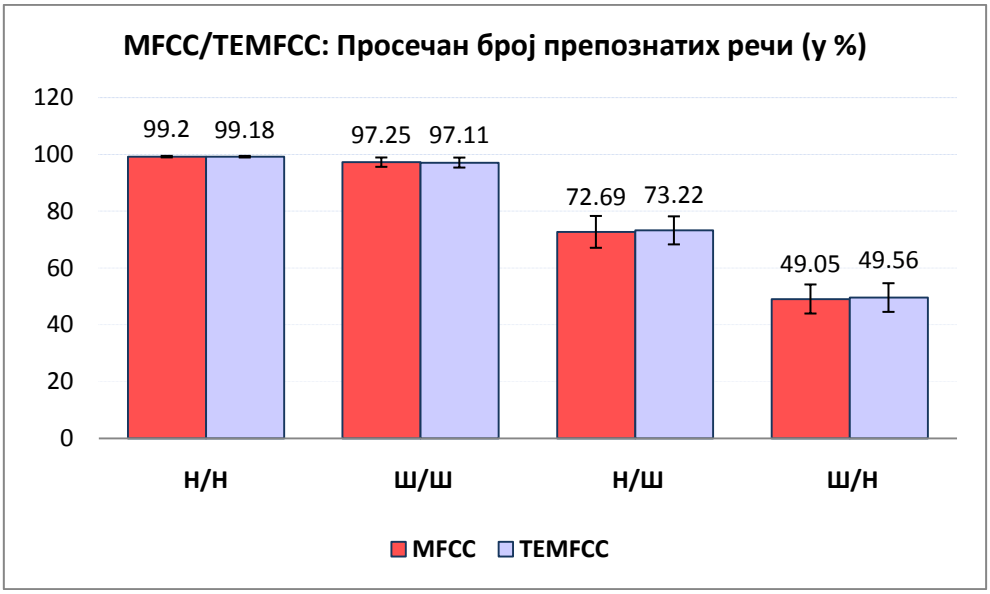
На сличан начин као и за MFCC векторско обележје и овде је очигледан велики утицај CMS нормализације. Нарочито се он огледа код неусаглашених сценарија где за сценарио „нормалан/шапат“ побољшање износи и преко 38%.

Поређење резултата на бази различитих параметара које садржи TEMFCC векторско обележје дато је на слици 6.18.



Слика 6.18 *Утицај врсте параметара на препознавање за TEMFCC обележје.*

Анализом добијених резултата могуће је поредити ова сродна MFCC и TEMFCC векторска обележја. Дијаграм на слици 6.19 даје упоредни приказ за векторе који се састоје од 24 параметра (кепстрални и делта кепстрални коефицијенти).



Слика 6.19 *Упоредна анализа препознавања за MFCC и TEMFCC обележја.*

Са дијаграма се уочава да коришћење TE оператора условљава побољшање препознавања од око 0,5% за оба неусаглашена сценарија. Код усаглашених сценарија разлика је на првој или другој децимали, тј. занемарљива.

6.2.4 РЕЗУЛТАТИ НА БАЗИ GFCC и TEGFCC ВЕКТОРСКИХ ОБЕЛЕЖЈА

Коришћење Gammatone филтера у процесу предобраде омогућава добијање GFCC векторског обележја [Marković et al., 2015], [Marković et al., 2017 b], а додавањем и нелинеарног Teager Energy оператора добијају се TEGFCC векторска обележја (као што је детаљно објашњено у потпоглављу 3.4).

Резултати добијени за сва четири сценарија са одговарајућим врстама параметара за GFCC обележје су приказани у табелама 6.21-6.24.

Табела 6.21 *GFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.78	99.56	99.78	99.78
Говорник 2	99.33	99.56	99.56	99.33
Говорник 3	98.89	98.89	98.89	98.89
Говорник 4	99.33	99.11	99.11	99.33
Говорник 5	98.89	99.33	99.33	99.33
Говорник 6	98.00	99.33	99.11	98.89
Говорник 7	96.22	97.11	97.11	97.11
Говорник 8	99.33	99.56	99.11	98.89
Говорник 9	99.33	98.89	98.89	98.89
Говорник 10	93.33	95.78	95.78	96.22
Ср. вред. ± Грешка	**98.24±1.24**	**98.71±0.78**	**98.67±0.77**	**98.67±0.69**

Табела 6.22 *GFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	95.11	96.00	96.00	96.00
Говорник 2	92.89	93.56	93.11	93.11
Говорник 3	95.11	95.33	95.78	95.78
Говорник 4	94.00	94.22	94.00	94.00
Говорник 5	86.67	87.56	87.78	87.78
Говорник 6	67.78	73.56	73.11	71.56
Говорник 7	81.11	82.67	82.22	82.00
Говорник 8	90.67	93.33	92.44	92.75
Говорник 9	87.11	90.00	89.56	88.50
Говорник 10	78.67	77.56	77.78	78.22
Ср. вред. ± Грешка	**86.91±5.47**	**88.38±4.91**	**88.18±4.92**	**87.97±5.11**

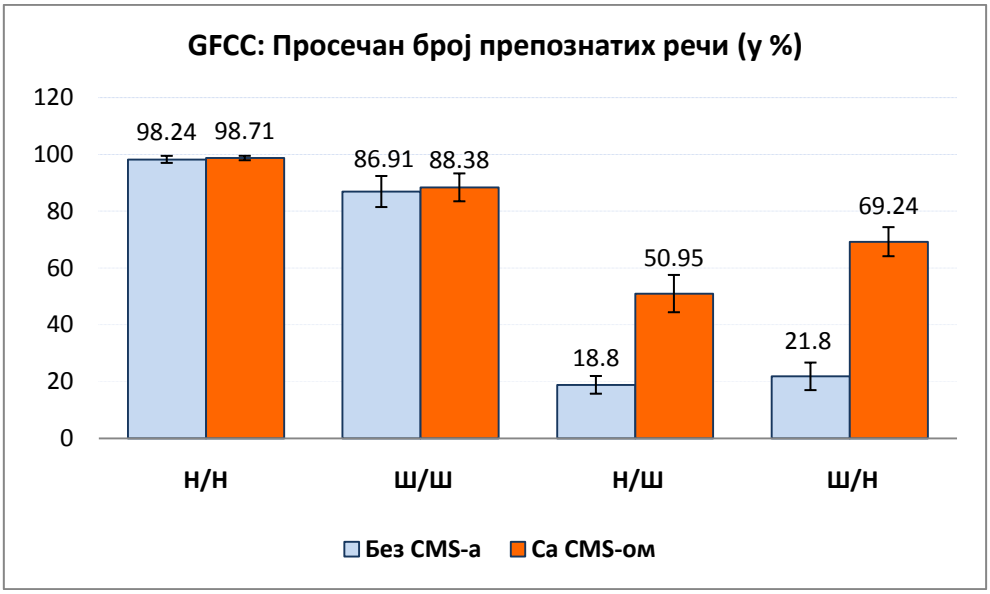
Табела 6.23 *GFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	23.56	56.89	56.44	56.44
Говорник 2	16.89	58.67	58.89	57.78
Говорник 3	10.00	38.44	38.67	37.78
Говорник 4	17.33	52.44	50.89	50.67
Говорник 5	17.78	54.44	53.33	53.11
Говорник 6	15.11	42.44	42.00	40.89
Говорник 7	24.00	54.22	54.67	54.89
Говорник 8	23.33	63.11	62.67	61.33
Говорник 9	25.33	59.11	58.89	58.67
Говорник 10	14.67	29.78	30.00	29.78
Ср. вред. ± Грешка	**18.80±3.12**	**50.95±6.57**	**50.65±6.47**	**50.13±6.47**

Табела 6.24 *GFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	21.78	78.00	77.11	77.11
Говорник 2	12.89	61.11	61.78	61.33
Говорник 3	17.56	67.56	68.44	67.78
Говорник 4	15.56	77.56	77.78	76.89
Говорник 5	20.22	62.89	62.67	61.56
Говорник 6	32.22	72.89	74.22	73.78
Говорник 7	33.11	65.33	65.78	64.89
Говорник 8	11.56	75.78	75.33	74.44
Говорник 9	30.22	77.11	77.78	77.78
Говорник 10	22.89	54.22	54.89	54.67
Ср. вред. ± Грешка	**21.80±4.85**	**69.24±5.12**	**69.58±5.00**	**69.02±5.04**

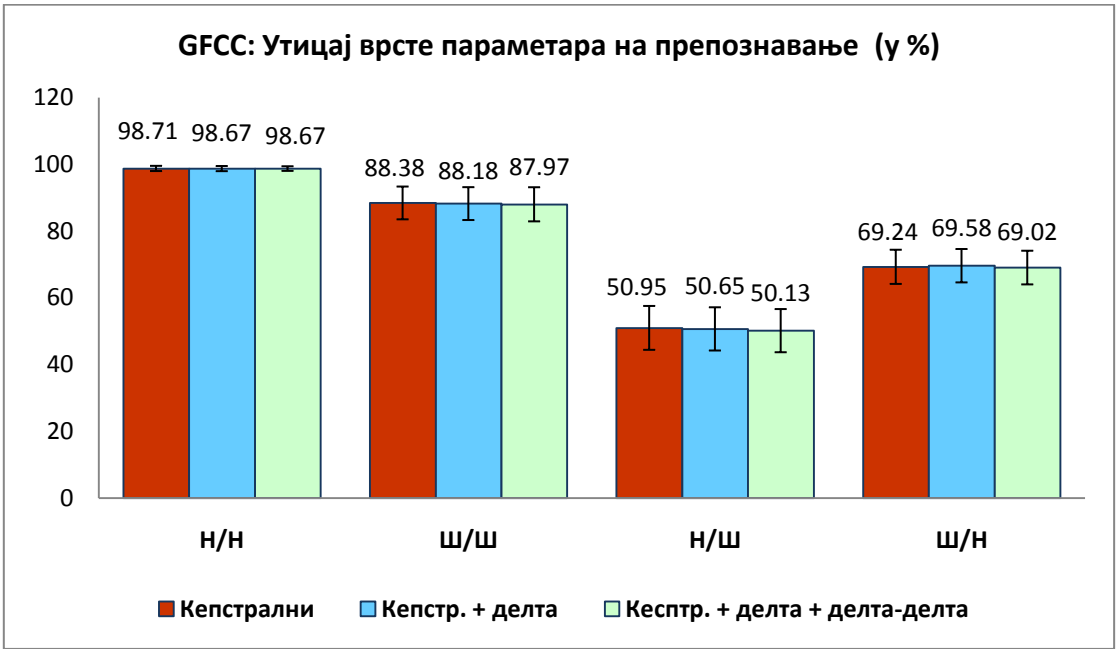
Коришћењем резултата из табеле 6.21-6.24 дат је дијаграм за сва четири сценарија и векторе састављене од кепстралних коефицијената и то за случај када није коришћена CMS нормализација и када јесте (слика 6.20).



Слика 6.20 *Резултати препознавања за GFCC обележје без и са CMS-ом.*

Анализа резултата са слике 6.20 показује да примена CMS-а даје побољшање препознавања од 0,5% за сценарио „нормалан/нормалан“ па до чак 47% за сценарио „шапат/нормалан“.

Утицај врсте параметара и дужине вектора на препознавање говора за ово векторско обележје приказан је на слици 6.21.



Слика 6.21 *Утицај врсте параметара на препознавање за GFCC обележје.*

На бази претходног дијаграма може се уочити да променом параметара и дужине вектора скоро да и нема промене у успешности препознавања. Разлике су на првој или другој децимали што је у рангу статистичке грешке.

Применом TE оператора на сигнал који пролази кроз Gammatone филтере добијају се TEGFCC векторска обележја. Резултати добијени за сва четири сценарија и за одговарајуће типове вектора овог обележја приказани су у табелама 6.25-6.28.

Табела 6.25 *TEGFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TEGFCC кепст. (без CMS-а)	TEGFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.56	99.33	99.78	99.78
Говорник 2	99.11	99.56	99.56	99.56
Говорник 3	99.33	99.78	99.44	99.44
Говорник 4	99.11	99.11	99.33	99.44
Говорник 5	98.89	99.44	99.56	99.33
Говорник 6	98.44	99.11	99.44	99.11
Говорник 7	96.33	98.00	97.78	98.00
Говорник 8	99.11	99.33	99.44	99.11
Говорник 9	99.33	98.89	98.89	98.89
Говорник 10	93.44	96.78	96.89	96.89
Ср. вред. ± Грешка	**98.27±1.20**	**98.93±0.56**	**99.01±0.58**	**98.96±0.54**

Табела 6.26 *TEGFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TEGFCC кепст. (без CMS-а)	TEGFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	94.89	96.11	96.22	96.11
Говорник 2	91.89	93.78	93.78	93.78
Говорник 3	95.44	94.89	94.89	94.78
Говорник 4	94.00	94.33	95.11	94.89
Говорник 5	87.78	88.78	88.78	88.44
Говорник 6	69.78	77.89	77.44	77.56
Говорник 7	81.33	81.44	82.11	82.33
Говорник 8	90.44	92.78	92.78	92.78
Говорник 9	87.33	90.56	90.33	90.33
Говорник 10	79.67	78.33	78.11	78.33
Ср. вред. ± Грешка	**87.26±5.06**	**88.89±4.37**	**88.96±4.44**	**88.93±4.37**

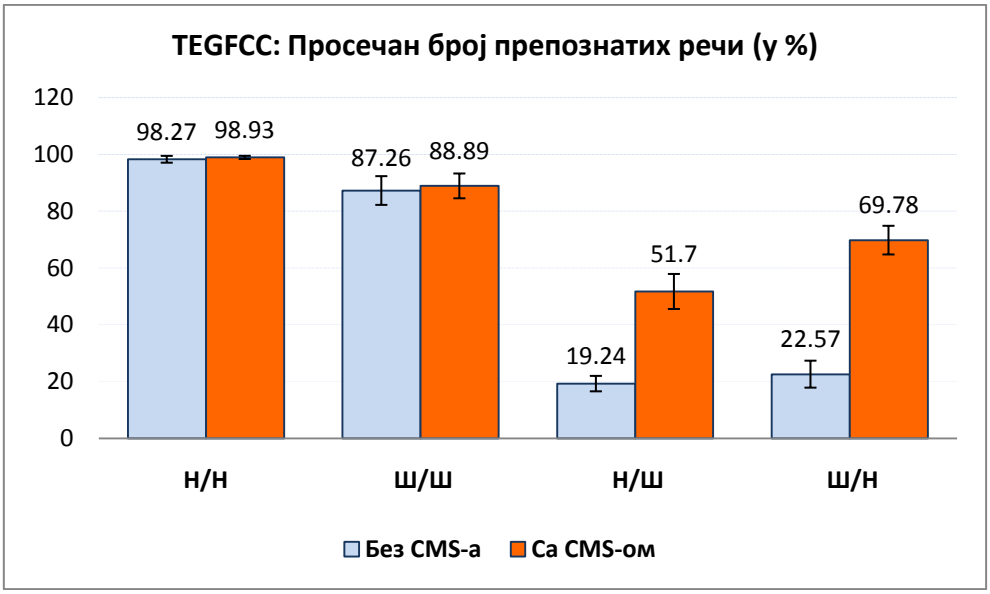
Табела 6.27 *TEGFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TEGFCC кепст. (без CMS-а)	TEGFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	24.11	57.11	57.44	57.33
Говорник 2	16.78	59.44	59.44	59.44
Говорник 3	12.33	40.22	41.44	41.56
Говорник 4	17.11	51.89	52.11	52.00
Говорник 5	18.56	53.44	53.22	53.11
Говорник 6	15.44	45.56	47.33	47.56
Говорник 7	22.89	55.11	54.89	54.56
Говорник 8	23.78	64.00	63.89	64.00
Говорник 9	25.11	58.89	58.78	58.56
Говорник 10	16.33	31.33	32.56	32.78
Ср. вред. ± Грешка	**19.24±2.72**	**51.70±6.17**	**52.11±5.82**	**52.09±5.76**

Табела 6.28 *TEGFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TEGFCC кепст. (без CMS-а)	TEGFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	22.56	78.11	77.11	77.89
Говорник 2	13.33	62.56	61.78	64.44
Говорник 3	18.44	67.56	68.44	67.11
Говорник 4	16.11	78.11	77.78	78.56
Говорник 5	22.22	63.56	62.67	64.44
Говорник 6	31.89	73.33	74.22	72.33
Говорник 7	33.22	64.89	65.78	64.89
Говорник 8	12.78	77.22	75.33	77.33
Говорник 9	32.00	77.33	77.78	77.56
Говорник 10	23.11	55.11	54.89	57.11
Ср. вред. ± Грешка	**22.57±4.75**	**69.78±5.05**	**70.29±5.00**	**70.17±4.68**

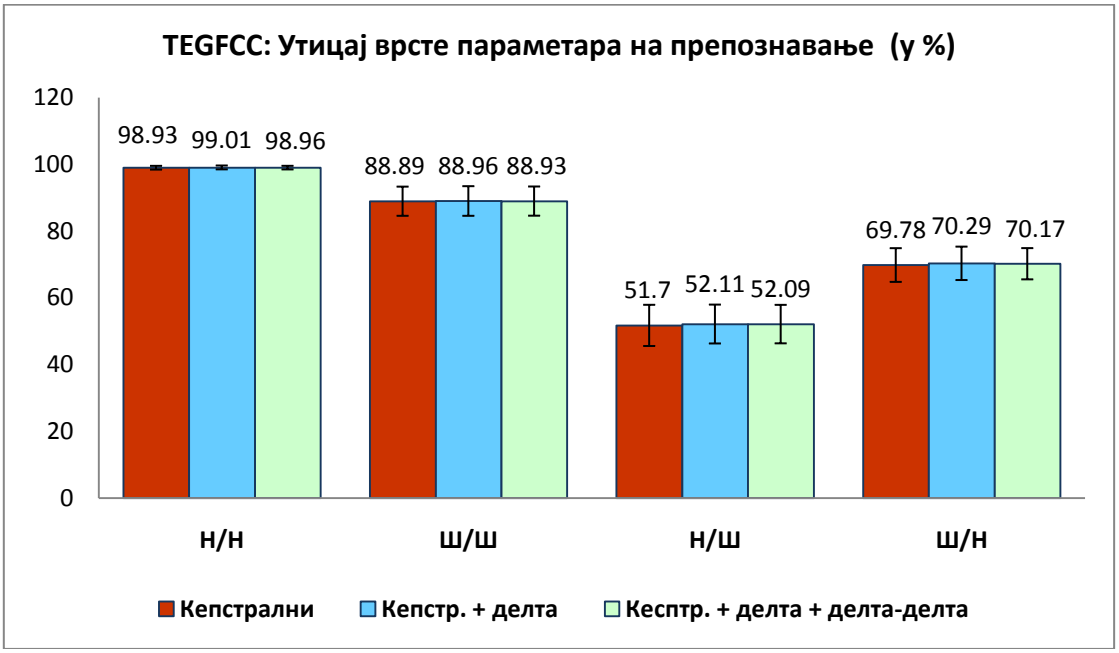
Коришћењем резултата приказаних у табелама 6.25-6.28 нацртан је дијаграм успешности препознавања за сва четири сценарија и векторе састављене од кепстралних коефицијената и то за случајеве без и са CMS нормализацијом (слика 6.22).



