УНИВЕРЗИТЕТ У НИШУ
ЕЛЕКТРОНСКИ ФАКУЛТЕТ

Бојан Д. Денић

ПРОЈЕКТОВАЊЕ КВАНТИЗЕРА ЗА ПРИМЕНУ У ОБРАДИ СИГНАЛА И НЕУРОНСКИМ МРЕЖАМА

ДОКТОРСКА ДИСЕРТАЦИЈА

Ниш, 2022.

UNIVERSITY OF NIŠ
FACULTY OF ELECTRONIC ENGINEERING

Bojan D. Denić

DESIGN OF QUANTIZERS FOR SIGNAL PROCESSING AND NEURAL NETWORKS APPLICATIONS

DOCTORAL DISSERTATION

Niš, 2022.

ЗАХВАЛНИЦА

Захваљујем се мом ментору проф. др Зорану Перићу на изузетној сарадњи како у току студија тако и током израде ове докторске дисертације. Срдачно се захваљујем и проф. др Владимиру Деспотовићу на личној и стручној подршци. Такође, најлепше се захваљујем проф.др Предрагу Петровићу на корисним саветима на личном и професионалном нивоу.

Захваљујем се и свим осталим члановима комисије на добронамерним коментарима и сугестијама у току завршне фазе ове дисертације.

Највећу и најзначајнију подршку током школовања добио сам од чланова моје породице и због тога им се најтоплије захваљујем.

Аутор

Подаци о докторској дисертацији

Ментор:	др Зоран Перић, редовни професор, Универзитет у Нишу, Електронски факултет
Наслов:	Пројектовање квантизера за примену у обради сигнала и неуронским мрежама
Резиме:	Скаларни квантизери су заступљени у многим напредним системима за процесирање и пренос сигнала, а њихов допринос се огледа у реализацији најзначајнијег корака у поступку дигитализације: дискретизације сигнала по амплитуди. С тим у вези, постоје оправдани разлози за развојем иновативних решења односно модела квантизера који имају мању комплексност а могу да пруже перформансе блиске стандардно коришћеним моделима квантизера уз доста краће време процесирања. Процес пројектовања квантизера за одређени тип сигнала је специфичан изазов, а у дисертацији је предложено више нових метода који су рачунски мање интензивни у односу на постојеће методе. Наиме, разматрано је пројектовање различитих типова квантизера како са малим и тако и са великим бројем нивоа, при чему се користе кодови са променљивом и са фиксном дужином кодних речи. Дисертација је организована тако да се бави развојем решења за кодовање телекомуникационих сигнала (говора) али и других типова сигнала попут параметара неуронске мреже. За кодовање говора предложен је већи број решења која припадају класи енкодера таласног облика. Основна карактеристика добијених решења јесте ниска комплексност, а настала су као резултат имплементације нових модела квантизера у непредиктивним и предиктивним техникама кодовања. Развијена решења имају за циљ да побољшају перформансе неких стандардизованих решења односно напреднијих решења исте или сличне комплексности. Тестирање је извршено на узорцима говора који су узети из познатих база података, а процена перфромански је извршена применом стандардних објективних мера. Такође, испитивано је слагање између експерименталних и теоријских резултата у циљу верификације исправности предложених решења. Поред тога, у дисертацији су предложена и решења заснована на скаларним квантизерима за компресију неуронских мрежа. Ова област истраживања је веома актуелна, а улога квантизације је нешто другачија него код говора и састоји се у обезбеђивању компромиса између перформанси и величине неуронске мреже. За *post-trening* квантизацију предложени су квантизери са малим бројем нивоа (нискорезолуциона квантизација), јер су они доста значајнији из угла компресије. Циљ је да се новим начином пројектовања квантизера унапреде перформансе квантоване неуронске мреже. Тестирање

	квантизера је извршено на више модела неуронских мрежа које се користе за класификацију слика (користе се познате базе података), а за процену перформанси се поред тачности предикције користи и SQNR. Заправо, тежило се ка утврђивању везе између ове две објективне мере што до сада није довољно истражено.
Научна област:	Електротехничко и рачунарско инжењерство
Научна дисциплина:	Телекомуникације
Кључне речи:	Квантизација, Компресија, Кодовање извора информација, Предиктивно кодовање, Адаптација, Говор, Неуронске мреже.
УДК:	(621.391+621.394.14):004
ЦЕРИФ класификација:	Т 121
Тип лиценце Креативне заједнице:	**CC BY-NC-ND**

Data on Doctoral Dissertation

Doctoral Supervisor:	dr Zoran Perić, full professor, University of Niš, Faculty of Electronic Engineering
Title:	Design of Quantizers for Signal Processing and Neural Networks Applications

Abstract:

Scalar quantizers are present in many advanced systems for signal processing and transmission, and their contribution is particular in the realization of the most important step in digitizing signals: the amplitude discretization. Accordingly, there are justified reasons for the development of innovative solutions, that is, quantizer models which offer reduced complexity, shorter processing time along with performance close to the standard quantizer models. Designing of a quantizer for a certain type of signal is a specific process and several new methods are proposed in the dissertation, which are computationally less intensive compared to the existing ones. Specifically, the design of different types of quantizers with low and high number of levels which apply variable and a fixed length coding, is considered.

The dissertation is organized in such a way that it deals with the development of coding solutions for standard telecommunication signals (e.g. speech), as well as other types of signals such as neural network parameters.

Many solutions, which belong to the class of waveform encoders, are proposed for speech coding. The developed solutions are characterized by low complexity and are obtained as a result of the implementation of new quantizer models in non-predictive and predictive coding techniques. The target of the proposed solutions is to enhance the performance of some standardized solutions or some advanced solutions with the same/similar complexity. Testing is performed using the speech examples extracted from the well-known databases, while performance evaluation of the proposed coding solutions is done by using the standard objective measures. In order to verify the correctness of the provided solutions, the matching between theoretical and experimental results is examined.

In addition to speech coding, in dissertation are proposed some novel solutions based on the scalar quantizers for neural network compression. This is an active research area, whereby the role of quantization in this area is somewhat different than in the speech coding, and consists of providing a compromise between performance and accuracy of the neural network. Dissertation strictly deals with the low-levels (low-resolution) quantizers intended for *post-training* quantization, since they are more significant regarding compression. The goal is to improve the performance of the quantized neural network by using the novel designing methods for quantizers. The proposed quantizers are applied to several neural network models used for image classification (some benchmark dataset are used), and

	as performance measure the prediction accuracy along with SQNR is used. In fact, there was an effort to determine the connection between these two measures, which has not been investigated sufficiently so far.
Scientific Field:	Electrical and Computer Engineering
Scientific Discipline:	Telecommunications
Key Words:	Quantization, Compression, Source coding, Predictive coding, Adaptation, Speech, Neural networks.
UDC:	(621.391+621.394.14):004
CERIF Classification:	T 121
Creative CommonsLicense Type:	**CC BY-NC-ND**

Садржај

Списак слика

Списак табела

1. Увод

Аналогно-дигитална конверзија је процес претварања аналогног сигнала у дискретни дигитални облик, који се може користити за даљу обраду микропроцесором. Ова конверзија претпоставља квантизацију улаза, што неизбежно уноси одређену количину грешке. Отуда је избор квантизера веома важан за добијање бољег квалитета сигнала, а такође је повезан са степеном компресије који се може постићи [1−12].

Квантизери се могу поделити у две категорије и то на скаларне и на векторске [1−11]. Ова класификација је направљена на основу тога да ли се квантује само један одмерак (скаларна квантизација) или одређени број одмерака истовремено (векторска квантизација). Кључни недостатак векторских квантизера јесте висока сложеност дизајна, па су из тог разлога скаларни квантизери бољи кандидати за реалне системе где је кашњење обраде критични параметар (нпр. кодовање говора). У овој дисертацији ће пажња доминантно бити усмерена на скаларне квантизере. Скаларним квантизером се заправо реална оса (опсег вредности сигнала) дели на одређени број непреклапајућих ћелија, а свака од тих ћелија је дефинисана репрезентационим (квантизационим) нивом [1–12]. Према томе, улазни подаци се мапирају у одговарајући репрезентациони ниво у зависности од тога којој ћелији припада улазна вредност. С обзиром на то да имају широк спектар примене, за скаларне квантизере се може рећи да играју важну улогу у обради и компресији сигнала. Од посебног значаја за ову дисертацију јесу примене у областима попут кодовања говора и компресије неуронских мрежа. Иако је област кодовања говора доста добро истражена и при томе је предложен велики број решења од којих су нека и практично реализована [13, 14], може се рећи да још увек постоји простор за нове доприносе у овој области, посебно у класи енкодера таласног облика. Заправо, предлагање нових решења које надмашују перформансе претходно развијених решења је представљао велики изазов за аутора. Са друге стране, значајно веће доприносе је

могуће остварити у области компресије неуронских мрежа имајајућу у виду чињеницу да је ова област недовољно истражена.

Говор припада класи временски променљивих сигнала [1−7]. Због тога се за ефикасну обраду говора препоручују адаптивне квантизационе шеме које спроводе фрејм по фрејм анализу и процесирање, где фрејм представља групу узастопних одмерака коначне дужине. Енкодери таласног облика се управо базирају на таквим шемама, а посебна пажња у дисертацији ће бити усмерена на две технике из ове класе енкодера: PCM (енг. Pulse Code Modulation) и делта модулацију (енг. Delta Modulation (DM)) односно њену варијанту адаптивну делта модлулацију (енг. Adaptive Delta Modulation (ADM)).

PCM је непредиктивна техника код које се кодују оригинални одмерци говора [1, 3, 8, 12]. За имплементацију модела скаларних квантизера у PCM-у користе се техника адаптације унапред и техника адаптације уназад [1, 8, 15]. Наведене две технике су развијене у циљу повећања степена робусности квантизера (пројектованог за неку специфичну вредност варијансе). Иначе, висока робусност омогућава квантизерима да остваре изузетно добре перформансе при процесирању временски променљивих сигнала (отпорни су на промену статистике сигнала). Техника адаптације унапред се користи чешће из разлога што омогућава боље перформансе (за око 1 dB) и отпорнија је на грешке у преносу, али захтева слање додатне информације до пријемника што није случај код технике адаптације уназад [1, 15]. У литератури је доступан већи број радова у којима је разматрана примена и процењивана ефикасност адаптације унапред како код квантизера са фиксном дужином кодних речи [16−19] тако и код квантизера са променљивом дужином кодних речи [20−22].

DM је предиктивна техника код које се разлика између тренутних вредности сигнала и њихових предвиђених вредности кодује коришћењем само једног бита [1, 8, 23]. Позната је и као поједностављена верзија DPCM-а (енг. Differental Pulse Code Modulation) [1, 3, 5, 8, 24−26]. Иницијална верзија DM-а, код које је корак квантизације константан, је позната под називом линеарна делта модулација (енг. Linear Delta Modulation (LDM)). Атрактивност LDM-а се огледа у релативно простој структури кодера и декодера и високом степену компресије.

ADM [1, 3, 8, 27−30] је побољшана верзија LDM-а, а предложена је како би се превазишли недостаци LDM-а који се односе на преоптерећење стрмином и грануларни шум. Како поменути недостаци настају као последица неадекватног избора корака квантизације, ADM покушава да прилагоди тај корак у складу са променама улазног сигнала. У литератури су до сада углавном предлагана тренутно адаптивна ADM решења (алгоритми), код којих се адаптација (ажурирање) корака квантизације или коефицијента предиктора врши на нивоу одмерка. Иако у својој оригиналној имплементацији ADM претпоставља употребу бинарног (два нивоа) квантизера, неки новији радови предлажу значајна побољшања перформанси коришћењем двобитне ADM [31−33] и дводигитне ADM [34], уз минимално повећање сложености. Такође, анализирани су и неки вишенивовски ADM системи [35, 36].

Квантизација је недавно постала предмет интересовања и у неуронским мрежама (NN). Савремене NN, нпр. развијене за потребе класификације слика [37], препознавање облика [38] или обраду говора [39], представљају комплексне архитектуре са великим бројем параметара који захтевају скупе рачунске и меморијске ресурсе. Са друге стране, висока комплексност NN-а може представљати лимитирајући фактор за примену у преносним и *edge* рачунарским уређајима са ограниченом меморијом и процесорском снагом, као и у сервисима где је кашњење критични параметар. Због тога се тежи ка смањивању капацитета (величине) NN-а, а управо је квантизација један од најчешће коришћених метода. Улога квантизације се огледа у мапирању параметера NN-а, који се обично чувају у формату са покретном тачком користећи велики број битова (32 бита), у формат са фиксном тачком користећи мањи број битова. Наравно, при томе постоји ограничење да квантована (редукована) NN пружи конкурентне перформансе у односу на NN пуне прецизности (енг. *full precision*). Утицај скаларне квантизације на перформансе квантоване NN је разматран у бројним радовима [40−54]. Када се квантизација изводи користећи 5 и више битова [40−43], показано је да квантована NN има занемарљиво мање перформансе у односу на NN пуне прецизности. Са смањењем резолуције, нпр. коришћењем 4 бита [44], 2 бита [45−47], тернарне (три нивоа) [48] или чак бинарне квантизације у екстремном

случају [49−54], примећена је одређена деградација перформанси али је остварен знатно већи степен компресије.

У овој дисертацији разматрано је унапређење постојећих модела квантизера у односу на стандардно коришћене квантизере у смислу смањења комплексности и скраћивања времена процесирања, што је изузетно важно за реалне апликације. Анализиран је већи број различитих модела квантизера (униформних и неуниформних) који користе кодове са променљивом и са фиксном дужином кодних речи, при чему је акценат већи на нискорезолуционој квантизацији (мали број нивоа) и њеној примени. Теоријски дизајн се искључиво заснива на статистичкој расподели сигнала која може бити ограничена или неограничена, а углавном се користе Лапласова и Гаусова расподела. Постоје бројна истраживања која показују да Лапласова расподела добро апроксимира дугорочну статистику говора, док је Гаусова расподела адекватнија за краткорочну статистику говора [1, 8, 55, 56]. Такође, ове расподеле се могу успешно користити и за статистичко моделовање параметара NN-а [57].

Допринос ове дисертације у области кодовања говора огледа се у развоју нових енкодера таласног облика (решења ниске комплексности), која имају за циљ да побољшају перформансе (у смислу квалитета реконструисаног сигнала и компресије) постојећих напреднијих решења сличне комплексности. Предложена решења добијена су као резултат имплементације нових модела квантизера у непредиктивним (PCM) и предиктивним (ADM) алгоритмима. Код оваквих система, квантизацијом се остварује компромис између економичног коришћења расположиве меморије и пропусног опсега са једне стране и што тачније реконструкције сигнала на пријему са друге стране. Такође, код ADM-а су имплементирани и нови модели предиктора који могу у значајној мери поправити перформансе стандардних типова предиктора уз мање трошкове преноса. Сви развијени енкодери су тестирани на сигналима који су екстраховани из познатих база података, а извршена је и упоредна анализа са решењима доступним у литератури при чему се користе стандардне објективне мере. Осим тога, у циљу верификације исправности предложених решења, испитивано је слагање између теоријских перформанси и перформанси постигнутих на реалном говору.

У оквиру дисертације су разматрана и решења за компресију NN-а. Квантизацијом се остварује компромис између перформанси (тачност предикције) и величине NN-а. Допринос дисертације се заснива на развоју и анализи нових модела квантизера који се од већине постојећих разликују у начину пројектовања (узимају у обзир статистику параметара NN-а). Исто тако, анализиране су перформансе квантоване NN добијене применом нових квантизера у односу на случајеве када се користе претходно развијени модели, а утврђивана је и веза између тачности предикције и SQNR-а.

Ова дисертација је сачињена од седам поглавља која су даље подељена на одељке и пододељке. Прво поглавље дисертације је намењено уводном делу. У другом поглављу дата је теоријска позадина скаларне квантизације односно дефинисан је скаларни квантизер и уведене су мере за процену перформанси квантизера, затим су представљени критеријуми који се користе при пројектовању а дата је и општа подела скаларних квантизера као и њихов кратак опис.

У трећем поглављу анализирани су Гаусови квантизери са променљивом дужином кодних речи, који имају мали број нивоа. Кодове са променљивом дужином кодних речи (односно ентропијске кодове) је пожељно користити, јер доприносе смањењу битске брзине и њеном приближавању ентропији којом је дефинисана доња граница за битску брзину (прва Шенонова теорема) [1, 2, 8−10]. Код ових кодова се мање вероватним симболима (нивоима) додељују краће кодне речи, док се дуже кодне речи додељују више вероватним симболима. Иако је развијен већи број ентропијских кодова, за примену код квантизера са малим бројем нивоа најпогоднији је Хафманов код [1, 2, 8−10, 58−62], јер може да омогући компактну (оптималну) дужину кодних речи. Поред класичног Хафмановог кода користе се и његове модификације, нпр. проширени Хафманов код кога се кодују блокови симбола [8−10]. У оквиру овог поглавља остварени су следећи доприноси: извршено је пројектовање асиметричног бинарног квантизера са проширеним Хафмановим кодом који у односу на (симетрични) Лојд-Макс бинарни квантизер има боље перформансе [63]; разматран је тернарни квантизер са Хафмановим кодом, при чему су предложена два нова метода за пројектовање који имају мању рачунску комплексност од Лојд-Макс алгоритма [64]; анализиран је утицај неприлагођења на варијансу у случају тернарне квантизације са

Хафмановим кодом и изведени су јако тачни апроксимативни изрази у затвореном облику за процену перформанси [64].

Четврто поглавље посвећено је PCM техници за кодовање говора где је предложено више нових решења како при мањим тако и при већим битским брзинама. Сва PCM решења реализована су применом технике адаптације унапред, а теоријско пројектовање квантизера је извршено и за неограничен и за ограничен Лапласов извор. Такође, развијени су и двомодни квантизери који алтернативно користе квантизере пројектоване за ограничен и неограничен извор. Доприноси овог поглавља су: изложен је нови неитеративни метод пројектовања оптималног бинарног квантизера и изведени су изрази за процену перформанси [65]; предложен је униформни квантизер са мртвом зоном са пет нивоа чији се излази кодују Хафмановим кодом а који даје боље перформансе од класичног униформног и неунифромног Лојд-Макс квантизера са истим бројем нивоа [66]; разматран је нови двомодни квантизер са уграђеним G.711 квантизером за високо-квалитетно кодовање сигнала који има способност да при нижој битској брзини оствари исти квалитет сигнала као и стандардизовани G.711 квантизер [67]; анализиран је и нови двомодни квантизер за компресију ограниченог Лапласовог извора који се базира на тернарним квантизерима са Хафмановим кодом [68].

У петом поглављу дисертације разматрани су нови алгоритми на бази ADM-а за кодовање говора. Развијена решења унапређују перформансе претходно реализованих вишебитних ADM система без уношења додатне комплексности, што је остварено поправљањем перформанси квантизера и предиктора који чине основне градивне блокове ових система. Користи се нови начин адаптације на нивоу фрејма, где се ажурирање параметара квантизера и предиктора врши једном по фрејму. За потребе ADM-а пројектован је већи број квантизера (и за Лапласов и за Гаусов извор) од којих већи удео чине квантизери са мањим бројем нивоа, а такође анализирани су и неки нови модели предиктора. Допринос овог поглавља огледа се у следећем: предложена је тернарна ADM (користи нови модел Лапласовог тернарног неуниформног квантизера и нови модел прекидачког линеарног предиктора првог реда) која даје веће перформансе од базичног ADM-а [69]; предложена је двобитна ADM (користи нови модел Лапласовог двобитног неуниформног квантизера и линеарни предиктор првог реда) која остварује

значајно боље перформансе у односу на своје еквиваленте из класе тренутно адаптивне ADM [70]; предложен је дводигитни ADM (користи нови модел Лапласовог неунифромног квантизера са шест нивоа и линеарни предиктор првог реда) чије су перформансе супериорније од дводигитног решења из класе тренутно адаптивне ADM [71]; предложен је нови вишенивовски ADM са G.711 квантизером који значајно поправља перформансе G.711 квантизера [72]; разматрана је још једна тернарна ADM (користи нови модел Гаусовог тернарног неуниформног квантизера и прекидачки предиктор другог реда) која представља ефикасније решење од тренутно адаптивног двобитног ADM-а [73]; разматран је двобитни ADM (користи нови модел Гаусовог двобитног униформног квантизера и нови тип предиктора ткзв. фракциони линеарни предиктор) а који је супериорнији од постојећих двобитних ADM алгоритама [74];

У шестом поглављу су представљена решења за компресију NN-а. Развијени квантизери се заснивају на статистици параметара NN-а који се квантују, за разлику од већине досадашњих решења која су субоптимална у том погледу (један од главних разлога за појаву деградације перформасни код квантоване NN). Намењени су за *post-training* квантизацију тежина односно за квантовање тежина обучене NN (од свих параметара NN-а удео тежина је највећи, па се из тог разлога ради компресија) а подаци за обучавање и тестирање NN-а се узимају из референтних база података. Теоријско пројектовање је урађено за Лапласову расподелу. Доприноси овог поглавља су: разматран је нови тип бинарног квантизера и извршено је побољшање перформанси и адаптација модела, а квантована NN остварује боље перформансе него када се користи иницијални модел [65]; предложен је нови двобитни униформни квантизер и урађена је адаптација, а квантована NN остварује веће перформансе него у случају примене других двобитних модела [75]; предложен је двобитни логаритамски компандинг квантизер који је бољи кандидат за квантовање тежина од најчешће коришћеног униформног квантизера а може се користити и као алтернатива адаптивним квантизерима [76].

Најзначајнији доприноси дисертације су наведени у закључку који је дат у седмом поглављу, а иза тога дат је списак коришћене литературе.

2. Теоријске основе скаларне квантизације

У овом поглављу биће представљене теоријске основе скаларне квантизације. Најпре ће бити дефинисани скаларни квантизер и одговарајуће евалуационе метрике као и два типа расподеле сигнала који су од интереса за ову дисертацију. Након тога ће бити представљени критеријуми који се користе при теоријском пројектовању квантизера. Биће речи и о типовима скаларних квантизера који се могу наћи у литератури.

2.1. Дефиниција скаларног квантизера и евалуационе метрике

Скаларни квантизер Q са нивоа N је одређен параметрима $t_0, t_1, ..., t_N$, који се називају праговима одлуке тако да је:

$$t_0 < t_1 < t_2 < \ldots < t_N \ , \tag{2.1}$$

где је $t_0 = -\infty$ и $t_N = +\infty$ и $\{t_0, t_1, ..., t_N\} \in$ и репрезентационим нивоима $Y = \{y_0, y_1, ..., y_N\} \in$ тако да је:

$$y_1 < y_2 < \ldots < y_N \ , \tag{2.2}$$

при чему $|Y| = N$ дефинише кодну књигу квантизера [1−12]. Квантизационе ћелије α_i, $i = 1,\ldots, N$, се изводе из прагова одлуке као $\alpha_i = (t_{i-1}, t_i]$, а свака ћелија α_i се представља репрезентационим нивоом $y_i \in \alpha_i$. Интервал $|x| \leq t_{N-1}$ (t_{N-1} се често означава са $t_{\max}$, а познат је под називом максимална амплитуда квантизера) се означава као грануларни регион а $\alpha_2,\ldots, \alpha_{N-1}$ су грануларне ћелије, док се интервал $|x| > t_{N-1}$ означава као регион прекорачења а α_1 и α_N су ћелије прекорачења. Ако вредност улазног сигнала x припада ћелији α_i онда се та вредност квантује нивоом y_i. Стога се скаларни квантизер може посматрати као функција Q: $R \rightarrow Y$ која мапира реалну вредност x у ниво y_i, односно $Q(x) = y_i$ ако и само ако је $x \in \alpha_i$.

Величине које се користе за процену перформанси квантизера јесу дисторзија D и битска брзина R [1−12]. Дисторзијом се мери грешка настала у процесу квантизације и састоји се из грануларне дисторзије D_g и дисторзије прекорачења D_o, тј. $D = D_g + D_o$. Ове две компоненте дисторзије одређују се помоћу следећих израза [1−3, 8, 11, 12]:

$$D_g = \sum_{i=2}^{N-1} \int_{t_{i-1}}^{t_i} \left(x - y_i\right)^2 p\left(x\right) dx \,, \tag{2.3}$$

$$D_o = \int_{-\infty}^{t_1} \left(x - y_1\right)^2 p\left(x\right) dx + \int_{t_N}^{\infty} \left(x - y_N\right)^2 p\left(x\right) dx \,, \tag{2.4}$$

где је $p(x)$ функција густине вероватноће (PDF) улазног сигнала. У дисертацији ће се доминантно користити Гаусова и Лапласова PDF са нултом средњом вредношћу и варијансом σ^2, које су дате изразима (2.5) и (2.6), респективно:

$$p\left(x,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma}\right\}, \tag{2.5}$$

$$p\left(x,\sigma\right) = \frac{1}{\sigma\sqrt{2}} \exp\left\{-\frac{\sqrt{2}\left|x\right|}{\sigma}\right\}. \tag{2.6}$$

Како последња два израза дефинишу парне и неограничене функције, то значи да ће бити разматрани симетрични квантизери.

SQNR, који заправо представља најчешће коришћену објективну меру перформанси, се дефинише као [1−3, 8, 11, 12]:

$$\mathrm{SQNR} = 10\log_{10}\left(\frac{\sigma^2}{D}\right), \tag{2.7}$$

и изражава се у децибелима (dB).

На нивое квантизера се могу имплементирати кодови са фиксном или променљивом дужином кодних речи. Када се користе кодови са фиксном дужином кодних речи и ако је N степен броја 2, тада се битска брзина дефинише као [1, 8−10]:

$$R = \log_2 N \text{ [bps]}, \tag{2.8}$$

иначе се одређује као:

$$R = \lceil \log_2 N \rceil \text{ [bps]}, \tag{2.9}$$

где је $\lceil x \rceil$ најближи цео број већи од x.

У случају примене кода са променљивом дужином кодних речи, битска брзина се одређује по следећој формули [1, 8−10]:

$$R = \sum_{i=1}^{N} l_i p(y_i) \text{ [bps]}, \tag{2.10}$$

где l_i и $p(y_i)$ означавају дужине кодних речи односно вероватноће које одговарају нивоима y_i, $i = 1, \ldots, N$.

2.2. Критеријуми при пројектовању скаларних квантизера

Проблем теоријског пројектовања квантизера је специфичан изазов, а зависи како од PDF сигнала који се квантује тако и од тога да ли је потребно оптимизовати перформансе за специфичну (односно референтну) варијансу или за опсег варијанси.

Када се квантизери пројектују за референтну варијансу, потребно је одабрати (оптимизовати) параметре квантизера (прагове одлуке и нивое) тако да буду испуњени унапред специфицирани критеријуми [1, 2]. Два најчешће коришћена критеријума за оптимизацију су:

1. оптимизација по дисторзији
2. оптимизација и по дисторзији и по брзини.

Оптимизацијом по дисторзији добија се дисторзијом-ограничен квантизер, код кога се прагови одлуке t_i и нивои y_i одређују тако да укупна дисторзија квантизера буде минимална односно SQNR максималан. Овакав тип оптимизације се

најчешће примењује када се користе кодови са фиксном дужином кодних речи, мада се примењује и код кодова са промењљивом дужином кодних речи.

Са друге стране, оптимизацијом и по дисторзији и по брзини добија се брзином-ограничен квантизер, код кога је циљ да се за дату битску брзину оствари што је могуће већи SQNR. Овај метод оптимизације се искључиво користи приликом примене кодова са промењљивом дужином кодних речи, а који имају за циљ да приближе битску брзину ентропији. Као што је поменуто, ентропијом је дефинисана доња граница за битску брзину а израчунава се као [1−3, 8−10]:

$$H = -\sum_{i=1}^{N} p(y_i)\log_2 p(y_i) \text{ [bps]}. \tag{2.11}$$

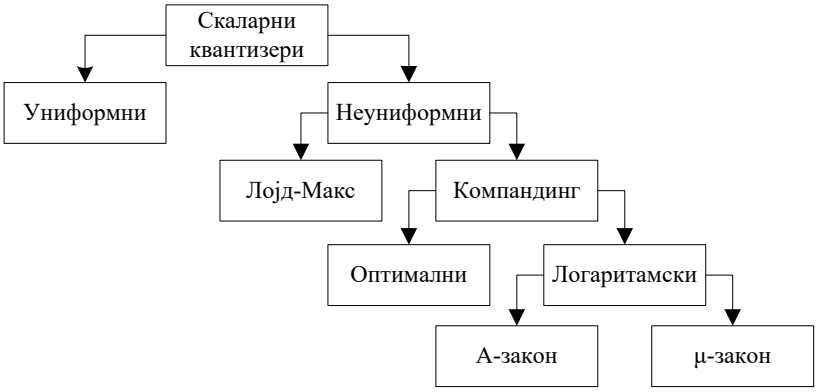
Квантизери пројектовани за једну варијансу се као такви најчешће не примењују при процесирању сигнала код којих се варијанса мења (нестационарни сигнали), јер могу бити недовољно ефикасни услед ниског нивоа робусности. Да би се поправила робусност квантизера и обезбедиле високе перформансе (SQNR) у опсегу варијанси од интереса углавном се користе два метода. Код првог метода ради се додатна оптимизација параметара квантизера, при чему се као мера перформанси користи максимални средњи SQNR дефинисан са [1, 2]:

$$\text{SQNR}_{av} = \frac{1}{m}\sum_{i=1}^{m}\text{SQNR}(\sigma_i), \tag{2.12}$$

где је m број појединачних варијанси σ_i из опсега у коме се врши оптимизација. Други метод, који је знатно ефикаснији али и сложенији од претходно поменутог метода, подразумева адаптацију квантизера о чему ће бити речи у наредним поглављима.