Слика 6.22 *Резултати препознавања за TEGFCC обележје без и са CMS-ом.*

И у овом случају потврђује се правило да примена CMS нормализације доприноси побољшању препознавања и то у износу од 0,6% (за сценарио „нормалан/нормалан“) па до чак 47% (за сценарио „шапат/нормалан“).

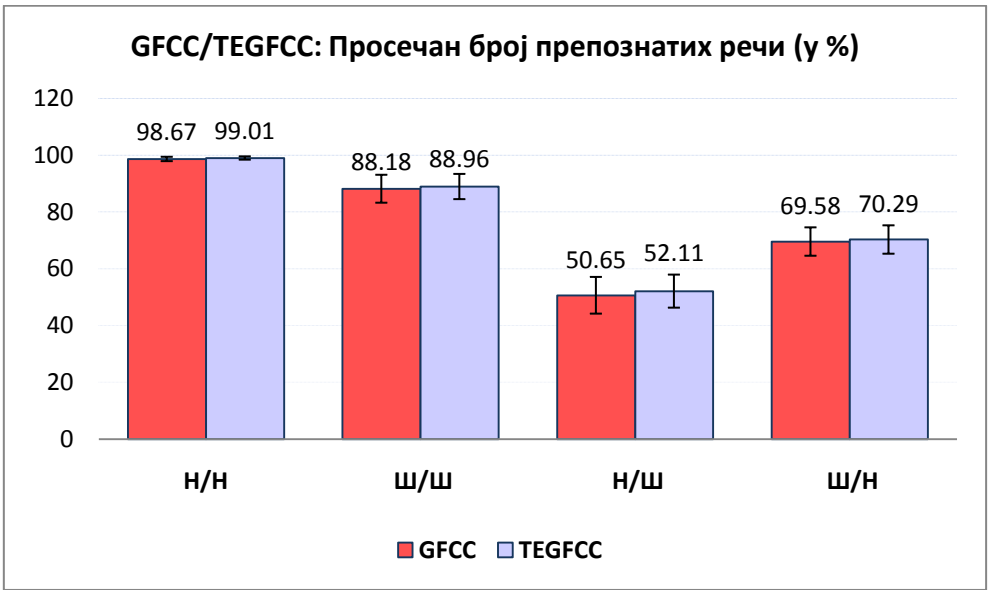
Утицаји различитих врста параметара (кепстрални, делта и делта-делта) дати су на слици 6.23 где је у свим случајевима примењена CMS нормализација.



Слика 6.23 *Утицај врсте параметара на препознавање за TEGFCC обележје.*

Анализом дијаграма са слике 6.23 може се уочити да су вредности препознавања врло сличне и да је највећа разлика од око 0,5% и то за сценарио „шапат/нормалан“.

На бази добијених резултата могуће је упоредити GFCC и TEGFCC векторска обележја. Поређење је извршено за векторе састављене од 24 параметра (12 кепстралних и 12 делта кепстралних коефицијената), а слика 6.24 даје њихове односе.



Слика 6.24 *Упоредна анализа препознавања за GFCC и TEGFCC обележја.*

Добијени дијаграм показује да се применом TE оператора добило побољшање препознавања у свим сценаријим и то од 0,6% (за сценарио „нормалан/нормалан“) до 1,5% (за сценарио „нормалан/шапат“).

6.2.5 РЕЗУЛТАТИ НА БАЗИ PLPCC и TEPLPCC ВЕКТОРСКИХ ОБЕЛЕЖЈА

Употребом перцептивне линеарне предикције током предобраде добија се PLPCC векторско обележје [Marković et al., 2016], а применом нелинеарног TE оператора добија се TEPLPCC векторско обележје (као што је детаљно објашњено у 3.5).

Резултати добијени за сва четири сценарија и за одговарајуће врсте вектора за PLPCC обележје су приказани у табелама 6.29-6.32.

Табела 6.29 *PLPCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	PLPCC кепст. (без CMS-а)	PLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.78	100	100	100
Говорник 2	99.33	99.78	99.56	99.56
Говорник 3	98.44	98.22	98.00	98.67
Говорник 4	99.33	99.33	99.33	99.11
Говорник 5	99.11	99.78	99.78	99.78
Говорник 6	98.00	99.56	98.89	99.11
Говорник 7	96.22	98.22	98.22	98.44
Говорник 8	98.67	98.44	98.44	98.44
Говорник 9	98.89	100	100	100
Говорник 10	92.67	98.44	99.11	99.11
Ср. вред. ± Грешка	**98.04±1.32**	**99.18±0.47**	**99.13±0.45**	**99.22±0.37**

Табела 6.30 *PLPCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	PLPCC кепст. (без CMS-а)	PLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	94.89	97.56	97.11	97.11
Говорник 2	96.89	98.22	98.00	98.00
Говорник 3	97.56	99.56	99.11	99.11
Говорник 4	94.67	97.56	98.00	98.00
Говорник 5	93.33	96.44	95.56	95.33
Говорник 6	77.11	93.78	93.78	93.33
Говорник 7	90.89	96.67	96.89	96.89
Говорник 8	86.48	95.56	95.56	94.89
Говорник 9	93.78	97.11	97.11	97.11
Говорник 10	81.33	89.33	88.89	88.89
Ср. вред. ± Грешка	**90.69±4.26**	**96.18±1.77**	**96.00±1.81**	**95.87±1.85**

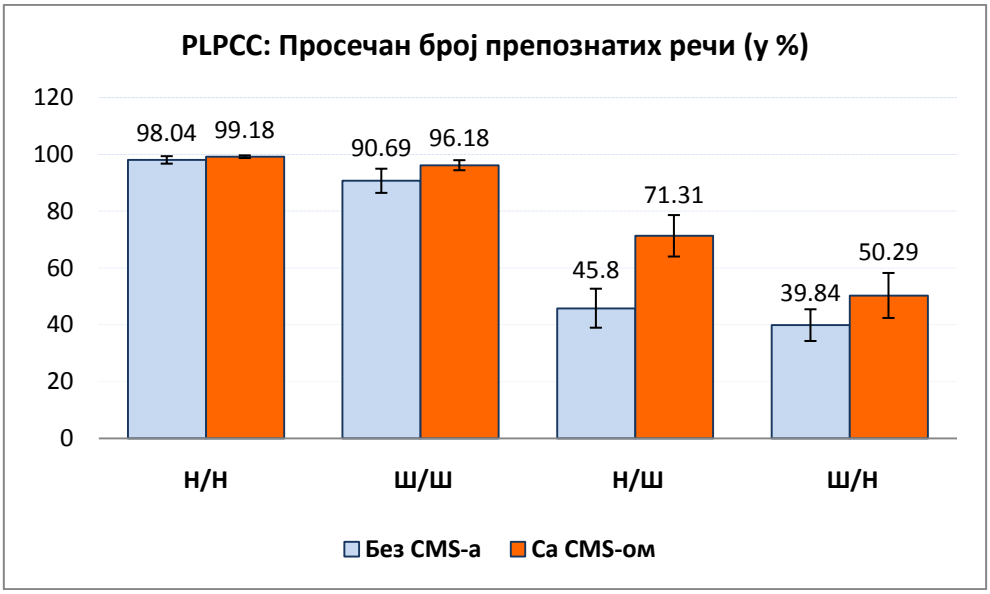
Табела 6.31 *PLPCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	PLPCC кепст. (без CMS-а)	PLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	56.00	80.22	79.78	79.11
Говорник 2	31.56	44.89	45.33	43.56
Говорник 3	49.78	86.67	86.89	87.33
Говорник 4	44.89	70.44	70.44	68.67
Говорник 5	40.44	62.89	62.00	60.22
Говорник 6	31.78	68.67	70.00	70.44
Говорник 7	47.78	72.67	71.78	72.00
Говорник 8	56.44	76.44	75.78	74.44
Говорник 9	64.00	82.22	83.11	81.33
Говорник 10	35.33	68.00	66.89	65.33
Ср. вред. ± Грешка	**45.80±6.87**	**71.31±7.30**	**71.20±7.33**	**70.24±7.60**

Табела 6.32 *PLPCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	PLPCC кепст. (без CMS-а)	PLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	54.44	63.78	63.78	63.78
Говорник 2	32.22	29.11	31.11	31.33
Говорник 3	43.78	61.78	63.11	63.78
Говорник 4	44.44	41.33	42.89	44.67
Говорник 5	42.00	40.00	42.00	41.78
Говорник 6	33.11	37.11	37.33	37.78
Говорник 7	44.00	52.44	53.33	56.00
Говорник 8	29.33	63.33	64.89	64.44
Говорник 9	48.44	63.56	65.33	64.44
Говорник 10	26.67	50.44	52.44	52.00
Ср. вред. ± Грешка	**39.84±5.59**	**50.29±7.93**	**51.62±7.84**	**52.00±7.70**

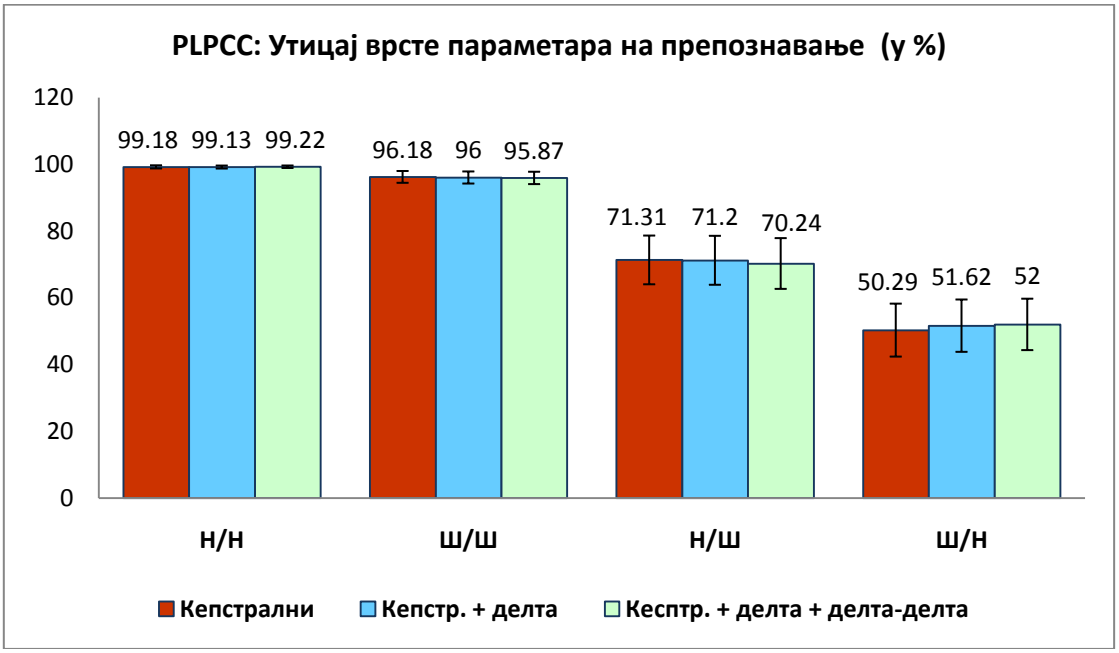
На основу табела 6.29-6.32 дат је дијаграм са резултатима препознавања за сва четири сценарија и векторе састављене од 12 кепстралних коефицијената за два случаја: када није и када јесте коришћена CMS нормализација (слика 6.25).



Слика 6.25 *Резултати препознавања за PLPCC обележје без и са CMS-ом.*

На основу добијеног дијаграма може се закључити да је CMS нормализација допринела побољшању препознавања у сваком од сценарија. При томе је ово повећање од око 1% (код сценарија „нормалан/нормалан“) до 25% (код сценарија „нормалан/шапат“).

Утицај врсте параметара и дужине вектора може се анализирати посматрањем дијаграма на слици 6.26.



Слика 6.26 *Утицај врсте параметара на препознавање за PLPCC обележје.*

Анализом добијених резултата са дијаграма на слици 6.26 може се утврдити да је успешност препознавања врло слична за све сценарије изузев за сценарио „шапат/нормалан“ где је коришћењем делта и делта-делта коефицијената добијено побољшање од око 1,7%.

Резултати препознавања за векторско обележје TEPLPCC и за раније наведене сценарије дати су у табелама 6.33-6.36.

Табела 6.33 *TEPLPCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TEPLPCC кепст. (без CMS-а)	TEPLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.78	99.78	100	99.78
Говорник 2	99.11	99.33	99.11	99.11
Говорник 3	98.00	98.44	98.67	98.67
Говорник 4	99.11	99.56	99.56	99.56
Говорник 5	99.11	99.33	99.33	98.89
Говорник 6	98.22	99.11	98.89	98.89
Говорник 7	96.00	98.22	98.00	98.22
Говорник 8	98.22	99.11	98.89	98.89
Говорник 9	98.89	100	100	100
Говорник 10	92.22	99.11	99.56	99.33
Ср. вред. ± Грешка	**97.87±1.39**	**99.20±0.34**	**99.20±0.39**	**99.13±0.33**

Табела 6.34 *TEPLPCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TEPLPCC кепст. (без CMS-а)	TEPLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	95.56	98.00	97.78	97.33
Говорник 2	96.67	97.56	97.56	97.56
Говорник 3	97.33	99.56	99.33	99.33
Говорник 4	95.11	97.78	96.89	96.67
Говорник 5	92.89	95.33	94.67	94.44
Говорник 6	71.11	94.22	92.67	92.22
Говорник 7	89.11	96.44	96.00	95.33
Говорник 8	87.33	96.00	94.89	94.89
Говорник 9	94.22	97.11	97.56	97.56
Говорник 10	80.44	88.44	88.22	87.11
Ср. вред. ± Грешка	**89.98±5.22**	**96.04±1.90**	**95.56±1.99**	**95.25±2.17**

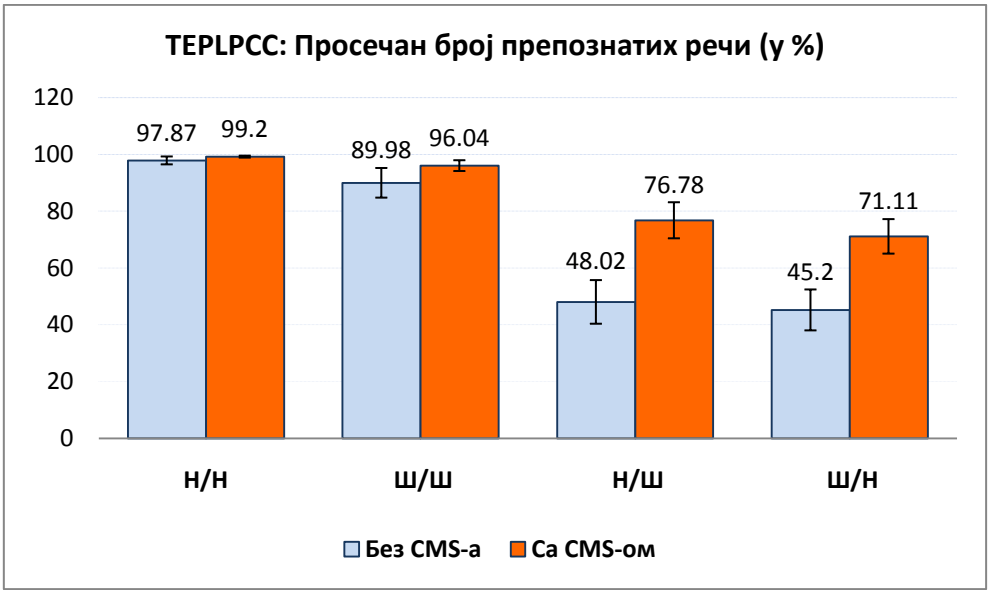
Табела 6.35 *TEPLPCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TEPLPCC кепст. (без CMS-а)	TEPLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	58.00	84.67	84.00	83.78
Говорник 2	46.00	56.00	54.00	52.67
Говорник 3	48.22	91.56	89.78	89.56
Говорник 4	40.67	76.22	76.22	75.11
Говорник 5	44.44	69.33	69.78	68.89
Говорник 6	30.44	76.00	76.44	75.56
Говорник 7	46.89	72.00	72.44	72.44
Говорник 8	60.22	81.56	80.89	80.22
Говорник 9	71.56	87.56	84.22	83.11
Говорник 10	33.78	72.89	71.33	71.56
Ср. вред. ± Грешка	**48.02±7.69**	**76.78±6.35**	**75.91±6.21**	**75.29±6.32**

Табела 6.36 *TEPLPCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TEPLPCC кепст. (без CMS-а)	TEPLPCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	53.11	80.67	80.67	80.00
Говорник 2	50.44	55.56	56.89	56.67
Говорник 3	49.11	90.00	89.33	89.56
Говорник 4	46.67	72.89	74.00	74.00
Говорник 5	56.89	67.11	67.33	68.22
Говорник 6	30.44	60.67	60.00	59.33
Говорник 7	49.33	66.22	68.22	68.22
Говорник 8	26.44	74.89	75.56	75.33
Говорник 9	58.67	73.56	73.33	73.11
Говорник 10	30.89	69.56	68.67	68.67
Ср. вред. ± Грешка	**45.20±7.21**	**71.11±6.08**	**71.40±5.88**	**71.31±5.91**

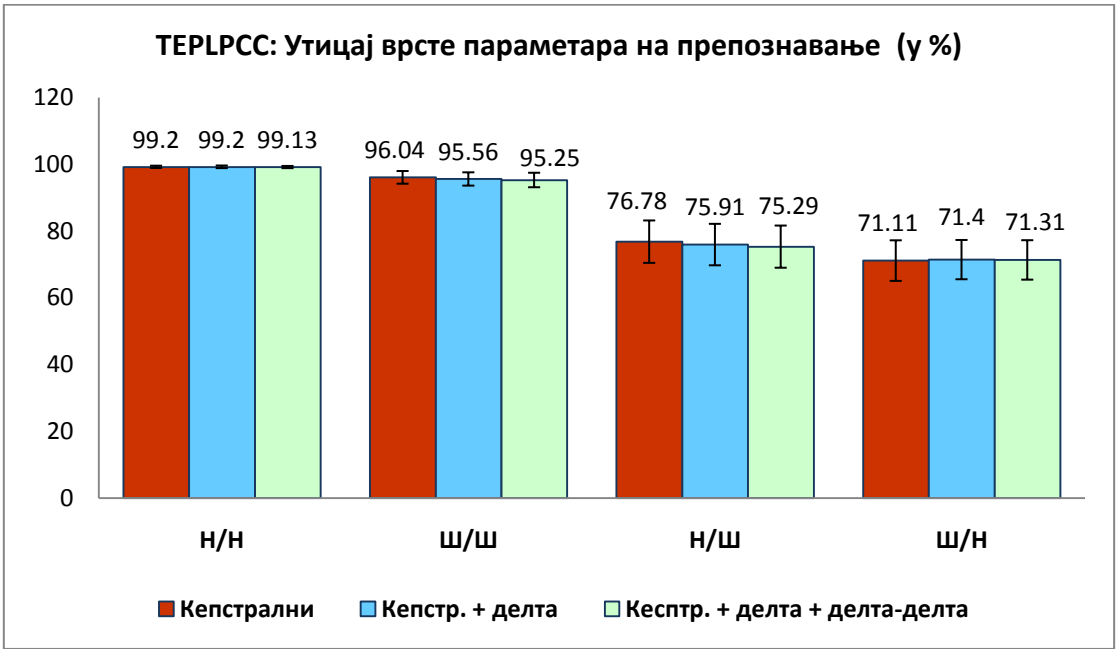
На основу резултата из табела 6.33-6.36 дат је дијаграм успешности препознавања за сва четири сценарија и векторе састављене од кепстралних коефицијената и то за случајеве без и са нормализацијом (слика 6.27).



Слика 6.27 *Резултати препознавања за TEPLPCC обележје без и са CMS-ом.*

Анализом добијеног дијаграма са слике 6.27 се може уочити да је применом CMS-а препознавање повећано за сваки сценарио, а да повећање износи од око 1,3% (за сценарио „нормалан/нормалан“) па све до око 29% (за сценарио „нормалан/шапат“).

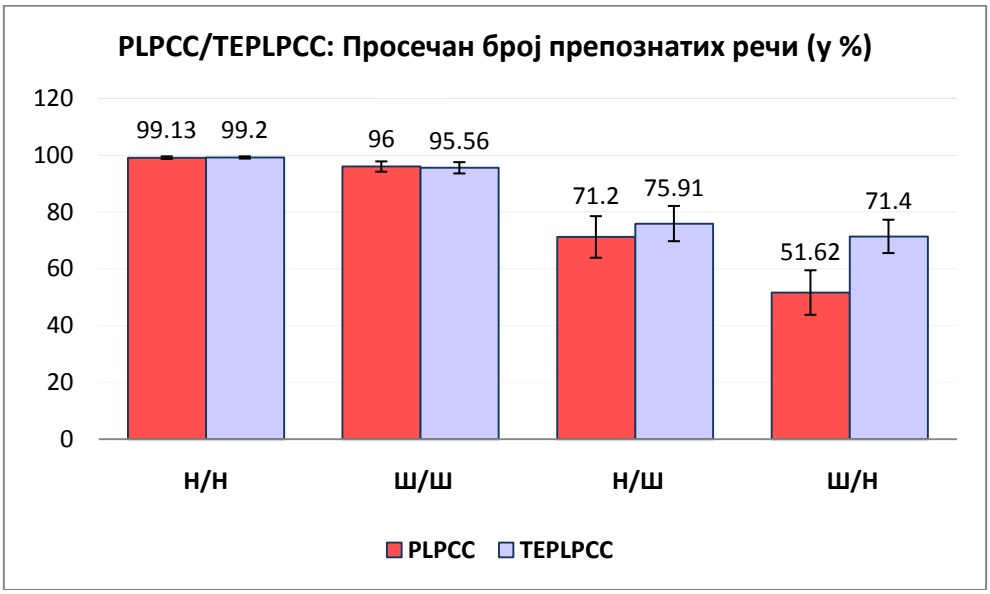
Утицај врсте параметара (кепстрални, делта и делта-делта) и дужине вектора (12, 24 и 36) може се анализирати на бази дијаграма са слике 6.28.



Слика 6.28 *Утицај врсте параметара на препознавање за TEPLPCC обележје.*

Са ове слике се може уочити да су сви резултати скоро идентични и да су једине разлике на првој или другој децимали што се може практично занемарити.

На бази добијених резултата могуће је поредити PLPCC и TEPLPCC векторска обележја. Поређење је извршено коришћењем вектора дужине 24 елемента који садржи кепстралне и делта кепстралне коефицијенте, а резултати поређења су дати на слици 6.29.



Слика 6.29 *Упоредна анализа препознавања за PLPCC и TEPLPCC обележја.*

На основу дијаграма са слике 6.29 може се уочити да за усаглашене сценарије („нормалан/нормалан“ и „шапат/шапат“) нема знатне разлике између ових обележја. Међутим, код неусаглашених сценарија TEPLPCC даје боље резултате од 5% па до 20% (код сценарија „шапат/нормалан“).

6.2.6 РЕЗУЛТАТИ НА БАЗИ RASTACC и TERASTACC ВЕКТОРСКИХ ОБЕЛЕЖЈА

Употребом RASTA нормализације на перцептивну линеарну предикцију током предобраде добијају се RASTACC векторска обележја [Marković et al., 2017 a], а применом нелинерног Teager Energy оператора добијају се TERASTACC векторска обележја (као што је детаљно објашњено у 3.6).

Резултати препознавања добијени за сва четири сценарија и за одговарајуће типове вектора са RASTACC векторским обележјем приказани су у табелама 6.37-6.40.

Табела 6.37 *RASTACC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	PLPCC кепст. (без RASTA)	PLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	99.78	99.33	99.33	99.33
Говорник 2	99.33	100	100	100
Говорник 3	98.44	97.33	97.33	97.56
Говорник 4	99.33	99.33	99.56	99.56
Говорник 5	99.11	99.78	99.78	99.56
Говорник 6	98.00	98.22	98.89	98.67
Говорник 7	96.22	98.89	98.89	98.67
Говорник 8	98.67	98.67	98.89	98.89
Говорник 9	98.89	99.11	99.33	99.33
Говорник 10	92.67	99.11	98.89	98.67
Ср. вред. ± Грешка	**98.04±1.32**	**98.98±0.48**	**99.09±0.46**	**99.02±0.43**

Табела 6.38 *RASTACC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	PLPCC кепст. (без RASTA)	PLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	94.89	98.44	98.44	98.22
Говорник 2	96.89	98.44	98.67	98.67
Говорник 3	97.56	98.89	98.89	98.67
Говорник 4	94.67	98.22	98.67	98.00
Говорник 5	93.33	96.67	96.67	96.67
Говорник 6	77.11	87.78	88.22	87.78
Говорник 7	90.89	95.33	95.78	95.56
Говорник 8	86.48	93.11	93.56	93.33
Говорник 9	93.78	98.44	98.44	98.22
Говорник 10	81.33	90.22	89.78	89.56
Ср. вред. ± Грешка	**90.69±4.26**	**95.55±2.44**	**95.71±2.44**	**95.47±2.46**

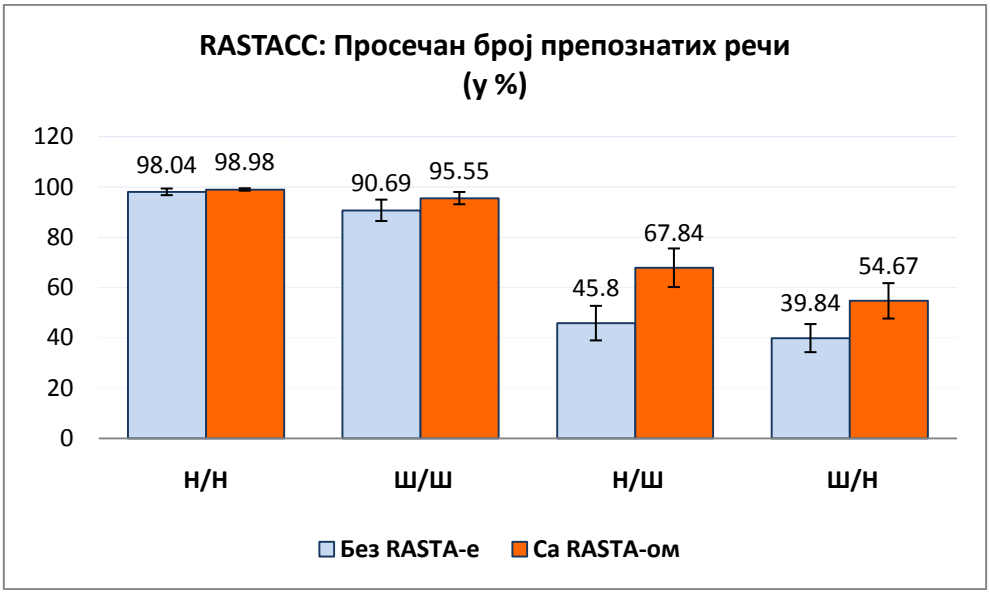
Табела 6.39 *RASTACC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	PLPCC кепст. (без RASTA)	PLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	56.00	76.89	77.56	77.78
Говорник 2	31.56	39.56	38.67	39.11
Говорник 3	49.78	84.44	85.78	86.44
Говорник 4	44.89	68.00	68.67	68.22
Говорник 5	40.44	57.56	55.78	56.22
Говорник 6	31.78	65.33	66.00	67.33
Говорник 7	47.78	70.89	70.44	70.67
Говорник 8	56.44	70.22	71.78	72.67
Говорник 9	64.00	77.78	81.56	81.78
Говорник 10	35.33	67.78	69.56	68.44
Ср. вред. ± Грешка	**45.80±6.87**	**67.84±7.68**	**68.58±8.34**	**68.87±8.32**

Табела 6.40 *RASTACC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	PLPCC кепст. (без RASTA)	PLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	54.44	70.44	69.78	69.56
Говорник 2	32.22	34.44	34.22	36.00
Говорник 3	43.78	64.89	64.67	66.89
Говорник 4	44.44	45.33	46.00	47.56
Говорник 5	42.00	45.56	46.44	47.56
Говорник 6	33.11	47.56	46.67	45.56
Говорник 7	44.00	54.89	54.89	57.33
Говорник 8	29.33	61.11	62.00	64.00
Говорник 9	48.44	66.44	68.00	68.22
Говорник 10	26.67	56.00	56.44	57.11
Ср. вред. ± Грешка	**39.84±5.59**	**54.67±7.05**	**54.91±7.12**	**55.98±7.07**

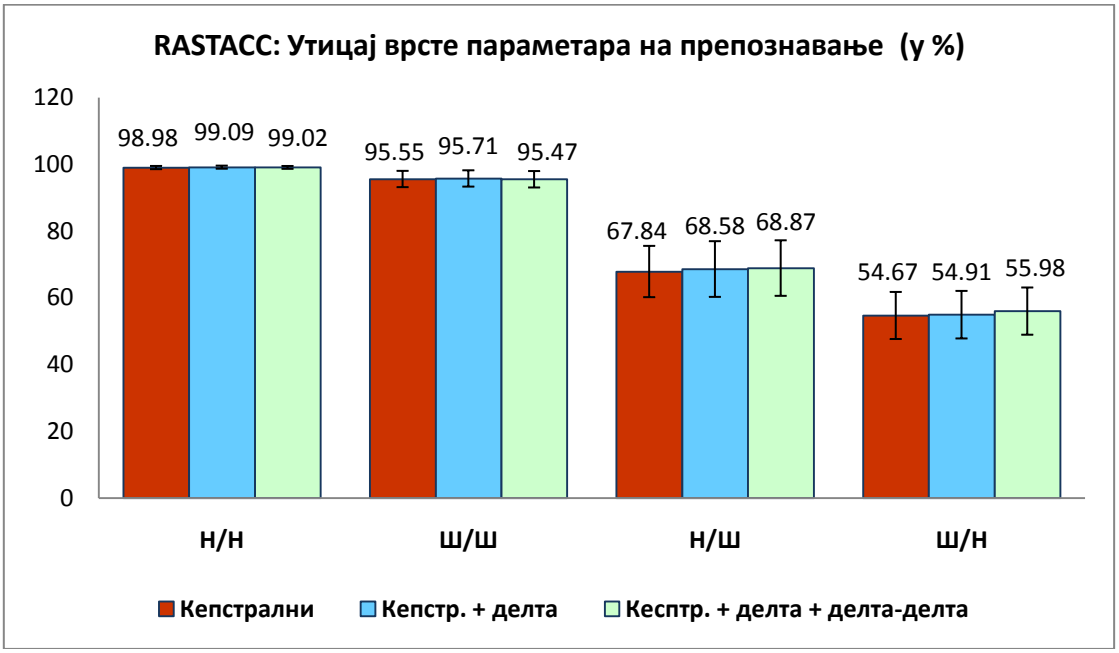
На основу табела 6.37-6.40 дат је дијаграм за резултате препознавања за сва четири сценарија и векторе састављене од кепстралних коефицијената и то за случајеве без и са RASTA нормализацијом. (слика 6.30).



Слика 6.30 *Резултати препознавања за PLPCC обележје без и са RASTA-ом.*

Са дијаграма се може уочити да применом нормализације RASTA у свим сценаријима дошло је до побољшања препознавања, од приближно 1% (за сценарио „нормалан/нормалан“) па до 22% (за сценарио „нормалан/шапат“).

Како утиче врста параметара и дужина вектора на успешност препознавања са RASTACC векторским обележјем приказано је на слици 6.31.



Слика 6.31 *Утицај врсте параметара на препознавање за RASTACC обележје.*

Са претходног дијаграма се може уочити да за усаглашене сценарије препознавање је исто за све три врсте параметара. Код неусаглашених сценарија постоји одређено побољшање увођењем делта и делта-делта кепстралних коефицијената које је нешто више од 1%.

Коришћењем нелинеарног Teager Energy оператора добија се TERASTACC векторско обележје. Резултати препознавања на бази ових обележја приказани су у табелама 6.41-6.44.

Табела 6.41 *TERASTACC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TEPLPCC кепст. (без RASTA)	TEPLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	99.78	99.78	99.56	99.56
Говорник 2	99.33	100	100	100
Говорник 3	98.44	98.44	98.67	98.89
Говорник 4	99.33	99.33	99.33	99.56
Говорник 5	99.11	99.78	99.56	99.56
Говорник 6	98.00	98.44	98.89	98.89
Говорник 7	96.22	99.11	98.89	98.67
Говорник 8	98.67	99.33	99.11	99.11
Говорник 9	98.89	99.11	100	99.78
Говорник 10	92.67	99.33	99.33	99.33
Ср. вред. ± Грешка	**98.04±1.32**	**99.26±0.33**	**99.33±0.28**	**99.34±0.27**

Табела 6.42 *TERASTACC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TEPLPCC кепст. (без RASTA)	TEPLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	94.89	98.44	98.22	98.00
Говорник 2	96.89	99.11	98.67	98.89
Говорник 3	97.56	99.11	99.33	99.33
Говорник 4	94.67	98.67	98.44	98.00
Говорник 5	93.33	95.78	95.78	95.33
Говорник 6	77.11	94.22	94.22	92.89
Говорник 7	90.89	95.78	96.00	96.00
Говорник 8	86.48	95.78	95.33	95.11
Говорник 9	93.78	98.44	97.78	97.78
Говорник 10	81.33	91.78	91.33	91.33
Ср. вред. ± Грешка	**90.69±4.26**	**96.71±1.52**	**96.51±1.53**	**96.27±1.63**

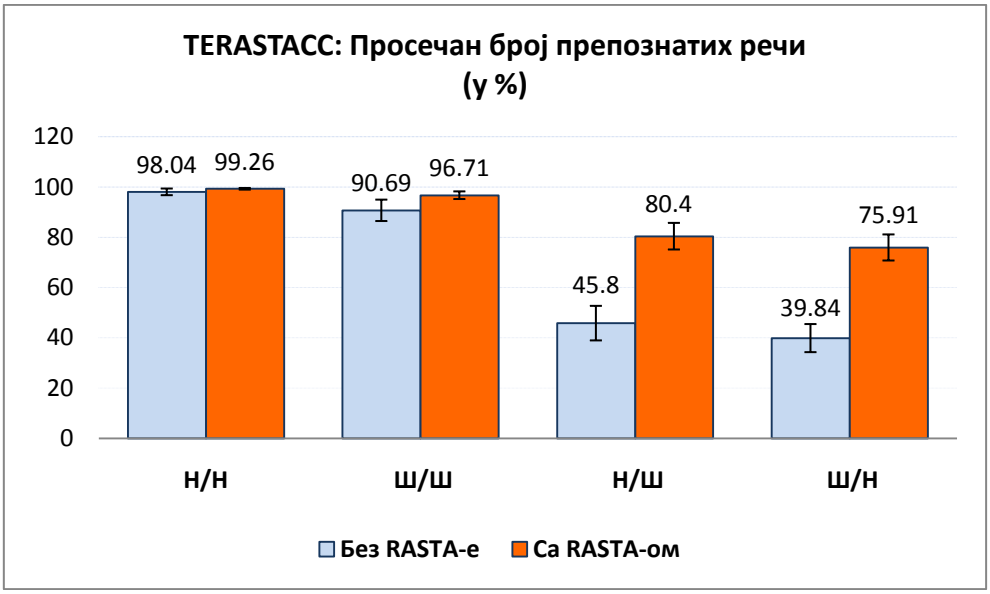
Табела 6.43 *TERASTACC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TEPLPCC кепст. (без RASTA)	TEPLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	56.00	89.56	90.00	89.78
Говорник 2	31.56	64.00	56.22	55.33
Говорник 3	49.78	92.89	93.33	92.67
Говорник 4	44.89	83.56	82.67	81.78
Говорник 5	40.44	73.56	71.56	70.44
Говорник 6	31.78	78.44	79.56	79.56
Говорник 7	47.78	76.44	78.22	77.33
Говорник 8	56.44	81.78	82.22	82.67
Говорник 9	64.00	87.56	88.44	88.00
Говорник 10	35.33	76.22	78.00	76.00
Ср. вред. ± Грешка	**45.80±6.87**	**80.40±5.29**	**80.02±6.55**	**79.36±6.69**