2.3. Врсте скаларних квантизера

Скаларне квантизере је могуће класификовати на начин како је приказано на слици 2.1.



Сл. 2.1 Основна подела скаларних квантизера.

Униформни квантизер. Униформни квантизер је најпростији тип скаларног квантизера и најчешће се користи када је једноставност апликације један од примарних циљева. Овој теми су посвећена обимна истраживања [1−3, 8, 77−80], где су током анализе посматрани различити аспекти и изведени вредни закључци. Код овог модела квантизера, грануларни регион [$-t_{\max}$, $t_{\max}$] је подељен на N еквидистантних ћелија а нивои се налазе на средини ћелија. Ширина ћелије односно корак квантизације Δ је дат изразом [1−3, 8]:

$$\Delta = \frac{2t_{\max}}{N}, \tag{2.13}$$

док се прагови одлуке и нивои рачунају на основу следећа два израза, респективно:

$$t_i = -t_{\max} + i\Delta,\ i = 0,...,N, \tag{2.14}$$

$$y_i = -t_{\max} + \left(i - \frac{1}{2}\right)\Delta,\ i = 1,...,N. \tag{2.15}$$

Из последња два израза се може закључити да је униформни квантизер са N нивоа у потпуности дефинисан са Δ (односно са $t_{\max}$) па се пројектовање овог квантизера своди на одређивање оптималне вредности за Δ. Униформни квантизер је пожељно користити за сигнале са униформном расподелом и тада се за $t_{\max}$ усваја вредност максималне вредности сигнала (дисторзија прекорачења тада не постоји). Осим тога, униформни квантизер се због своје једноставности и солидних перформанси може користити и за процесирање сигнала са неком

неуниформном (неограниченом и ограниченом) расподелом (тада се дисторзија прекорачења узима у обзир).

Неуниформни квантизер. Неуниформни квантизер је први избор за сигнале са неуниформном расподелом и стога је овом моделу посвећена значајна пажња у литератури [1−3, 8, 81, 82]. Параметри овог квантизера се одређују у зависности од PDF сигнала, што имплицира да квантизационе ћелије нису еквидистантне односно да се нивои не налазе на средини ћелија. Дакле, неуниформним квантизером је извршена подела реалне осе тако да се вредности сигнала које се чешће јављају квантују финије од вредности које се јављају ређе. Неуниформни квантизер је могуће пројектовати на два начина: Лојд-Макс алгоритмом и компандинг техником.

Лојд-Макс алгоритам. Лојд-Макс алгоритам је познати итеративни метод за пројектовање дисторзијом-ограничених квантизера, када је позната PDF сигнала. Пројектовање се изводи кроз следеће кораке [1−3, 8, 18, 83]:

1. Иницијализација прагова одлуке $t_i^{(0)}$, $i = 1\ldots, N\text{-}1$ ($t_0 = -\infty$ и $t_N = \infty$) и нивоа $y_i^{(0)}$, $i = 1\ldots, N$. На основу ових параметара процењује се почетна вредност дисторзије $D^{(0)}$.
2. Нове вредности прагова одлуке и нивоа се одређују као:

$$y_i^{(j+1)} = \frac{\int_{x_{i-1}^{(j)}}^{x_i^{(j)}} xp(x)dx}{\int_{x_{i-1}^{(j)}}^{x_i^{(j)}} p(x)dx}, \quad j = 0,1,\ldots , \tag{2.16}$$

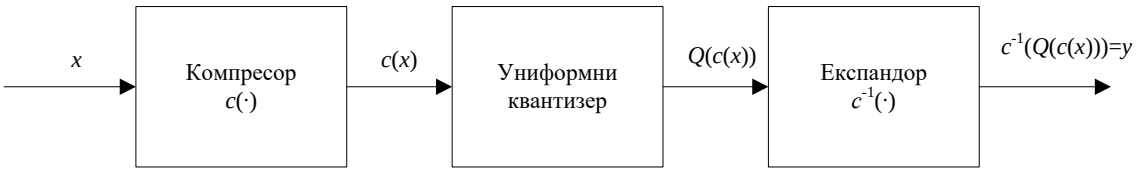
$$x_i^{(j+1)} = \frac{y_i^{(j+1)} + y_{i-1}^{(j+1)}}{2}, \quad j = 0,1,\ldots . \tag{2.17}$$

   Последња два израза су позната као правило центроида и правило најближег суседа, респективно [1−3, 8].
3. Алгоритам се прекида када је следећа неједнакост испуњена:

$$\frac{\left|D^{(j+1)} - D^{(j)}\right|}{D^{(j)}} \le \varepsilon = 0.005 \ . \tag{2.18}$$

Препорука је да се Лојд-Макс алгоритам користи за пројектовање квантизера са мањим бројем нивоа (нпр. $N \le 32$), јер за већи број нивоа постаје временски захтеван.

Компандинг техника. Као алтернатива Лојд-Макс алгоритму за пројектовање неуниформних квантизера са великим бројем нивоа користи се компандинг техника. Квантизери реализовани овом техником називају се компандинг квантизери. Градивни блокови сваког компандинг квантизера су компресор, униформни квантизер и експандор (слика 2.2) [1−3]. Компресор врши трансформацију вредности сигнала x компресорском функцијом $c(\cdot)$ која је нелинеарна (појачава знатно више мале вредности сигнала које се јављају са већом вероватноћом од већих вредности сигнала), а на чијем излазу се добија сигнал $c(x)$. $c(x)$ се униформно квантује и добија се квантовани сигнал $Q(c(x))$ који се затим доводи на улаз експандора. Експандор, након примене инверзне компресорске функције $c^{-1}(\cdot)$ над квантованим сигналом, генерише на свом излазу реконструисани сигнал $y = c^{-1}(Q(c(x)))$.



Сл. 2.2. Структура компандинг квантизера.

Пројектовање компандинг квантизера се углавном своди на одређивање компресорске функције. У зависности од типа примењене компресорске функције компандинг квантизери се деле на оптималне [16, 84] и логаритамске компандинг квантизере [85−88]. Основна предност оптималних компандинг квантизера је та што за референтну варијансу могу да дају перформансе (SQNR) које су изузетно блиске Лојд-Макс квантизеру, а главни недостатак је мала робусност. С друге стране, логаритамски компандинг квантизери, за референтну варијансу, имају нешто мањи SQNR од оптималних компандинг квантизера, али су зато изузетно робустни. Логаритамски компандинг квантизери се широко користе у кодовању говора, а стандардизоване су две верзије: са A законом компресије и μ законом компресије [1−3, 89]. У оквиру ове дисертације разматраће се неколико решења на бази компресорске функције са μ законом, која је дата са [1]:

$$c(x) = \frac{t_{\max}}{\ln(1+\mu)} \ln\left(1+\mu\frac{|x|}{t_{\max}}\right)\operatorname{sgn}(x), \quad 0 \le x \le t_{\max}\,, \tag{2.19}$$

где је μ фактор компресије (μ = 255 код америчког стандарда за PCM), а $t_{\max}$ одговара максималној амплитуди квантизера.

3. Пројектовање Гаусових скаларних квантизера са променљивом дужином кодних речи за мале битске брзине

У овом поглављу биће представљена два решења за компресију Гаусовог извора, која се заснивају на примени Хафмановог кода као и проширеног Хафмановог кода. Прво ће бити разматран асиметрични бинарни квантизер са проширеним Хафмановим кодом. До сада се радило пројектовање симетричног бинарног квантизера, док асиметрични модел треба да понуди мању битску брзину уз малу деградацију у SQNR-у. Разматраће се и тернарни квантизер са Хафмановим кодом, за чије пројектовање ће бити предложена два нова метода, итеративни и неитеративни, а показаће се да су рачунски су мање интензивни од Лојд-Макс алгоритма. Неитеративни метод заснива се на апроксимацији *Q*-функције и представља важан део овог поглавља, за кога ће бити показано да даје перформансе јако блиске Лојд-Макс алгоритму. На крају овог поглавља биће изложена анализа тернарног квантизера са Хафмановим кодом у присуству неприлагођења на варијансу, што је јако важно с обзиром на то да се овај ефекат често среће у пракси. За процену перформанси у овом случају биће изведени јако тачни изрази на бази апроксимације *Q*-функције.

3.1 Хафманов код

Хафманов код је развијен од стране Дејвида Хафмана (енг. David Huffman) и један је од најчешће коришћених кодова из класе ентропијских кодова [8−10, 58−62]. Хафманов код је заправо оптимални префиксни код, а могуће га је применити и за кодовање нивоа квантизера (за редукцију битске брзине). Да би Хафманов код могао да се имплементира неопходно је познавати вероватноће нивоа (симбола). Нека $p(y_i)$ означава вероватноћу нивоа y_i, $i = 1, 2,..., N$. За претпостављену $p(x)$, вероватноће нивоа се израчунавају на следећи начин [1, 8]:

$$p(y_1) = \int_{-\infty}^{t_1} p(x)dx\,,\ p(y_i) = \int_{t_{i-1}}^{t_i} p(x)dx,\ i = 2,...,N-1,\ p(y_N) = \int_{t_{N-1}}^{\infty} p(x)dx\,. \quad (3.1)$$

Код овог кода, поступак одређивања дужине кодних речи и конструкција кодног стабла се укратко може описати помоћу следећих корака [8−10]:

1. Сортирати вероватноће нивоа у опадајућем поретку, а вероватноће посматрати као чворове.
2. Спровести итеративни поступак. У свакој итерацији се проналазе два чвора са најмањим вероватноћама и повезују се у нови чвор чија је вероватноћа једнака збиру вероватноћа одабрана два чвора. Поступак се наставља све док се не повежу два чвора у нови чвор чија вероватноћа износи 1.
3. Одређивање кодних речи. Кодна реч за сваки ниво се одређује полазећи од корена стабла (чвор са вероватноћом 1) и грана које се стичу у њему, при чему се горњој грани додељује 0 а доњој грани 1. Процес додељивања се наставља улево док се не покрију све могуће гране. Кодна реч се формира од нула и јединица које се налазе на путу од корена стабла до чвора који одговара одређеном нивоу.

Када се знају вероватноће $p(y_i)$ и дужине кодних речи l_i онда се користи израз (2.10) да би се израчунала Хафманова битска брзина.

3.2 Проширени Хафманов код

Проширени Хафманов код се односи на технику кодовања којом се одређују оптималне дужине кодних речи за блокове од два или више симбола [8−10]. Наиме, груписањем симбола у блокове могуће је у неким случајевима додатно смањити битску брзину и још више је приближити ентропији извора. Нека је N број симбола дискретног извора а S величина блока. Број симбола проширеног извора односно број блокова износи N^S, а њихове вероватноће се одређују као [8]:

$$p_{i,j,\ldots,k} = p(y_i)\,p(y_j)\ldots p(y_k), i = 1,\ldots,N,\ j = 1,\ldots,N,\ldots,k = 1,\ldots,N \quad . \tag{3.2}$$

Када су вероватноће познате, спроводи се исти поступак за одређивање дужине кодних речи и конструкцију кодне књиге као у одељку 3.1. Средња битска брзина се у овом случају рачуна као [8]:

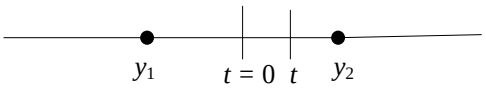
$$\bar{R} = \sum_{i=1}^{3}\sum_{j=1}^{3}\cdots\sum_{k=1}^{3} p_{i,j,\ldots k} l_{i,j,\ldots k} \,, \tag{3.3}$$

где је $l_{ij,..,k}$ кодна реч за блок чија је вероватноћа $p_{ij,..,k}$, а број битова по симболу је дат са [8]:

$$R = \frac{\bar{R}}{S} \,. \tag{3.4}$$

3.3 Асиметрични бинарни квантизер са проширеним Хафмановим кодом

Асиметрични бинарни квантизер који се овде разматра предложен је у раду [63]. На слици 3.1 је дат приказ овог модела, а дефинисан је са три параметра: променљивим прагом одлуке t и нивоима y_1 и y_2.



Сл. 3.1 Асиметрични бинарни квантизер.

Нивои квантизера одређују се из услова центроида (израз (2.16)). За Гаусову PDF (израз (2.5)) и $\sigma^2 = \sigma_{\text{ref}}^2 = 1$ добија се:

$$y_1 = \frac{\int_{-\infty}^{t} x^2 p(x)\,dx}{\int_{-\infty}^{t} p(x)\,dx} = \frac{1/2 - \frac{t}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\} + Q(t)}{1 - Q(t)} \,, \tag{3.5}$$

$$y_2 = \frac{\int_{t}^{\infty} x^2 p(x)\,dx}{\int_{t}^{\infty} p(x)\,dx} = \frac{t}{\sqrt{2\pi}}\frac{\exp\left\{-\frac{t^2}{2}\right\}}{Q(t)} + 1, \tag{3.6}$$

а са $Q(\cdot)$ означена је Гаусова Q-функција дата са [90−101]:

$$Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp\left\{-\frac{v^2}{2}\right\}dv\,. \tag{3.7}$$

Из израза (3.5) и (3.6) је јасно да нивои квантизера нису симетрични.

Дисторзија овог модела квантизера се рачуна по формули:

$$\begin{aligned} D&=\int_{-\infty}^{t}(x-y_1)^2p(x)dx+\int_t^{\infty}(x-y_2)^2p(x)dx\\ &=1-\exp\{-t^2\}\left(\frac{1}{1-Q(t)}-\frac{1}{Q(t)}\right) \end{aligned}\quad , \tag{3.8}$$

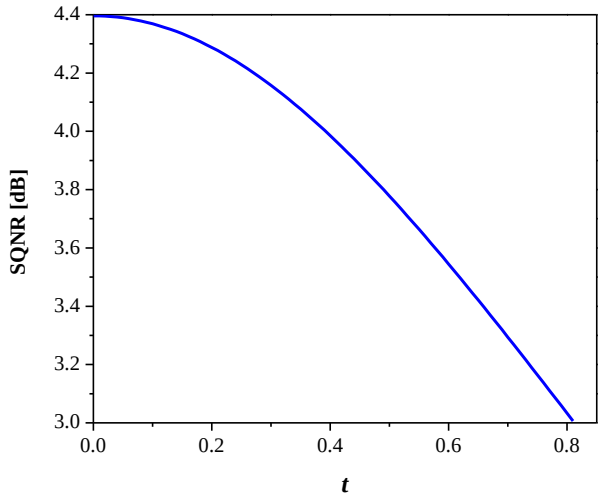
док је SQNR = $-10\log_{10}D$.

На квантизер се примењује проширени Хафманов код [8]. Вероватноће нивоа (симбола) се за Гаусову PDF одређују као:

$$p(y_1)=\int_{-\infty}^{t}p(x)dx=1-Q(t)\,, \tag{3.9}$$

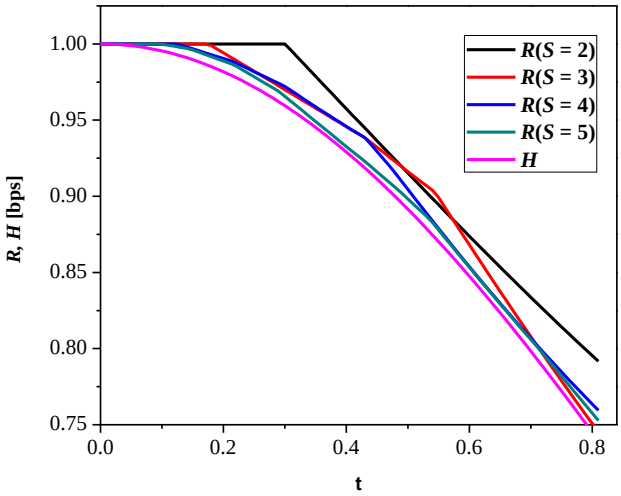
$$p(y_2)=\int_t^{\infty}p(x)dx=Q(t)\,. \tag{3.10}$$

Како је $N = 2$, онда је број блокова 2^S. Вероватноће ових блокова се одређују изразом (3.2), а $\overline{R}$ и R се рачунају на основу (3.3) и (3.4), респективно.



Сл. 3.2 Дефинисање опсега у коме се врши избор оптималне вредности за t.

Изрази (3.8)−(3.10) показују да перформансе зависе од t. За $t = 0$ добија се (симетрични) бинарни Лојд-Макс квантизер (SQNR = 4.44 dB и R = 1 bps) [1−3, 8]. Међутим, овде је циљ да се за дато S (величина блока) изврши оптимизација асиметричног бинарног квантизера и по дисторзији и по брзини. Оптимало t се у овом случају тражи у опсегу $t \in (0, 0.81)$, што одговара сценарију када оптимални SQNR (4.44 dB) опадне до вредности 3.01 dB (SQNR за бинарни Лојд-Макс квантизер за Лапласов извор), као што је приказано на слици 3.2.



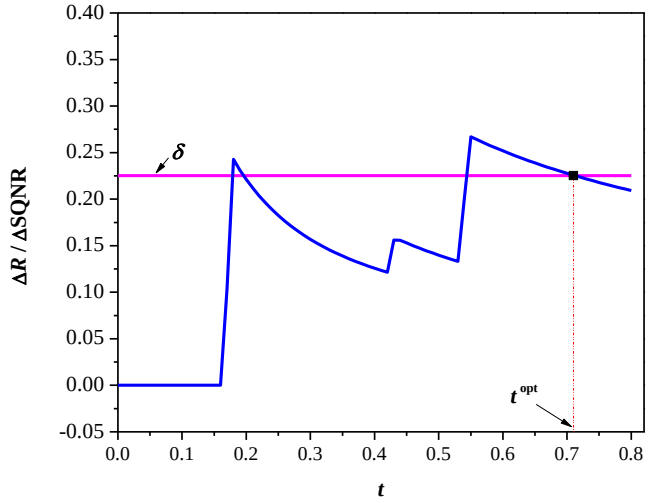
Сл. 3.3 Ентропија H и битска брзина R (израчуната за неколико вредности S) у функцији од t.

На слици 3.3 су приказане вредности за ентропију H и битску брзину R (за блокове од два, три, четири и пет симбола) израчунате за различите вредности t (опсег за t је исти као на слици 3.2). Са ове слике се види да је проширеним Хафмановим кодом могуће R приближити ка H, а најбољи резултати добијени су за S = 5 али је тада комплексност кодера већа у односу на остале разматране случајеве. Због тога се, узимајући у обзир постигнуте вредности R (у односу на H) и комплексност проширеног Хафмановог кодера, као компромисно решење усваја S = 3.

За дато S, оптимално t се бира на основу следећег критеријума:

$$\frac{\Delta R}{\Delta \mathrm{SQNR}} \geq \frac{\Delta R^{e}}{\Delta \mathrm{SQNR}^{e}} = \delta, \tag{3.11}$$

где је $\Delta R/\Delta$SQNR нагиб криве R(SQNR) који се тражи између две узастопне вредности t ($t \in (0, 0.81)$) а $\Delta R^e/\Delta\text{SQNR}^e$ је теоријски очекивана вредност нагиба која се у овом случају тражи између Лојд-Макс квантизера са $N = 0$ (SQNR = 0 dB, $R = 0$ bps) и $N = 2$ (SQNR = 4.44 dB, $R = 1$ bps) нивоа ($\delta = 0.2252$).



Сл. 3.4 Одређивање оптималне вредности за t на основу критеријума (3.11) ($S = 3$).

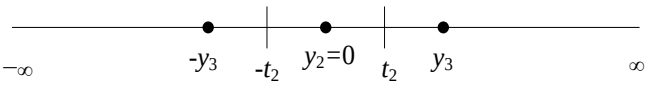
На слици 3.4 дато је $\Delta R/\Delta$SQNR у функцији од t. Као што се може уочити, критеријум (3.11) је испуњен за неколико вредности t (видети тачке пресека са δ) али се за оптималну вредност усваја $t = t^{\text{opt}} = 0.71$ јер је тада битска брзина мања него у осталим тачкама пресека (слика 3.3).

Дакле, предложеним асиметричним моделом ($t = 0.71$, SQNR = 3.3 dB, $R = 0.81$ bps) остварена је уштеда у битској брзини са малим губитком у SQNR-у у поређењу са бинарним Лојд-Макс квантизером.

3.4 Неуниформни тернарни квантизер са Хафмановим кодом

У овом одељку описано је решење на бази тернарне квантизације које је предложено у раду [64], а представиће се два нова метода за пројектовање дисторзијом-ограниченог квантизера: итеративни и неитеративни.

Пројектовање итеративним методом. Симетрични тернарни квантизер је приказан на слици 3.5, а у потпуности је дефинисан прагом одлуке t_2 и нивоом y_3. Дакле, потребно је одредити вредности наведена два параметра тако да се оствари најмања могућа вредност дисторзије. Заправо, показаће се да је t_2 кључни параметар при пројектовању, па се предложени итеративни метод заснива на одређивању оптималне вредности за t_2.



Сл. 3.5 Симетрични тернарни квантизер.

Нивои тернарног квантизера кодују се Хафановим кодом. С тим у вези, вероватноће нивоа морају бити познате. За Гаусову PDF (израз (2.5)) вероватноће се рачунају као:

$$p(-y_3)=p(y_3)=\int_{t_2}^{\infty}p(x)dx=Q(t_2), \tag{3.12}$$

$$p(y_2)=2\int_{0}^{t_2}p(x)dx=1-2Q(t_2), \tag{3.13}$$

док се Хафманова битска брзина рачуна према изразу (2.10).

Дисторзија тернарног квантизера је дата са:

$$D=2\int_{0}^{t_2}x^2p(x)dx+2\int_{t_2}^{\infty}(x-y_3)^2p(x)dx, \tag{3.14}$$

а може да се напише и у следећем облику:

$$D=2\int_{0}^{t_2}x^2p(x)dx+2\int_{t_2}^{\infty}x^2p(x)dx-4y_3\int_{t_2}^{\infty}xp(x)dx+2y_3^2\int_{t_2}^{\infty}p(x)dx. \tag{3.15}$$

Ако се y_3 одреди из услова центроида (израз (2.16)):

$$y_3 = \int_{t_2}^{\infty} xp(x)dx \Big/ \int_{t_2}^{\infty} p(x)dx \ , \tag{3.16}$$

онда важи следећа једнакост:

$$y_3 \int_{t_2}^{\infty} p(x)dx = \int_{t_2}^{\infty} xp(x)dx \, . \tag{3.17}$$

Користећи (3.17) и имајући у виду да је $2\int_0^{t_2} x^2 p(x)dx + 2\int_{t_2}^{\infty} x^2 p(x)dx = \sigma_{\text{ref}}^2 = 1$, израз (3.15) након сређивања постаје:

$$D = 1 - 2y_3^2 \int_{t_2}^{\infty} p(x)dx = 1 - 2y_3^2 Q(t_2) \, . \tag{3.18}$$

Из (3.16) се у случају Гаусове PDF за y_3 добија:

$$y_3 = \exp\left\{-\frac{t_2^2}{2}\right\} \Big/ Q(t_2), \tag{3.19}$$

а применом у (3.18) долази се до следећег израза за дисторзију:

$$D = 1 - \frac{1}{\pi} \frac{\exp\left\{-t_2^2\right\}}{Q(t_2)} \, . \tag{3.20}$$

Из (3.20) се лако уочава да на D утиче само t_2. Оптимално t_2 је дефинисано следећом лемом.

Лема 3.1. За дисторзијом-ограничен тернарни Гаусов квантизер t_2 се може одредити на основу следећег итеративног правила:

$$t_2^{(i)} = \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{\left(t_2^{(i-1)}\right)^2}{2}\right\} \Big/ Q\left(t_2^{(i-1)}\right). \tag{3.21}$$

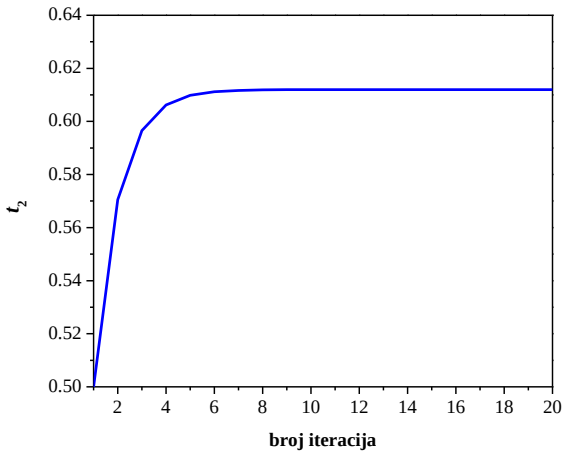
Доказ. Из услова $\partial D / \partial t_2 = 0$ добија се следећа интегрална једначина:

$$2t_2 Q(t_2) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t_2^2}{2}\right\}. \tag{3.22}$$

Решавањем (3.22) по t_2 добија се:

$$t_2 = \frac{1}{2\sqrt{2\pi}} \exp\left\{-\frac{t_2^2}{2}\right\} / Q(t_2), \tag{3.23}$$

што указује на то да се оптимална вредност за t_2 може добити итеративно, чиме је лема доказана.



Сл. 3.6 Итеративни процес за одређивање t_2 дефинисан са (3.21), за $t_2^{(0)} = 0.5$.

На слици 3.6 је дато t_2 у функцији од броја итерација, када је почетна вредност подешена на $t_2^{(0)} = 0.5$. t_2 након свега неколико итерација конвергира ка вредности 0.612, а ова вредност се савршено поклапа са вредношћу у табели 3.1 која је добијена Лојд-Макс алгоритмом. То показује да је итеративни процес (3.21) јако ефикасан.

Табела 3.1. Поређење резултата добијених предложеним итеративним методом и Лојд-Макс алгоритмом

	t_2	y_3	**SQNR [dB]**	R **[bps]**
Предложени	0.612	1.224	5.78	1.54
Лојд-Макс	0.612	1.224	5.78	1.54

Предложени итеративни метод је рачунски мање интензиван од Лојд-Макс алгоритма. Израз (3.21) показује да је за тачно одређивање t_2 потребан одређени број итерација и нумерички прорачун Q-функције (типично се представља збиром експоненцијалних чланова). Под претпоставком да m означава број чланова који се користе у сумирању а l број итерација, рачунска комплексност C (број аритметичких операција) предложеног метода је $C = l \cdot m + 1$. Са друге стране, рачунска комплексност Лојд-Макс алгоритама (за проблем тернарне квантизације) је $C = l_1(m+4)$, где l_1 означава број итерација ($l_1 > l$). Поређењем се утврђује да је рачунска комплексност смањена за $l_1(m+4) / (l \cdot m+1)$ пута.

Пројектовање неитеративним методом. Дати метод се заснива на примени апроксимације Q-функције (алтернатива за Q-функцију), а има за циљ да додатно поједностави процес пројектовања. Предлаже се јако тачан апроксимативни израз за t_2 као и јако тачни апроксимативни изрази за процену перформанси (SQNR и R) тернарног квантизера.

За разматрани квантизер користи се следећа апроксимација Q-функције [92]:

$$F(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2+a}} \exp\left(-\frac{x^2}{2}\right), 1 \le a < 2\,, \tag{3.24}$$

која дефинише класу апроксимација које су горње границе за $x > \sqrt{\left(\sqrt{4a+1} + 2a^2 - 2a - 1\right)/(4-2a)}$. Главни задатак јесте да се одреди најпогоднија вредност параметра a. Следећа лема даје везу између t_2 и a.

Лема 3.2. За дисторзијом-ограничен тернарни Гаусов квантизер праг одлуке се може апроксимирати са $t_2^A = \sqrt{a/3}$.