Табела 6.44 *TERASTACC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TEPLPCC кепст. (без RASTA)	TEPLPCC кепст. (са RASTA-ом)	Кепст. + Δ (са RASTA-ом)	Кепст. + Δ + ΔΔ (са RASTA-ом)
Говорник 1	54.44	85.78	85.78	83.56
Говорник 2	32.22	64.89	63.11	63.78
Говорник 3	43.78	91.33	91.33	90.89
Говорник 4	44.44	80.22	79.33	78.67
Говорник 5	42.00	72.44	72.89	71.56
Говорник 6	33.11	65.33	65.56	65.33
Говорник 7	44.00	72.22	72.89	73.78
Говорник 8	29.33	74.89	76.22	76.22
Говорник 9	48.44	78.89	80.22	79.78
Говорник 10	26.67	73.11	74.00	72.44
Ср. вред. ± Грешка	**39.84±5.59**	**75.91±5.19**	**76.13±5.30**	**75.60±5.06**

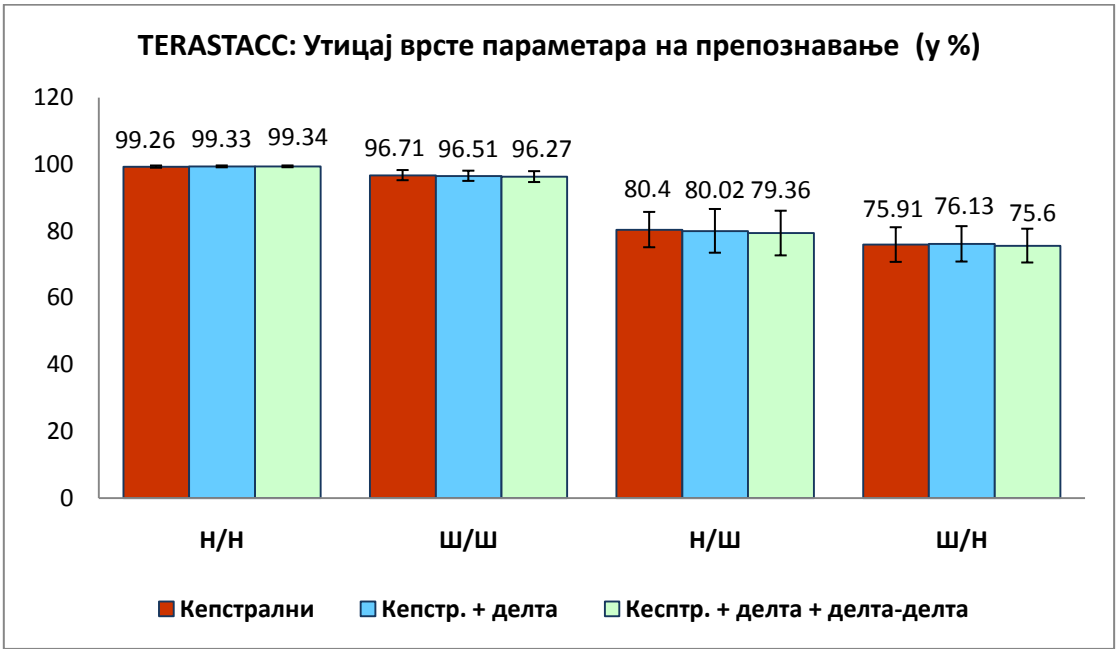
Коришћењем табела 6.41-6.44 дефинисан је дијаграм који даје резултате препознавања за сва четири сценарија и векторе састављене од кепстралних коефицијената и то у случају када није и када јесте коришћена RASTA нормализација (слика 6.32).



Слика 6.32 *Резултати препознавања за TEPLPCC обележје без и са RASTA-ом.*

На бази приказаних резултата може се закључити да коришћење RASTA нормализације утиче на повећање успешности препознавања од 1,2% (за сценарио „нормалан/нормалан“) па до 36% (за сценарио „шапат/нормалан“).

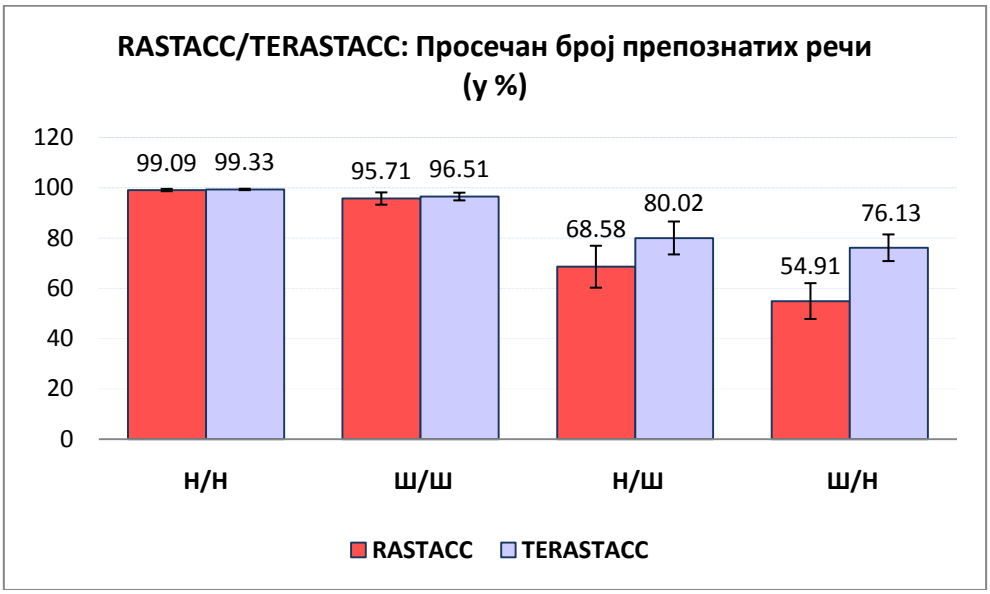
Утицај врсте параметара (кепстрални, делта и делта-делта) на успешност препознавања дат је на слици 6.33.



Слика 6.33 *Утицај врсте параметара на препознавање за TERASTACC обележје.*

Са дијаграма на слици 6.33 уочава се да нема знатног побољшања увођењем делта и/или делта-делта параметара и да је то побољшање испод 0,5%.

На основу добијених резултата могуће је поредити RASTACC и TERASTACC векторска обележја. Слика 6.34 даје односе успешности препознавања при чему се користио вектор састављен од 24 коефицијента (12 кепстралних и 12 делта кепстралних).



Слика 6.34 *Упоредна анализа препознавања за RASTACC и TERASTACC обележја.*

Базирано на претходном дијаграму може се уочити да примена ТЕ оператора утиче на повећање препознавања и то од 1,2% (код сценарија „нормалан/нормалан“) па до 21% (за сценарио „шапат/нормалан“).

У циљу евалуације заједничког утицаја RASTA и CMS нормализације извршена су тестирања базе говорних узорака са овом комбинацијом. На говорни сигнал се најпре применила RASTA нормализација (на сличан начин као на слици 3.19, коришћењем PLP), а затим и CMS нормализација (пре блока за делта и делта-делта коефицијенте, као на слици 3.17). Добијени резултати нису довели до успешнијег препознавања, већ донекле и до деградације, па због тога овде нису експлицитно наведени.

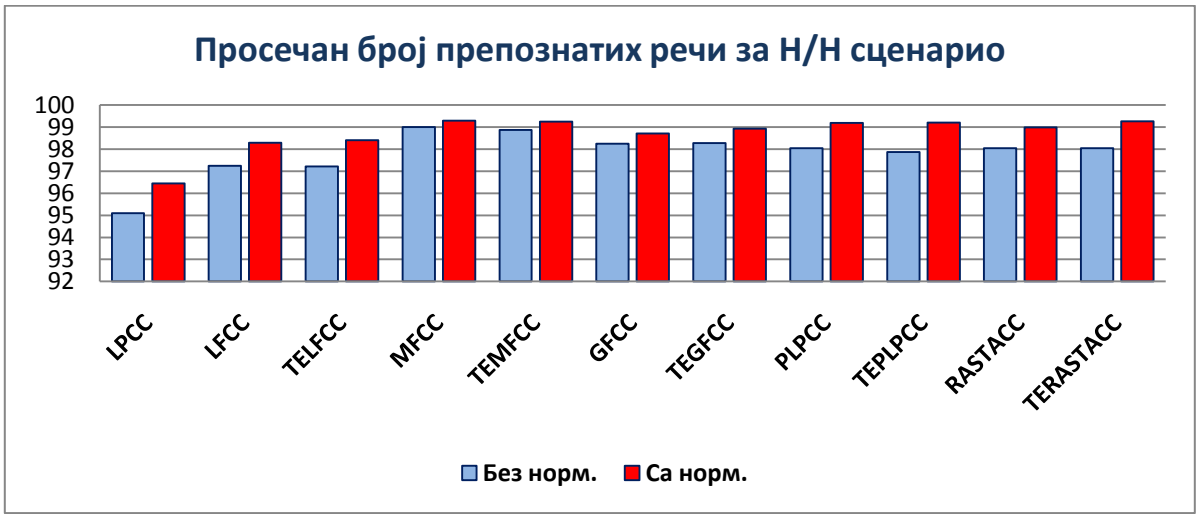
6.3 УПОРЕДНА АНАЛИЗА И ДИСКУСИЈА РЕЗУЛТАТА

На бази горе наведених резултата могуће је извршити њихову упоредну анализу и одредити која векторска обележја дају најбоље резултате за сваки од сценарија. Такође, може се анализирати потенцијално постојање комбинованог решења тј. формирање хибридне конфигурације која би омогућила оптимално решење за одговарајуће сценарије.

6.3.1 УПОРЕДНА АНАЛИЗА ВЕКТОРСКИХ ОБЕЛЕЖЈА

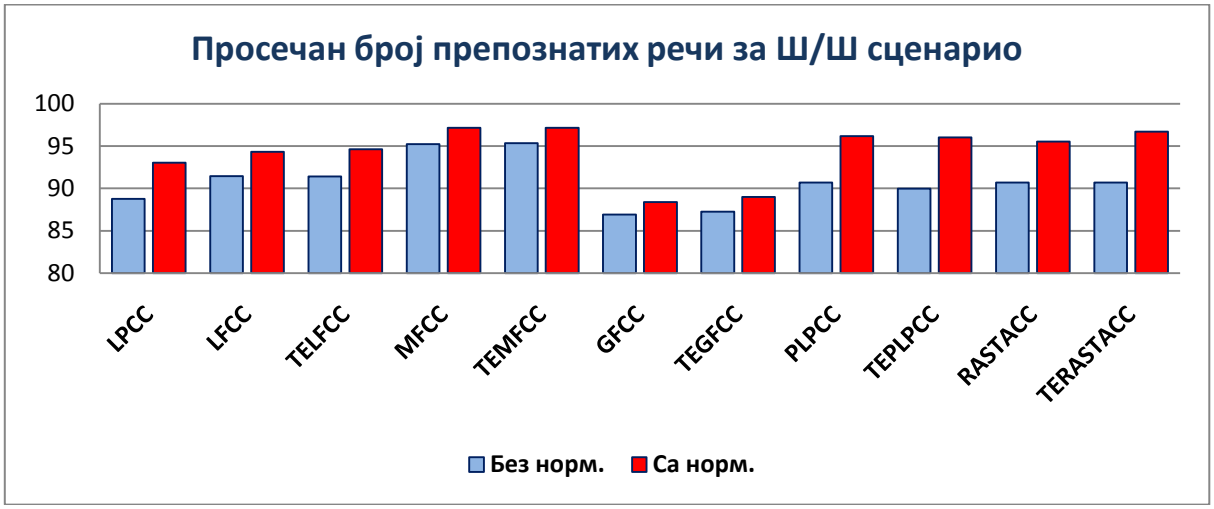
Са аспекта примењене нормализације могу се поредити одговарајући резултати. Разматрана су векторска обележја: LPCC, LFCC, TELFCC, MFCC, TEMFCC, GFCC, TEGFCC, PLPCC, TEPLPCC, RASTACC и TERASTACC. Вектори су састављени од 12 кепстралних коефицијената при

чему први репрезент је без, а други са нормализацијом. Слике 6.35-6.38 дају дијаграме успешности препознавања за сценарије: „нормалан/нормалан“ (Н/Н), „шапат/шапат“ (Ш/Ш), „нормалан/шапат“ (Н/Ш) и „шапат/нормалан“ (Ш/Н) респективно.



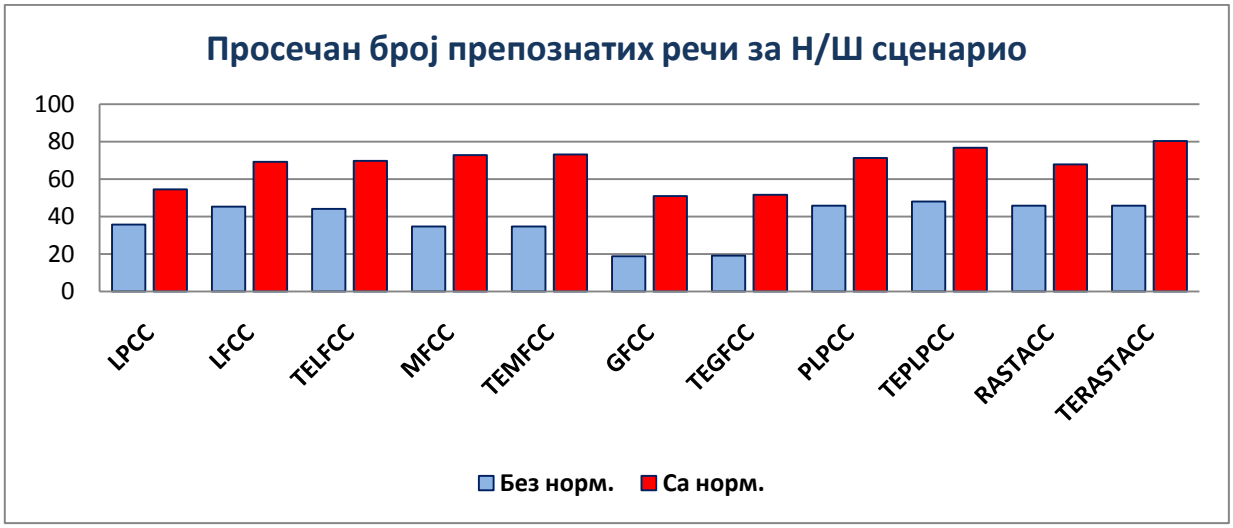
Слика 6.35 *Просечан број препознатих речи за Н/Н сценарио без и са нормализацијом.*

На основу дијаграма са слике 6.35 може се закључити да се најбољи резултати препознавања постижу са MFCC, TERASTACC, TEMFCC, TEPLPCC и PLPCC векторским обележјима и да је успешност препознавања око 99,20%. У свим разматраним случајевима коришћење нормализације знатно доприноси повећању препознавања.



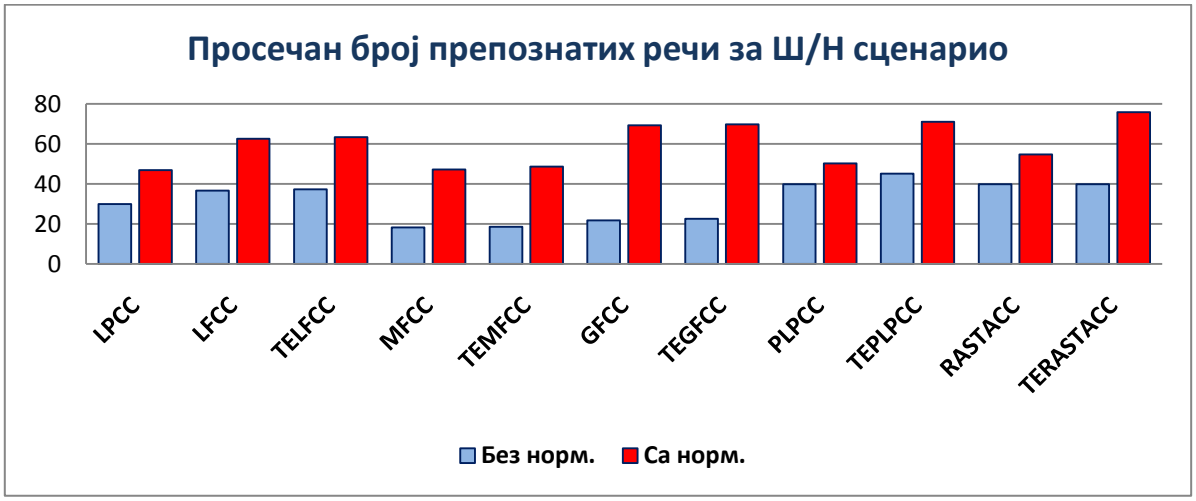
Слика 6.36 *Просечан број препознатих речи за Ш/Ш сценарио без и са нормализацијом.*

На бази дијаграма са слике 6.36 може се уочити да се најбољи резултати за сценарио Ш/Ш добијају са обележјима MFCC, TEMFCC, TERASTACC, PLPCC и TEPLPCC и њихове вредности су око 96-97%. Примена нормализације је знатно поправила успешност препознавања. Дакле, ради се о истој врсти „успешних“ обележја као и за Н/Н сценарио.



Слика 6.37 *Просечан број препознатих речи за Н/Ш сценарио без и са нормализацијом.*

За сценарио Н/Ш и даље иста векторска обележја, као што је напред описано, дају најбоље резултате који су од 70-80%. Међутим, између њих се посебно истиче TERASTACC обележје које омогућава препознавање од 80,4%. Тренд без и са нормализацијом је и даље исти као код претходних сценарија.