Доказ. Да би се ово доказало користи се интегрална једначина дата изразом (3.22). Заменом $Q(x)$ са $F(x)$ добија се:

$$2t_2^A \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\left(t_2^A\right)^2+a}} \exp\left\{-\frac{\left(t_2^A\right)^2}{2}\right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\left(t_2^A\right)^2}{2}\right\}, \tag{3.25}$$

где је са t_2^{A} означена апроксимативна вредност за t_2. Сређивањем последњег израза долази се до:

$$t_2^A = \sqrt{a/3}\,, \tag{3.26}$$

што доказује лему.

Апроксимативна вредност нивоа (означена је са y_3^{A}) добија се из правила најближег суседа (2.17), што даје $y_3^{\mathrm{A}} = 2 \cdot t_2^{\mathrm{A}}$.

Предложени метод има рачунску комплексност $C = 2$. Сходно томе, рачунска комплексност је смањена за $l_1(m+4) / 2$ пута у односу на Лојд-Макс алгоритам, а добијени резултат је бољи и у односу на претходно описан итеративни метод.

Конкретна нумеричка вредност за a из израза (3.26) одређује се на основу следећег критеријума:

$$S(a_j) = \underset{a_j}{\arg\min} \left\{ \frac{1}{M} \left[\sum_{i=1}^{M} \frac{|F(x(i), a_j) - Q(x(i))|}{Q(x(i))} \right] \right\}, \tag{3.27}$$

$$a_j = 1 + \frac{j}{M_1}, j = 1, \ldots, M_1, \tag{3.28}$$

$$x(i) = x(1) + i \frac{x(M) - x(1)}{M}, i = 1, \ldots, M. \tag{3.29}$$

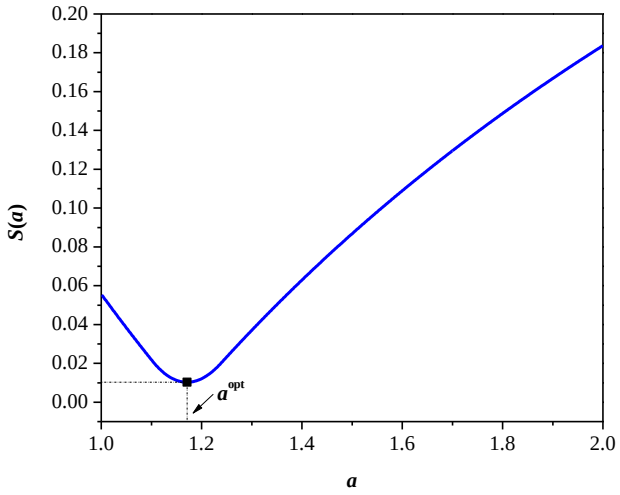
Другим речима, a се бира тако да средња релативна грешка $S(a_j)$ при апроксимацији Q-функције буде минимимална у опсегу аргумената [$x(1) = x^{\mathrm{low}}$, $x(M) = x^{\mathrm{up}}$]. Границе опсега x^{low} и x^{up} се одређују из горње [98] (означена са је $F^{[98]}(x)$) и доње границе за Q–функцију [90] (означена је са $F^{[90]}(x)$), респективно:

$$F^{[98]}(x) = \frac{(1 - \exp(-Cx)) \exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi} x}, \; C = \sqrt{\pi/2}\,, \tag{3.30}$$

$$F^{[90]}(x) = \frac{1}{\sqrt{2\pi}} \frac{x^2 + b - 1}{x(x^2 + b)} \exp\left(-\frac{x^2}{2}\right), b \in R\,. \tag{3.31}$$

$F^{[90]}(x)$ је доња граница за Q-функцију за $x > \sqrt{b(b-1)/(3-b)}$ [90]. Конкретно, доња граница $x^{\mathrm{low}} = log2/C$ добија се као решење једначине (3.22) када се уместо $Q(x)$ користи $F^{[98]}(x)$. Слично, горња граница $x^{\mathrm{up}} = \sqrt{2 - b}$, b = 1.3659 добија се као резултат решавања једначине (3.22) када се $Q(x)$ замени са $F^{[90]}(x)$.

У тако утврђеном интервалу аргумената, за $a \in [1, 2)$ израчунава се $S(a)$ (израз (3.27)). Зависност $S(a)$ од a приказана је на слици 3.7, а $S(a)$ постиже минимум за за $a = a^{\mathrm{opt}} = 1.1587$. Тиме су специфицирани t_2^A (видети лему 3.2) и y_3^A.



Сл. 3.7 Избор параметра a из апроксимације (3.24) за проблем тернарне квантизације.

Дакле, апроксимација Q-функције (3.24) се своди на следећи облик:

$$F(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x^2+1.1587}} \exp\left\{-\frac{x^2}{2}\right\}. \tag{3.32}$$

Заменом (3.26) и (3.32) у (3.20) добија се једноставан апроксимативни израз за дисторзију у функцији од a^{opt}:

$$D^A = 1 - \sqrt{\frac{8a^{\text{opt}}}{3\pi} \exp\left\{-\frac{a^{\text{opt}}}{3}\right\}}, \; a^{\text{opt}} = 1.1587, \tag{3.33}$$

док се за SQNR добија:

$$\text{SQNR}^A = 10\log_{10}\left(\frac{1}{1-\sqrt{\frac{8a^{\text{opt}}}{3\pi}\exp\left\{-\frac{a^{\text{opt}}}{3}\right\}}}\right), \; a^{\text{opt}} = 1.1587. \tag{3.34}$$

Такође, може се извести и апроксимативни израз за R. Наиме, за усвојену вредност параметра a (a = 1.1587) се уз помоћ леме 3.2 и израза (3.12) и (3.13) може показати да је $p(y_2{}^A) > p(y_3{}^A) = p(-y_3{}^A)$. То значи да Хафманов кодер нивоу $y_2{}^A$ додељује једнобитну кодну реч а преосталим нивоима додељује двобитне кодне речи. На основу (2.10), за битску брзину се добија:

$$R^A = p(y_3^A) + 4p_3(y_3^A) = 1 + 2Q(t_2^A). \tag{3.35}$$

Заменом (3.26) и (3.32) у (3.35) добија се следећи апроксимативни израз:

$$R^A = 1 + \sqrt{\frac{3}{2\pi a}\exp\left\{-\frac{a^{\text{opt}}}{3}\right\}},\ a^{\text{opt}} = 1.158. \tag{3.36}$$

Анализа теоријских резултата. Тачност изведених формула за праг одлуке (3.26), SQNR (3.34) и битску брзину (3.36) поредиће се у односу на тачне вредности ових параметара (табела 3.1) али и у односу на вредности које су добијене применом још неколико апроксимација Q-функције које имају сличан аналитички облик као и предложена апроксимација (3.32) (табела 3.2):

$$F^{[100]}(x) = \frac{\left(1-\exp\left\{-\frac{Ax}{\sqrt{2}}\right)\right)\exp\left\{-\frac{x^2}{2}\right\}}{\sqrt{2\pi}Bx},\ \text{A} = 1.98, B = 1.135, \tag{3.37}$$

$$F^{[96]}(x) = \frac{1}{\sqrt{2\pi}}\frac{\pi}{\sqrt{2\pi}+\pi x}\exp\left\{-\frac{x^2}{2}\right\}, \tag{3.38}$$

$$F^{[96]}(x) = \frac{1}{2}\exp\left\{-\frac{x^2}{2}\right\}. \tag{3.39}$$

Табела 3.2 Апроксимативне вредности за праг одлуке, дисторзију и битску брзину добијене применом различитих апроксимација Q–функције

Апроксимација	$\boldsymbol{t^A}$	$\boldsymbol{D^A}$	$\boldsymbol{R^A}$
$\boldsymbol{F^{[100]}(x)}$	$\frac{\sqrt{2}}{A}\log\left(\frac{2}{2-B}\right)$	$1-\frac{4\log\left(\frac{2}{2-B}\right)\exp\left(-\frac{\log^2\left(\frac{2}{2-B}\right)}{A^2}\right)}{A\sqrt{\pi}}$	$1 + A\frac{\exp\left(-\frac{\log^2\left(\frac{2}{2-B}\right)}{A^2}\right)}{2\sqrt{\pi}\log\left(\frac{2}{2-B}\right)}$
$\boldsymbol{F^{[98]}(x)}$	$\sqrt{2/\pi}\log(2)$	$1-\frac{4\log(2)\exp\left(-\frac{log^2(2)}{\pi}\right)}{\pi}$	$1+\frac{\exp\left(-\frac{\log^2(2)}{\pi}\right)}{2\log(2)}$
$\boldsymbol{F^{[96]}(x)}$	$\frac{1}{\sqrt{2\pi}}$	$1-\frac{2\exp\left(-\frac{1}{4\pi}\right)}{\pi}$	$1+\exp\left(-\frac{1}{4\pi}\right)$
$\boldsymbol{F^{[96]}(x)}$	$\sqrt{2/\pi}$	$1-\frac{4}{\pi}\exp\left(-\frac{1}{\pi}\right)$	$1+\frac{1}{2}\exp\left(-\frac{1}{\pi}\right)$

Као мера за поређење користи се релативна грешка (RE) [90−101]. За поменуте параметре RE се дефинише као:

$$\delta_{t_2} = \frac{|t_2^A - t_2|}{t_2},\ \delta_{\text{SQNR}} = \frac{|\text{SQNR}^A - \text{SQNR}|}{\text{SQNR}},\ \delta_R = \frac{|R^A - R|}{R}. \tag{3.40}$$

Табела 3.3 Релативне грешке у процентима (%) добијене за различите апроксимације Q – функције

Релативна грешка	$F^{[100]}(x)$	$F^{[98]}(x)$	$F^{[96]}(x)$	$F^{[96]}(x)$	$F(x)$
δ_{t2} **[%]**	2.18	9.63	30.37	34.81	**1.59**
δ_{SQNR} **[%]**	3.46	14.67	56.97	46.59	**2.45**
δ_{R} **[%]**	1.07	5.10	11.48	24.86	**0.72**

Вредности за RE за одговарајуће параметре сумиране су у табели 3.3. RE је најмања када се користи предложена апроксимација Q-функције (3.32) (максимална вредност за RE иде до 2.45 %). То показује да су изведени изрази (3.26), (3.34) и (3.36) јако тачни.

3.5 Тернарна квантизација за Гаусов изор у случају неприлагођења на варијансу

Неприлагођење на варијансу је ефекат који се често јавља у пракси и односи се на сценарио када се варијанса за коју је квантизер пројектован разликује од варијансе сигнала који се квантује [1, 2, 8]. Због тога је од изузетне важности испитати перформансе квантизера у присуству овог ефекта. У литератури је до сада испитиван утицај овог ефекта на перформансе различитих типова квантизера, где је уочен негативан утицај на перформансе [102−105]. За случај тернарног квантизера предложиће се јако тачни апроксимативни изрази за процену перформанси (SQNR и R) на бази апроксимације Q-функције.

Разматра се ситуација када је тернарни квантизер пројектован за варијансу $\sigma_{\text{ref}}^2 = 1$ (процес пројектовања је дат у претходном одељку), а примењује се за квантовање Гаусовог сигнала са варијансом σ^2 при чему важи да је $\sigma^2 \neq \sigma_{\text{ref}}^2 = 1$. Дисторзија тернарног квантизера се у овом случају процењује као:

$$D(\sigma) = 2\int_{0}^{t_2(\sigma_{\text{ref}})} x^2 p(x,\sigma)dx + 2\int_{t_2(\sigma_{\text{ref}})}^{\infty} \left(x - y_3(\sigma_{\text{ref}})\right)^2 p(x,\sigma)dx$$
$$= \sigma^2\left(1 - \frac{8t_2}{\sigma^2}\left(\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{t_2^2}{2\sigma^2}\right\} - t_2 Q\left(\frac{t_2}{\sigma}\right)\right)\right), \qquad (3.41)$$

где се $t_2 = t_2(\sigma_{\text{ref}})$ и $y_3 = y_3(\sigma_{\text{ref}}) = 2 \cdot t_2$ односе на параметре одређене за $\sigma_{\text{ref}}^2 = 1$, а SQNR је дат са:

$$\text{SQNR}(\sigma) = 10\log_{10}\left(\frac{\sigma^2}{D(\sigma)}\right) = 10\log_{10}\left(\frac{1}{1-\frac{8t_2}{\sigma^2}\left(\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{t_2^2}{2\sigma^2}\right\}-t_2 Q\left(\frac{t_2}{\sigma}\right)\right)}\right). \tag{3.42}$$

Као што се види, SQNR зависи од σ. Вероватноће нивоа такође зависе од σ:

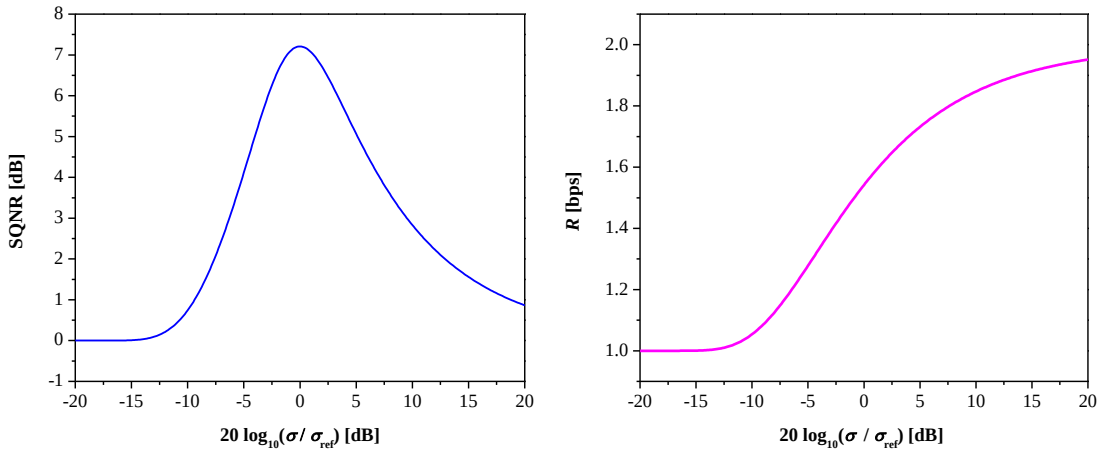
$$p(-y_3) = p(y_3) = \int_{t_2}^{\infty} p(x,\sigma)dx = Q\left(\frac{t_2}{\sigma}\right), \tag{3.43}$$

$$p(y_2) = 2\int_0^{t_2} p(x,\sigma)dx = 1 - 2Q\left(\frac{t_2}{\sigma}\right), \tag{3.44}$$

као и битска брзина која је дата са:

$$R(\sigma) = 1 + 2Q\left(\frac{t_2}{\sigma}\right). \tag{3.45}$$

На слици 3.8 приказане су перформансе (SQNR и R) тернарног квантизера ($\sigma_{\text{ref}}^2 = 1$) када се ефекат неприлагођења на варијансу јавља, где се види снажан утицај овог ефекта на перформансе. На пример, са слике 3.8-а) се види да SQNR опада за $\sigma^2 \neq \sigma_{\text{ref}}^2$ а да за $\sigma^2 = \sigma_{\text{ref}}^2 = 1$ (случај када прилагођење на варијансу постоји) постиже максималну вредност.



а) б)

Сл. 3.8 Перформансе тернарног квантизера ($\sigma_{\text{ref}}^2 = 1$) у присуству ефекта неприлагођења на варијансу: а) SQNR и б) R.

Ако се у изразима (3.42) и (3.45) за t_2 уместо тачне користи апроксимативна вредност из предходног одељка $t_2^A = \sqrt{a/3}$ (a = 1.1587), добија се:

$$\mathrm{SQNR}'(\sigma) = 10\log_{10}\left(\frac{1}{1-\frac{8}{\sigma^2}\sqrt{\frac{a}{3}}\left(\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{a}{6\sigma^2}\right\}-\sqrt{\frac{a}{3}}Q(x')\right)}\right), \tag{3.46}$$

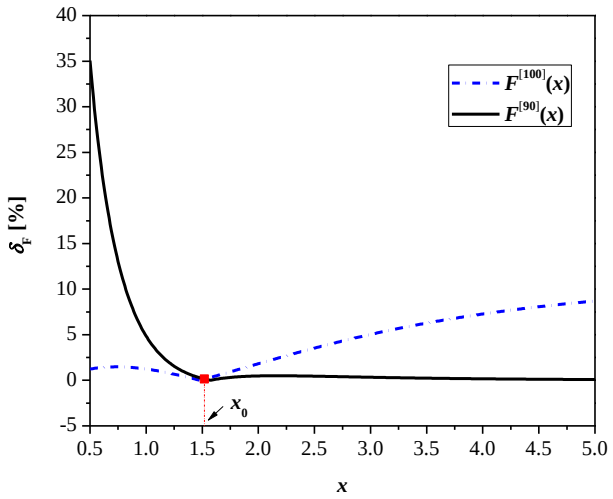
$$R'(\sigma) = 1 + 2Q(x'), \tag{3.47}$$

где је $x' = \sqrt{a/3}/\sigma$.

У последња два израза фигурише Q-функција, па није могуће извести егзактне изразе за SQNR и R. Као алтернатива Q-функцији предложена је следећа композитна апроксимација:

$$F(x) = \begin{cases} F^{[100]}(x) = \frac{\left(1-\exp\left\{-\frac{Ax}{\sqrt{2}}\right\}\right)\exp\left\{-\frac{x^2}{2}\right\}}{\sqrt{2\pi}Bx}, \text{ако је } x < x_0 \\ F^{[90]}(x) = \frac{1}{\sqrt{2\pi}}\frac{x^2+b-1}{x(x^2+b)}\exp\left\{-\frac{x^2}{2}\right\}, \text{другде} \end{cases}, \tag{3.48}$$

где x_0 означава границу између одговарајућих опсега аргумената у којима се одређена апроксимација примењује, а A, B и b су параметри одговарајућих апроксимација које се оптимизују за дати проблем. Идеја за композитном апрксимацијом настала је из разлога што се аргумент Q-функције $\sqrt{\frac{a}{3}}/\sigma$ мења у широком опсегу (слика 3.8), па се тако $F(x)^{[100]}$ користи за ниже вредности аргумента а $F(x)^{[90]}$ за веће вредности аргумента.



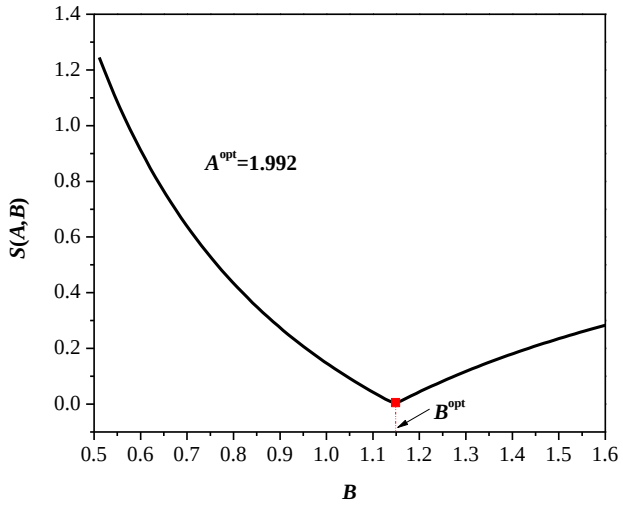
Сл. 3.9 Избор x_o у зависности од RE, за $F^{[100]}(x)$ (A = 1.98 и B = 1.135) и $F^{[90]}(x)$ (b = 2.2).

На слици 3.9 приказано је једно решење за избор x_0 које се заснива на RE. За $F^{[100]}(x)$ са параметрима A = 1.98 и B = 1.135 и $F^{[90]}(x)$ са параметром b = 2.2 израчуната је RE у опсегу $x \in (0.5, 5)$, а граница између опсега аргумената у којима је једна апроксимација боља од друге (у смислу нижих RE вредности) је одређена са x_0 = 1.51.

Када је x_0 познато, A и B из $F^{[100]}(x)$ се оптимизују тако да средња RE у опсегу $x \in (0, x_0 = 1.51)$ буде минимална:

$$S(A,B) = \underset{A,B}{\arg\min} \left\{ \frac{1}{M} \left[\sum_{i=1}^{M} \frac{|F^{[100]}(x(i),A,B) - Q(x)|}{Q(x)} \right] \right\}. \tag{3.49}$$

Нумеричком оптимизацијом добија се да је $A = A^{\text{opt}} = 1.992$ и $B = B^{\text{opt}} = 1.149$, као што показује слика 3.10.



Сл. 3.10 Избор оптималних вредности за A и B за апроксимацију $F^{[100]}(x)$ у интервалу $x \in (0, x_0)$.

Са друге стране, параметар b из апроксимације $F^{[90]}(x)$ се оптимизује за опсег $x \geq x_0$. Оптимална вредност се одређује из услова $x_0 = 1.51 = \sqrt{b(b-1)/(3-b)}$ шт даје $b = b^{\text{opt}} = 2.0525$.

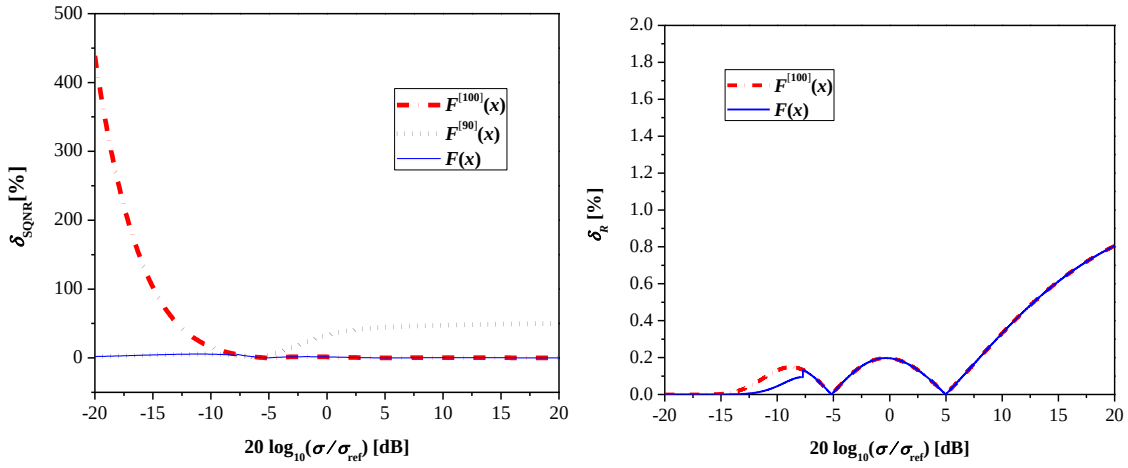
Коначно, применом (3.48) у (3.46) и (3.47) изводе се следећи апроксимативни изрази за SQNR и R:

$$\mathrm{SQNR}^{A}=\begin{cases}10\log_{10}\left(\dfrac{1}{1-\frac{8}{\sigma^2}\sqrt{\frac{a}{3}}\left(\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{a}{6\sigma^2}\right\}-\sqrt{\frac{a}{3}}F^{[100]}(x')\right)}\right), x'<1.51\\ 10\log_{10}\left(\dfrac{1}{1-\frac{8}{\sigma^2}\sqrt{\frac{a}{3}}\left(\frac{\sigma}{\sqrt{2\pi}}\exp\left\{-\frac{a}{6\sigma^2}\right\}-\sqrt{\frac{a}{3}}F^{[90]}(x')\right)}\right), \text{другде}\end{cases}, \tag{3.50}$$

$$R^{A}=\begin{cases}1+2F^{[100]}(x'), & x'<1.51\\ 1+2F^{[90]}(x'), \text{другде}\end{cases}. \tag{3.51}$$

Анализа нумеричких резултата. Поређење изведених апроксимативних формула за SQNR и R се врши са тачним (изрази (3.42) и (3.45)) али и са другим доступним формулама за скаларну квантизацију [99].

Слика 3.11 илуструје RE вредности за предложене формуле (опсег варијанси је исти као на слици 3.8). У циљу поређења, приказане су и RE вредности за случајеве када се у изразима (3.42) и (3.45) засебно користе компоненете композитне апроксимације $F^{[100]}(x)$ и $F^{[90]}(x)$. У табели 3.4 дата је средња и максимална вредност за δ_{SQNR} и δ_{R}. Анализом приказаних резултата долази се до закључка да су предложене апроксимативне формуле (када се користи композитна апроксимација (3.48)) јако тачне.



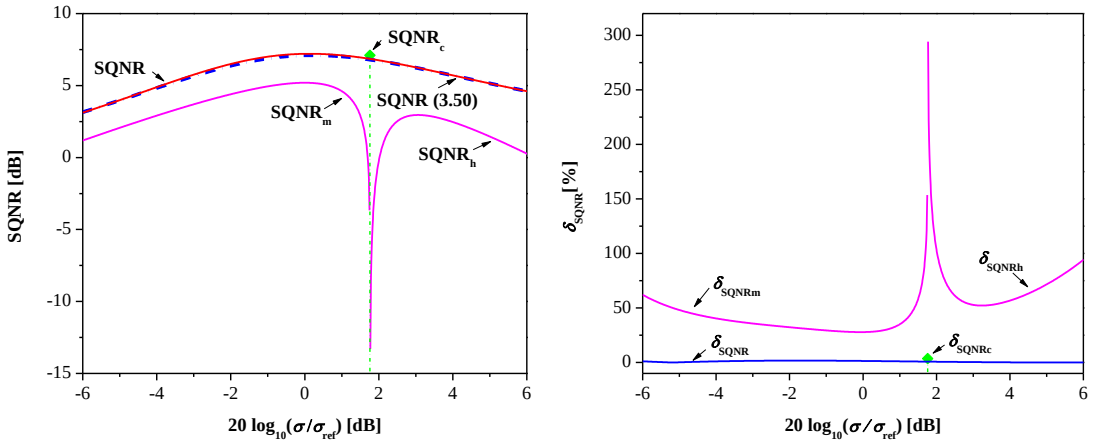
а) б)

Сл. 3.11 Релативне грешке у широком опсегу улазних варијанси за: а) SQNR и б) R.

Табела 3.4 Средња и максимална RE за SQNR и R, за различите апроксимације Q-функције

Апроксимација	**Средња δ_{SQNR} [%]**	**максимална δ_{SQNR} [%]**	**средња δ_R [%]**	**максимална δ_R [%]**
$F^{[100]}(x)$	36.77	438.61	0.22	0.81
$F^{[90]}(x)$	26.79	50.00	46.62	288.76
$F(x)$	**1.78**	**5.58**	**0.21**	**0.81**

Изведена апроксимативна формула (3.50) је упоређена и са постојећим формулама за SQNR [99]. Конкретно, из [99] су узете у обзир формуле када постоји благо ($SQNR_m$), критично ($SQNR_c$) и велико ($SQNR_h$) неприлагођење на варијансу. SQNR криве као и RE израчунате у том случају илустроване су на слици 3.12. Предложена SQNR формула је знатно ефикаснија од $SQNR_m$ и $SQNR_h$ јер се код њих уочава велико одступање од тачних вредности, што има за последицу да је RE велико (слика 3.12-б)). У поређењу са $SQNR_c$ где је уочена RE од свега 3.7 %, изведена SQNR формула је тачнија јер тада RE износи 0.81 %.



а) б)

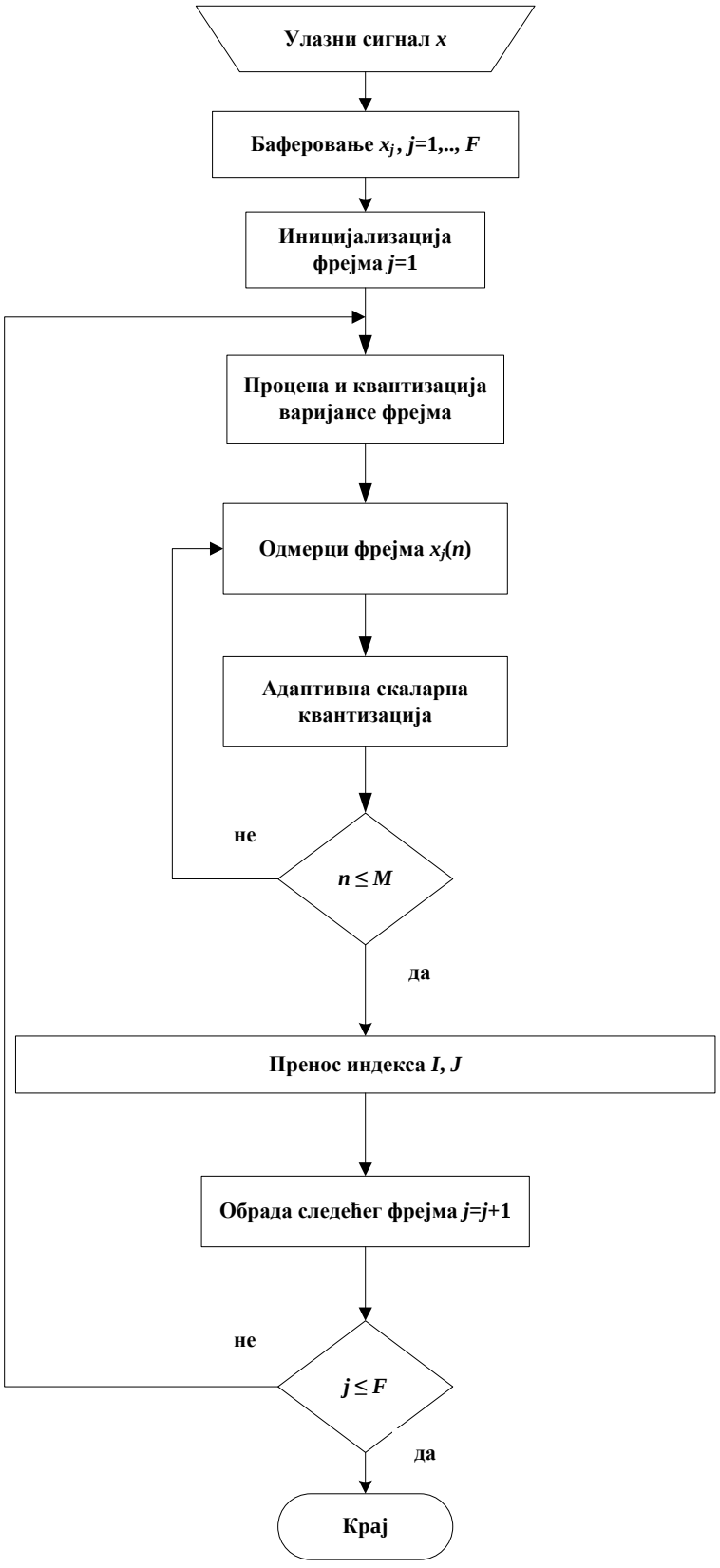
Сл. 3.12 а) Поређење са SQNR формулама из рада [99] и б) δ_{SQNR} у посматраном опсегу варијансе.

4. Пројектовање квантизера за PCM кодовање говорног сигнала

У овом поглављу дисертације биће представљена решења на бази PCM технике за компресију (мало N) и за висококвалитетно (велико N) кодовање говора. Код ових система врши се кодовање оригиналних вредности говора. Имплементација нових модела квантизера, који су теоријски пројектовани за Лапласов извор, у PCM-у врши се применом технике адаптације унапред. На почетку ће пажња бити посвећена квантизерима са малим бројем нивоа односно нискорезолуционој квантизацији. Биће предложен бинарни PCM кодек који имплементира бинарни квантизер оптимално пројектован за Лапласов извор. Затим ће се бити анализиран PCM кодек са униформним квантизером са мртвом зоном са пет нивоа и Хафмановим кодом, а биће показано да овакав модел квантизера може да пружи боље перформансе од класичног униформног и неуниформног Лојд-Макс квантизера са истим бројем нивоа. Након тога, биће представљен PCM кодек са двомодним квантизером за висококвалитетно кодовање говора, где примењени квантизер наизменично комбинује два неуниформна квантизера са истим бројем нивоа али различитим грануларним регионима. Биће показано да овај тип кодека може да постигне исти квалитет сигнала при мањој битској брзини од G.711 квантизера. На крају овог поглавља биће представљен PCM кодек са двомодним квантизером за компресију говора, а имплементирани квантизер се састоји из два тернарна квантизера са Хафмановим кодом пројектована за ограничен Лапласов извор.

4.1 Техника адаптације унапред

Адаптација унапред је техника кодовања која се заснива на фрејм по фрејм анализи и обради сигнала, где фрејм представља групу узастопних одмерака сигнала [1−3, 8, 15]. Основна идеја састоји се у томе да се варијанса фрејма процени пре него што се изврши процес квантизације и да се сходно томе изврши подешавање параметара квантизера (прагова одлуке и нивоа). Информација о варијанси фрејма се затим преноси до декодера као додатна информација.



Сл. 4.1 Дијаграм тока за технику адаптације унапред.

Процес адаптације квантизера се може представити у облику алгоритма чији је дијаграм тока дат на слици 4.1. Алгоритам подразумева следеће кораке:

1. ***Баферовање фрејма.*** У бафер се смешта група одмерака улазног сигнала $x_j(n)$, $n = 1,\ldots, M,\ \ j = 1,\ldots, F$, где је M величина фрејма (такође означава капацитет бафера), j је индекс фрејма а F је укупан број фрејмова.
2. ***Процена и квантизација варијансе фрејма.*** За j-ти фрејм, варијанса σ_j^2 се процењује као [1−3, 8]:

$$\sigma_j^2 = \frac{1}{M}\sum_{n=1}^{M} x_j^2(n)\,,\ j = 1,\ldots,F\,, \tag{4.1}$$

при чему се претпоставља да фрејм има нулту средњу вредност. Квантизација овог параметра се врши помоћу лог-униформног квантизера (Q_{LU}), који заправо представља униформни квантизер у логаритамском домену. Дакле, варијанса у логаритамском домену V_j [dB] = $20\cdot\log_{10}(\sigma_j)$ се помоћу квантизера Q_{LU} квантује на једну од L могућих вредности [16−22]:

$$Q_{LU}:\ \ V_j = V_i \ \Big|\ V_i = V_{\min} + \frac{2i-1}{2}\frac{B}{L}\,,\ i = 1,\ldots,L\,, \tag{4.2}$$

где је са B[dB] = V_{max}[dB] - V_{min}[dB] дефинисан динамички опсег сигнала, а V_{max} и V_{min} означавају максималну и минималну вредност варијансе фрејма, респективно. L-ти ниво се кодује са $R_{LU} = \log_2 L$ битова и преноси се индексом J до декодера као додатна информација.

3. ***Адаптивна квантизација.*** Квантована вредност варијансе (односно ниво квантизера Q_{LU}) се даље користи за скалирање параметара неадаптивног (пројектованог за σ_{ref}^2) квантизера са N нивоа и као резултат тога добија се адаптивни квантизер. Дакле, скуп прагова одлуке $\boldsymbol{T}^a(g_j)$ и скуп нивоа $\boldsymbol{Y}^a(g_j)$ адаптивног квантизера се одређују на следећи начин:

$$\boldsymbol{T}^a\left(g_j\right) = g_j \cdot \boldsymbol{T}\left(\sigma_{ref}\right), \tag{4.3}$$

$$\boldsymbol{Y}^a\left(g_j\right) = g_j \cdot \boldsymbol{Y}\left(\sigma_{ref}\right), \tag{4.4}$$

где $\boldsymbol{T}(\sigma_{ref})$ и $\boldsymbol{Y}(\sigma_{ref})$ означавају скупове прагова одлуке и репрезентационих нивоа квантизера пројектованог за σ_{ref}^2, респективно, док је g_j фактор скалирања дат са:

$$g_j = 10^{V_i/20}. \tag{4.5}$$

Након што се дефинише адаптивни квантизер за *j*-ти фрејм, сваки од *M* одмерака *j*-тог фрејма пролази кроз квантизер и кодује се одговарајућим кодним речима (може се користи код са фиксном или са промењљивом дужином кодних речи) и преноси се индексом *I* до декодера.

4. ***Реконструкција сигнала.*** Реконструкција одмерака *j*-тог фрејма сигнала се врши у декодеру на основу индекса *I* и *J*.
5. ***Понављајти претходне кораке док сви фрејмови не буду обрађени*.**

Битска брзина за адаптивни квантизер се израчунава на следећи начин:

$$R = R_N + \frac{\log_2 L}{M} \quad [\text{bps}], \tag{4.6}$$

где је R_N битска брзина неадаптивног квантизера и зависи од примењеног кода а други члан је додатна информација.

Мере за процену перформанси при кодовању говора. За процену ефикасности адаптивног модела квантизера (PCM кодека) на говору користи се сегментни SQNR (SQNR_{seg}), тј. SQNR се израчунава за сваки фрејм сигнала и затим усредњује [1−10]:

$$\text{SQNR}_{\text{seg}} = \frac{1}{F}\sum_{j=1}^{F} 10\log_{10}\left(\frac{\sigma_j^2}{D_j}\right), \tag{4.7}$$

где је σ_j^2 дато изразом (4.1) а D_j је дисторзија за *j*-ти фрејм говора и дата је са:

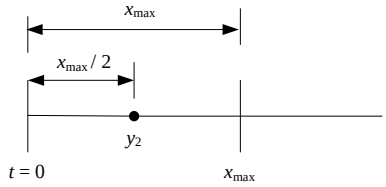
$$D_j = \frac{1}{M}\sum_{n=1}^{M}\left(x(n) - y(n)\right)^2, \tag{4.8}$$

где $x_j(n)$ и $y_j(n)$ означавају оригиналне и реконструисане одмерке говора, респективно.

4.2 Бинарни PCM кодек

У овом одељку представљено је ниско-резолуционо PCM решење на бази бинарне квантизације, а које је анализирано у раду [65].

Опис и пројектовање квантизера. Бинарни симетрични квантизер ($y_1 = -y_2$) је приказан на слици 4.2, а x_{max} је помоћни параметар при пројектовању. Између x_{max} и y_2 постоји следећа веза: $y_2 = t_{max} / 2$.



Сл. 4.2 Симетрични бинарни квантизер.

Грануларна дисторзија и дисторзија прекорачења бинарног квантизера, за претпостављену Лапласову PDF (израз (2.6)) и $\sigma^2 = \sigma_{ref}^2 = 1$, се израчунавају као:

$$D_g = 2\int_0^{x_{max}} \left(x - \frac{x_{max}}{2}\right)^2 p(x)\,dx = 1 - \frac{x_{max}}{\sqrt{2}} + \frac{x_{max}^2}{4} - \left(1 + \frac{x_{max}}{\sqrt{2}} + \frac{x_{max}^2}{4}\right)\exp\left\{-\sqrt{2}x_{max}\right\}, \qquad (4.9)$$

$$D_o = 2\int_{x_{max}}^{\infty} \left(x - \frac{x_{max}}{2}\right)^2 p(x)\,dx = \left(1 + \frac{x_{max}}{\sqrt{2}} + \frac{x_{max}^2}{4}\right)\exp\left(-\sqrt{2}x_{max}\right), \qquad (4.10)$$

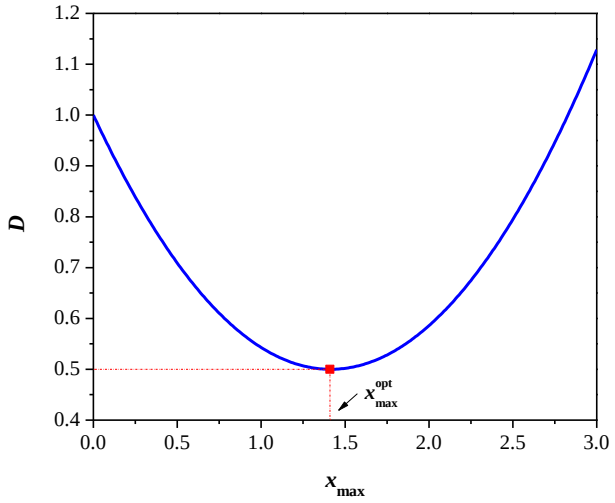
док је укупна дисторзија дата са:

$$D = D_g + D_o = 1 - \frac{x_{max}}{\sqrt{2}} + \frac{x_{max}^2}{4}\ . \qquad (4.11)$$

Лако се уочава да је D функција од x_{max}. За дисторзијом-ограничен квантизер, оптимално x_{max} се одређује као решење једначине $\partial D / \partial x_{max} = 0$, што даје:

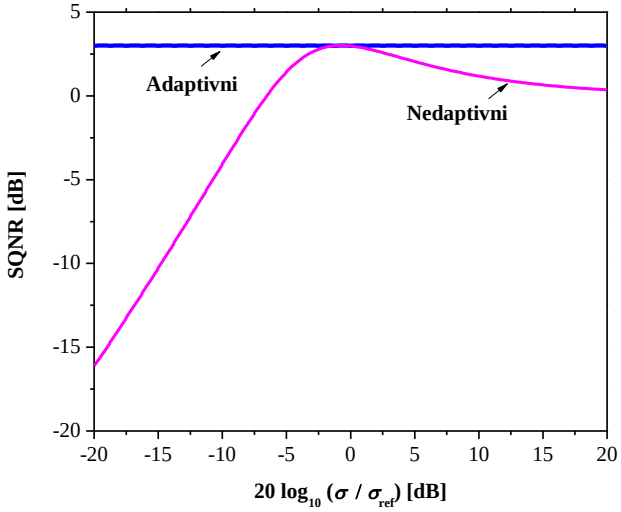
$$\partial D / \partial x_{max} = 0 \Rightarrow x_{max} = x_{max}^{opt} = \sqrt{2}\,. \qquad (4.12)$$

Претходно добијени резултат се може верификовати помоћу слике 4.3. Коначно, за ниво дисторзијом-ограниченог бинарног квантизера добија се $y_2 = x_{\max} / 2 = 1/\sqrt{2}$.



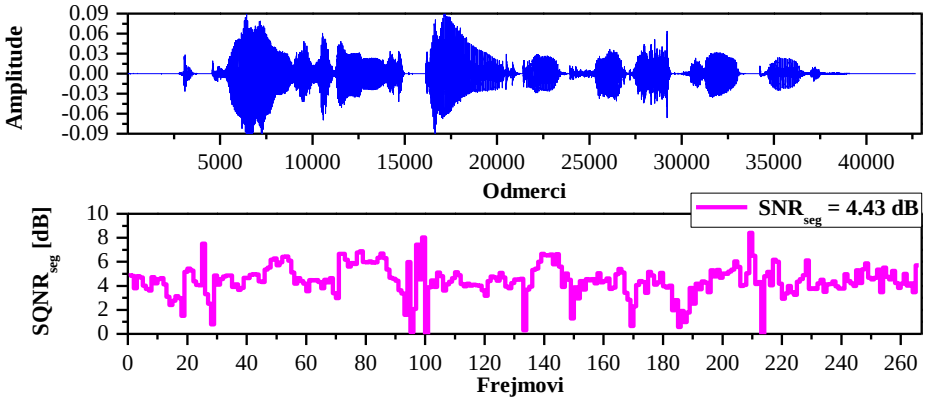
Сл. 4.3 D у зависности од $x_{\max}$ за Лапласов бинарни квантизер.

Теоријске перформансе бинарног PCM кодека. Након имплементације бинарног квантизера у техници адаптације унапред (одељак 4.1) добија се бинарни PCM кодек (односно адаптивни бинарни квантизер). Перформансе (SQNR) овог кодека када се користи Q_{LU} са L = 32 нивоа дате су на слици 4.4. У поређењу са неадаптивним бинарним квантизером даје доста боље перформансе (скоро константан SQNR у целом посматраном опсегу варијансе).



Сл. 4.4 Перформансе бинарног PCM кодека (Q_{LU} са L = 32 нивоа).

Примена на говорни сигнал. Говор трајања 3s (фреквенција одмеравања је 16 kHz) екстрахован из базе *Harvard Psychoacoustic Sentences* [106] (реченица на енглеском језику изговорена од стране женског говорника) се користи као тест сигнал. На слици 4.5 је за предложени бинарни PCM кодек приказан SQNR по фрејмовима величине 10 ms (M = 160) као и сегментни SQNR ($SQNR_{seg}$ = 4.43 dB), када се користи Q_{LU} са 32 нивоа. Добијени резулати су у сагласности са теоријским резулатима на слици 4.4, што потврђује исправност теоријског модела за PCM кодек.



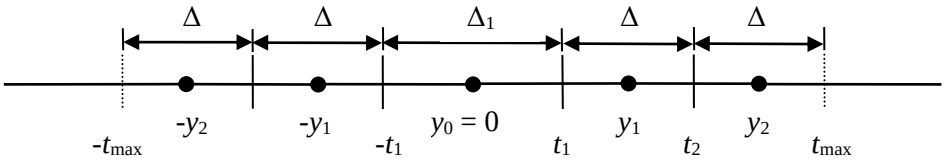
Сл. 4.5 Оригинални говор и SQNR по фрејмовима дужине 10 ms за бинарни PCM кодек.

4.3 PCM кодек са униформним квантизером са мртвом зоном

У овом одељку разматра се још једно ниско-резолуционо PCM решење, које се базира на примени униформног квантизера са мртвом зоном и променљивом дужином кодних речи, а анализирано је у раду [66].

Опис и пројектовање квантизера. Симетрични униформни квантизер са мртвом зоном са непарним бројем нивоа (N = 5) дат је слици 4.6 а састоји се из два дела:

1. мртве зоне ширине Δ_1, дефинисане у интервалу $(-t_1, t_1)$ која обухвата ниво $y_0 = 0$;
2. униформног дела (са кораком Δ) дефинисаног у интервалу $(-t_{max}, -t_1) \cup (t_1, t_{max})$ који обухвата преостала четири нивоа.



Сл. 4.6 Униформни квантизер са мртвом зоном са пет нивоа.

Дисторзија D једнака је збиру дисторзије у мртвој зони D_{DZ} и дисторзије у региону изван мртве зоне D_{OP}. За Лапласову PDF и $\sigma^2 = \sigma_{\text{ref}}^2 = 1$, D_{DZ} и D_{OP} се рачунају коришћењем следећих израза:

$$D_{DZ} = 2\int_0^{t_1} x^2 p(x)dx = 1-\left(1+t_1\left(\sqrt{2}+t_1\right)\right)\exp\left\{-\sqrt{2}t_1\right\}, \tag{4.13}$$

$$D_{OP} = 2\int_{t_1}^{t_1+\Delta} \left(x-y_1\right)^2 p(x)dx + 2\int_{t_1+\Delta}^{\infty} \left(x-y_2\right)^2 p(x)dx\ . \tag{4.14}$$

За разлику од класичног униформног квантизера код кога се нивои налазе на средини ћелија, код овог модела се нивои у униформном делу одређују из услова центроида (израз (2.16)), што даје:

$$y_1 = \frac{\int_{t_1}^{t_1+\Delta} xp(x)dx}{\int_{t_1}^{t_1+\Delta} p(x)dx} = \frac{\exp\left\{-\sqrt{2}t_1\right\}\left(1+\sqrt{2}t_1\right)-\exp\left\{-\sqrt{2}\left(t_1+\Delta\right)\right\}\left(1+\sqrt{2}\left(t_1+\Delta\right)\right)}{\sqrt{2}\left(\exp\left\{-\sqrt{2}t_1\right\}-\exp\left\{-\sqrt{2}\left(t_1+\Delta\right)\right\}\right)}, \tag{4.15}$$

$$y_2 = \frac{\int_{t_1+\Delta}^{t_1+2\Delta} xp(x)dx}{\int_{t_1+\Delta}^{t_1+2\Delta} p(x)dx} = \frac{\exp\left\{-\sqrt{2}\left(t_1+\Delta\right)\right\}\left(1+\sqrt{2}\left(t_1+\Delta\right)\right)-\exp\left\{-\sqrt{2}\left(t_1+2\Delta\right)\right\}\left(1+\sqrt{2}\left(t_1+2\Delta\right)\right)}{\sqrt{2}\left(\exp\left\{-\sqrt{2}\left(t_1+\Delta\right)\right\}-\exp\left\{-\sqrt{2}\left(t_1+2\Delta\right)\right\}\right)}. \tag{4.16}$$

Заменом (4.15) и (4.16) у (4.14) и решавањем интеграла добија се:

$$D_{OP} = \frac{1}{2}\frac{\exp\left\{-\sqrt{2}\left(t_1+\Delta\right)\right\}}{\left(\exp\left\{\sqrt{2}\Delta\right\}-1\right)^2}\left(\exp\left\{3\sqrt{2}\Delta\right\}+2\Delta^2-2\left(1+\Delta^2\right)\exp\left\{2\sqrt{2}\Delta\right\}+\left(1+2\Delta^2\right)\exp\left\{\sqrt{2}\Delta\right\}\right). \tag{4.17}$$

SQNR се рачуна као $\text{SQNR} = -10\log_{10}\left(D_{DZ}+D_{OP}\right)$.

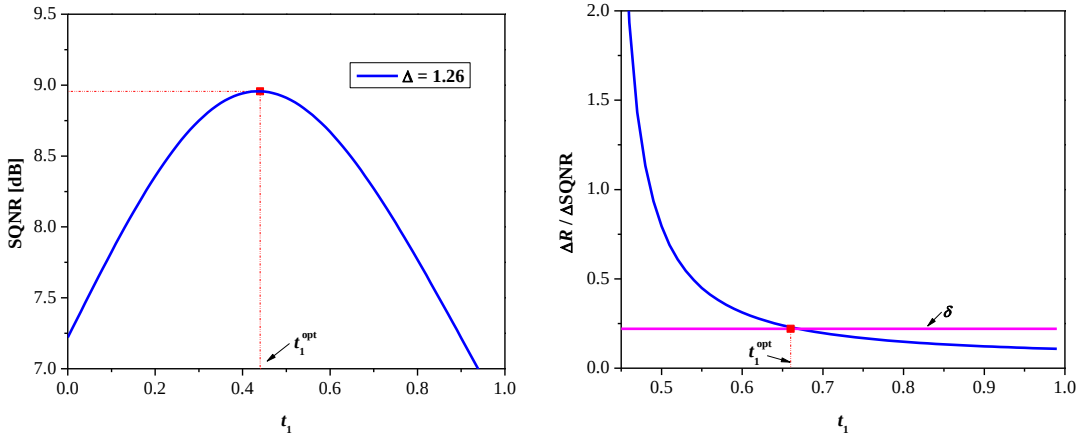
На нивое овог квантизера примењује се Хафманов код. Хафманова битска брзина се израчунава према изразу (2.10), при чему се вероватноће нивоа за Лапласову PDF одређују као:

$$p(y_0) = 2\int_0^{t_1} p(x)dx = 1 - \exp\left\{-\sqrt{2}t_1\right\}, \tag{4.18}$$

$$p(y_1) = p(-y_1) = \int_{t_1}^{t_1+\Delta} p(x)dx = \frac{1}{2}\exp\left\{-\sqrt{2}t_1\right\}\left(1 - \exp\left\{-\sqrt{2}\Delta\right\}\right), \tag{4.19}$$

$$p(y_2) = p(-y_2) = \int_{t_1+\Delta}^{\infty} p(x)dx = \frac{1}{2}\exp\left\{-\sqrt{2}\left(t_1+\Delta\right)\right\}. \tag{4.20}$$

Изрази (4.13) и (4.17)-(4.20) показују да перформансе овог квантизера зависе од t_1 и Δ. Вредности ових параметара се бирају у зависности од примењеног критеријума оптимизације односно од тога да ли се пројектује дисторзијом-ограничен или брзином-ограничен квантизер.



а) б)

Сл. 4.7 Избор вредности за t_1 и Δ за униформни квантизер са мртвом зоном: а) дисторзијом-ограничен и б) брзином-ограничен квантизер.

За дисторзијом-ограничен квантизер усваја се t_1 = 0.4457 и Δ = 1.26, као што показује слика 4.6-а). Са друге стране, модификацијом дисторзијом-ограниченог квантизера добијен је брзином-ограничен квантизер са параметрима t_1 = 0.6685 и Δ = 1.26 (слика 4.6-б)). Наиме, код овог модела униформни део је исти као код дисторзијом-ограниченог квантизера (Δ = 1.26), а оптимално t_1 је добијено уз

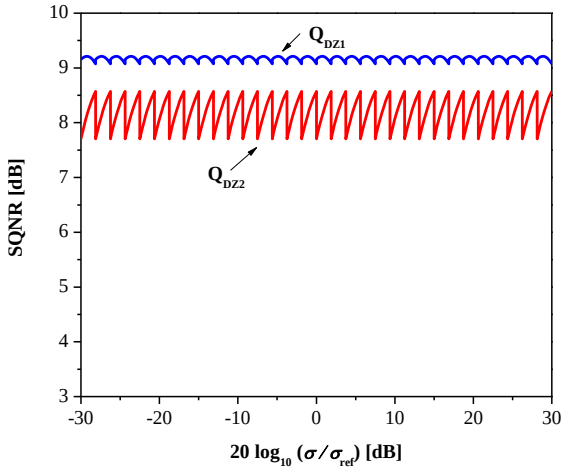
помоћ критеријума (3.11) при чему је опсег за t_1 од 0.4457 до 1 док је δ = 0.2203 (нагиб између Лојд-Макс квантизера са 2 и 4 нивоа јер је добијено R испод 2 bps).

Табела 4.1 Постигнуте перформансе предложеног квантизера у односу на класични униформни и Лојд-Макс квантизер са пет нивоа

	Δ	Δ_1	$p(y_1)= p(-y_1)$	$p(y_2)=p(-y_2)$	$p(y_0)$	D	SQNR[dB]	R[bps]
Q_{DZ1}	1.26	0.8914	0.0448	0.2214	0.4676	0.1272	8.9556	1.9331
Q_{DZ2}	1.26	1.3370	0.0327	0.1616	0.6115	0.1442	8.4096	1.6808
Q_{LM}	-	-	0.0561	0.2200	0.4478	0.1198	9.2152	2
Q_U	1	-	0.0599	0.1866	0.5069	0.1332	8.7560	1.9194

У табели 4.1 дати су детаљи за дисторзијом-ограничен (означен је са Q_{DZ1}) и брзином-ограничен (означен је са Q_{DZ2}) униформни квантизер са мртвом зоном. У овој табели су дати и детаљи за Лојд-Макс (означен је са Q_{LM}) и униформни квантизер (означен је са Q_U) са истим бројем нивоа и Хафмановим кодом. Као што се и очекивало, Q_{DZ1} постиже мало већи SQNR у поређењу са Q_{DZ2}, али Q_{DZ2} има мању битску брзину. Поређењем перформанси са Q_{LM} види се да и Q_{DZ1} и Q_{DZ2} имају мало мањи SQNR али омогућавају уштеду у битској брзини, а притом су и једноставнији су за реализацију. У односу на Q_U, Q_{DZ1} има приближно исту битску брзину а остварује већи SQNR, док Q_{DZ2} даје мању битску брзину уз прибижно исти SQNR.

Теоријске перформансе PCM кодека. Разматрани PCM кодек настао је као резултат имплементације униформног квантизера са мртвом зоном у техници адаптације унапред (одељак 4.1).



Сл. 4.8 SQNR у широком опсегу варијансе за разматрани PCM кодек са униформним квантизером са мртвом зоном (Q_{LU} са L = 32 нивоа).

Теоријске перформансе (SQNR) кодека за две верзије квантизера су приказане на слици 4.7, где се види да је SQNR скоро константан у целом посматраном опсегу варијансе а у случају примене Q_{DZ2} примећују се мало веће варијације у SQNR-у.

Примена на говорни сигнал. Као тест сигнал користи се говор са око милион одмерака (реченице на српском језику изговорене од стране мушког говорника), који је одмераван на 16 kHz. Добијене перформансе (сегментни SQNR и битска брзина) за предложени PCM кодек са квантизером Q_{DZ2} (Q_{LU} са L = 32 нивоа) и различито M су приказане у табели 4.2.

Табела 4.2 Перформансе предложеног PCM кодека са квантизером Q_{DZ2} на тест говорном сигналу за различите величине фрејма

M	**80**	**160**	**240**	**320**
SQNR [dB]	9.1815	9.0453	8.9826	8.9385
***R* [bps]**	2.0591	2.0278	2.0174	2.0122

Из табеле се види да је за M = 80 добијен мало већи SQNR него у осталим случајевима. То је заправо очекивани исход, јер се тада параметри квантизера Q_{DZ2} чешће ажурирају. Иначе, добијене $SQNR_{seg}$ вредности се у доброј мери слажу са теоријским вредностима на слици 4.8.