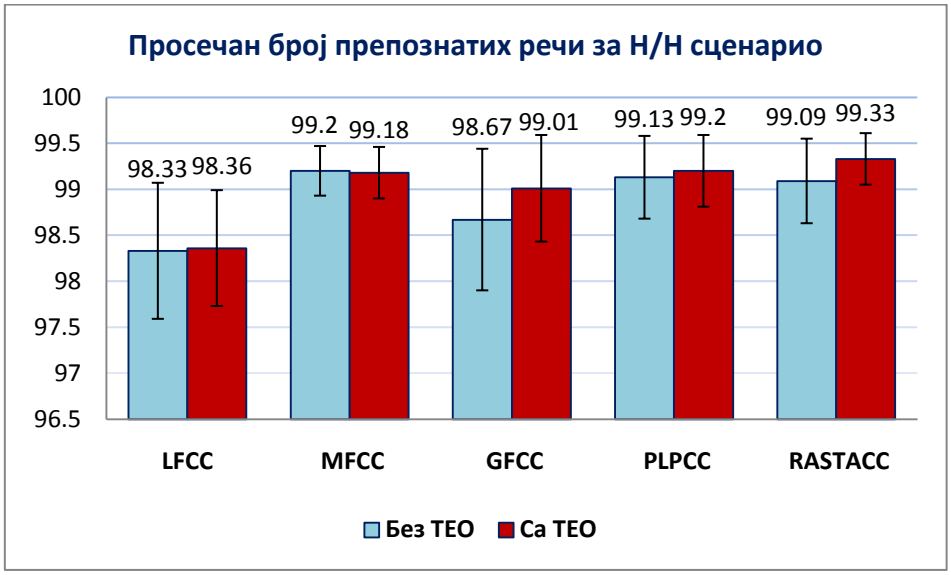


Слика 6.38 *Просечан број препознатих речи за Ш/Н сценарио без и са нормализацијом.*

Као што се са дијаграма на слици 6.38 може уочити најбољи су резултати препознавања за Ш/Н сценарио када се користе TERASTACC, TEPLPCC, TEGFCC и GFCC векторска обележја. Ради се о успешности препознавања од 70-76%. Посебно се истиче TERASTACC векторско обележје које омогућава препознавање од скоро 76%. Примена нормализације знатно доприноси побољшању препознавања.

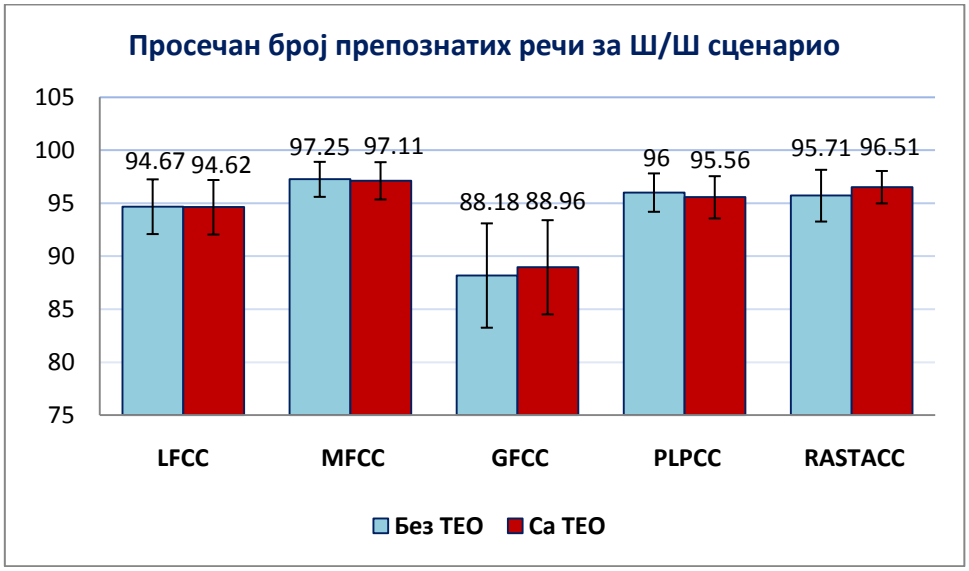
Утицај Teager Energy оператора може се размотрити посматрањем експерименталних резултата за она векторска обележја на која се може применити овај оператор, а то су LFCC, MFCC, GFCC, PLPCC и RASTACC. На сликама 6.39-6.41 дати су дијаграми успешности препознавања за тип вектора који се састоји од 24 коефицијента (12 кепстралних и 12 делта кепстралних) примењеног на сва четири сценарија. Разлог избора овог типа вектора је што

он у просеку даје најбоље резултате у односу на остале типове (само кепстрални или кепстрални са делта и делта-делта кепстралним коефицијентима).



Слика 6.39 *Просечан број препознатих речи за Н/Н сценарио без и са ТЕ оператором.*

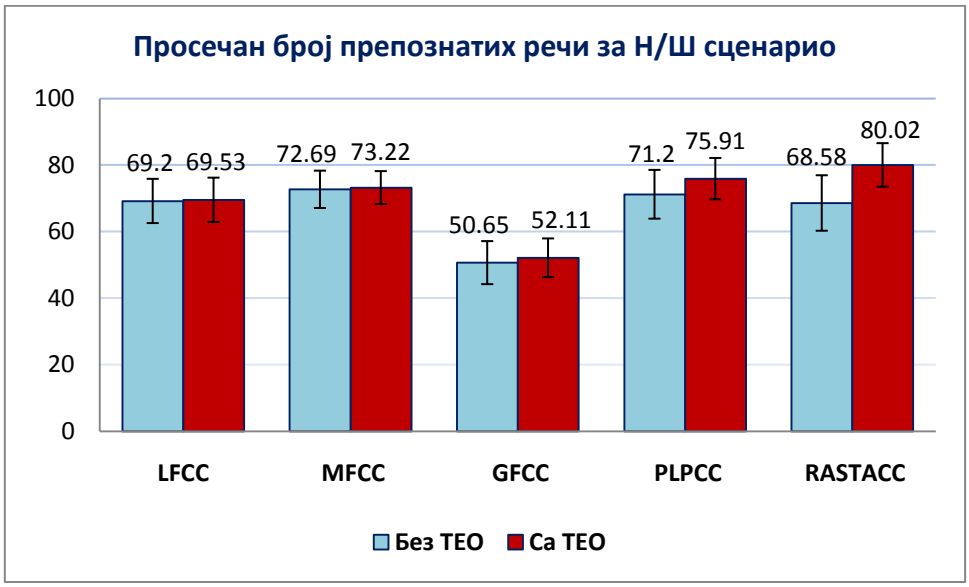
На основу претходног дијаграма може се уочити да у скоро свим случајевима коришћење ТЕ оператора доводи до повећања препознавања. Изузетак је MFCC векторско обележје, али је и код њега разлика у препознавању скоро занемарљива (на другој децимали) када се не користи или када се користи овај оператор. Најбољи резултат се постигао са обележјем TERASTACC и износи 99,33%.



Слика 6.40 *Просечан број препознатих речи за Ш/Ш сценарио без и са ТЕ оператором.*

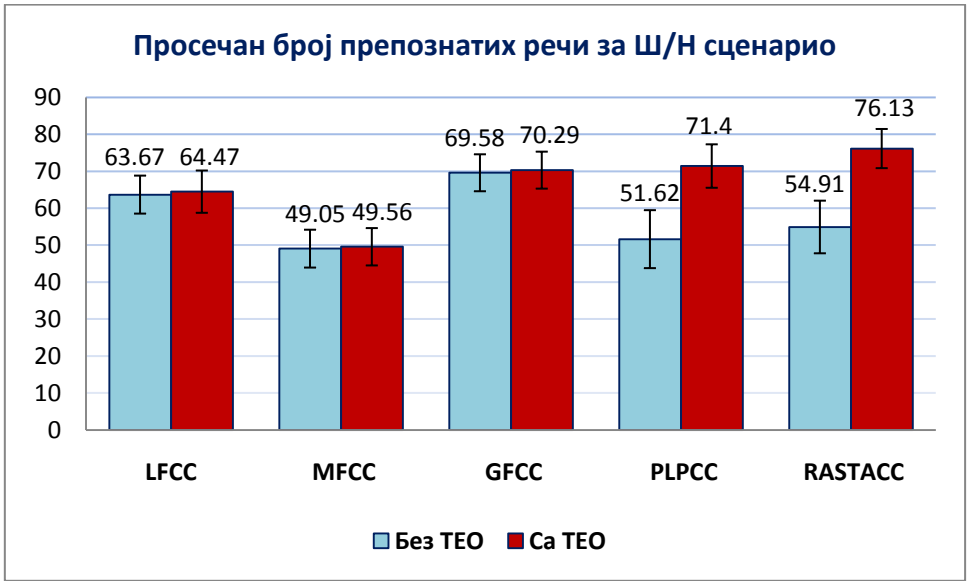
На бази добијеног дијаграма са слике 6.40 за Ш/Ш сценарио може се закључити да у неким случајева нема знатног побољшања применом ТЕ оператора и највеће побољшање је код GFCC и RASTACC обележја, али и код њих овај допринос је мањи од 1%. Углавном су

резултати слични и разлике на првој или другој децимали. Најбољи резултат (са TEO) је добијен за TEMFCC обележје и износи 97,11%.



Слика 6.41 *Просечан број препознатих речи за Н/Ш сценарио без и са ТЕ оператором.*

За Н/Ш врсту неусаглашеног сценарија се може уочити да је у свим случајевима примена ТЕ оператора допринела повећању препознавања. Посебно је потребно истаћи да је векторско обележје типа TERASTACC показало висок ниво препознавања у односу на све остале и то од 80,02%.



Слика 6.42 *Просечан број препознатих речи за Ш/Н сценарио без и са ТЕ оператором.*

На основу дијаграма који је приказан на слици 6.42 може се уочити да је у свим случајевима када је примењен ТЕ оператор за Ш/Н сценарио дошло до побољшања препознавања. Посебно висок ниво препознавања је за векторска обележја типа TERASTACC, TEPLPCC и TEGFCC. Велики скок препознавања (и преко 20%) се може уочити код векторских

обележја PLPCC и RASTACC када се на њих примени TE оператор. Најбољи резултати су поново са TERASTACC обележјем и износе 76,13%.

6.3.2 ДИСКУСИЈА РЕЗУЛТАТА

Анализом појединачних сценарија може се утврдити следеће:

1) За **сценарио „нормалан/нормалан“** (Н/Н) најбољи резултат препознавања остварен је помоћу TERASTACC векторског обележја. Његова вредност износи око 99,3%. После овог векторског обележја по успешности препознавања следе: TEPLPCC, MFCC, TEMFCC, PLPCC, RASTACC, TEGFCC, GFCC, TELFCC, LFCC и LPCC. Овај сценарио је дао најбоље резултате препознавања што је и било очекивано с обзиром на услове и начин снимања и препознавања.
2) За **сценарио „шапат/шапат“** (Ш/Ш) најбољи резултат од око 97,3% је добијен помоћу MFCC обележја. Затим са становишта резултата следе обележја: TEMFCC, TERASTACC, PLPCC, TEPLPCC, RASTACC, LFCC, TELFCC, LPCC, GFCC и TEGFCC. И овај сценарио се истакао са високим успехом препознавања. И код њега су се, као и код Н/Н, издвојила одређена обележја (MFCC, TEMFCC, TERASTACC, PLPCC и TEPLPCC) са приближно истим успехом препознавања.
3) За **сценарио „нормалан/шапат“** (Н/Ш) најбољи успех препознавања омогућило је TERASTACC векторско обележје у износу од преко 80% са вектором састављеним од кепстралних коефицијената. По успешности следе обележја: TEPLPCC, TEMFCC, MFCC, PLPCC, TELFCC, LFCC, RASTACC, LPCC, TEGFCC и GFCC. Овде се пре свих истичу векторска обележја са TE оператором (TERASTACC, TEPLPCC и TEMFCC), а такође и обележја која су била успешна код усаглашених сценарија (MFCC и PLPCC).
4) За **сценарио „шапат/нормалан“** (Ш/Н) најбољи резултат се постигао такође са TERASTACC векторским обележјем. Његова вредност препознавања је преко 76%. Затим следе векторска обележја: TEPLPCC, TEGFCC, GFCC, TELFCC, LFCC, RASTACC, PLPCC, TEMFCC, MFCC и LPCC. И у овом случају векторска обележја са TE оператором (TERASTACC, TEPLPCC и TEGFCC) предњаче над осталима, али се може уочити да су TEGFCC и TEMFCC заменили места са аспекта овог неусаглашеног сценарија.

На бази горе изнетих резултата уочава се да је обележје TERASTACC доминантно и за усаглашене („нормалан/нормалан“ и „шапат/шапат“) и за неусаглашене („нормалан/шапат“ и „шапат/нормалан“) сценарије. Такође, и остала обележја која користе TE оператор дају у просеку боље резултате. Од класичних обележја (без TE оператора) посебно се истиче MFCC за усаглашене сценарије, а GFCC за сценарио „шапат/нормалан“.

Анализом утицаја нормализације (слике 6.33-6.36) може се закључити да је у свим случајевима дошло до побољшања успешности препознавања. Максимални случајеви побољшања, гледано по сценаријима су:

1) За **сценарио „нормалан/нормалан“** скок од око 1,3% (за LPCC и TEPLPCC векторска обележја),

2) За **сценарио „шапат/шапат“** скок од око 6% (за TEPLPCC и TERASTACC векторска обележја),

3) За **сценарио „нормалан/шапат“** скок од 35-38% (за TERASTACC, MFCC и TEMFCC векторска обележја) и

4) За **сценарио „шапат/нормалан“** скок од 36-47% (за TERASTACC, GFCC и TEGFCC векторска обележја).

Може се закључити да се применом нормализације добија знатно побољшање успешности препознавања за неусаглашене сценарије, док код усаглашених већи је добитак за „шапат/шапат“ него за „нормалан/нормалан“ сценарио, што је и очекивано.

Када је у питању примена Teager Energy оператора онда се могу извести следећи закључци:

1) за **сценарио „нормалан/нормалан“** и скоро свим случајевима даје побољшање које није велико (испод 0,5%). Међутим, с обзиром да је у овом сценарију успешност препознавања око 99% и то је значајан успех. Једино код TEMFCC нема знатног побољшања (у односу на MFCC), али је и овде разлика успешности на другој децимали.

2) за **сценарио „шапат/шапат“** у највећем броју случајева овај оператор даје побољшање, али се поново показало да је за MFCC резултат мало бољи (на првој децимали) него за TEMFCC. Што се тиче LFCC и TELFCC разлика између њих је на другој децимали. Највеће побољшање се добија са векторским обележјем TERASTACC и износи око 1% (у односу на RASTACC),

3) за **сценарио „нормалан/шапат“** примена TE оператора у свим случајевима доноси побољшање и то у распону од 0,3% (код TELFCC) па до око 12% (код TERASTACC). За овај сценарио TE оператор има значајан допринос.

4) за **сценарио „шапат/нормалан“** примена TE оператора такође знатно побољшава успешност препознавања. У свим случајевима се јавља повећање и оно износи од 0,3% (код TELFCC) па чак до 21% (код TERASTACC). Коришћење овог оператора за сценарио „шапат/нормалан“ овде долази до пуног изражаја.

Може се закључити да постоји оправдана потреба за коришћењем TE оператора у препознавању говора, а посебно за неусаглашене сценарије. Пошто TE оператор успешно описује нагле промене турбулентног кретања у вокалном тракту (које су посебно изражене током продукције шапата), то се испоставља да је његова употреба веома корисна и са апекта повећања успешности препознавања мултимодалног говора врло пожељна.

На бази статистичке анализе која је подразумевала постављање нивоа поузданости на 95% одређене су границе грешке (Margins of Errors) које су заједно са средњом вредношћу дале интервале поузданости. Границе грешке су представљене у табелама и на дијаграмима. Анализирањем интервала поузданости за векторе дужине 24 параметра може се закључити да су грешке најмање за сценарио „нормалан/нормалан“, а затим следе сценарији „шапат/шапат“, „шапат/нормалан“ и „нормалан/шапат“ респективно. За сценарио „нормалан/нормалан“ ови интервали се крећу од ±0,27 (за MFCC векторско обележје) до

±1,1 (за LPCC векторско обележје). За сценарио „шапат/шапат“ интервали се крећу од ±1,53 (за TERASTACC векторско обележје) до ±4,92 (за GFCC векторско обележје). Код неусаглашених сценарија интервали су знатно већи. За „нормалан/шапат“ они су од ±4,94 (за TEMFCC векторско обележје) до ±8,34 (за RASTACC векторско обележје). Код сценарија „шапат/нормалан“ интервали поузданости су од ±4,53 (за LPCC векторско обележје) до ±7,84 (за PLPCC векторско обележје). Као што је и очекивано, онде где је успешност препознавања већа ту је грешка мања, а интервали краћи.

Анализом дужине вектора која може бити 12 (кепстралних коефицијената), 24 (12 кепстралних плус 12 делта кепстралних) и 36 (12 кепстралних, 12 делта и 12 делта-делта коефицијената)параметара закључује се да у највећем броју случајева и у свим сценаријима вектор од 24 параметра даје задовољавајуће резултате. Вектор од 36 параметара у више наврата није дао боље резултате од оног са 24 параметра. Они су лошији од дела процента па до пар процената (посебно код неусаглашених сценарија). Разлоге треба тражити у томе што су говорни узорци снимани у условима минималног позадинског шума (а познато је да делта-делта коефицијенти дају добре резултате при утицају шумне средине), а такође и да грешка заокруживања приликом рачунања дугачких вектора добија адитивни карактер. Стога је препорука коришћење вектора од 24 параметра.

Парцијални резултати добијени коришћењем HMM метода (Прилог А) су у сагласности са онима који су добијени DTW методом и потврђују оправданост коришћења нормализације и TE оператора.

На бази спроведених експеримената и добијених резултата несумњив је закључак да се применом нормализације добија знатно побољшање препознавања за све сценарије и векторска обележја, а применом нелинеарног TE оператора добијају се такође одређена побољшања која су посебно изражена код неусаглашених сценарија. Међу свим обележјима посебно се истакло TERASTACC које је дало значајне резултате за све, а посебно за „нормалан/шапат“ и „шапат/нормалан“ сценарије.

7. ЗАКЉУЧАК

Овим радом је показано да се препознавање мултимодалног говора значајно може поправити коришћењем нормализације и применом нелинеарног Teager Energy оператора. Примена нормализације утиче на усклађивање спектралних разлика између модалитета говора, а такође и на смањење варијација које настају током изговора у истом моду, док коришћење Teager Energy оператора добро описује турбулентно кретања ваздуха унутар вокалног тракта које је једна од важних појава при шапату. Разматран је систем који је зависан од говорника, а за препознавање коришћена метода динамичког усклађивања узорака (DTW) позната по брзом и ефикасном раду.