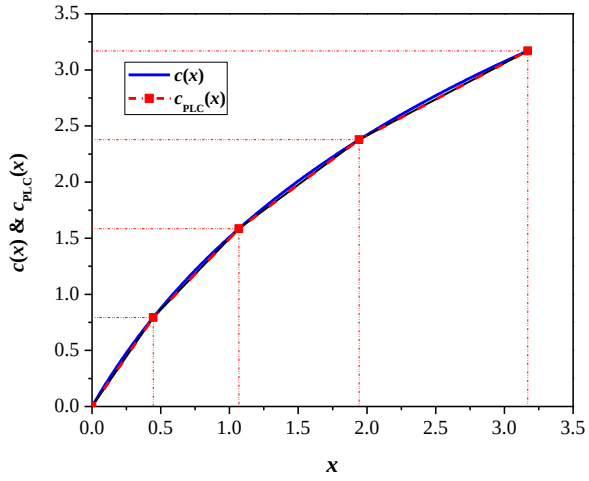
4.4 PCM кодек на бази двомодне квантизације

У претходним одељцима анализирани су PCM кодеци који користе Лапласове квантизере за мале битске брзине. У овом одељку акценат је на једном PCM кодеку за високо-квалитетно кодовање говора који користи двомодни скаларни квантизер (DMSQ) а који је предложен у раду [67].

Опис DMSQ-а. DMSQ комбинује два део-по-део линеарна компандинг квантизера (PLCSQ) са истим бројем нивоа али са различитим гарнуларним регионима (ткзв. ограничен и неограничен PLCSQ). Основна идеја састоји се у томе да се у оквиру DMSQ-а обезбеди што чешћи избор ограниченог PLCSQ-а, јер

је на тај начин могуће постићи веће перформансе у односу на случај када се користи само неограничени PLCSQ. У наставку је дат опис PLCSQ-а.

Разматрани PLCSQ. PLCSQ са N нивоа је симетричан и заснива се на линеаризацији нелинеарне компресорске функције $c(x)$ дате изразом (2.19) са сплајн функцијом првог реда [107]. Наиме, позитивни део опсега [0, $t_{\max}$] функције $c(x)$ ($c(x)$ има исти опсег као и PLCSQ) дели се на L сегмената а сваки сегмент се апроксимира линеарном правом (полином првог степена). На слици 4.8 дат је приказ нелинеарне компресорске функције $c(x)$ и апроксимације добијене сплајн функцијом (означена је са $c_{\text{PLC}}(x)$) са $L = 4$ сегмента када је $\mu = 2.87$ и $t_{\max} = 3.17$.



Сл. 4.9. Линеаризација нелинеарне компресорске функције $c(x)$ са сплајн функцијом првог реда $c_{\text{PLC}}(x)$ са $L = 4$ сегмента ($\mu = 2.87$ и $t_{\max} = 3.17$).

Дакле, опсег PLCSQ-а се дели на L неједнаких сегмената а сваки сегмент је подељен униформно на $n = N / 2L$ ћелија. Границе сегмената се одређују као:

$$t_i = \frac{t_{\max}}{\mu}\left(\left(1+\mu\right)^{\frac{i}{L}} - 1\right),\ i = 0,1,...,L, \tag{4.21}$$

а ширина ћелије унутар i-тог сегмента као:

$$\Delta_i = \frac{2t_{\max}L}{\mu N}\left(\left(1+\mu\right)^{\frac{i}{L}} - \left(1+\mu\right)^{\frac{i-1}{L}}\right),\ i = 1,...,L. \tag{4.22}$$

Прагови одлуке j-те ћелије унутар i-тог сегмента, t_{ij}, $i = 1,...,L$, $j = 1,...,n$, одређују се као [1]:

$$c_{\mathrm{PLC}}\left(t_{i,j}\right)=\frac{i-1}{L}t_{\max}+2j\frac{t_{\max}}{N}\Rightarrow t_{i,j}=c_{\mathrm{PLC}}^{-1}\left(\frac{i-1}{L}t_{\max}+2j\frac{t_{\max}}{N}\right)=t_{i-1}+j\Delta_i\,, \tag{4.23}$$

док се *j*-ти ниво унутар *i*-те ћелије (y_{ij}) одређује као [1]:

$$c_{\mathrm{PLC}}\left(y_{i,j}\right)=\frac{i-1}{L}t_{\max}+\frac{\left(2j-1\right)}{N}t_{\max}\Rightarrow y_{i,j}=c_{\mathrm{PLC}}^{-1}\left(\frac{i-1}{L}t_{\max}+\frac{\left(2j-1\right)}{N}t_{\max}\right)=t_{i-1}+\frac{\left(2j-1\right)}{2}\Delta_i\,. \tag{4.24}$$

Дисторзија PLCSQ-а састоји се из грануларне дисторзије D_g и дисторзије прекорачења D_o. Да би се D_g и D_o процениле користе се следећи изрази [1]:

$$D_g=2\sum_{i=1}^{L}\frac{\Delta_i^2}{12}\int_{t_{i-1}}^{t_i}p\left(x,\sigma\right)dx\,, \tag{4.25}$$

$$D_o=2\int_{t_{\max}}^{\infty}\left(x-y_{L,n}\right)^2\,p\left(x,\sigma\right)dx\,, \tag{4.26}$$

где $y_{L,n}$ означава *n*-ти ниво унутар *L*-тог сегмента PLCSQ-а (израз (4.24)). За Лапласову PDF и $\sigma^2=\sigma_{\mathrm{ref}}^2=1$ изводе се следећи изрази за D_g и D_o:

$$\begin{aligned}D_g=\frac{L^2t_{\max}^2}{3N^2\mu^2}\sum_{i=1}^{L}\left(1+\mu\right)^{\frac{2(i-1)}{L}}\left(\left(1+\mu\right)^{\frac{1}{L}}-1\right)^2\times\exp\left\{-\frac{\sqrt{2}t_{\max}}{\mu}\left(\left(1+\mu\right)^{\frac{i-1}{L}}-1\right)\right\}\\ \times\left(1-\exp\left\{-\frac{\sqrt{2}t_{\max}}{\mu}\left(\left(1+\mu\right)^{\frac{i}{L}}-\left(1+\mu\right)^{\frac{i-1}{L}}\right)\right\}\right)\end{aligned}\,, \tag{4.27}$$

$$D_o=\left(\left(\frac{\left(1+\mu\right)t_{\max}\,L\left(1-\left(1+\mu\right)^{-\frac{1}{L}}\right)}{\mu N}+\frac{1}{\sqrt{2}}\right)+\frac{1}{2}\right)\times\exp\left\{-\sqrt{2}t_{\max}\right\}. \tag{4.28}$$

Нивои PLCSQ-а се кодују кодним речима фиксне дужине па је битска брзина дефинисана са $R=\log_2 N=\log_2(2\cdot L\cdot n)$ bps. У процесу кодовња се један бит користи за знак (позитивни или негативни део PLCSQ-а), док се *i*-ти сегмент ком припада вредност улазног одмерка кодује природним бинарним кодом са $\log_2 L$ битова, а ниво на који се вредност одмерка квантује се такође кодује природним бинарним кодом са $\log_2 n$ битова.

Специјалан случај PLCSQ-а. За $\mu = \mu^{\mathrm{G.711}} = 255$ и $N = 256$ ($L = 8$ и $n = 16$) PLCSQ је еквивалентан квантизеру који је имплементиран у препоруци G.711 [89], а познат је и под називом G.711 квантизер. Важна особина овог квантизера јесте да је сваки наредни сегмент дупло већи од претходног. Пошто максимална амплитуда квантизера није специфицирана у препоруци G.711, $t_{\max}{}^{\mathrm{G.711}}$ се одређуе као [108]:

$$t_{\max}^{G.711} = \sigma_{\mathrm{ref}} \frac{1}{\sqrt{2}} \log\left(\frac{3\mu^{G.711} N^2}{\log^2\left(\mu^{G.711}+1\right)} \right). \tag{4.29}$$

Пројектовање DMSQ. DMSQ се користи за процесирање фрејмова, а за дати фрејм алтернативно користи ограничен и неограничен PLCSQ (G.711 квантизер) са истим N. Због тога је потребно пренети и индекс коришћеног PLCSQ-а, једном по фрејму. Ограничени PLCSQ се пројектује за сигнале ограничене по амплитуди односно за фрејмове чије вредности се налазе унутар његовог грануларног региона, док се неограничени PLCSQ пројектује за остале фрејмове.

Дисторзија DMSQ-а се дефинише на следећи начин [16, 17, 109]:

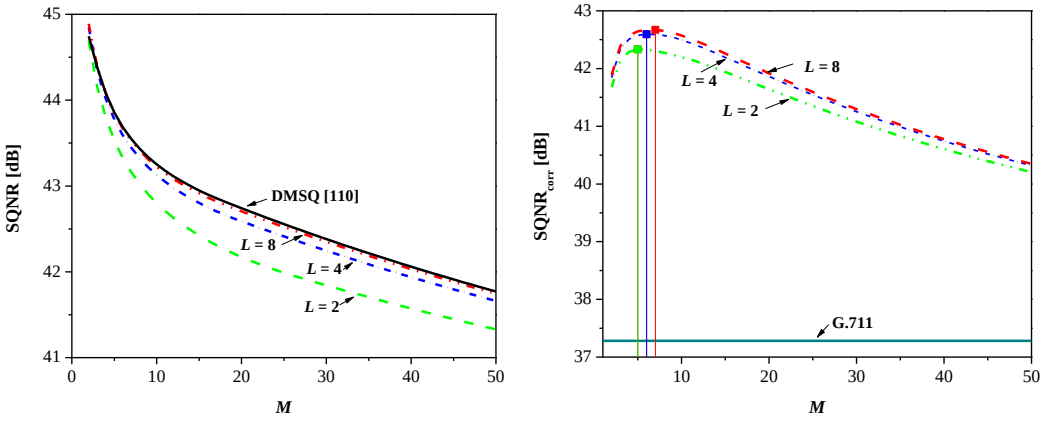
$$D = P^r D^r + \left(1 - P^r\right) D^{ur}, \tag{4.30}$$

где је D^r грануларна дисторзија ограниченог PLCSQ-а (израз (4.27)), D^{ur} означава укупну дисторзију неограниченог PLCSQ-а (збир грануларне дисторзије (4.27) и дисторзије прекорачења (4.28)), а P^r је дато са:

$$P^r = \left(2\int_0^{t_{\max}^r} p(x)\,dx \right)^M = \left(1 - \exp\left\{-\sqrt{2}t_{\max}^r\right\}\right)^M, \tag{4.31}$$

и представља вероватноћу избора ограниченог PLCSQ-а чија је максимална амплитуда $t_{\max}{}^r$. За дато M (дужина фрејма) D зависи од $t_{\max}{}^r$ and μ^r, с обзиром да је G.711 квантизер у потпуности дефинисан. Оптималне вредности за ове параметре се одређују из услова:

$$\frac{\partial D}{\partial t_{\max}^r} = 0 \;\Rightarrow\; t_{\max}^r = t_{\max}^{r,\mathrm{opt}};\; \frac{\partial D}{\partial \mu^r} = 0 \;\Rightarrow\; \mu^r = \mu^{r,\mathrm{opt}}. \tag{4.32}$$



Сл. 4.10 Перформансе DMSQ-а за оптимално пројектован ограничен PLCSQ у функцији од M: а) SQNR и б) $\mathrm{SQNR}_{\mathrm{corr}}$.

Због додатне информације која се јавља, битска брзина за DMSQ је:

$$R_{\mathrm{dm}} = \log_2 N + \frac{1}{M}. \tag{4.33}$$

Дакле, M је веома важан параметар јер утиче и на P^r (тј. на SQNR) и на R_{dm}.

На слици 4.10-а) приказана је зависност SQNR-а од M за DMSQ са $N = 256$ нивоа када је ограничени PLCSQ оптимално пројектован за различито L ($t_{\max}{}^r$ and μ^r су одређени према (4.42)). Када PLCSQ користи $c_{\mathrm{PLC}}(x)$ са $L = 8$ сегмената, перформансе DMSQ-а конвергирају ка горњој граници која је дата у [110]. Да би се одредило оптимално M у обзир се узимају и перформансе G.711 квантизера ($R = 8$ bps). DMSQ има битску брзину већу за $1/M$ bps (израз (4.43)), па се дефинише:

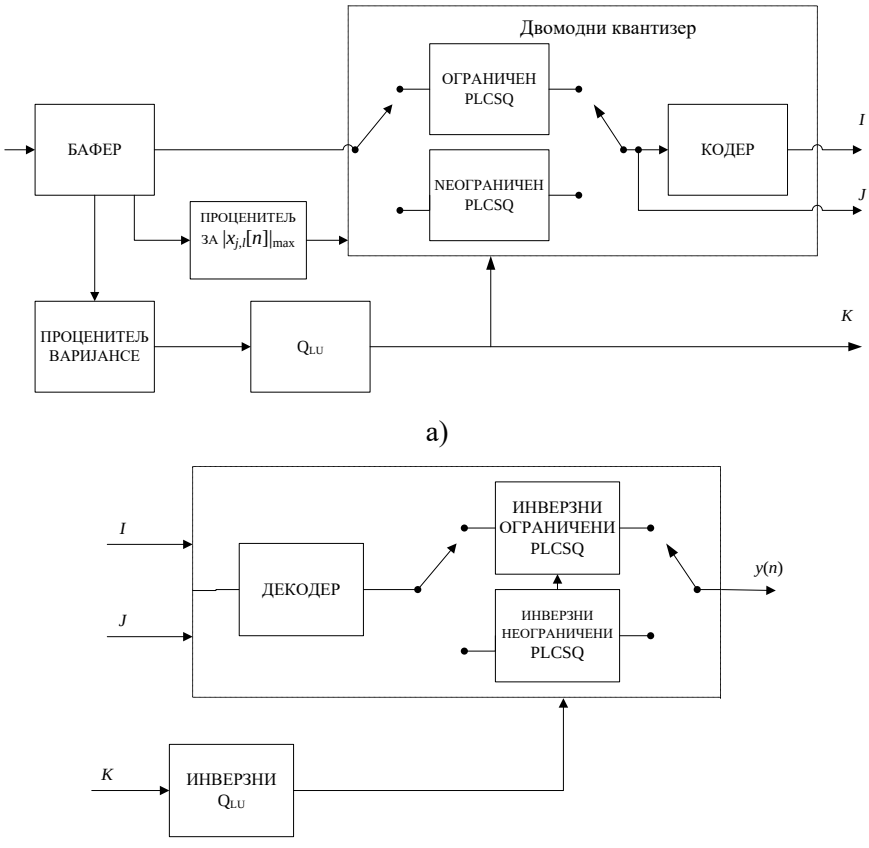
$$\mathrm{SQNR}_{\mathrm{corr}} = \mathrm{SQNR} - \frac{6\mathrm{dB}}{M}, \tag{4.34}$$

јер SQNR са једним додатним битом расте за око 6 dB код високорезолуционе квантизације [1, 2, 8].

Зависност $\mathrm{SQNR}_{\mathrm{corr}}$ од M је дата на слици 4.10-б). Оптимално M се бира тако да добитак у односу на G.711 квантизер буде највећи, што зависи од дизајна ограниченог PLCSQ-а. Конкретно, када је ограничени PLCSQ реализован са $L = 2$, $L = 4$ и $L = 8$ сегмената, оптималне вредности за DMSQ су $M = 5$ ($t_{\max}{}^{r,\mathrm{opt}} = 2.92$,

$\mu^{r,\mathrm{opt}} = 2.27$), $M = 6$ ($t_{\max}{}^{r,\mathrm{opt}} = 3.17$, $\mu^{r,\mathrm{opt}} = 2.87$) и $M = 7$ ($t_{\max}{}^{r,\mathrm{opt}} = 3.32$, $\mu^{r,\mathrm{opt}} = 3.21$), респективно. При томе, DMSQ даје преко 5 dB већи SQNR од G.711 квантизера. За тако дефинисане параметре, ограничени PLCSQ се у оквиру DMSQ-а бира са вероватноћама од $P^r = 0.922$ када је $L = 2$, $P^r = 0.934$ када је $L = 4$ и $P^r = 0.938$ када је $L = 8$, што је јако повољно.

PCM кодек са двомодним квантизером. Блок шема кодера и декодера за ово PCM решење је дата на слици 4.11. Процесирање сигнала се врши применом фрејм/подфрејм логике, где се адаптација ограниченог и неограниченог PLCSQ-а врши на нивоу фрејма а прилагођење на максималну амплитуду (избор између два PLCSQ-а) се врши на нивоу подфрејма. Дакле, направљена је мала модификација у односу на поступак из одељка 4.1.



Сл. 4.11 Блок шема PCM кодека са двомодним квантизером: а) кодер и б) декодер.

Укратко, дати PCM кодек ради на следећи начин. Текући, j-ти фрејм дужине M_0 одмерака се дели на подфрејмове $x_{j,l}[n]$, $l = 1,\ldots, M_0/M$, $n = 1,\ldots, M$, где l означава индекс подфрејма унутар j-тог фрејма, а процесирање се врши са једним од два

доступна PLCSQ-а. Прво се врши процена максималне апсолутне амплитуде подфрејма $|x_{j,l}[n]|_{\max}$. Ако је $|x_{j,l}[n]|_{\max} \leq t_{\max}{}^{r}(g_j)$ бира се адаптивни ограничени PLCSQ, а у супротном бира се адаптивни G.711 квантизер. Иначе, адаптација PLCSQ-а се врши на начин како је описано у одељку 4.1.

Битска брзина за овај тип PCM кодека је дата са:

$$R = R_{\text{dm}} + \frac{\log_2 L}{M_0} \text{ [bps]}. \tag{4.35}$$

Анализа теоријских резултата. У табели 4.3 сумиране су вредности за SQNR_{av} (израз (2.12)) и за R за разматрани PCM кодек (N = 256) када се број нивоа квантизера Q_{LU} мења, а посматра се опсег варијанси [-20dB, 20dB] око референтне $\sigma_{\text{ref}}{}^2 = 1$ а величина фрејма је $M_0 = 240$.

Табела 4.3 Вредности за SQNR_{av} и R за PCM кодек са двомодним квантизером (N = 256) за различито L (M_0 = 240)

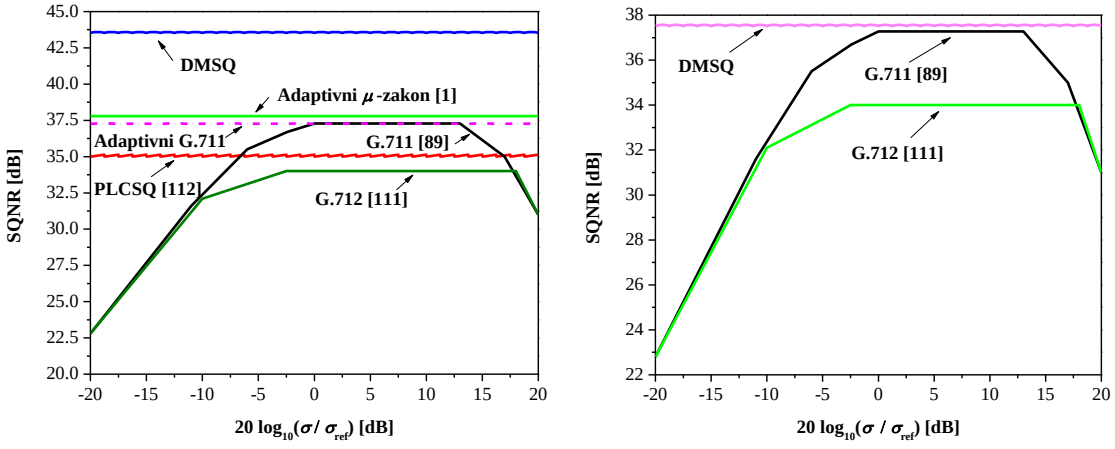
L (R_L[bps])	SQNR_{av} [dB]	R [bps]
2 (1)	36.9920	8.1708
4 (2)	41.8159	8.1750
8 (3)	43.2025	8.1792
16 (4)	43.4956	8.1833
32 (5)	**43.5670**	**8.1875**
64 (6)	43.5848	8.1917
128 (7)	43.5893	8.1958
256 (8)	43.5904	8.2000

L се бира помоћу следећег критеријума:

$$\frac{\text{SQNR}_{av}\left(N, M_0, L\right) - \text{SQNR}_{av}\left(N, M_0, L/2\right)}{R\left(N, M_0, L\right) - R\left(N, M_0, L/2\right)} \geq 6\frac{\text{dB}}{\text{bit}}. \tag{4.36}$$

Односно, L (тј. R_L) треба повећавати све док се SQNR_{av} повећава за 6 dB или више са порастом броја битова за један. На основу (4.36), из табеле 4.3 бира се L = 32 (R_L = 5 bps).

Слика 4.12-а) пореди перформансе предложеног PCM кодека са двомодним квантизером (N = 256, M_0 = 240, M = 6, Q_{LU} са L = 32 нивоа) у односу на препоруке G.711 [89] и G.712 (дефинише минималну SQNR вредност у системима за високо

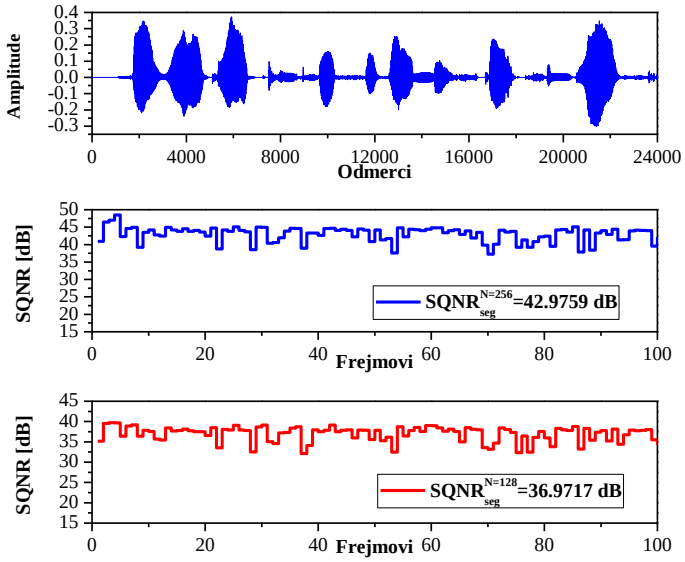


а) б)

Сл. 4.12 Перформансе PCM кодека са двомодним квантизером (M_0 = 240, M = 6, Q_{LU} са L = 32 нивоа) у односу на друга PCM решења за: а) N = 256 (R = 8.1785 bps) и б) N = 128 (R = 7.1785 bps).

-квалитетни пренос говора) [111] али и постојећа PCM решења из рада [1] (adaptivni μ-zakon [1])] и рада [112] (PLCSQ [112]). Јасно се види да предложени PCM кодек у потпуности задовољава наведене препоруке, а у исто време остварује добитак од 5.77 dB и 8.5 dB у односу решења из [1] и [112], респективно, и 6 dB у односу на PCM са G.711 квантизером (adaptivni G.711). Добитак је остварен по цену повећане битске брзине за 0.1666 bps (1 бит по подфрејму). Слика 4.12-б) показује да је могуће задовољити обе наведене препоруке и када PCM кодек користи DMSQ са N = 128 нивоа односно при брзини од R = 7.1785 bps, што је за 0.8125 bps мање него код PCM кодека са G.711 квантизером.

Примена на говорни сигнал. Тест говорни сигнал је екстрахован из TIMIT базе [113] а има дужину од 24 000 одмерака (фреквенција одмеравања је 8 kHz). SQNR на нивоу фрејма за PCM кодек са двомодним квантизером са 128 и 256 нивоа је приказан на слици 4.13, а добијени сегментни SQNR се јако добро слаже са теоријским резултатима са слике 4.12.



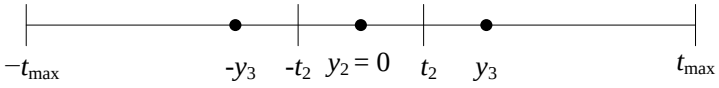
Сл. 4.13 SQNR по фрејмовима дужине 240 одмерака за PCM са двомодним квантизером са N = 128 и N = 256 нивоа (M = 6, Q_{LU} са L = 32 нивоа).

4.5 PCM кодек са двомодним ограниченим квантизерима

У овом одељку ће бити речи о PCM кодеку са двомодним квантизером који користи два тернарна квантизера са Хафмановим кодом пројектована за ограничен извор (ограничен квантизер), а представљен је у раду [68].

Опис тернарног DMSQ. Овај DMSQ се састоји од два ограничена тернарна квантизера означена са Q_{R1} и Q_{R2} који су дефинисани параметрима t_{max1} и t_{max2} ($t_{max1} < t_{max2}$), респективно, а избор између расположивих квантизера дефинисан је посебним правилом. Опис ограниченог тернарног квантизера дат је у наставку.

Опис тернарног квантизера. Симетрични (ограничени) тернарни квантизер Q_R је приказан на слици 4.14, а за кодовање нивоа користи се Хафманов код.



Сл. 4.14 Модел ограниченог тернарног квантизера.

Са $t_{\max}$ означена је максимална амплитуда квантизера која се поклапа са максималном амплитудом сигнала који се моделује са ограниченим Лапласовим извором чија је PDF дата са [1, 2]:

$$p(x,\sigma)=\frac{1}{\sqrt{2}\sigma}\exp\left\{-\frac{\sqrt{2}|x|}{\sigma}\right\}\Big/\left(1-\exp\left\{-\frac{\sqrt{2}t_{\max}}{\sigma}\right\}\right). \tag{4.37}$$

Дисторзија D има само грануларну компоненту D_g јер у овом случају дисторзија прекорачења D_o не постоји. D_g је дата са:

$$D(\sigma)=D_g(\sigma)=2\int_0^{t_2}x^2p(x,\sigma)dx+2\int_{t_2}^{t_{\max}}(x-y_3)^2\,p(x,\sigma)dx\,. \tag{4.38}$$

SQNR се дефинише на следећи начин:

$$\mathrm{SQNR}(\sigma)=10\log_{10}\left(\frac{P_{sor}(\sigma)}{D(\sigma)}\right), \tag{4.39}$$

где је $P_{sor}(\sigma)$ снага извора дата са [1, 2]:

$$P_{sor}(\sigma)=2\int_0^{t_{\max}}x^2p(x,\sigma)dx\quad. \tag{4.40}$$

Вероватоће нивоа ограниченог тернарног квантизера се одређују као:

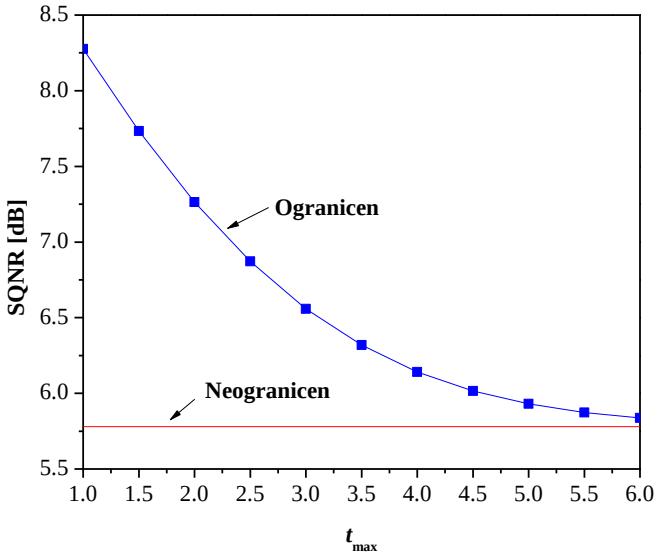
$$p(y_3)=p(-y_3)=\int_{t_2}^{t_{\max}}p(x,\sigma)dx=\frac{1}{2}-\left(\exp\left\{-\frac{\sqrt{2}t_2}{\sigma}\right\}-1\right)/2\left(\exp\left\{-\frac{\sqrt{2}t_{\max}}{\sigma}\right\}-1\right), \tag{4.41}$$

$$p(y_2)=2\int_0^{t_2}p(x,\sigma)dx=\left(\exp\left\{-\frac{\sqrt{2}t_2}{\sigma}\right\}-1\right)\Big/\left(\exp\left\{-\frac{\sqrt{2}t_{\max}}{\sigma}\right\}-1\right), \tag{4.42}$$

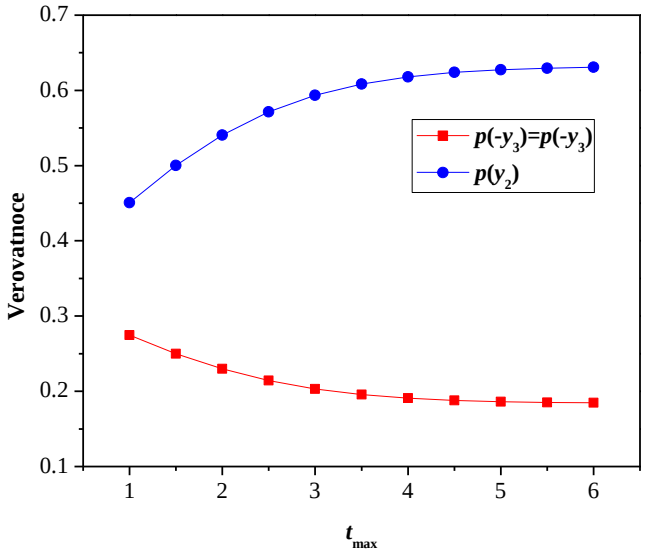
а за прорачун битске брзине користи се израз (2.10).

Разматрани квантизер се за $\sigma^2=\sigma_{\mathrm{ref}}^2=1$ пројектује уз помоћ Лојд-Макс алгоритма (одељак 2.3), што значи да се t_2 и y_3 за дато $t_{\max}$ одређују итеративно. На слици 4.15 су приказане добијене SQNR вредности за различито $t_{\max}$. У поређењу са

тернарним квантизером оптимално пројектованим за неограничен Лапласов извор (израз (2.6)) који је означен са Q_{UR}, Q_R даје већи SQNR посебно за мање вредности t_{max}. Слика 4.16 даје зависност вероватноћа од t_{max}. $p(y_2)$ има веће вредности од $p(y_3)$ и $p(-y_3)$, што значи да Хафманов кодер нивоу y_2 додељује кодну реч **'0'** а преосталим нивоима кодне речи '**10**' и '**11**'.



Сл. 4.15 SQNR за ограничени тернарни квантизер добијен за различите вредности t_{max}.

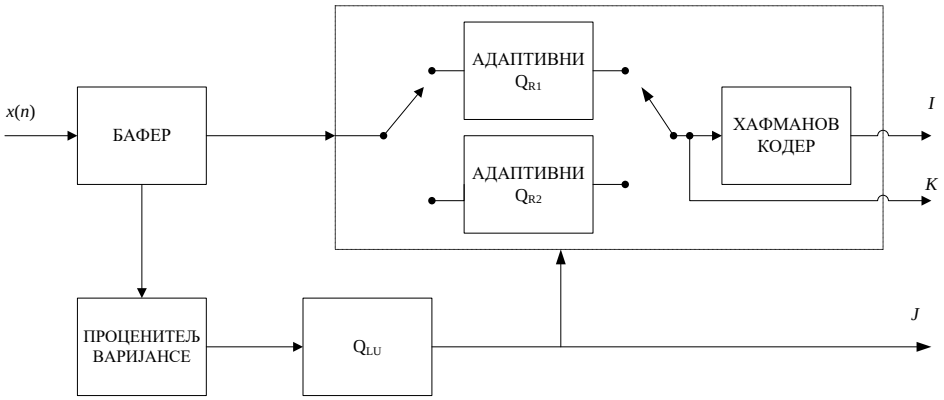


Сл. 4.16 Вероватноће у функцији од t_{max} за ограничени тернарни квантизер.

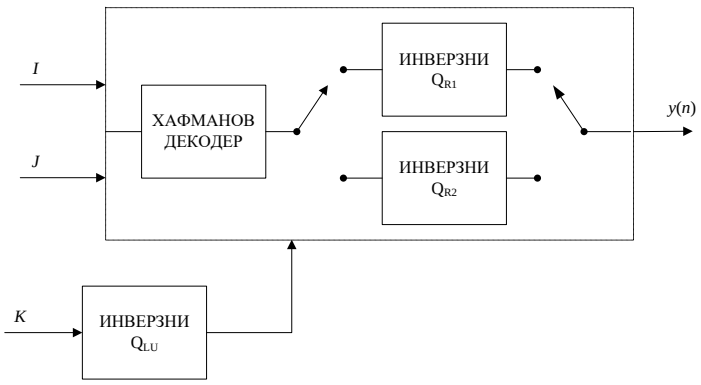
PCM кодек са тернарним DMSQ. На слици 4.17 дата је блок шема за разматрани PCM кодек. Адаптација тернарних квантизера Q_{R1} и Q_{R2} се одвија на начин како је описано у одељку 4.1, а селекција за *j*-ти фрејм се врши у зависности од тога који квантизер остварује мању вредност дисторзије. Информација о коришћеном квантизеру се кодује са једним битом и индексом *K* шаље до декодера. Битска брзина је дата са:

$$R = R_t + \frac{1+\log_2 L}{M_0} \text{ [bps]}, \tag{4.43}$$

где је са R_t означена битска брзина за ограничени тернарни квантизер.



а)

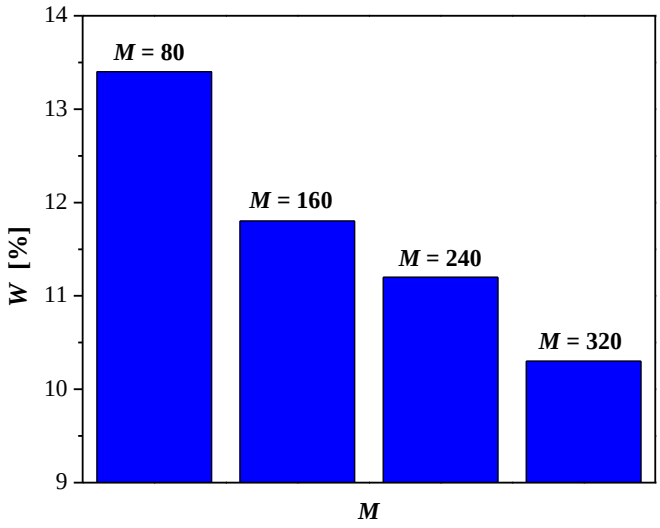


б)

Сл. 4.17 Блок шема PCM кодека са тернарним DMSQ: а) кодер и б) декодер.

Примена на говорни сигнал. За процену тежина *w* која дефинише удео Q_{R1} у оквиру DMSQ-а користи се тренинг секвенца од приближно милион одмерака говора одмераваним на 16 kHz (реченице изговорене на српском језику од стране

мушког говорника). На слици 4.18 су дати резултати процене за различито M. w се благо смањује како се M повећава, што значи да се Q_{R1} ређе користи за дуже фрејмове.



Сл. 4.18 Проценат фрејмова за које се користи Q_{R1} у зависности од M.

Говор од 66 500 одмерака (фреквенција одмеравања је 16 кHz) који није био укључен у тренинг секвенцу се користи као тест сигнал. Перформансе описаног PCM кодека се на тест сигналу процењују се помоћу еквивалентног SQNR-а:

$$\mathrm{SQNR} = w \cdot \mathrm{SQNR}_1 + (1-w)\mathrm{SQNR}_2 , \tag{4.44}$$

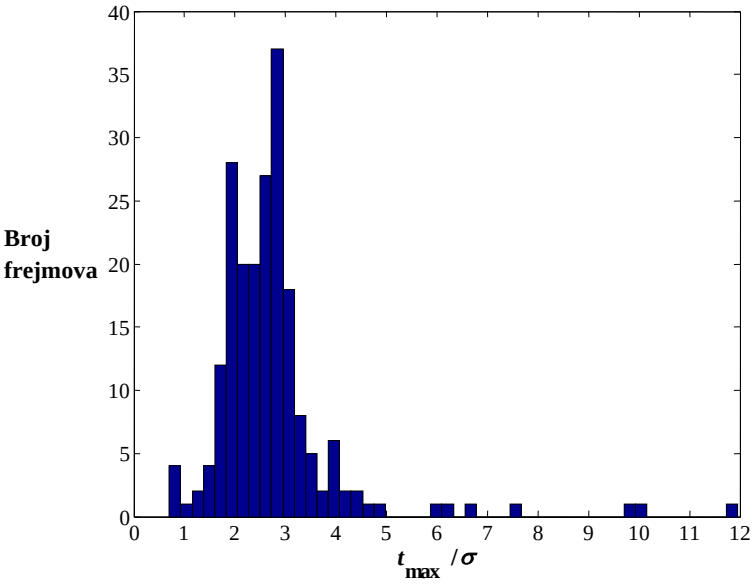
где SQNR_1 и SQNR_2 означавају SQNR (израз (4.7)) добијен за адаптивне Q_{R1} и Q_{R2}, респективно.

Табела 4.4 даје SQNR и R за овај PCM кодек ($t_{\max 1} = 1.1$, $t_{\max 2} = 3$, Q_{LU} са $L = 32$ нивоа), за различито M. Параметри тернарног DMSQ су подешени на те вредности због тога што тада PCM кодек постиже највећи SQNR. Заиста, ако се погледа слика 4.19 где је дата статистичка расподела за $t_{j,\max}/\sigma_j$ (максимална вредност фрејма нормализована са варијансом фрејма) види се да за велику већину фрејма важи $1.5 < t_{j,\max}/\sigma_j < 3$, при чему највише има фрејмова код којих је $t_{j,\max} / \sigma_j = 3$. PCM кодек са квантизером Q_{UR} као и са Лојд-Макс квантизерима са два ($Q_{N=2}$) и четири нивоа ($Q_{N=4}$) користи се за поређење. Најбољи SQNR се добија за $M = 80$. Када се упореди са PCM кодеком који користи Q_{UR}, предложени кодек постиже за

1.43 dB већи SQNR. Осим тога, у односу на PCM кодек са $Q_{N=2}$ остварен је за 3.26 dB већи SQNR а у поређењу са PCM кодеком са $Q_{N=4}$ постигнут је скоро исти SQNR уз мању битску брзину за 0.562 bps.

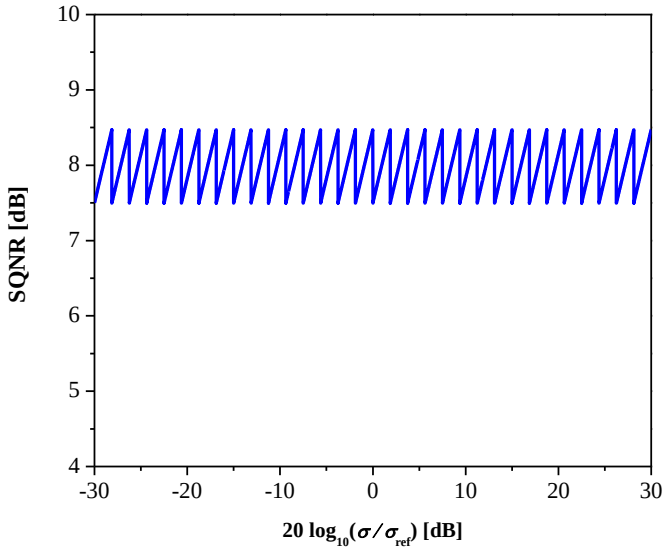
Табела 4.4 Перформансе предложеног PCM кодека и других сличних PCM решења, за различите дужине фрејма

	Предложени PCM кодек (t_{max1}=1.1,t_{max2}=3)		**PCM са Q_{UR}**		**PCM са $Q_{N=2}$**		**PCM са $Q_{N=4}$**	
	SQNR [dB]	***R* [bps]**	**SQNR [dB]**	***R* [bps]**	**SQNR [dB]**	***R* [bps]**	**SQNR [dB]**	***R* [bps]**
***M*=80**	7.856	1.500	6.417	1.429	4.597	1.062	7.857	2.062
***M*=160**	7.465	1.459	6.279	1.398	4.385	1.031	7.872	2.031
***M*=240**	7.309	1.446	6.389	1.387	4.318	1.021	7.972	2.021
***M*=320**	7.167	1.439	6.346	1.382	4.227	1.016	7.928	2.016



Сл. 4.19 Хистограм t_{max}/σ за фрејмове од 320 одмерака.

На слици 4.20 дат је теоријски SQNR за разматрани PCM кодек (w = 0.132, t_{max1} = 1.1, t_{max2} = 3, Q_{LU} са L = 32 нивоа). Теоријски SQNR се добро подудара са резултатима из табеле 4.4.



Сл. 4.20. SQNR у широком динамичком опсегу варијансе улазног сигнала за PCM кодек са тернарним DMSQ (t_{max1} = 1.1, t_{max2} = 3, Q_{LU} са L = 32 нивоа).

5. Пројектовање квантизера за алгоритме засноване на ADM за говорни сигнал

У овом поглављу дисертације биће описани нови предиктивни алгоритми на бази ADM-а, где се врши кодовање сигнала грешке предикције (разлика између оригиналних и предвиђених вредности сигнала) а не оригиналних вредности сигнала као у претходном поглављу. Алгоритми користе адаптацију на нивоу фрејма, што је новина у односу на постојећа ADM решења. На почетку ће бити представљени ADM алгоритми за Лапласов извор. У оквиру тога биће анализирана тернарна ADM са новим тернарним квантизером са Хафмановим кодом и прекидачким линеарним предиктором (LP) првог реда, која значајно поправља перформансе базичног ADM алгоритма. Затим ће бити описана двобитна ADM са оптималним двобитним квантизером и адаптивним LP-ом првог реда, која унапређује перформансе напреднијих решења из класе тренутно адаптивне ADM. Након тога биће разматран дводигитни ADM са новим неунформним квантизером са шест нивоа и адаптивним LP-ом првог реда који има боље перформансе од постојећих дводигитних ADM кодека. Последње решење за Лапалсов извор које ће бити представљено односи се на вишенивовску ADM са адаптивним G.711 квантизером и прекидачким LP-ом првог реда, којим је могуће поправити перформансе неадаптивног односно адаптивног G.711 квантизера. За Гаусов извор биће предложена два ADM алгоритма. Најпре ће бити предложена тернарна ADM са новим тернарним квантизером који примењује Хафманов код и прекидачким LP-ом другог реда, а која може да да веће перформансе од неких популарних ADM решења. На крају ће бити анализиран ADM са новим двобитним униформним квантизером и фракционим линеарним предиктором, којим је могуће остварити значајна побољшања неких напреднијих алгоритама сличне комплексности из класе тренутно адаптивне ADM.

5.1 Линеарни предиктори првог и другог реда

У овом одељку ће бити дат опис модела LP-а првог и другог реда, с обзиром да се доминантно користе при анализи предложених ADM алгоритама.

Предикција у обради сигнала је математички модел где се будуће вредности дискретног (по времену) сигнала процењују на основу претходних вредности сигнала [1−8, 122−124]. Код LP-а првог реда користи се историја од једног одмерка за предикцију тренутног одмерка сигнала [1, 3, 8]:

$$x_p(n) = a \cdot x(n-1), \tag{5.1}$$

где је a коефицијент LP-а. Грешка предикције се дефинише као разлика између стварне $x(n)$ и процењене вредности одмерка $x_p(n)$:

$$e(n) = x(n) - x_p(n) = x(n) - a \cdot x(n-1). \tag{5.2}$$

За процену перформанси предиктора користи се средња квадратна грешка предикције σ_e^2 [1, 3, 8]. За LP првог реда σ_e^2 је дата са:

$$\begin{aligned} \sigma_e^2 &= E\left[\left(x(n) - x_p(n)\right)^2\right] \\ &= \sigma_x^2\left(1 - 2a\rho_1 + a^2\right) \end{aligned}, \tag{5.3}$$

где је $E[\cdot]$ математичко очекивање, σ_x^2 је варијанса сигнала а ρ_1 је коефицијент корелације првог реда. Оптимално a се добија као решење једначине $\partial\sigma_e^2 / \partial a = 0$, што даје [1, 3, 8]:

$$a = \rho_1. \tag{5.4}$$

Заменом (5.4) у (5.3) добија се минимална средња квадратна грешка предикције:

$$\sigma_{e,\min}^2 = \sigma_x^2\left(1 - a^2\right). \tag{5.5}$$

За LP првог реда, добитак предикције G је дат са [1, 3, 8]:

$$G = 10\log_{10}\left(\frac{\sigma_e^2}{\sigma_x^2}\right) = 10\log_{10}\left(\frac{1}{1-a^2}\right). \tag{5.6}$$

Код LP-а другог реда користи се историја од два одмерка за предикцију тренутног одмерка сигнала [1, 3, 8]:

$$x_p(n) = a_1 \cdot x(n-1) + a_2 \cdot x(n-2), \tag{5.7}$$

где су a_1 и a_2 коефицијенти LP-а. Грешка предикције је дата са:

$$e(n) = x(n) - x_p(n) = x(n) - a_1 \cdot x(n-1) - a_2 \cdot x(n-2), \tag{5.8}$$

а σ_e^2 за LP другог реда је:

$$\begin{aligned} \sigma_e^2 &= E\left[\left(x[n] - \hat{x}[n]\right)^2\right] \\ &= \sigma_x^2\left(1 + a_1^2 + a_2^2 - 2a_1\rho_1 - 2a_2\rho_2 + 2a_1a_2\rho_1\right) \end{aligned}, \tag{5.9}$$

где је ρ_2 је коефицијент корелације другог реда. Оптимални коефицијенти a_1 и a_2 се израчунавају као [1, 3, 8]:

$$a_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2},\ a_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}. \tag{5.10}$$

Заменом (5.10) у (5.9) добија се:

$$\sigma_{e,\min}^2 = \sigma_x^2\left(1 - \sum_{i=1}^{2} a_i\rho_i\right), \tag{5.11}$$

а добитак предикције у овом случају је [1, 3, 8]:

$$G = 10\log_{10}\left(\frac{\sigma_e^2}{\sigma_x^2}\right) = 10\log_{10}\left(\frac{1}{1 - \sum_{i=1}^{2} a_i\rho_i}\right). \tag{5.12}$$

5.2 Евалуационе метрике код кодовања говора

Сви алгоритми ће бити тестирани на говорном сигналу. Као објективна мера перформанси користи се сегментни SNR [1, 3]:

$$\mathrm{SNR} = \frac{1}{F}\sum_{j=1}^{F} 10\log_{10}\left(\frac{1/M\sum_{n=1}^{M} x_j^2(n)}{1/M\sum_{n=1}^{M}\left(x_j(n)-y_j(n)\right)}\right), \tag{5.13}$$

где су $x_j(n)$ оригинални одмерци а $y_j(n)$ реконструисани одмерци j-тог фрејма. Код предиктивних алгоритама као што је ADM, SNR се може разложити на две компоненте, SQNR и G који се рачунају као [1, 3]:

$$\mathrm{SQNR} = \frac{1}{F}\sum_{j=1}^{F} 10\log_{10}\left(\sum_{i=1}^{M} e_j^2(n)/\sum_{i=1}^{M}\left(e_j(n)-e_{q,j}(n)\right)\right), \tag{5.14}$$

$$G = \frac{1}{F}\sum_{j=1}^{F} 10\log_{10}\left(\sum_{i=1}^{M} x_j^2(n)/\sum_{i=1}^{M} e_j^2(n)\right), \tag{5.15}$$

где $e_j(n)$ означава одмерке сигнала грешке предикције а њихове квантоване вредности (нивои квантизера) су означене са $e_{q,j}(n)$.

5.3 ADM алгоритми за кодовање Лапласовог извора

5. 3.1 Тернарна ADM

Овај ADM алгоритам је описан у раду [69], а базира се на тернарном квантизеру са промењливом дужином кодних речи и прекидачким LP-ом првог реда.

Опис и пројектовање тернарног квантизера. Симетрични тернарни квантизер са Хафановим кодом је већ описан и анализиран у одељку 3.4, за Гаусову PDF. Сада ће пројектовање (одређивање параметара t_2 и y_3) бити урађено за Лапласову PDF и $\sigma^2 = \sigma_{\mathrm{ref}}^2 = 1$.

Вероватноће нивоа датог квантизера се за Лапласову PDF израчунавају помоћу следећих формула:

$$p(-y_3) = p(y_3) = \int_{t_2}^{\infty} p(x)dx = \frac{1}{2}\exp\left\{-\sqrt{2}t_2\right\}, \tag{5.16}$$

$$p(y_2) = 2\int_{0}^{t_2} p(x)dx = 1 - \exp\left\{-\sqrt{2}t_2\right\}. \tag{5.17}$$

и зависе од t_2. Дисторзија D се рачуна као:

$$D = 2\int_{0}^{t_2} x^2 p(x)dx + 2\int_{t_2}^{\infty} (x - y_3)^2 p(x)dx. \tag{5.18}$$

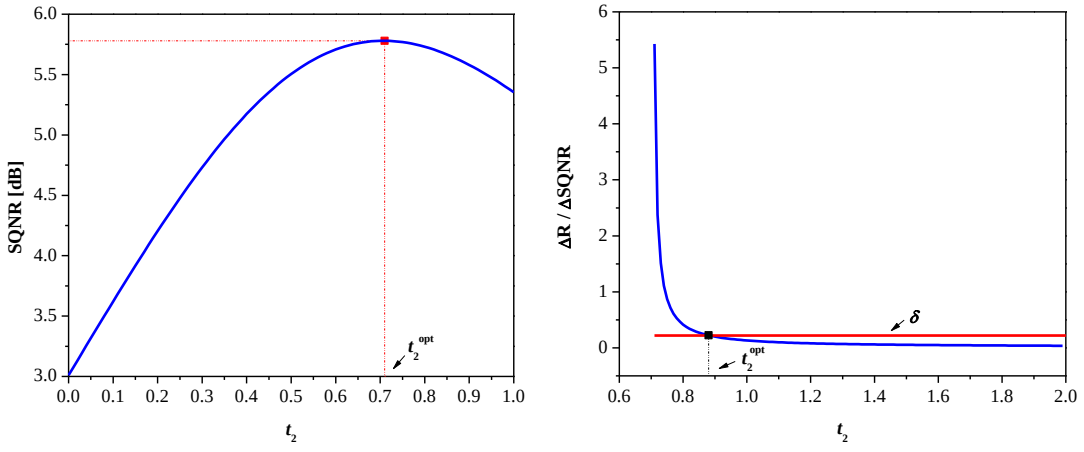
Ако се y_3 одреди из правила центроида (израз (2.16)):

$$y_3 = \frac{\int_{t_2}^{\infty} xp(x)dx}{\int_{t_2}^{\infty} p(x)dx} = t_2 + \sqrt{2}, \tag{5.19}$$

онда се израз за дисторзију своди на:

$$D = 1 - \left(\frac{1}{2} + \sqrt{2}t_2 + t_2^2\right)\exp\left\{-\sqrt{2}t_2\right\}, \tag{5.20}$$

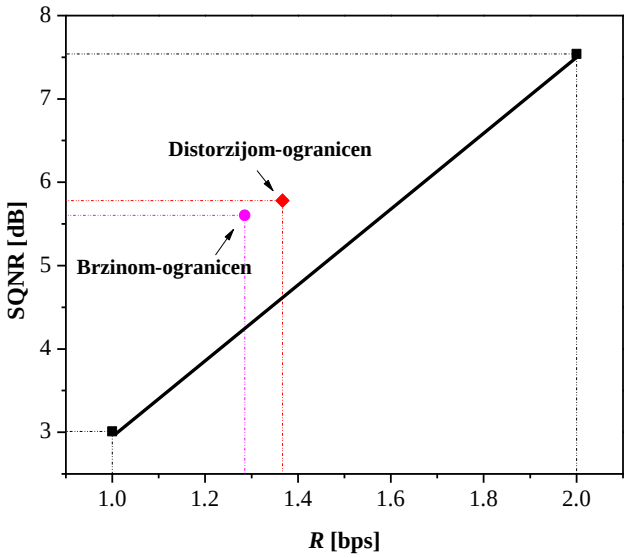
и такође зависи од t_2. Дакле, t_2 је кључни параметар а циљ је да се пројектују дисторзијом-ограничен и брзином-ограничен тернарни квантизер.



а) б)

Сл. 5.1 Избор оптималне вредности за t_2: а) оптимизација по дисторзији и б) оптимизација и по дисторзији и по брзини.

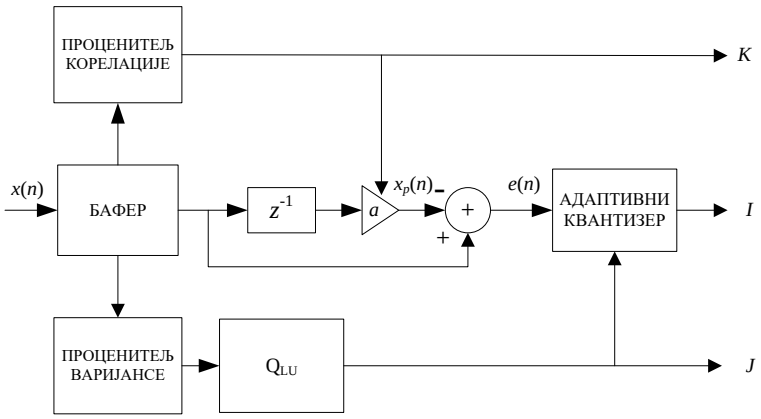
На слици 5.1 су дате оптималне вредности за t_2, за два примењена критеријума оптимизације. За дисторзијом-ограничен квантизер бира се вредност $t_2 = t_2^{opt} = 0.71$, јер је тада SQNR највећи (слика 5.1-а)). За брзином-ограничен квантизер (слика 5.1-б)) оптимално t_2 ($t_2^{opt} = 0.887$) је изабрано применом критеријума (3.11), при чему се користио опсег $t_2 \in (0.71, 2)$ док је $\delta = 0.2208$ (нагиб између Лојд-Макс кванизера са 2 (SQNR = 3.01 dB, R = 1 bps) и 4 нивоа (SQNR = 7.54 dB, R = 3 bps) [1−3, 8]).



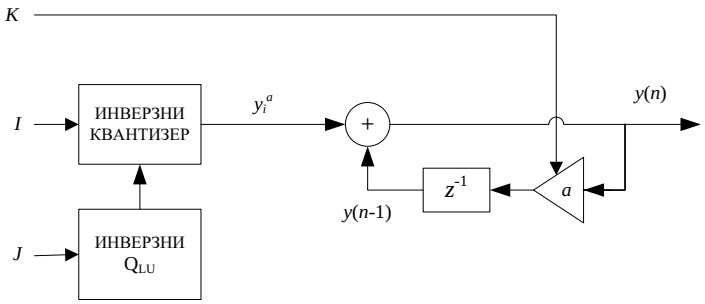
Сл. 5.2 Перформансе предложеног тернарног квантизера у поређењу са Лојд-Макс квантизером.

Слика 5.2 показује да су дисторзијом-ограничен (SQNR = 5.78 dB, R = 1. 3664 bps) и брзином-ограничен (SQNR = 5.60 dB, R = 1.2852 bps) тернарни квантизер боља решења од Лојд-Макс квантизера са истом битском брзином, јер могу да пруже већи SQNR за више од 1 dB.

Опис тернарне ADM. Блок шема кодера и декодера је дата на слици 5.3, а користе се адаптивни тернарни квантизер (адаптација се врши на варијансу фрејма) и прекидачки предиктор са два фиксна LP-а првог реда дефинисана са a_1 и a_2. Селекција између одговарајућих LP-а се за дати фрејм врши на основу коефицијента корелације ρ.



а)



б)

Сл. 5.3 Предложена тернарна ADM: а) кодер и б) декодер.

То је систем је са отвореном повратном спрегом чији се принцип рада може описати следећим корацима:

1. ***Баферовање фрејма.*** Исто као у кораку 1 алгоритма из одељка 4.1.
2. ***Процена и квантизација варијансе фрејма.*** Исто као у кораку 2 алгоритма из одељка 4.1.
3. ***Процена коефицијента корелације и одабир коефицијента предикције.*** За j-ти фрејм коефицијент корелације ρ_j се процењује као [1, 3, 8]:

$$\rho_j = \frac{\sum_{i=1}^{M-1} x_j(n)\, x_j(n+1)}{\sum_{i=1}^{M} x_j^2(n)}. \qquad (5.21)$$

Ако је $\rho_j < 0.8$ фрејм се класификује као слабо корелисан и у оквиру прекидачког предиктора бира се коефицијент a_1, иначе се фрејм сматра високо корелисаним и бира се коефицијент a_2. У циљу коректног

декодовања сигнала потребно је пренети информацију о изабраном LP-у једном по фрејму, користећи један бит (индекс *K*).

4. ***Одређивање сигнала грешке предикције.*** Сигнал разлике $e_j(n)$ се добија као: $e_j(n)=x_j(n)-a_k \cdot x_j(n\text{-}1)$, $k = 1, 2$.
5. ***Адаптивна тернарна квантизација.*** Адаптација тернарног квантизера се ради исто као у кораку 3 алгоритма из одељка 4.1, при чему је фактор скалирања дефинисан са [1]:

$$g_j = 10^{V_i/20}\sqrt{1-a_k}\ ,\ k=1,2\,. \tag{5.22}$$

   Адаптивним квантизером се квантује $e_j(n)$ и добија се квантовани сигнал грешке предикције $e_{q,j}(n) = y_i^a$ (индекс *I*).
6. ***Реконструкција сигнала.*** Декодер на основу индекса *I*, *J* и *K* декодује (реконструише) одмреке оригиналног фрејма. Реконструисани сигнал $y_j(n)$ за *j*-ти фрејм се добија као:

$$y_j(n) = a_k \cdot y_j(n-1) + y_i^a \operatorname{sgn}\left(e_j(n)\right),\ k=1,2,\ i=2,3\,. \tag{5.23}$$

7. ***Понављајти претходне кораке док сви фрејмови не буду обрађени.***

Битска брзина за тернарни ADM алгоритам је:

$$R_{\mathrm{DM}} = R + \frac{1+\log_2 L}{M}\ [\mathrm{bps}], \tag{5.24}$$

где R означава битску брзину тернарног квантизера док други члан представља додатну информацију.