7.1. ПРЕГЛЕД РЕЗУЛТАТА

У овим истраживањима разматрано је једанаест векторских обележја. На једно од њих (LPCC - Linear Prediction Cepstral Coefficients), због његове природе, није примењен Teager Energy оператор, док на следећа векторска обележја је примењен овај оператор: LFCC (Linear Frequency Cepstral Coefficients), MFCC (Mel Frequency Cepstral Coefficients), GFCC (Gammatone Filterbank Cepstral Coefficients), PLPCC (Perceptual Linear Prediction Cepstral Coefficients) и RASTACC (RelAtive SpecTrA Cepstral Coefficients). Тако су добијена и одговарајућа векторска обележја типа: TELFCC, TEMFCC, TEGFCC, TEPLPCC и TERASTACC респективно. Извршен је детаљни опис добијања свих векторских обележја и дати дијаграми тока у којима је идентификован сваки од блокова обраде. Тако су истакнути процеси кроз које сигнал пролази као што су преемфазис, формирање рамова и преклапање, прозоровање, примена брзе Фуријеове трансформације, отежавање на бази различитих скала, отежавање на бази гласности, нормализација, добијање делта и делта-делта кепстралних коефицијената и слично.

Детаљно је описана и база говорних узорака Whi-Spe која је коришћена. Дати су елементи базе, начин снимања, листа говорника и карактеристике опреме која се при снимању користила. У експериментима је коришћено свих 10.000 узорака говора (од којих су пола репрезенти шапата, а пола нормалног говора). Такође је била и подједнака заступљеност полова (по пет женских и мушких говорника).

Акустичка обележја су репрезентована помоћу одговарајућих вектора кепстралних коефицијената. При томе су у истраживању коришћена четири типа вектора и то:

- вектори састављени од 12 кепстралних коефицијената без нормализације,
- вектори састављени од 12 кепстралних коефицијената са нормализацијом,

- вектори састављени од 24 коефицијента (12 кепстралних и 12 делта) са нормализацијом и
- вектори састављени од 36 коефицијената (12 кепстралних, 12 делта и 12 делта-делта) са нормализацијом.

За свако од поменутих векторских обележја развијен је софтверски модул коришћењем софтверског пакета MATLAB. Ови софтверски модули су на бази варијација улазних аргумената генерисали различите типове вектора. Тако је број различитих софтверских модула био једанаест, а укупан број варијација 44, па је број различитих вектора који су генерисани током истраживања био 44.000.

Анализирана векторска обележја су детаљно приказана и то:

- табеларно - по сваком од четири сценарија (Н/Н, Ш/Ш, Н/Ш и Ш/Н) и за сваког говорника,
- дијаграмима - где је вршено поређење за сва четири сценарија са и без нормализације,
- дијаграмима - где је вршено поређење за различите типове вектора (кепстрални, делта и делта-делта) на којима је примењена нормализација и
- дијаграмима - где је вршено поређење одговарјућих вектора са и без примењеног ТЕ оператора.

Сумарни резултати анализирани по сценаријима дати су кроз следећи преглед:

1) за „**нормалан/нормалан“** сценарио најбоља успешност препознавања је са TERASTACC векторским обележјем и износи 99,3%, а затим следе TEPLPCC, MFCC итд,
2) за „**шапат/шапат“** сценарио најбољи резултат је добијен са MFCC векторским обележјем и износи 97,3%, а затим следе TEMFCC, TERASTACC итд,
3) за „**нормалан/шапат“** сценарио најефикасније је TERASTACC векторско обележје са препознавањем од око 80%, а затим су по успешности TEPLPCC, TEMFCC итд и
4) за „**шапат/нормалан“** сценарио по резултатима се поново истиче TERASTACC векторско обележје са успешношћу од 76%, а потом следе TEPLPCC, TEGFCC итд.

Утицај нормализације такође је детаљно анализиран и приказан кроз одговарајућа побољшања препознавања:

1) за сценарио **„нормалан/нормалан“** максимални скок је од око 1,3% (за LPCC и TEPLPCC векторска обележја),
2) за сценарио „**шапат/шапат“** максимални скок је од око 6% (за TEPLPCC и TERASTACC векторска обележја),
3) за сценарио **„нормалан/шапат“** максимални скок је од 35-38% (за TERASTACC, MFCC и TEMFCC векторска обележја) и

4) за сценарио **„шапат/нормалан“** максимални скок је од 36-47% (за TERASTACC, GFCC и TEGFCC векторска обележја).

Утицај Teager Energy оператора такође је приказан у овом раду, а сумарно се уочава:

1) За сценарио **„нормалан/нормалан“** у скоро свим случајевима TE даје побољшање које није велико (испод 0,5%). Међутим, с обзиром да је у овом сценарију успешност препознавања око 99% и то је солидан успех. Једино код TEMFCC нема знатног побољшања (у односу на MFCC), али и овде је разлика на другој децимали;
2) За сценарио **„шапат/шапат“** у највећем броју случајева овај оператор даје побољшање, али се поново показало да је за MFCC резултат мало бољи (на првој децимали) него за TEMFCC. Што се тиче LFCC и TELFCC разлика између њих је на другој децимали. Највеће побољшање се добија са векторским обележјем TERASTACC и износи око 1% (у односу на RASTACC);
3) За сценарио **„нормалан/шапат“** примена TE оператора у свим случајевима доноси побољшање и то у распону од 0,3% (код TELFCC) па до око 12% (код TERASTACC). За овај сценарио TE оператор има значајан допринос;
4) За сценарио **„шапат/нормалан“** примена TE оператора такође знатно побољшава успешност препознавања. У свим случајевима оно је евидентно и износи од 0,3% (код TELFCC) па чак до 21% (код TERASTACC). Коришћење овог оператора за сценарио „шапат/нормалан“ овде је дошло до пуног изражаја.

Анализирана је и дужина вектора и њен утицај на препознавање. У обзир су узети вектори састављени од 12, 24 и 36 параметара. Показало се да вектор састављен од 24 параметра (12 кепстралних и 12 делта кепстралних коефицијената) даје задовољавајуће резултате. Иако је вектор састављен од 36 параметара по својој природи робуснији он није у већини случајева био бољи од оног састављеног од 24 параметра. Разлози за то могу бити што су ови узорци снимани у условима потиснутог амбијенталног шума (па делта-делта доприноси нису дошли до изражаја) као и могућност акумулираног доприноса грешке заокруживања приликом рачунања дугачких вектора јер су делта-делта коефицијенти врло мали. Стога је за оваква истраживања препорука да се користе вектори од 24 параметра.

7.2. ДОПРИНОС ДИСЕРТАЦИЈЕ

Развијање система за препознавање говора је стални императив који за циљ има да се омогући успешно препознавање у околностима када говор одступа од нормалног. Због тога је посебно актуелно истраживање мултимодалних облика говора, међу којима значајно место заузима шапат. Проналажење оптималних векторских обележја, којима се описују акустичке особине сигнала, могу утицати на побољшање резултата препознавања. У складу са тим, доприноси докторске дисертације огледају се у следећем:

- Креиран је одређени скуп векторских обележја (њих једанаест) и на бази њих извршена анализа успешности препознавања мултимодалног говора.

- Размотрени су различити сценарији (усаглашени и неусаглашени) и различити типови вектора (састављени од кепстралних, делта и делта-делта кепстралних коефицијената) и њихов утицај на успешност препознавања.
- Доказано је (теоретски и експериментално) да нормализација директно утиче на препознавање мултимодалног говора са посебним акцентом на неусаглашене сценарије. Применом нормализације (CMS и/или RASTA) успешност препознавања је знатно побољшана.
- Показало се да нелинерани Teager Energy оператор (који успешно апроксимира турбулентно кретања ваздуха у вокалном тракту) даје значајне резултате у побољшању препознавања мултимодалног говора са посебним акцентом на сценарије где учествује шапат.
- Примена нелинеарног Teager Energy оператора омогућила је добијање нових векторских обележја. Предложени су нови алгоритми и добијена нова векторска обележја као што су TEPLPCC и TERASTACC.
- Показало се да са новим векторским обележјима који садрже Teager Energy оператор (а посебно са TERASTACC) резултати успешности препознавања добијају знатно побољшање.
- Експериментално је доказано да комбинација нормализације и Teager Energy оператора даје значајна побољшања. Дата је препорука да за препознавање шапата ова комбинација буде обавезно заступљена у одговарајућим алгоритмима.
- На бази поређења свих векторских обележја по свим сценаријима и за све типове вектора предложена је одговарајућа градацијска скала. Она представља основу за избор најпогоднијих обележја мултимодалног говора.

Резултати добијени на основу ових истраживања могу се користити за даљу анализу и креирање нових алгоритама са циљем унапређења препознавања мултимодалног говора.

7.3. МОГУЋНОСТИ ДАЉИХ ПРАВАЦА ИСТРАЖИВАЊА

У поступку даљег унапређења система за препознавање мултимодалног говора могу се идентификовати одређене смернице будућих истраживања и то:

- значајно проширење говорне базе Whi-Spe на бар 100 говорника при чему би полови били подједнако заступљени,
- истраживање успешности препознавања говора зависно од пола на основу проширене базе и одређивање оптималног векторског обележја према полу говорника,
- спровођење анализе препознавања мултимодалног говора независно од говорника,
- анализа утицаја променљивог односа сигнал-шум на препознавање мултимодалног говора,
- примена метода за „добар“ и „лош“ шапат на снимљену говорну базу,

- коришћење других познатих метода и софтверских пакета (Kaldi,...) за анализу препознавања ове врсте говора,
- добијање хибридних решења која би објединила одговарајућа векторска обележја са циљем максимизације успешности препознавања мултимодалног говора.

Треба очекивати да би се применом горе поменутих смерница добио робустан систем за препознавање мултимодалног говора, а током таквих истраживања дошло би се и до нових идеја за унапређење одређених алгоритама.

ЛИТЕРАТУРА

[Acquah, 2010] Acquah H. (2010).“Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship”. J. Develop. Agri. Econ. 2(1), 001–006, 2010.

[Ahmadi et al., 2008] Ahmadi F., McLoughlin I.V., Sharifzadeh H.R. (2008). “Analysis-by synthesis method for whisper-speech reconstruction," in Proc. of IEEE APCCAS, 2008, pp. 1280-1283.

[Atal, 1974] Atal B. (1974). "Automatic recognition of speakers from their voices", Proc. IEEE 64, pp. 460-475, 1974.

[Bellman, 1957] Bellman R.E. (1957). “Dynamic Programming”, Princeton University Press, Princeton, New Jersey, USA, 1957.

[Benesty et al., 2008] Benesty J., Sondhi M.M., Huang Y. (Eds.) (2008). “Springer Handbook of Speech Processing”, Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-49125-5, 2008.

[Boll, 1979] Boll S.F. (1979). “Suppression of acoustic noise in speech using spectral subtraction”, IEEE, Vol. 27, No. 2, pp. 113-120, 1979.

[Boril, Hansen, 2010] Boril H. and Hansen J.H.L. (2010). “Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1379–1393, August 2010.

[Bou-Ghazle, Hansen, 2000] Bou-Ghazle S.E. and Hansen J.H.L. (2000). “A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress”, vol. 8, No. 4, July, IEEE Trans Speech and Audio, 2000.

[Catford, 1977] Catford J.C. (1977). “Fundamental problems in phonetics”, Edinburgh University Press, Edinburgh, 1977.

[Cheng et al., 2005] Cheng O., Abdulla W. and Salcic Z. (2005). “Performance evaluation of front end algorithms for robust speech recognition,” in *Proc. ISSPA,* 2005, pp.711-714 *Speech Communication*, 45, pp.129-152, 2005.

[Davis, Mermelstein, 1980] Davis S.B. and Mermelstein P. (1980). “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 4, pp. 357–366, Aug. 1980.

[De Veth, Boves, 1998] De Veth J. and Boves L. (1998). “Channel Normalization Techniques for Automatic Speech Recognition over the Telephone”, Speech Communication, 1998, 25, pp. 149-164.

[Deitel et al., Deitel H.M., Deitel., P.J. and Neito T.R. (1999). “Visual Basic 6 how to

1999] program", Prentice Hall, New Jarsy, 1999.

[Dimitriadis et al., 2005] Dimitriadis D., Maragos P., Potamianos A. (2005). "Auditory Teager Energy Cepstrum Coefficients for Robust Speech Recognition", Proc. of European Speech Processing Conference, Lisbon, Portugal, 2005.

[Douglas-Cowie et al., 2003] Douglas-Cowie, E., Campbell, N., Cowie, R. and Roach, P. (2003). "Emotional speech: Towards a new generation of database." Speech Communication, Vol. 40, pp. 33-60, 2003.

[Eklund, Traunmuller, 1996] Eklund I., Traunmuller H. (1996). "Comparative study of male and female whispered and phonated versions of the long vowels of Swedish", *Phonetica,* 1996, 54, pp. 1–21.

[Fan, Hansen, 2010] Fan X., Hansen J.H.L. (2010). "Acoustic analysis for speaker identification of whispered speech", in *IEEE ICASSP'10*, 2010, pp. 5046-5049.

[Fan, Hansen, 2011] Fan X., Hansen J.H.L. (2011). "Speaker identification within Whispered Speech Audio Stream," *IEEE Transactions on Audio, Speech and Language Processing*, 19(5), 2011, pp. 1408-1421.

[Furui, 1981] Furui S. (1981). "Cepstral analysis technique for automatic speaker verification," IEEE Transactions on Acoustics, Speech and Signal Processing, 1981, vol. 29, no. 2, pp. 254–272.

[Gales, 1996] Gales M.F. (1996). "The Generation And Use Of Regression Class Trees For MLLR Adaptation", Cambridge University Engineering Department, Tech. Rep., 1996.

[Gales, Woodland, 1996] Gales M., Woodland P. (1996). "Mean and variance adaptation within the MLLR framework", Comput. Speech Lang. 10(4), 249–264 (1996).

[Galić et al., 2011] Galić J., Jovičić S., Marković B. (2011). "Uporedna analiza tri modela za automatsko prepoznavanje govora pri šapatu", Zbornik apstrakta konferencije TAKTONS 2011, str. 12, Novi Sad, Srbija, 2011.

[Galić et al., 2013 a] Galić J., Popović M., Marković B., Grozdić Đ.T., Jovičić S.T. (2013). „Primjena skrivenih Markovljevih modela u prepoznavanju govora u šapatu", Zbornik radova konferencije INFOTEH 2013, str. 387-390, Jahorina, Republika Srpska, 2013.

[Galić et al., 2013 b] Galić J., Jovičić S., Grozdić Đ., Marković B. (2013). "The influence of feature vector selection on performance of automatic recognition of whispered speech", Proceedings Speech and Language 2013, 4th International Conference on Fundamental and Applied Aspects of Speech and Language, pp. 258-264, Belgrade, October 25-26, 2013.

[Galić et al., 2014 a] Galić J., Jovičić S.T., Grozdić Đ. and Marković B. (2014). "HTK-Based Recognition of Whispered Speech", A. Ronzhin et al. (Eds.): SPECOM 2014, LNAI 8773, pp. 251–258, 2014, Springer International Publishing Switzerland, 2014.

[Galić et al., 2014 Galić J., Jovičić S.T., Grozdić Đ., Marković B. (2014). „Constrained Lexicon

b] Speaker Dependent Recognition of Whispered Speech“, Zbornik radova konferencije INDEL 2014, pp. 180-184, Banja Luka, Republika Srpska, 2014.

[Gavidia-Ceballos, 1995] Gavidia-Ceballos L. (1995). “Analysis and Modeling of Speech for Laryngeal Pathology Assessment,” Department of Biomedical Engineering, Duke University, Durham, North Carolina, 1995.

[Gavidia-Ceballos, Hansen, 1996] Gavidia-Ceballos L., Hansen J.H.L. (1996). “Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection,” IEEE Trans Biomedical Engineering, vol.43, No.4, Apr 1996.

[Georgogiannis, Digalakis, 2012] Georgogiannis A., Digalakis V. (2012). “Speech Emotion Recognition Using Non-Linear Teager Energy Based Features in Noisy Environments,” EURASIP, 2012, pp. 2045–2049.

[Ghaffarzadegan et al., 2015] Ghaffarzadegan S., Boril H., Hansen J.H.L. (2015). “Generative modeling of pseudo-target domain adaptation samples for whispered speech recognition”, *IEEE ICASSP’15, 2015*, pp. 5024-5028.

[Ghaffarzadegan et al., 2016] Ghaffarzadegan S., Boril H., Hansen J.H.L. (2016). “Generative modeling of pseudo-whisper for robust whispered speech recognition”, *IEEE/ACM* Transactions on audio, speech, and language processing, vol. 24, No. 10, October 2016, pp. 1705-1720.

[Glasberg, Moore, 1990] Glasberg B., Moore B. (1990). “Derivation of auditory filter shapes from notched-noise data”, *Hearing Research*, 1990, Vol. 47, pp. 103-108.

[Grozdić et al., 2012] Grozdić Đ.T., Marković B., Galić J., Jovičić S.T. (2012). „Primena neuralnih mreža u prepoznavanju govora u šapatu“, Zbornik radova 20tog Telekomunikacionog foruma TELFOR 2012, str. 728-731, Beograd, 20-22. novembra 2012.

[Grozdić et al., 2013 a] Grozdić Đ., Marković B., Galić J., Jovičić S., Furundžić D. (2013). “Neural Network-Based recognition of whispered speech”, Proceedings Speech and Language 2013, 4th International Conference on Fundamental and Applied Aspects of Speech and Language, pp. 223-229, Belgrade, October 25-26, 2013.

[Grozdić et al., 2013 b] Grozdić Đ.T., Marković B., Galić J., Jovičić S.T. (2013). „Application of Neural Networks in Whispered Speech Recognition“, TELFOR Journal, Vol. 5, No. 2, 2013, pp. 103-106.

[Grozdić et al., 2013 c] Grozdić Đ.T., Jovičić S.T., Galić J. and Marković B. (2013). “Experiments in whisper recognition using neural networks”, In: S.T. Jovičić, M. Subotić, M. Sovilj (Eds.): VERBAL COMMUNICATION QUALITY, Interdisciplinary Research, II; CUŽA, Belgrade, ISBN 978-86-81879-46-7, pp. 91-110, 2013.

[Grozdić et al., 2014] Grozdić Đ., Jovičić S., Galić J., Marković B. (2014). „Application of inverse filtering in enhancement of whisper recognition“, Zbornik radova konferencije NEUREL 2014, str. 157-161, Beograd, Srbija, 2014.

[Grozdić et al., Grozdić Đ.T., Šumarac-Pavlović D., Jovičić S.T., Galić J. and Marković B. (2015).

2015] „Komparacija tehnika normalizacije kepstralnih koeficijenata u automatskom prepoznavanju šapata”, Zbornik radova 59. Konferencije ETRAN, ETRAN 2015, str. AKI1.8.1-5, Srebrno jezero, Srbija 8-11. jun, 2015.