Примена на говорни сигнал. Тренинг секвенца од приближно 3 минута говора одмераваног на фреквенцији од 16 kHz (реченице изговорене на српском језику од стране мушког говорника) користи се за одређивање коефицијената прекидачког предиктора. a_1 се добија као средња вредност коефицијента корелације у случају слабо корелисаних фрејмова ($\rho < 0.8$), док се a_2 добија као средња вредност у случају високо корелисаних фрејмова. За $M = 80$ добијено је $a_1 = 0.23$ и $a_2 = 0.95$.

Као тест сигнал користи се говор од 66500 одмерака који није био укључен у тренинг секвенцу. Користи се Q_{LU} са $L = 32$ нивоа.

Табела 5.1 Перформансе добијене на тест говорном сигналу за класичну, тернарну и двобитну ADM (M = 80)

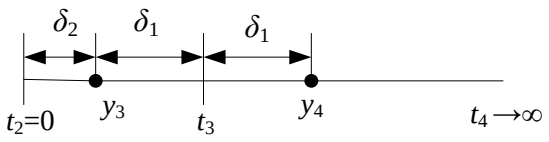
M = 80	Класична ADM (бинарни Лојд-Макс)	Тернарна ADM (дисторзијом-ограничен)	Тернарна ADM (брзином-ограничен)	Двобитна ADM (двобитни Лојд-Макс)
SNR [dB]	10.978	15.274	14.747	15.529
R_{DM} [bps]	1.062	1.441	1.355	2.062

У табели 5.1 су приказане перформансе (SNR и R) за предложену тернарну ADM када је M = 80. Као што се и очекивало, мало бољи SNR је остварен код ADM-а са дисторзијом-ограниченим квантизером. У сврху поређења дате су и перформансе за класични ADM и двобитни ADM, који користе одговарајуће адаптивне Лојд-Макс квантизере и фиксни LP првог реда (a = 0.95). Из табеле се види да тернарна ADM значајно поправља SNR класичне ADM (добитак је већи од 4 dB), а остварује приближно исти SNR и мању битску брзину (последица примене Хафмановог кода) у односу на двобитни ADM.

5.3.2 Двобитни ADM

У овом одељку представљен је ADM алгоритам из рада [70], који примењује оптимални двобитни неунифромни квантизер и LP првог реда.

Опис и пројектовање двобитног квантизера. Симетрични неуниформни квантизер са N = 4 нивоа ($R = \log_2 N$ = 2 bps) је приказан на слици 5.4.



Сл. 5.4 Симетрични неуниформни квантизер са N = 4 нивоа.

Пројектовање се врши за Лапласову PDF и $\sigma^2 = \sigma_{ref}^2 = 1$ а користи се нови итеративни метод који је једноставнији од Лојд-Макс алгоритма [1, 2]. δ_i на слици 5.4 означава офсет а дефинише се као растојање између одговарајућег нивоа и

доњег прага одлуке ($\delta_1 = y_4 - t_3$, $\delta_2 = y_3 - t_2 = y_3$). Штавише, δ_i (i =1, 2) у потпуности дефинишу разматрани квантизер јер се његови параметри добијају као:

$$t_3 = \delta_1 + \delta_2 \ , \ y_3 = \delta_2 \, , y_4 = 2\delta_1 + \delta_2 \ . \tag{5.25}$$

Дисторзија датог двобитног квантизера је:

$$D = 2 \int_{t_2=0}^{t_3} \left(x - y_3\right)^2 p\left(x\right) dx + 2 \int_{t_3}^{\infty} \left(x - y_4\right)^2 p\left(x\right) dx \ . \tag{5.26}$$

Теорема 5.1. Оптимални двобитни неуниформни Лапласов квантизер се може пројектовати коришћењем следећег итеративног правила:

$$\delta_2^{(i+1)} = \frac{1}{\sqrt{2}} - \sqrt{2} \exp\left\{-\left(1 + \sqrt{2}\delta_2^{(i)}\right)\right\}, \ i = 0,1,... \ . \tag{5.27}$$

Доказ. Из правила центроида (2.16) се за Лапласову PDF добија:

$$y_3 = \frac{\int_0^{t_3} xp\left(x\right) dx}{\int_0^{t_3} p\left(x\right) dx} = \frac{1}{\sqrt{2}} - \frac{t_3 \exp\left\{-\sqrt{2}t_3\right\}}{1 - \exp\left\{-\sqrt{2}t_3\right\}} \ , \tag{5.28}$$

$$y_4 = \frac{\int_{t_3}^{\infty} xp\left(x\right) dx}{\int_{t_3}^{\infty} p\left(x\right) dx} = t_3 + \frac{1}{\sqrt{2}} \ . \tag{5.29}$$

Према основној дефиницији офсета δ_1 и израза (5.29) следи да је $\delta_1 = 1/\sqrt{2}$. На основу (5.25) важи да је $t_3 = 1/\sqrt{2} + \delta_2$, док је $y_3 = \delta_2$. Заменом ових једнакости у (5.28) и додатним сређивањем долази се до израза:

$$\delta_2 = \frac{1}{\sqrt{2}} - \sqrt{2} \exp\left(-\left(1 + \sqrt{2}\delta_2\right)\right) \ , \tag{5.30}$$

који се може решити итеративно, чиме је доказ завршен.

Последица 5.1. Укупна дисторзија оптималног двобитног Лапласовог квантизера је:

$$D = \delta_2^2 . \tag{5.31}$$

Доказ. Израз (5.26) може да се напише у облику:

$$\begin{aligned} D = 2\int_{t_2=0}^{t_3} x^2 p(x)dx - 4y_3\int_{t_2=0}^{t_3} xp(x)dx + 2y_3^2\int_{t_2=0}^{t_3} p(x)dx \\ +2\int_{t_3}^{\infty} x^2 p(x)dx - 4y_4\int_{t_3}^{\infty} xp(x)dx + 2y_4^2\int_{t_3}^{\infty} p(x)dx \end{aligned} . \tag{5.32}$$

Користећи изразе (5.28) и (5.29) и затим користећи чињеницу да је $2\int_0^{\infty} x^2 p(x)dx = \sigma_{\text{ref}}^2 = 1$, добија се следећи израз за дисторзију:

$$D = 1 - y_3^2 p(y_3) - y_4^2 p(y_4) , \tag{5.33}$$

где су $p(y_3)$ и $p(y_4)$ вероватноће нивоа y_3 и y_4, респективно, дате са:

$$p(y_3) = 2\int_0^{t_3} p(x)dx = \frac{1}{2}\left(1 - \exp\left\{-\sqrt{2}t_3\right\}\right) , \tag{5.34}$$

$$p(y_4) = \int_{t_3}^{\infty} p(x)dx = \frac{1}{2}\exp\left\{-\sqrt{2}t_3\right\} . \tag{5.35}$$

Заменом (5.25) у (5.34) и (5.35) и даљом применом у (5.33) добија се:

$$D = 1 - \delta_2^2 - 2\left(\sqrt{2}\delta_2 + 1\right)\exp\left(-\left(1 + \sqrt{2}\delta_2\right)\right) . \tag{5.36}$$

Коначно, заменом (5.30) у (5.36) за дисторзију се добија [125]:

$$D = \delta_2^2 , \tag{5.37}$$

што закључује доказ.

Табела 5.2 Детаљи за двобитни неуниформни квантизер који је пројектован новим итеративним методом ($\sigma_{\text{ref}}^2 = 1$)

δ_1	δ_2	t_3	y_3	y_4	SQNR [dB]	R [bps]
0.7071	0.4198	1.1269	0.4198	1.834	7.54	2

Детаљи за овако пројектован квантизер су дати у табели 5.2. Перформансе добијене Лојд-Макс алгоритмом [1, 2, 8] се у потпуности поклапају са вредностима из ове табеле.

Опис двобитне ADM. Дијаграм тока за предложени ADM са адаптивним двобитним квантизером и адаптивним LP-ом првог реда је приказан на слици 5.5. Алгоритам се извршава у следећим корацима:

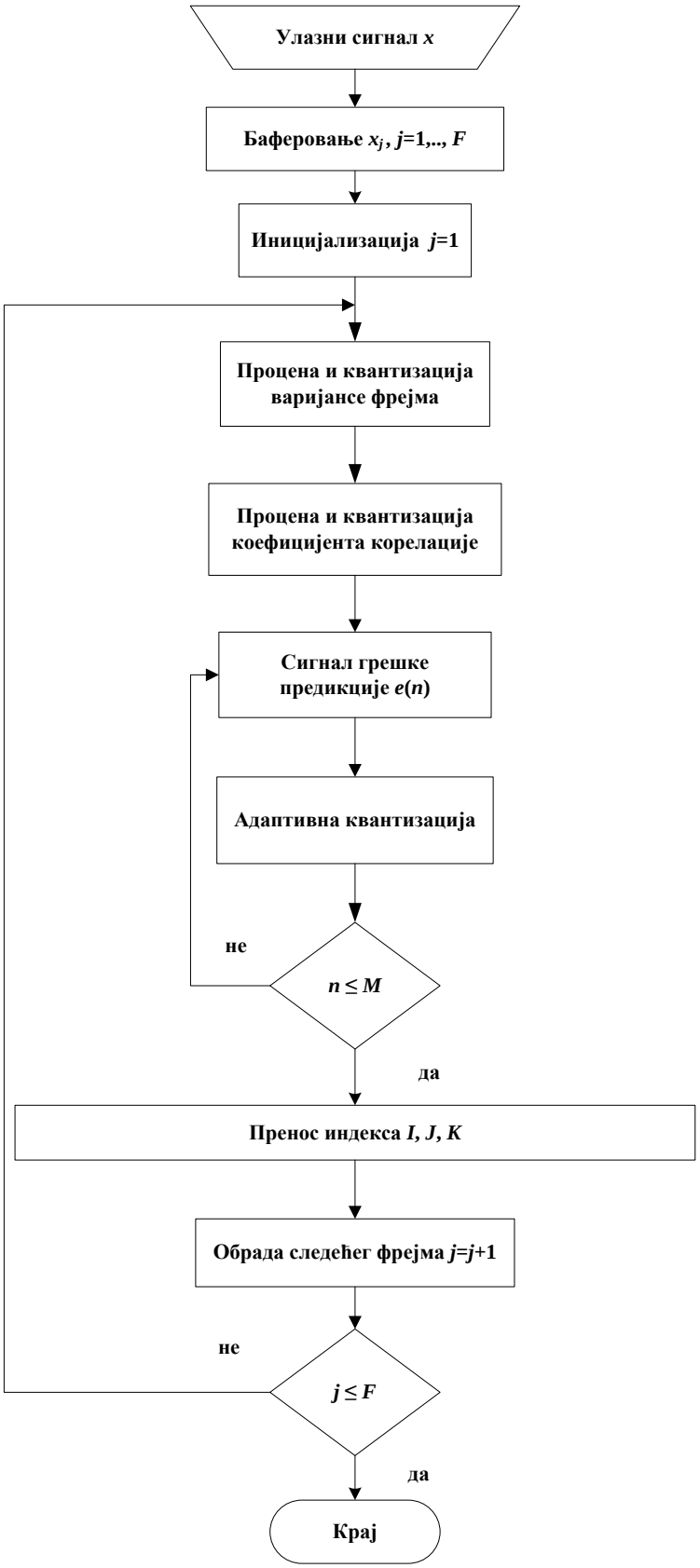
1. ***Баферовање фрејма.*** Исто као у кораку 1 алгоритма из одељка 4.1.
2. ***Процена и квантизација варијансе фрејма.*** Исто као у кораку 2 алгоритма из одељка 4.2.
3. ***Процена и квантизација коефицијента корелације (предикције).*** За j-ти фрејм, коефицијент корелације ρ_j ($a_j = \rho_j$ за LP првог реда [1, 3, 8]) се процењује помоћу израза (5.21). Пошто је информација о LP коефицијенту потребна на пријемном крају (као и у локалном декодеру) како би декодовање било исправно, ρ_j се квантује униформним квантизером Q_ρ са N_ρ нивоа. Q_ρ је дефинисан на следећи начин:

$$Q_\rho: \quad a_j \in \rho_l \mid \rho_l = \rho_{\min} + (2l-1)\frac{\Delta^\rho}{2},\ l = 1,\ldots,N_\rho,\ \Delta^\rho = \frac{\rho_{\max} - \rho_{\min}}{N_\rho}, \tag{5.38}$$

   где су $\rho_{\max}$ и $\rho_{\min}$ максимална и минимална процењена вредност коефицијента корелације, респективно. Квантована вредност ρ_l се преноси једном по фрејму (тј. коефицијент предиктора се подешава једном по фрејму) са $R_\rho = \log_2 N_\rho$ битова (индекс K).
4. ***Одређивање сигнала грешке предикције.*** Сигнал грешке предикције се дефинише као $e_j(n)=x_j(n)-y_j(n)$, где је $x_j(n)$ вредност одмерка j-ог фрејма, а $y_j(n)$ је његова реконструисана вредност добијена у локалном декодеру:

$$y_j(n) = \rho_l \cdot y_j(n-1) + y_i^a \operatorname{sgn}\left(e_j(n)\right), \tag{5.39}$$

   где се $y_i^a = g_j \cdot y_i$ (σ_{ref}), $i = 3, 4$, односи на ниво адаптивног двобитног квантизера.



Сл. 5.5 Блок дијаграм предложеног двобитногADM алгоритма.

5. ***Адаптивна двобитна квантизација.*** Параметри адаптивног двобитног квантизера се добијају као у кораку 3 алгоритма из одељка 4.1, с тим што је сада фактор скалирања дат са [1]:

$$g_j = 10^{V_i/20} \cdot \sqrt{1-\rho_l^2} \ . \tag{5.40}$$

$e_j(n)$ пролази кроз адаптивни двобитни квантизер на чијем излазу се добија квантовани сигнал грешке предикције $e_{q,j}(n) = y_i^a$ (индекс I).

6. ***Понављајти претходне кораке док сви фрејмови не буду обрађени*.**

Битска брзина за овај ADM алгоритам је одређена са:

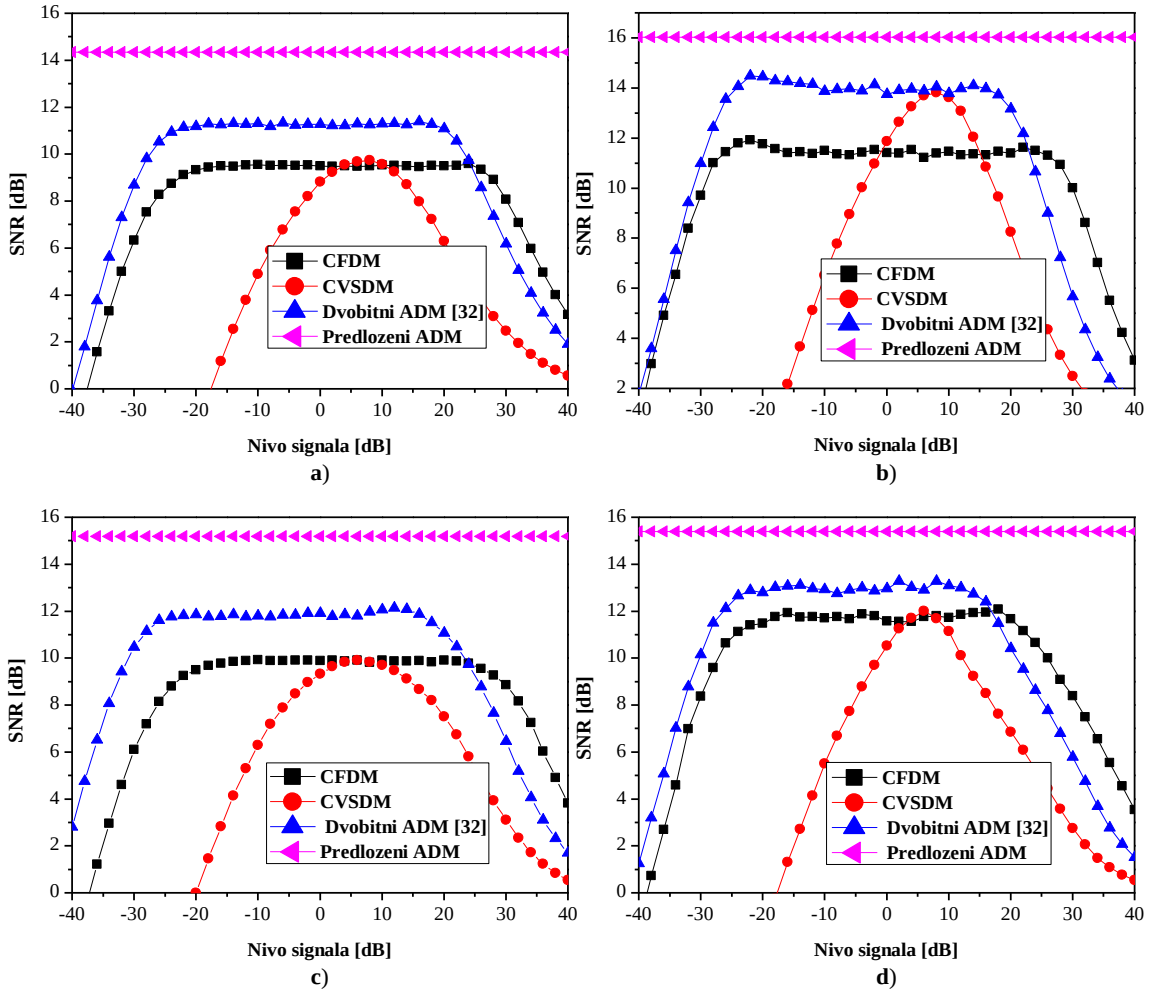
$$R_{\text{DM}} = 2 + \frac{R_{\text{LU}} + R_\rho}{M} [\text{bps}] \ . \tag{5.41}$$

Примена на говорни сигнал. Користе се четири различита тест говорна сигнала (реченице изговорене на америчком енглеском језику) чији су детаљи приказани у табели 5.3.

Табела 5.3 Основне информације о коришћеним тест говорним сигналима

говорник	фреквенција одмеравања [Hz]	трајање [s]	број изговорених реченица
мушки 1	22050	9	2
мушки 2	22050	6	1
женски 1	22050	9	2
женски 2	22050	4	1

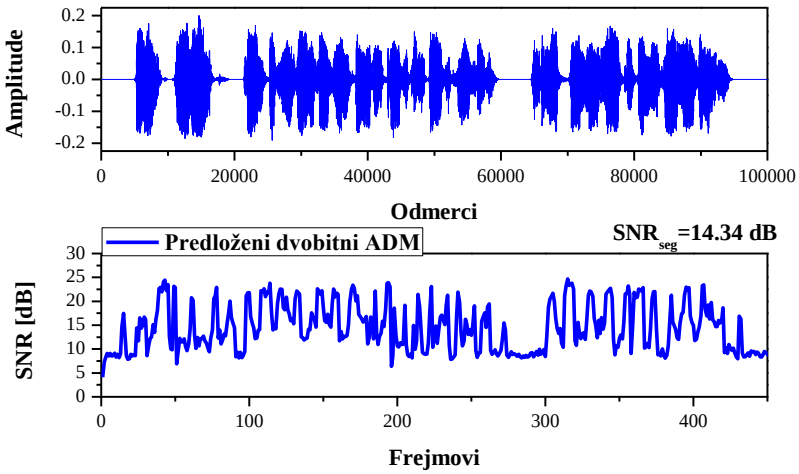
Слика 5.6, где је дат SNR у функцији од нивоа улазног сигнала, пореди перформансе предложеног двобитног ADM-а (Q_{LU} са L = 32 нивоа и Q_ρ са N_ρ = 32 нивоа, величина фрејма је 20 ms) са решењима из класе тренутно адаптивних кодека: двобитни ADM (α = 1.1, β = 1.8 и γ = 1.2) [32], CFDM (α = 1.1) и CVSDM (β = 0.9) [1, 8], при истој излазној битској брзини од 22.05 kbps. Резултати за мушке говорнике су представљени на слици 5.6-а) и слици 5.6-б), док се преостале две слике односе на женске говорнике. Разматрани двобитни ADM је доста ефективнији од осталих ADM алгоритма, јер не само да постиже већи максимални SNR већ има и шири динамички опсег (константан SNR у целом посматаном опсегу нивоа сигнала).



Сл. 5.6 Перформансе за CFDM (α = 1.1), CVSDM (β = 0.9), двобитни ADM (α = 1.1, β = 1.8, γ = 1.2) [32] и предложени двобитни ADM (Q_{LU} са L = 32 нивоа и Q_{ρ} са N_{ρ} = 32 нивоа, дужина фрејма од 20 ms) при излазној битској брзини од 22.05 kbps за различите говорнике: а) мушки 1, б) мушки 2, ц) женски 1 и д) женски 2.

На пример, у случају говорника '**мушки 1'** (слика 5.6-а)), предложени двобитни ADM даје преко 3 dB већи максимални SNR од двобитног решења [32] и преко 5 dB у односу на CFDM и CVSDM [1, 8].

SNR по фрејмовима дужине 20 ms за предложени двобитни ADM у случају говорника **'мушки 1'** је дат на слици 5.7. Веће вредности SNR-а су добијене у области активног говора, а сегментни SNR има вредност SNR_{seg} = 14.34 dB.

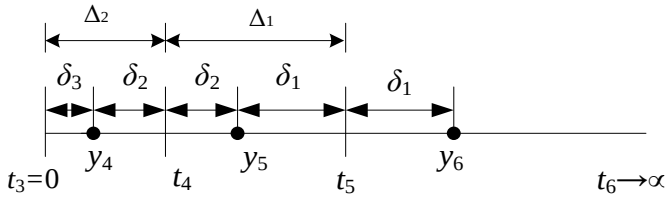


Сл. 5.7 Оригинални говорни сигнал (мушки 1) и SNR за различите фрејмове дужине 20 ms за разматрани двобитни ADM (Q_{LU} са L = 32 нивоа и Q_ρ са N_ρ = 32 нивоа).

5.3.3 Дводигитни ADM

Разматра се дводигитни ADM алгоритам из рада [71] који имплементира адаптивни неуниформни квантизер са шест нивоа и адаптивни LP првог реда.

Опис и пројектовање квантизера. Слика 5.8 приказује модел симетричног неуниформног квантизера са N = 6 нивоа, а пројектују се две верзије (за Лапласов извор и σ^2 = 1): дисторзијом-ограничен и брзином-ограничен квантизер. С тим у вези, неопходно је дефинисати дисторзију и битску брзину.



Сл. 5.8 Симетрични неуниформни квантизер са N = 6 нивоа.

Дисторзија се процењује помоћу следећег израза:

$$D = 2\int_{0}^{t_4}(x-y_4)^2\,p(x)\,dx + 2\int_{t_4}^{t_5}(x-y_5)^2\,p(x)\,dx + 2\int_{t_5}^{\infty}(x-y_6)^2\,p(x)\,dx\ . \qquad (5.42)$$

За кодовање нивоа користи се енкодер на чијем излазу се генеришу кодне речи промењиве дужине. Наиме, y_4 се кодује са једним битом док се y_5 и y_6 кодују са два бита а један бит се користи за знак. Вероватноће нивоа су дате са:

$$p(y_4) = p(-y_4) = \int_0^{t_4} p(x)dx = \frac{1}{2}\left(1 - \exp\left\{-\sqrt{2}t_4\right\}\right), \tag{5.43}$$

$$p(y_5) = p(-y_5) = \int_{t_4}^{t_5} p(x)dx = \frac{1}{2}\left(\exp\left\{-\sqrt{2}t_4\right\} - \exp\left\{-\sqrt{2}t_5\right\}\right), \tag{5.44}$$

$$P(y_6) = P(-y_6) = \int_{t_5}^{\infty} p(x)dx = \frac{1}{2}\exp\left\{-\sqrt{2}t_5\right\}, \tag{5.45}$$

а користећи израз (2.10) за битску брзину се добија:

$$R = 4 \cdot p(y_4) + 6 \cdot \left(p(y_5) + p(y_6)\right). \tag{5.46}$$

За пројектовање дисторзијом-ограниченог квантизера користи се поједностављени Лојд-Макс алгоритам (захтева мањи број итерација од класичног Лојд-Макс алгоритма) [61]. При томе, потребно је дефинисати неке екстерне параметре са слике 5.8:

- Офсет δ_i је већ дефинисан у претходном одељку (растојање између одговарајућег нивоа и доњег прага одлуке). Такође важи да је:

$$\delta_i = t_{6-i} - y_{6-i} \ , \ \ i = 1,2,3. \tag{5.47}$$

- Ширина квантизационе ћелије $\Delta_i = t_{6-i} - t_{6-i-1} (\Delta_0 \to \infty)$. Осим тога, важи да је $\Delta_{i-1} = \delta_{i-1} + \delta_i$, или еквивалентно:

$$\delta_i = \Delta_{i-1} - \delta_{i-1} \ , \ \ i = 1,2,3. \tag{5.48}$$

Са слике 5.8 се јасно види да δ_i у потпуности дефинишу предложени квантизер (прагови одлуке одређују се као $t_4 = \delta_2 + \delta_3$ и $t_5 = \delta_1 + 2\delta_2 + \delta_3$, док се нивои одређују као $y_4 = \delta_3$, $y_5 = 2\delta_2 + \delta_3$ и $y_6 = 2\delta_1 + 2\delta_2 + \delta_3$). Поступак одређивања δ_i дат је у наставку.

На основу (2.16) долази се до следећих израза за нивое:

$$y_6 = t_5 + \frac{1}{\sqrt{2}}, \tag{5.49}$$

$$y_{6-i} = t_{5-i} + \frac{1}{\sqrt{2}} + \frac{\Delta_i}{1-e^{\sqrt{2}\Delta_i}}. \tag{5.50}$$

Комбиновањем (5.48) и (5.50) добија се следећа једначина:

$$\Delta_i \left(1 - \frac{1}{1-e^{\sqrt{2}\Delta_i}} \right) = \delta_i + \frac{1}{\sqrt{2}}, \tag{5.51}$$

која показује да се Δ_i може одредити нумерички ако се зна офсет δ_i.

Процес добијања δ_i тече на следећи начин. На основу израза (5.47) и (5.49) следи да је $\delta_1 = 1/\sqrt{2}$. Заменом δ_1 у (5.51) и нумеричким решавањем добија се Δ_1, а на основу (5.48) добија се δ_2. Када се зна δ_2, онда се из (5.51) одређује Δ_2, а из (5.48) добија се δ_3. Добијене су следеће нумеричке вредности за δ_i: $\delta_1 = 1/\sqrt{2}$, $\delta_2 = 0.4198$ и $\delta_3 = 0.2998$.

Друга верзија квантизера добија се модификацијом претходно пројектованог дисторзијом-ограниченог квантизера. Наиме, брзином-ограничен квантизер се састоји од унутрашњег дела ($-t_4$, t_4) који садржи два нивоа $-y_4$ и y_4 и спољашњег дела који садржи четири нивоа $-y_6$, $-y_5$, y_5 и y_6. t_4 дефинише границу између унутрашњег и спољашњег дела и пројектовање овог типа квантизера се своди се на одређивање оптималне вредности за t_4.