[Grozdić et al., 2017] Grozdić Đ., Jovičić S., Šumarac-Pavlović D., Galić J., Marković B. (2017). “Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition”, Advances in Electrical and Computer Engineering, Vol. 17. Number 1, 2017, pp 21-26.

[Hansen, 1988] Hansen J.H.L. (1998). "Analysis and Compensation of Stressed and Noisy Speech with Application to Robust Automatics Recognition", Ph.D. Thesis, Georgia Inst. Tech. July, 1988.

[Hansen, Patil, 2007] Hansen J.H.L., Patil S. (2007). "Speech Under Stress: Analysis, Modeling and Recognition ", C. Muller (Ed.): Speaker Classification I, LNAI 4343, pp. 108–137, Springer-Verlag Berlin Heidelberg, 2007.

[Hermansky, 1986] Hermansky H. (1986). “Perceptual linear predictive (PLP) analysis of speech”, J. Acoust. Soc. Am. 87(4), 1738-1752 (1986).

[Hermansky, Morgan, 1994] Hermansky H. and Morgan N. (1994). "RASTA processing of speech", IEEE Trans. on Speech and Audio Proc., vol. 2, no. 4, pp. 578-589, Oct. 1994.

[HTK] web страница: http://htk.eng.cam.ac.uk/ посећена 30. 3. 2017.

[Ito et al., 2005] Ito T., Takeda K., Itakura F. (2005). “Analysis and Recognition of Whispered speech,” *Speech Communication*, 45, 2005, pp. 129-152.

[Jou et al., 2005] Jou S.C., Schultz T. and Waibel A. (2005). “Whispery speech recognition using adapted articulatory features”, *ICASSP-05.* (2005). Paper SP-P15.12.

[Jovičić, 1998] Jovičić S.T. (1998). “Formant feature differences between whispered and voiced sustained vowels,” ACUSTICA - Acta Acoustica, 84(4), pp. 739-743 (1998).

[Jovičić, 1999] Jovičić S.T. (1999). “Govorna komunikacija – fiziologija, psihoakustika i percepcija“, Nauka, Beograd, 1999.

[Jovičić et al., 2004] Jovičić S.T., Kašić Z., Đorđević M., Rajković M. (2004). “Serbian emotional speech database: design, processing and evaluation”, *SPECOM-2004*, St. Petersburg, Russia, 2004, pp. 77-81.

[Jovičić et al., 2008] Jovičić S.T., Punišić S., Šarić Z. (2008). “Time-frequency detection of stridence in fricatives and affricates”, *Int. Conf. Acoustics’08,* Paris, 2008, pp. 5137-5141.

[Jovičić, Šarić, 2008] Jovičić S.T., Šarić Z.M. (2008). „Acoustic analysis of consonants in whispered speech”, *Journal of Voice*, 22(3), 2008, pp. 263-274.

[Kaiser, 1990 a] Kaiser J.F. (1990). "On a simple algorithm to calculate the 'energy' of a signal*", Proc. IEEE ICASSP’90,* Albuquerque, New Mexico, pp. 381-384, April 1990.

[Kaiser, 1990 b] Kaiser J.F. (1990). "On Teager's energy algorithm and its generalization to

continuous signals", *Proc. 4th IEEE Digital Signal Proc. Workshop, Mohonk* (New Paltz), NY, September 1990.

[KALDI] web страница: http://kaldi-asr.org/ посећена 30. 3. 2017.

[Kondoz, 1994] Kondoz A.M. (1994). "Digital Speech Coding for Low Bit Rate Communication Systems", John Wiley & Sons, 1994.,

[Kostek, 1999] Kostek B. (1999). "Soft Computing in Acoustics, Applications of Neural Network, Fuzzy Logic and Rough Sets of Musical Acoustics", Springer-Verlag, Berlin, 1999.

[Kozierski et al., 2016] Kozierski P., Sadalla T., Drags S., Dobrowski , A., Horla D. (2016), *Kaldi toolkit in Polish whispery speech recognition*, Przeglad Elektrotechniczny, **R.92**, 11, 301–304.

[Lass et al., 1976] Lass N.J., Hughes K.R., Bowyer M.D., Waters L.T., Bourne V.T. (1976). "Speaker sex identification from voiced, whispered and filtered isolated vowels", *J Acoust Soc Am.,* 1976;59:675-678.

[Lee, Rose, 1996] Lee L., Rose R.C. (1996). "Speaker normalization using efficient frequency warping procedure," in *Proc. ICASSP*, 1996, pp. 353–356.

[Ljubić et al., 2014] Ljubić G., Surudžić M., Marković B. (2014). „Realizacija DTW algoritma korišćenjem MATALB-a", „Tehnika i praksa", Broj 12, 2014, VŠTSS Čačak, str. 125-132.

[Marković, 2002 a] Marković B.R. (2002). "Grafički interfejs za poređenje govornih signala korišćenjem DTW metoda", Zbornik radova DOGS2002, Bečej, 16-17. maj 2002.

[Марковић, 2002] Марковић Б.Р. (2002). "WiseHMM – Графички интерфејс за обучавање скривених Марковљевих модела (HMM) континуалног типа", Zbornik radova XLVI za ETRAN, Vol. II, str. 325-328, Banja Vrućica-Teslić, 2002.

[Marković, 2002 b] Marković B. (2002). "Upotreba skrivenih Markovljevih modela u prepoznavanju govora", Interdisciplinarni časopis "Nauka Tehnika Bezbednost", broj 2, str. 3-23, Institut bezbednosti, Beograd, decembar 2002.

[Марковић, 2004] Марковић Б.Р. (2004). "Функција „Позивање говором" у мобилној телефонији", магистарски рад, Електротехнички факултет, Универзитет у Београду, Београд, 2004.

[Марковић, 2005] Марковић Б.Р. (2005). "WiseEdit 1.0 – софтверски пакет за аквизицију говорних узорака у облику wave фајлова", Zbornik radova XLVI za ETRAN, tom II, str. 388-391, Budva, 5-10. juna 2005.

[Marković, Marković, 2008] Marković G., Marković B. (2008). "Vizuelni DTW kao nastavno sredstvo za poređenje govornih uzoraka", Tehnika i informatika u obrazovanju, TIO '08, str. 409-415, Tehnički fakultet, Čačak, 9-11. maja.

[Marković**,** Marković B.**,** Luković G. (2012). „WiseWave 1.4 – softver za analizu uticaja prozorovanja na prepoznavanje izolovano izgovorenih reči", Zbornik radova

Luković, 2012] 56. Konferencije za ETRAN, str. AK1.6-1-4, Zlatibor, 11-14. juna 2012.

[Marković et al., 2013 a] Marković B., Jovičić S.T., Galić J., Grozdić Đ. (2013). "Whispered Speech Database: Design, Processing and Application", 16th International Conference, I. Habernal and V. Matousek (Eds.): TSD 2013, LNAI 8082, Springer-Verlag Berlin Heidelberg, pp. 591-598. (2013).

[Marković et al., 2013 b] Marković B.**,** Galić J., Grozdić Đ., Jovičić S.T. (2013). "Application of DTW method for whispered speech recognition", Proceedings Speech and Language 2013, *4th International Conference on Fundamental and Applied Aspects of Speech and Language*, pp. 308-315, Belgrade, October 25-26, 2013.

[Marković et al., 2013 c] Marković B.**,** Lacmanović D., Tatović M., Marković G. (2013). „Analiza uticaja vrste kepstralnih koeficijenata na prepoznavanje izolovano izgovorenih reči srpskog jezika", „Tehnika i praksa", Broj 9, 2013, VŠTSS Čačak, str. 97-104.

[Marković et al., 2013 d] Marković B.**,** Mitić M., Pajkić M., Milenković N., Marković G. (2013). „Analiza uticaja globalnog ograničenja CE2-1 na prepoznavanje izolovano izgovorenih reči srpskog jezika", „Tehnika i praksa", Broj 10, 2013, VŠTSS Čačak, str. 147-153.

[Marković**,** Grozdić, 2014] Marković B.R., Grozdić Đ.T. (2014). „The LPCC-DTW Analysis for Whispered Speech Recognition", Proceedings of 1st International Conference of Electrical, Electronic and Computer Engineering, IcETRAN 2014, pp. AKI1.1.1-4, Vrnjačka Banja, Serbia, June 2-5, 2014.

[Marković et al., 2015] Marković B.R.**,** Jovičić S.T., Galić J. and Grozdić Đ.T. (2015). „Recognition of the Multimodal Speech Based on GFCC Features", Proceedings of 2nd International Conference of Electrical, Electronic and Computer Engineering, IcETRAN 2015, pp. AKI1.3.1-5, Silver Lake, Serbia, June 8-11, 2015.

[Marković et al., 2016] Marković B.R., Jovičić S.T., Mijić M., Galić J. and Grozdić Đ.T. (2016). „Recognition of Whispered Speech Based on PLP Features and DTW Algorithm", Proceedings of 3rd International Conference of Electrical, Electronic and Computer Engineering, IcETRAN 2016, pp. AKI1.3.1-5, Zlatibor, Serbia, June 2016.

[Marković et al., 2017 a] Marković B.R., Stevanović G., Jovičić S.T., Mijić M., Galić J. and Grozdić Đ.T. (2017). „Recognition of Normal and Whispered Speech Based on RASTA Filtering and DTW Algorithm", Proceedings of 4th International Conference of Electrical, Electronic and Computer Engineering, IcETRAN 2017, Kladovo, Serbia, Jun 05-08, pp. AKI2.8.1-4.

[Marković et al., 2017 b] Marković B., Galić J., Grozdić Đ, S T. Jovičić and Mijić M. (2017). „Whispered Speech Recognition Based on Gammatone Filter Cepstral Coefficients", Journal of Communication Technolgy and Electronics, 2017, Vol. 62, No. 11, pp. 1255-1261.

[Marković et al., 2018] Marković B.R**.**, Galić J., Mijić M. (2018). "Application of Teager Energy Operator on Linear and Mel Scales for Whispered Speech Recognition",

Archives of Acoustics, 2018, Vol. 43, No. 1, pp. 3-9.

[Mathur et al., 2012] Mathur A., Reddy S.M. and Hegde R.M. (2012). “Significance of parametric spectral ratio methods in detection and recognition of whispered speech” EURASIP Journal on Advances in Signal Processing 2012, 2012:157

[MATLAB] web страница: https://www.mathworks.com/ посећена 30. 3. 2017.

[Matsuda, Kasuya, 1999] Matsuda M., Kasuya H. (1999). ”Acoustic nature of the whisper”, *Proc. Eurospeech 99,* 1, 1999, pp. 137-140.

[Matsuda et al., 2000] Matsuda M., Mori H., Kasuya H. (2000). “Formant structure of whispered vowels”, *J Acoust Soc Japan,* 2000;56(7):477-487.

[McLoughlin, 2007] McLoughlin I.V. (2007). “Line spectral pairs", Signal Processing Journal, 2007, pp. 448-467.

[Miller, Nicely, 1955] Miller G.A. and Nicely P.E. (1955). "An analysis of perceptual confusions among some English consonants“, J. Acoust. Soc. Am., 27, 338- 352.

[Mitrović et al., 2012] Mitrović R., Živanović I., Radeljić I., Marković B. (2012). “Korišćenje asinhronog dinamičkog programiranja u rešavanju problema optimalne staze“, Tehnika i informatika u obrazovanju, TIO 2012, 4. internacionalna konferecija, Zbornik, str. 193-199, Tehnički fakultet, Čačak, 1-3. jun 2012.

[Petterson, 1992] Patterson R.D., Robinson K., Holdsworth J., McKeown D., Zhang C., and Allerhand M.H. (1992). “Complex sounds and auditory images,” In *Auditory Physiology and Perception*, (Eds.) Y Cazals, L. Demany, K.Horner, Pergamon, Oxford, 1992, pp. 429-446.

[PRAAT] web страница: http://www.fon.hum.uva.nl/praat/ посећена 30. 3. 2017.

[Rabiner, Juang, 1993] Rabiner L., Juang B-H. (1993). *Fundamentals of speech recognition*, (Prentice Hall, New Jersey) (1993).

[Rostolland, 1982 a] Rostolland D. (1982). "Acoustic Features of Shouted Voice Part I", Acustica, Vol. 50 pp.118-125, 1982.

[Rostolland, 1982 b] Rostolland D. (1982). "Phonetic Structure of Shouted Voice Part II", Acustica, Vol. 51 pp.80-89, 1982.

[Rubin et al., 2004] Rubin A.D., Praneetvatakul V., Gherson S., Moyer C.A., Sataloff R.T. (2004). “Laryngeal hyperfunction during whispering: reality or myth?” *Journal of Voice,* 20, 2004, pp. 121–127.

[Sakoe, Chiba, 1978] Sakoe H. and Chiba S. (1978). „Dynamic programming optimization for spoken word recognition”, IEEE Trans. Acoustics, Speech, Signal Proc., pp 43-49, 1978.

[Sandberg et al., 2010] Sandberg J., Scherer R., Hess M., Muller F. (2010). “Whispering – A Single-Subject Study of Glottal Configuration and Aerodynamics”, *Journal of Voice,* Vol. 24. No. 5, 2010, pp. 574-584.

[Sharifzadeh et al., Sharifzadeh H.R., McLoughlin I.V., Ahmadi F. (2008). “Regeneration of speech

2008] in voice-loss patients," in Proc. of ICBME, vol. 23, 2008, pp. 1065-1068.

[Sharifzadeh et al., 2009] Sharifzadeh H.R., McLoughlin I.V. and Ahamdi F. (2009). "Voiced Speech from Whispers for Post-Laryngectomised Patients". *IAENG International Journal of Computer Science*, 36:4, IJCS_36_4_13 (Advance online publication: 19 November 2009).

[Sharifzadeh et al., 2012] Sharifzadeh H.R., McLoughlin I.V., Russell M.J. (2012). „A Comprehensive Vowel Space for Whispered Speech", *J Voice,* 2012;26(2):e49-e56.

[Симић, Остојић, 1996] Симић Р., Остојић Б. (1996). „Основи фонологије српског књижевног језика", Универзитетски уџбеници 22, Универзитет у Београду, 1996.

[Smith, 2016] Smith D.R.R. (2016). "Speaker-Sex Discrimination for Voiced and Whispered Vowels at Short Duration", i-perception, SAGE, Oct.2016, pp 1-13.

[Solomon et al., 1989] Solomon N.P., McCall G.N., Trosset M.W., Gray W.C. (1989). "Laryngeal configuration and constriction during two types of whispering", *J Speech Hearing Research.* 1989;32:161-174.

[Subotić et al., 2013] Subotić M., Čabarkapa N., Vojnović M. (2013). "Fricative characteristics", Verbal Communication Quality, Interdisciplinary Research, II, Eds. Jovičić S.T., Subotić M., ISBN 978-86-81879-46-7, LAAC, IEPSP, Belgrade, 2013, pp. 5-21.

[Sundberg et al., 2010] Sundberg J., Scherer R., Hess M., and Muller F. (2010). "Whispering—A Single-Subject Study of Glottal Configuration and Aerodynamics", Journal of Voice, 2010; Vol. 24, Issue 5, pp 574-584.

[Tartter, 1986] Tartter V.C. (1986). "What's in a whisper?", J. Acoust. Soc. Am. 86, 1678 (1986).

[Tartter, 1991] Tartter V.C. (1991). "Identifiability of vowels and speakers from whispered syllables", *Percept Psychophys.* 1991, 49, 365-372.

[Thomas, 1969] Thomas I.B. (1969). "Perceived pitch of whispered vowels," Journal of the Acoustical Society of America, vol. 46, 1969, pp. 468-470.

[Tohkura, 1986] Tohkura Y. (1986). "A Weighted Cepstral Distance Measure for Speech Recognation", ICASSP 86, Tokyo, pp. 761-764.

[Tritschler, Gopinath, 1999] Tritschler A., Gopinath R. (1999). "Improved speaker segmentation and segments clustering using the Bayesian information criterion." in *Proc. Eurospeech,* vol. 2*,* (Citeseer, 1999), pp. 679–682.

[Tsunoda et al., 1997] Tsounoda K., Ohta Y., Soda Y., Niimi S. and Hirosi H. (1997). "Laryngeal adjustment in whispering", Annual of Otology, Rhinology and Laryngology, Vol. 106. pp. 41-43, (1997).

[Tsunoda et al., 2012] Tsunoda K., Sekimoto S., and Baer T. (2012). "Brain Activity in Aphonia after a Coughing Episode: Different Brain Activity in Healthy Whispering and Pathological Aphonic Conditions", Journal of Voice, 2012, Vol. 26, No. 5, pp. 668.e11-668.e13.

[Wikipedia] web страница: https:/en.wikipedia.org/wiki/Glotis посећена 25. 9. 2017.

[Wilson, 1998] Wilson J.B. (1998). "A Comparative Analysis of Whispered and Normally Phonated Speech Using An LPC-10 Vocoder", RADC, Final Report TR-75-264, 1998, pp. 739-743.

[Zhang, Hansen, 2007] Zhang C., Hansen J.H.L. (2007). “Analysis and classification of Speech Mode: Whisper through Shouted,” Interspeech 2007, 2007, pp. 2289-2292.

[Zhang, Hansen, 2011] Zhang C., and Hansen J.H.L. (2011). ”Whisper-Island Detection Based on Unsupervised Segmentation With Entropy-Based Speech Feature Processing”, *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4), 2011, pp. 883 – 894.

ПРИЛОЗИ

У овом делу дати су прилози који се састоје од резултата препознавања мултимодалног говора базираних на HMM методи, биографији и изјавама аутора.

ПА: РЕЗУЛТАТИ НА БАЗИ HMM МЕТОДЕ[1]

Коришћењем HTK алата добијени су одговарајући HMM резултати за векторска обележја LFCC, TELFCC, MFCC, TEMFCC [Marković et al., 2018] и GFCC [Marković et al., 2017 b].

Сви вектори су генерисани на начин као што је описано у поглављу 3 и идентични су онима који су коришћени за експерименте са DTW методом. Фонетском транскрипцијом свака реч је издељена на скуп монофона, а сваки монофон описан са по 5 стања. HMM модел који је употребљен за ове експерименте подразумевао је кретање „с-лева-на-десно“ без прескакања стања. За почетни модел је коришћен „flat-start“ тј. узете су средње вредности и варијансе свих узорака из обуке и свим стањима додељене исте иницијалне вредности. Број циклуса реестаимације је био лимитиран на 5. У фази тестирања употребљен је Витербијев алгоритам.

ПА.1 HMM РЕЗУЛТАТИ СА LFCC ВЕКТОРСКИМ ОБЕЛЕЖЈЕМ

У табелама ПА.1-4 дати су резултати препознавања са одговарајућом средњом вредношћу и границом грешке за ниво поузданости од 95%.

Табела ПА.1 *LFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	98.6	98.6	99.6	98.8
Говорник 2	98.6	99	99.8	99.8
Говорник 3	94.8	98.8	98.6	99.4
Говорник 4	100	99.6	100	99.6
Говорник 5	98.8	99	99.6	99.2
Говорник 6	97.6	98.6	99	99
Говорник 7	93.2	96.8	98.6	98.6
Говорник 8	99.00	99.2	99.4	99.6
Говорник 9	96.6	98	99.2	99.6
Говорник 10	97.4	97.4	99.4	99.6
Ср. вред. ± Грешка	**97.46±1.29**	**98.5±0.53**	**99.32±0.29**	**99.32±0.25**

[1] Ови резултати су добијени у сарадњи са колегом Јованом Галићем, студентом докторских студија на ЕТФ-у

Табела ПА.2 *LFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	96.6	98.4	98.4	98.6
Говорник 2	95.4	97.8	99.2	99.2
Говорник 3	97.8	98.4	99.6	99.8
Говорник 4	99	98.4	99.2	99
Говорник 5	97.2	97.2	96.6	97.4
Говорник 6	96.8	98.4	99	98
Говорник 7	94.6	97.8	98.8	98.8
Говорник 8	95.4	96.8	99.2	98.8
Говорник 9	94.6	97.6	99.2	99.2
Говорник 10	94.4	97	98.2	98.4
Ср. вред. ± Грешка	**96.18±0.96**	**97.78±0.39**	**98.74±0.53**	**98.72±0.42**

Табела ПА.3 *LFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	47.24	81.6	88.8	87.84
Говорник 2	12.76	59.48	73.52	71.64
Говорник 3	18.36	70.72	72.44	72.24
Говорник 4	31.32	69.16	71.88	72.92
Говорник 5	28.96	69.04	73	75.4
Говорник 6	45.04	77.48	85.6	81.96
Говорник 7	27.28	75.84	81.96	78.04
Говорник 8	17.08	57.92	71.8	72.68
Говорник 9	38.36	81.68	83.36	84.24
Говорник 10	16.72	68.96	76.96	78.32
Ср. вред. ± Грешка	**28.31±7.59**	**71.19±5.09**	**77.93±3.98**	**77.53±3.48**

Табела ПА.4 *LFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	LFCC кепст. (без CMS-а)	LFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	24.76	75.52	85.12	86.24
Говорник 2	5.76	40.12	52.96	54.32
Говорник 3	16.48	59.32	69.44	74.04
Говорник 4	17.88	45.52	50.2	52.08
Говорник 5	17.8	64.44	66.16	68.96
Говорник 6	40.2	68.16	74.48	72.88
Говорник 7	23.96	59.92	69.24	68.92
Говорник 8	18.08	56.52	69.68	70
Говорник 9	35.24	75.68	82.76	82.4
Говорник 10	27.4	71	76.52	78.08
Ср. вред. ± Грешка	**22.76±6.15**	**61.62±7.41**	**69.66±7.01**	**70.79±6.76**

ПА.2 HMM РЕЗУЛТАТИ СА TELFCC ВЕКТОРСКИМ ОБЕЛЕЖЈЕМ

На сличан начин резултати за TELFCC дати су табелама ПА.5-8.

Табела ПА.5 *TELFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	98.8	98.4	97.8	98
Говорник 2	98	97.8	97.8	98.2
Говорник 3	96.8	98.6	99	99
Говорник 4	98.8	99	98.4	98.8
Говорник 5	99.4	99.2	99.6	99.2
Говорник 6	96.2	96.6	97.4	98.2
Говорник 7	92	94.6	97.8	98.6
Говорник 8	98.8	99	98.8	98.6
Говорник 9	95.2	97.6	98.4	99.8
Говорник 10	97.8	97.8	99	99.2
Ср. вред. ± Грешка	**97.18±1.40**	**97.86±0.86**	**98.4±0.43**	**98.76±0.34**

Табела ПА.6 *TELFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	96.8	97.8	97.6	97.4
Говорник 2	96.4	97.2	99.2	98.6
Говорник 3	98	99.2	99.2	99.4
Говорник 4	99	97.4	98.2	99
Говорник 5	97.2	96.6	97	97.6
Говорник 6	96.4	96.2	97.6	97.6
Говорник 7	95.6	94.8	98.4	98
Говорник 8	95.4	96	98	98.6
Говорник 9	93.8	97	99.2	99.2
Говорник 10	94.8	94.8	97.4	98.2
Ср. вред. ± Грешка	**96.34±0.95**	**96.7±0.83**	**98.18±0.50**	**98.36±0.44**

Табела ПА.7 *TELFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	33.68	79.6	80.8	78.72
Говорник 2	17.08	63.28	75.48	76.2
Говорник 3	17.92	69.16	70.32	75.28
Говорник 4	13.32	63.92	72.56	73.2
Говорник 5	16.04	65.44	68.04	71.8
Говорник 6	35.92	74.88	77.28	76.08
Говорник 7	38.4	74.76	81.52	79.88
Говорник 8	23.24	69.04	68.16	71.84
Говорник 9	32.16	77.2	82.12	81.08
Говорник 10	31.08	76.16	78.32	76.6
Ср. вред. ± Грешка	**25.89±5.80**	**71.34±3.67**	**75.46±3.36**	**76.07±1.98**

Табела ПА.8 *TELFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TELFCC кепст. (без CMS-а)	TELFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	31.64	72.16	78.4	78.92
Говорник 2	7.76	42.52	50.8	49.24
Говорник 3	14.88	51.6	60.2	57.72
Говорник 4	8.04	51.24	52.48	52.92
Говорник 5	15.52	58.68	64.64	62.52
Говорник 6	49.96	68.08	71.68	72.8
Говорник 7	47.44	63.04	63.68	65.64
Говорник 8	25.08	57.64	60.88	61.76
Говорник 9	35.64	73.76	78.68	78.12
Говорник 10	35.16	70.76	75.16	72.64
Ср. вред. ± Грешка	**27.11±9.51**	**60.95±6.46**	**65.66±6.23**	**65.23±6.36**

ПА.3 HMM РЕЗУЛТАТИ СА MFCC ВЕКТОРСКИМ ОБЕЛЕЖЈЕМ

За MFCC векторско обележје и HMM метод резултати су приказани у табелама ПА.9-12.

Табела ПА.9 *MFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	98	99.6	99.4	99.6
Говорник 2	99.8	99.6	100	100
Говорник 3	95.8	98.8	99	99.2
Говорник 4	99.6	99.4	99.4	99.6
Говорник 5	99.4	99	100	99.8
Говорник 6	98.4	96.6	99	99.6
Говорник 7	95.6	97.6	99.2	99.4
Говорник 8	99	99.8	98	98.2
Говорник 9	98.4	98.6	99.2	99.4
Говорник 10	97.6	99.4	99.6	99.6
Ср. вред. ± Грешка	**98.16±0.91**	**98.84±0.63**	**99.28±0.36**	**99.44±0.30**

Табела ПА.10 *MFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	97.2	97.2	99	99.2
Говорник 2	94	98.2	98.6	98.8
Говорник 3	95.4	99.2	99.4	98.8
Говорник 4	98.4	97.2	99	98.8
Говорник 5	95.4	96.4	99.2	97.8
Говорник 6	95.6	96.8	98.6	97.4
Говорник 7	95.4	98	99	99.2
Говорник 8	95	96.2	98.4	99.4
Говорник 9	95.6	97.4	99.6	99.2
Говорник 10	94.4	94.8	98.2	98.8
Ср. вред. ± Грешка	**95.64±0.80**	**97.14±0.75**	**98.9±0.28**	**98.74±0.40**

Табела ПА.11 *MFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	24.44	72.72	86.84	88.96
Говорник 2	9.36	58.56	56.32	53.24
Говорник 3	11.6	65.84	73.6	73.8
Говорник 4	12.76	68.24	67.64	61.64
Говорник 5	12.44	63.28	72.88	72.52
Говорник 6	11.16	60.36	59.32	56.32
Говорник 7	20	78.2	84.76	84.28
Говорник 8	9.8	73	78.52	79.32
Говорник 9	23.32	75.2	80.52	74.68
Говорник 10	9.2	62.32	69.88	72.64
Ср. вред. ± Грешка	**14.41±3.64**	**67.77±4.18**	**73.03±6.27**	**71.74±7.18**

Табела ПА.12 *MFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	MFCC кепст. (без CMS-а)	MFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	41.8	69.4	87.96	90.04
Говорник 2	13.08	52.76	61.72	62.72
Говорник 3	11.96	40.84	56.68	61.12
Говорник 4	16.36	57.8	75.72	75.68
Говорник 5	26.84	72.96	83.64	83.36
Говорник 6	16.52	60.84	79.56	77.92
Говорник 7	20.24	72.92	83.36	87.12
Говорник 8	17.72	80.64	92.44	92.6
Говорник 9	22.64	73.88	85.04	87.04
Говорник 10	22.6	61.32	70.04	72.6
Ср. вред. ± Грешка	**20.98±5.35**	**64.34±7.39**	**77.62±7.18**	**79.02±6.83**

ПА.4 HMM РЕЗУЛТАТИ СА TEMFCC ВЕКТОРСКИМ ОБЕЛЕЖЈЕМ

На сличан начин дати су и резултати за TEMFCC векторско обележје у табелама ПА.13-16.

Табела ПА.13 *TEMFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.4	99.4	99.8	100
Говорник 2	99	98.8	100	100
Говорник 3	96.2	100	99.2	99.8
Говорник 4	99.6	99.6	99.8	99.6
Говорник 5	99.6	99.8	100	100
Говорник 6	96.6	97	98.8	99.8
Говорник 7	94.8	97.2	98.6	99.4
Говорник 8	99.2	100	98.4	99.4
Говорник 9	98.6	99	99.4	99.6
Говорник 10	98.6	99	99.8	99.8
Ср. вред. ± Грешка	**98.16±1.04**	**98.98±0.67**	**99.38±0.37**	**99.74±0.14**

Табела ПА.14 *TEMFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	98.8	98	99	99.2
Говорник 2	98	98.6	99.2	99.4
Говорник 3	98.2	99.2	99.8	99.8
Говорник 4	99.4	97	99.2	99.2
Говорник 5	97	97	98.4	99
Говорник 6	95.4	97.4	98.8	99
Говорник 7	95.6	97.4	99.2	99
Говорник 8	97.4	96.6	99	99
Говорник 9	96.8	96.8	99.8	99.8
Говорник 10	95.4	95.6	99.6	99.4
Ср. вред. ± Грешка	**97.2±0.88**	**97.36±0.64**	**99.2±0.27**	**99.28±0.20**

Табела ПА.15 *TELMCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	13	73.72	86.92	89.16
Говорник 2	6.88	55.32	68.2	73.24
Говорник 3	6.8	62.6	78.04	81.92
Говорник 4	9.24	67.04	74.08	71.8
Говорник 5	10.2	62.96	75	77.4
Говорник 6	11.48	59.92	63.36	63.48
Говорник 7	18.12	76.64	83.12	81.36
Говорник 8	8.76	69.6	78.04	79.76
Говорник 9	15.76	75.68	82	81.16
Говорник 10	10.48	59.2	68.88	70.08
Ср. вред. ± Грешка	**11.07±2.27**	**66.27±4.61**	**75.76±4.57**	**76.94±4.56**

Табела ПА.16 *TEMFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	TEMFCC кепст. (без CMS-а)	TEMFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	9.08	66.28	87.24	91.56
Говорник 2	8.24	44.76	59.68	61.2
Говорник 3	8.44	41.48	63.16	69.92
Говорник 4	9.36	53.52	73.68	75.88
Говорник 5	9.8	67.88	84.36	84.68
Говорник 6	12.56	56.4	76.16	80.56
Говорник 7	20.12	68.44	85.36	84.84
Говорник 8	12.12	79.24	90.72	88.48
Говорник 9	13.6	70.16	85.28	85.72
Говорник 10	16.92	56.32	68.16	69.04
Ср. вред. ± Грешка	**12.02±2.35**	**60.45±7.67**	**77.38±6.83**	**79.19±6.03**

ПА.5 HMM РЕЗУЛТАТИ СА GFCC ВЕКТОРСКИМ ОБЕЛЕЖЈЕМ

За векторска обележја базирана на Gammatone филтрима резултати су приказани у табелама ПА.17-20.

Табела ПА.17 *GFCC: резултати препознавања за сценарио „нормалан/нормалан“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	99.40	97.60	99.40	99.6
Говорник 2	97.60	97.20	97.20	96.8
Говорник 3	93.80	95.40	98.60	98.2
Говорник 4	98.80	98.20	99.00	99.8
Говорник 5	98.60	96.60	99.40	98.8
Говорник 6	93.00	96.00	96.60	96.8
Говорник 7	93.00	92.80	93.80	94.00
Говорник 8	98.00	96.80	97.40	99.4
Говорник 9	93.00	94.60	96.80	97.8
Говорник 10	96.60	96.80	98.40	97.8
Ср. вред. ± Грешка	**96.18±1.66**	**96.20±0.98**	**97.66±1.06**	**97.90±1.08**

Табела ПА.18 *GFCC: резултати препознавања за сценарио „шапат/шапат“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	95.00	94.80	96.20	96.8
Говорник 2	93.60	95.80	95.60	92.6
Говорник 3	95.60	93.60	96.40	97.0
Говорник 4	96.60	95.20	95.00	94.0
Говорник 5	92.40	88.60	90.80	90.2
Говорник 6	93.20	90.40	90.00	89.2
Говорник 7	90.80	90.60	90.60	90.0
Говорник 8	92.80	92.60	92.80	91.4
Говорник 9	87.20	84.80	90.80	89.2
Говорник 10	88.20	91.20	91.60	90.4
Ср. вред. ± Грешка	**92.54±1.89**	**91.76±2.10**	**92.98±1.59**	**92.08±1.83**

Табела ПА.19 *GFCC: резултати препознавања за сценарио „нормалан/шапат“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	27.24	68.68	69.20	68.08
Говорник 2	14.92	44.44	51.48	51.48
Говорник 3	11.44	39.12	50.12	45.28
Говорник 4	16.80	53.76	52.64	39.28
Говорник 5	14.92	49.36	58.32	52.12
Говорник 6	19.56	43.08	37.40	27.56
Говорник 7	29.96	60.20	60.76	60.04
Говорник 8	11.80	49.16	48.52	40.08
Говорник 9	24.32	56.00	60.84	56.12
Говорник 10	10.20	40.84	47.40	47.32
Ср. вред. ± Грешка	**18.12±4.30**	**50.46±5.78**	**53.67±5.52**	**48.74±7.16**

Табела ПА.20 *GFCC: резултати препознавања за сценарио „шапат/нормалан“*

Врста вектора/ Говорник	GFCC кепст. (без CMS-а)	GFCC кепст. (са CMS-ом)	Кепст. + Δ (са CMS-ом)	Кепст. + Δ + ΔΔ (са CMS-ом)
Говорник 1	51.36	79.92	85.20	84.64
Говорник 2	19.28	41.96	53.84	60.52
Говорник 3	19.80	61.40	70.08	75.24
Говорник 4	34.76	60.72	65.36	59.48
Говорник 5	28.80	76.64	78.48	75.88
Говорник 6	34.20	54.32	63.04	59.16
Говорник 7	34.28	57.76	66.28	62.08
Говорник 8	35.76	60.00	71.76	67.76
Говорник 9	30.28	58.04	73.20	68.16
Говорник 10	29.68	45.72	64.64	65.76
Ср. вред. ± Грешка	**31.82±5.60**	**59.65±7.29**	**69.19±5.40**	**67.87±5.22**

ПРИЛОГ Б: БИОГРАФИЈА

Бранко Марковић је рођен 3. јануара 1966. године у Сивчини, општина Ивањица. Основну школу је завршио у Слатини, општина Чачак, а гимназију у Чачку.

На Електротехничком факултету у Београду дипломирао је на смеру Телекомуникације 1992. године са просечном оценом 9.13, а на дипломском 10. Исте године уписао је постдипломске студије на смеру „Дигитални пренос информација“. Након 10 година проведених на раду у иностранству (Канада и САД) магистарску тезу под називом „Функција ‘Позивање говором’ у мобилној телефонији“ одбранио је 2004. године под менторством проф. др Слободана Т. Јовичића. На магистарским студијама положио је све испите са оценом 10.

Бранко Марковић је радну каријеру започео у Институту „Михајло Пупин“ у Београду где је у периоду 1992 - 1993. био ангажован на развоју мрежних уређаја мултиплексера и модема. Од 1993. до 2002. године радио је у Канади и САД-у у областима софтвера за телекомуникационе компаније: Bell Canada, Architel и Nortel Networks, на пројектима као што су Video Conferencing Manager, NYNEX OLP Project, AIMS, ASAP, Telezone, OMS и други.

Почев од 2004. године па до данас, Бранко Марковић ради као предавач на Високој школи техничких струковних студија Чачак где је ангажован за групу предмета „Информациони системи и технологије“. Креирао је одређен број нових предмета за потребе ове школе као што су Рачунарске мреже, Интернет технологије (Основне студије) као и Софтверски алати, Мултимедијалне комуникације и Вишеслојна софтверска архитектура (Специјалистичке студије). Био је ментор већег броја дипломских и специјалистичких радова. Аутор је уџбеника/скрипти за одређен број предмета.

ПРИЛОГ В: ИЗЈАВЕ АУТОРА

Изјава о ауторству

Име и презиме аутора Бранко Марковић
Број индекса /

Изјављујем

да је докторска дисертација под насловом

„Анализа обележја у говорном сигналу за потребе препознавања мултимодалног говора"

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 31. 10. 2017.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора ___Бранко Марковић__
Број индекса _____/_____
Студијски програм _____/_____
Наслов рада ___“Анализа обележја у говорном сигналу за потребе препознавања мултимодалног говора“_____
Ментор ___др Миомир Мијић, редовни професор Електротехничког факултета, Универзитет у Београду_____

Потписани___Бранко Марковић_____

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду.**
Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.
Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, ___31. 10. 2017. г.___

_____

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:
„Анализа обележја у говорном сигналу за потребе препознавања мултимодалног говора“
која је моје ауторско дело.
Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.
Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.
1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)
(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 31. 10. 2017. г.

_____

1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца

не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.