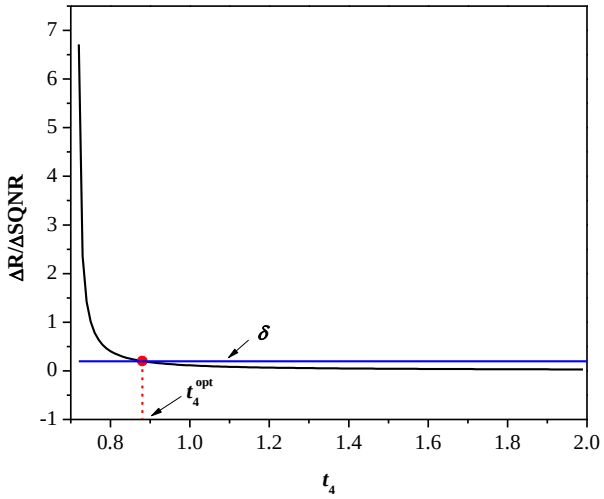
Ниво y_4 се одређује из правила центроида (2.16), што даје:

$$y_4 = \frac{\int_0^{t_4} xp(x)dx}{\int_0^{t_4} p(x)dx} = \frac{1}{\sqrt{2}} + \frac{t_4}{1-\exp\left\{\sqrt{2}t_4\right\}}, \tag{5.52}$$

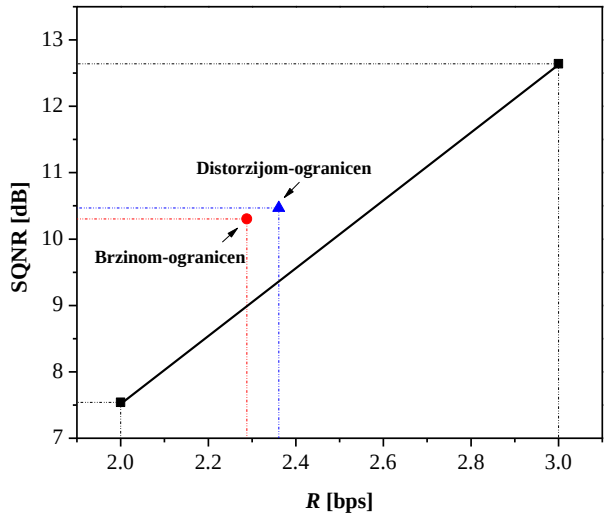
а параметри у спољашњем делу квантизера се налазе као:

$$t_5 = t_4 + \delta_2,\ y_5 = t_4 + \delta_1 + \delta_2,\ y_6 = t_4 + \delta_1 + 2\delta_2. \tag{5.53}$$

Као и у ранијим случајевима, пројектовање брзином-ограниченог квантизера се врши уз помоћ критеријума (3.11). Оптимално t_4 се тражи у опсегу $t_4 \in (t_4^*, 2)$, где је $t_4^* = \delta_2 + \delta_3 = 0.7196$ вредност прага код дисторзијом-ограниченог квантизера, док је $\delta = 0.1961$ (нагиб између Лојд-Макс квантизера са $N = 4$ (SQNR = 7.54 dB, $R = 2$ bps) и $N = 8$ нивоа (SQNR =12.64 dB, $R = 3$ bps) [1, 2, 8]). Са слике 5.9 се види да је тражена вредност $t_4 = t_4^{opt} = 0.745$.



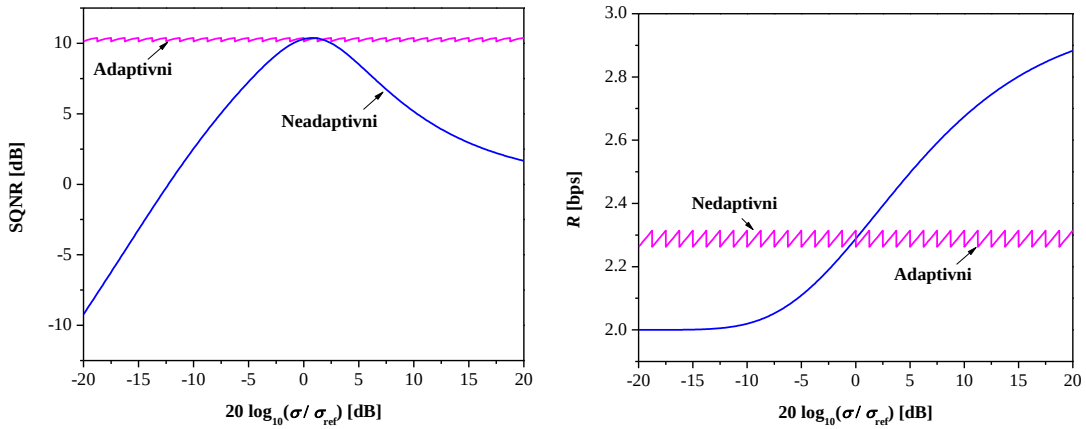
Сл. 5.9 Избор оптималне вредности за t_4 за брзином-ограничен неуниформни квантизер са шест нивоа.



Сл. 5.10 Перформансе предложеног квантизера са шест нивоа у поређењу са Лојд-Макс квантизером.

На слици 5.10 су приказане постигнуте вредности за SQNR и R за обе верзије разматраног квантизера. Види се да брзином-ограничен квантизер у односу на дисторзијом-ограничен квантизер остварује мању битску брзину уз мимимални губитак у SQNR-у. Такође, обе верзије предложеног квантизера постижу за више од 1 dB већи SQNR од Лојд-Макс квантизера са истом битском брзином, и због тога су бољи кандидати за практичну примену.

Перформансе (SQNR и R) неадаптивног и адаптивног брзином-ограниченог квантизера су демонстриране на слици 5.11. Из приложених резултата се види да адаптивни квантизер има већу робусност.



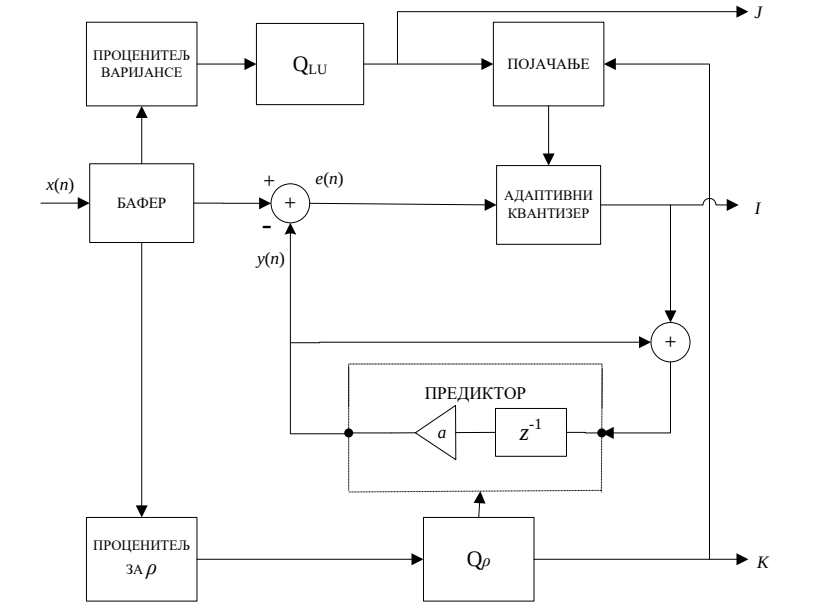
а) б)

Сл. 5.11 Перформансе адаптивног и неадаптивног квантизера са шест нивоа (брзином-ограничен квантизер): а) SQNR и б) R.

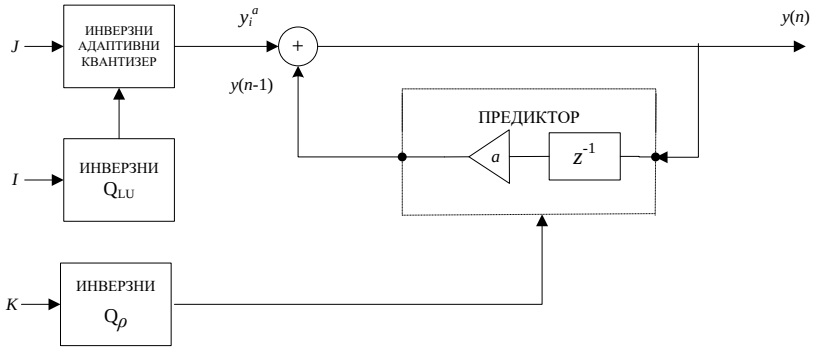
Опис дводигитне ADM. Слика 5.12 даје приказ блок шеме кодера и декодера за ADM конфигурацију са адаптивним неуниформним квантизером са шест нивоа и адаптивним LP-ом првог реда. Принцип рада се може описати истим корацима као у одељку 5.3.2, с том разликом што се сада користи квантизер са шест уместо са четири нивоа. За овај конкретан случај, укупна битска брзина се рачуна по формули:

$$R_{\mathrm{DM}} = R + \frac{R_{\mathrm{LU}} + R_{\rho}}{M}\,\mathrm{bps}, \tag{5.54}$$

где је R битска брзина квантизера дефинисана изразом (5.46).



а)

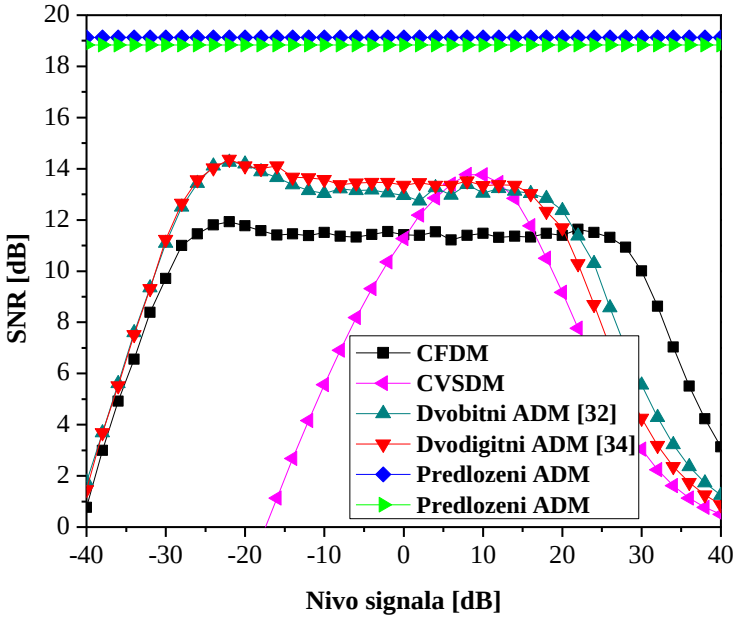


б)

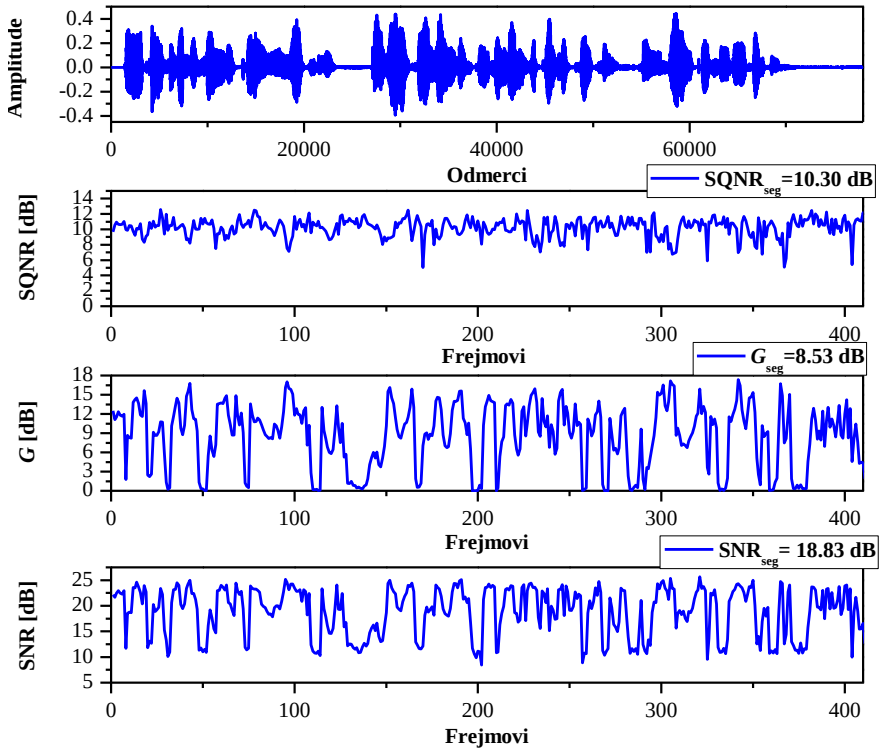
Сл. 5.12 Предложена дводигитна ADM: а) кодер и б) декодер.

Примена на говорни сигнал. Користи се тест говорни сигнал од 8 секунди који је одмераван на фреквенцији од 22050 Hz (реченице на америчком енглеском језику изговорене од стране мушког говорника).

Слика 5.13 даје SNR за предложени двобитни ADM (Q_{LU} са L = 32 нивоа, Q_ρ са N_ρ = 32 нивоа, величина фрејма је 20 ms). Нешто већи SNR добијен је за случај када ADM користи дисторзијом-ограничен квантизер. У сврху поређења користе се дводигитни ADM (α = 1.1, β = 2.2) [34], двобитни ADM (α = 1.1, β = 1.8, γ = 1.2) [32], CFDM (α = 1.1) и CVSDM (β = 0.9) [1, 8]. Остварена је већа робусност али и значајно побољшање у SNR-у: за око 5 dB у односу на алгоритме [32] и [34], за око 8 dB у односу на CFDM и за око 5 dB у односу на CVSDM.



Сл. 5.13 SNR за различите ADM конфигурације за излазну битску брзину од 22050 bps.



Сл. 5.14 Оригинални говорни сигнал и SQNR, *G* и SNR по фрејмовима говора дужине 20 ms за предложени дводигитни ADM са брзином-ограниченим квантизером (Q_{LU} са L = 32 нивоа и Q_ρ with N_ρ = 32 нивоа).

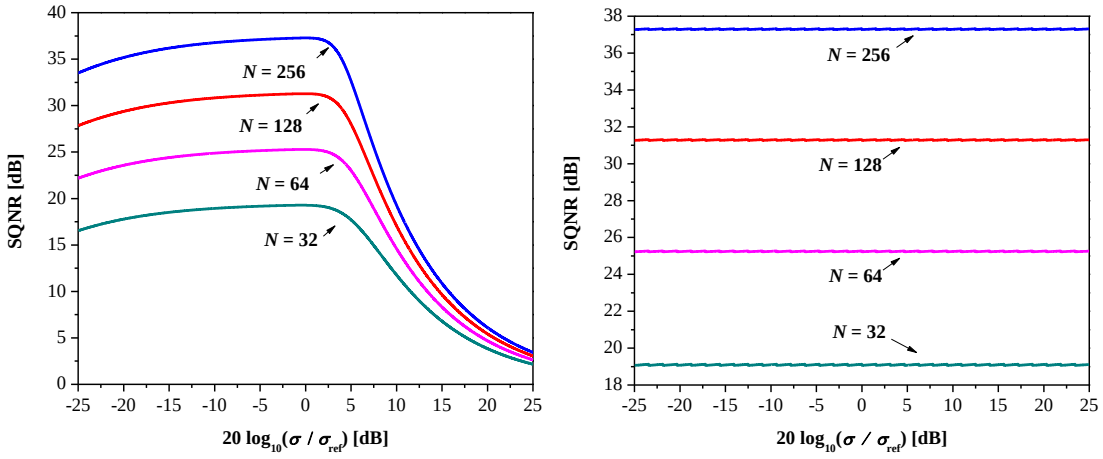
На слици 5.14 дат је SNR, SQNR и G на нивоу фрејма за дводигитни ADM са брзином-ограниченим квантизером, за величину фрејма од 20 ms. Постигнути сегментни SQNR ($SQNR_{seg}$ = 10.3 dB) се веома добро поклапа са теоријским резултима за адаптивни квантизер на слици 5.11.

5.3.4 Вишенивовска ADM

Овај одељак описује једно ADM решење које се базира на примени G.711 квантизера [89] и прекидачког LP-а првог реда, а које је предложено у раду [72].

Опис и пројектовање квантизера. Детаљан опис G.711 квантизера (PLCQ са параметрима μ = 255, N = 256, L = 8, m = 16, R = 8 bps) је већ дат у одељку 4.4 где су изведени изрази за процену перформанси ((4.37) и (4.38)). Овде ће бити анализиране перформансе неадативног и адаптивног PLCQ-а и за број нивоа мањи од 256, односно за 32, 64 и 128 нивоа.

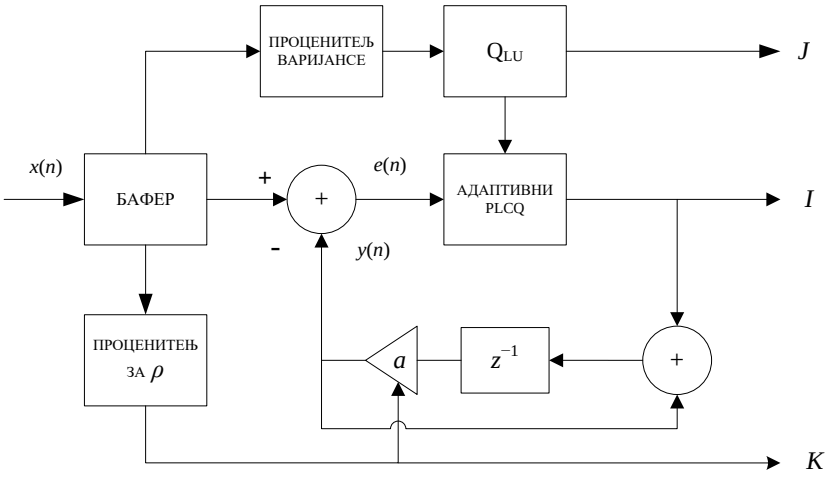
Слика 5.15-а) приказује SQNR за неадаптивни PLCQ за различито N у опсегу варијанси [-25 dB, 25 dB] у односу на σ_{ref}^2 = 1. Неадаптивни PLCQ је робустан, нарочито за $\sigma^2 < \sigma_{ref}^2$. Адаптацијом PLCQ-а ниво робусности је значајно повећан (за свако N), као што се може видети на слици 5.15-б).



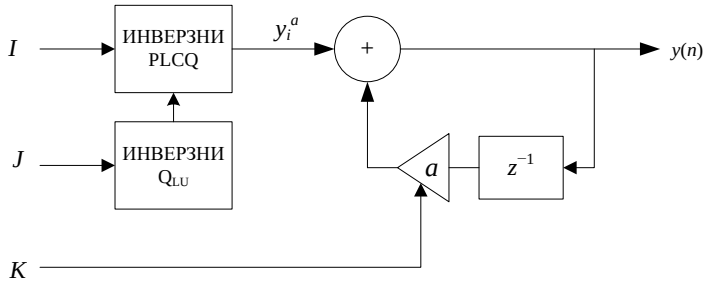
а) б)

Сл. 5.15 SQNR у широком опсегу варијансе улазног сигнала када се N мења за: а) неадаптивни PLCQ (μ = 255) и б) адаптивни PLCQ (μ = 255, Q_{LU} за L = 32 нивоа).

Опис вишенивовске ADM. Овај ADM систем користи адаптивни PLCQ и прекидачки LP првог реда из одељка 5.3.1, а блок шема кодера и декодера се може видети на слици 5.16.



а)



б)

Сл 5.16 Предложена вишенивовска ADM: а) кодер и б) декодер.

Следећи кораци описују принцип рада дате вишенивовске ADM:

1. ***Баферовање фрејма.*** Исто као у кораку 1 алгоритма из одељка 4.1.
2. ***Процена и квантизација варијансе фрејма.*** Исто као у кораку 2 алгоритма из одељка 4.1.
3. ***Процена коефицијента корелације и одабир коефицијента предикције.*** Исто као у кораку 3 алгоритма из пододељка 5.3.1.
4. ***Одређивање сигнала грешке предикције.*** За j-ти фрејм, сигнал грешке предикције је дат са: $e_j(n)=x_j(n)-y_j(n)$, где $x_j(n)$ означава вредност одмерка док је $y_j(n)$ је његова реконструисана вредност дата са:

$$y_j(n) = a_k \cdot y_j(n-1) + y_i^a(n)\,\mathrm{sgn}(e_j(n)),\ k = 1,2, \tag{5.55}$$

где су y_i^a нивои адаптивног PLCQ-а.

5. ***Адаптивна квантизација.*** Параметри фиксног PLCQ се одређују према (4.23) и (4.24). Адаптација PLCQ-а за *j*-ти фрејм се ради као у кораку 3 алгоритма из одељка 4.1 а фактор скалирања је дат са (5.22). Након тога се сваки одмерак сигнала грешке предиције квантује са адаптивним PLCQ-ом и квантована вредност се преноси до декодера са $R = \log_2 N$ bps (индекс *I*).
6. ***Реконструкција сигнала.*** У декодеру се врши реконструкција одмерака оригиналног фрејма на основу индекса *I*, *J* и *K*.
7. ***Понављајти претходне кораке док сви фрејмови не буду обрађени*.**

Битска брзина за вишенивовски ADM систем је:

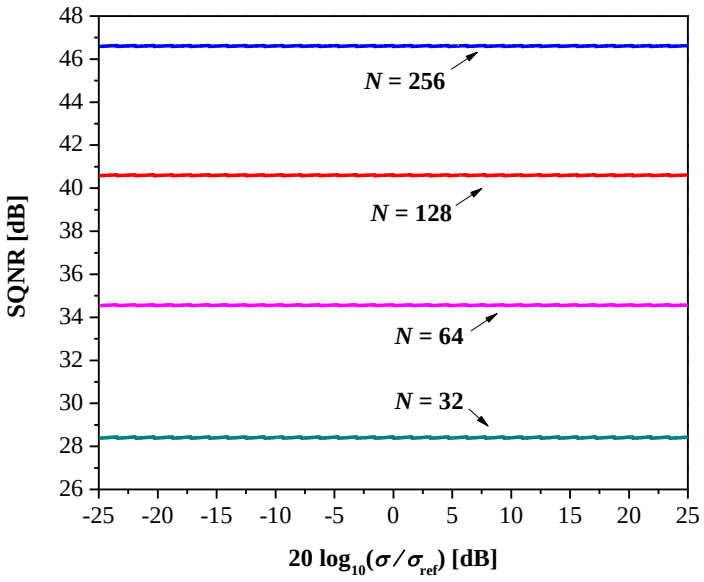
$$R_{\mathrm{DM}} = \log_2 N + \frac{1+\log_2 L}{M}\ [\mathrm{bps}]. \tag{5.56}$$

Анализа теоријског модела вишенивовске ADM. Из литературе је познато да је проценат тишине у говору обично око 25 % [4], стога се усваја да је $w = 0.25$ што заправо дефинише удео слабо корелисаних фрејмова. Суседни одмерци говора су виско корелисани са коефицијентом корелације близу један [1−10], па се за звучне фрејмове усваја $a_2 = 0.97$. Поред тога, усваја се $a_1 = 0.3$ за слабо корелисане фрејмове. За прекидачки предиктор, еквивалентни добитак се рачуна као:

$$G_{\mathrm{eq}} = wG_1 + (1-w)G_2, \tag{5.57}$$

где су G_1 и G_2 дефинисани са (5.6) и означавају добитке предикције у случају слабо корелисаних и високо корелисаних фрејмова, респективно.

Теоријски SNR за дати ADM систем и различит број нивоа PLCQ-а је приказан на слици 5.17, а резултати показују да је остварен добитак од око 10 dB у односу на адаптивни PLCSQ (за свако *N*) са слике 5.15-б).



Сл. 5.17 SNR за вишенивовску ADM када се број нивоа N квантизера PLCQ мења (μ = 255 и Q_{LU} са L = 32 нивоа).

Примена на говорни сигнал. Тест говорни сигнал се састоји од 66 500 одмерака а одмераван је на фреквенцији од 16 kHz (реченица изговорена на српском језику од стране мушког говорника). Из Табеле 5.4, где су сумиране перформансе за предложену ADM и адаптивни PLCQ (PCM кодек са PLCQ-ом) за различите дужине фрејма, јасно се види да ADM постиже за приближно 10 dB већи SNR. Остварено побољшање добија на значају посебно за N = 256, будући да је за тај број нивоа PLCQ стандардизован [89]. Такође, резултати за SNR из ове табеле се у великој мери слажу са теоријским резултатима са слика 5.15 и 5.17.

Табела 5.4. Перформансе за предложени вишенивовски ADM (Q_{LU} са L = 32 нивоа) и адаптивни PLCQ за различити број нивоа и различите величине фрејма

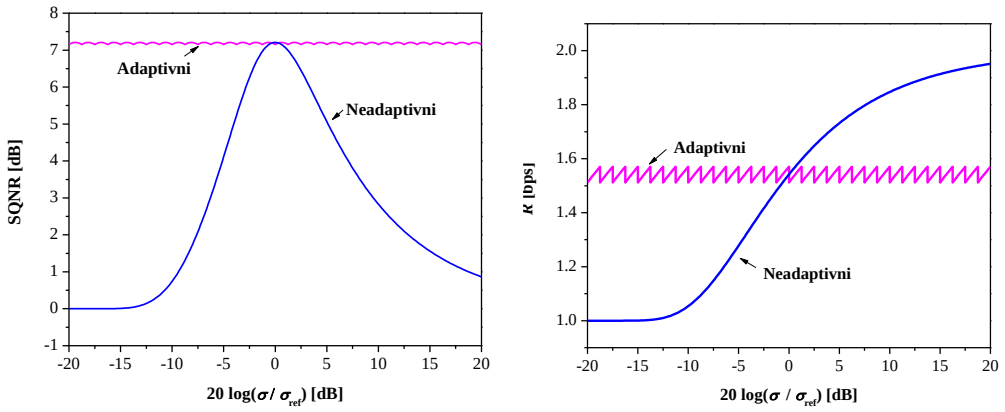
	N = 32			**N = 64**			**N= 128**			**N = 256**		
M	**SNR [dB]**	**SQNR [dB]**	**R_{DM} [bps]**	**SNR [dB]**	**SQNR [dB]**	**R_{DM} [bps]**	**SNR [dB]**	**SQNR [dB]**	**R_{DM} [bps]**	**SNR [dB]**	**SQNR [dB]**	**R_{DM} [bps]**
80	29.00	19.19	5.075	35.14	25.28	6.075	41.08	31.32	7.075	46.97	37.40	8.075
160	28.98	19.12	5.037	35.06	25.21	6.037	41.00	31.27	7.037	46.96	37.35	8.037
240	29.03	19.08	5.025	35.02	25.19	6.025	40.96	31.31	7.025	46.96	37.38	8.025
320	28.78	19.00	5.019	34.83	25.07	6.019	40.75	31.14	7.019	46.72	37.17	8.019

5.4 Алгоритми за кодовање Гаусовог извора

5.4.1 Тернарна ADM

У овом делу дата је ADM конфигурација предложена у раду [73], која се заснива на примени адаптивног тернарног квантизера и прекидачког LP-а другог реда. За разлику од пододељка 5.3.1 где је разматрана тернарна ADM са отвореном повратном спрегом, овде се разматра ADM систем са затвореном повратном спрегом.

Опис и пројектовање квантизера. Користи се дисторзијом-ограничен тернарни квантизер са Хафмановим кодом који је већ пројектован за Гаусов извор у одељку 3.2 (t_2 = 0.612 и y_3 = 1.224). Перформансе (SQNR и R) овог модела пре и након адаптације се могу видети на слици 5.18.

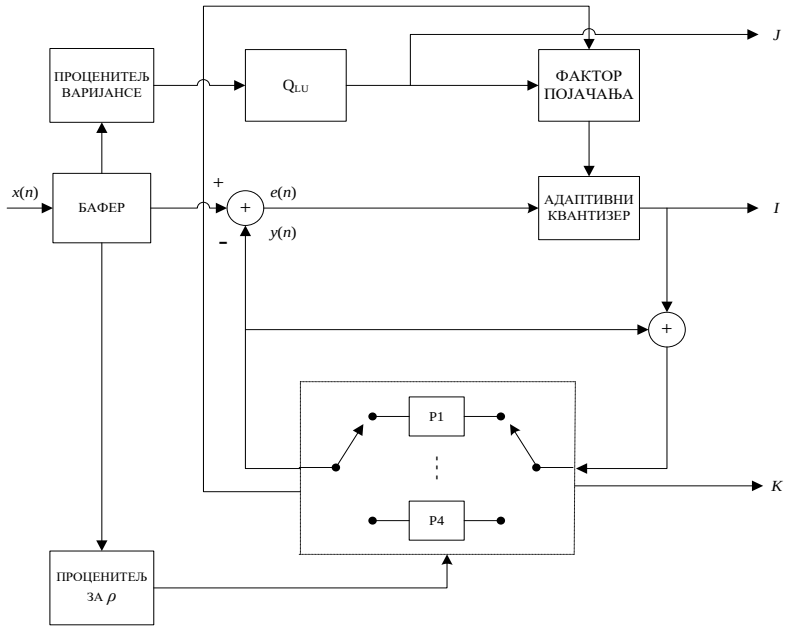


а) б)

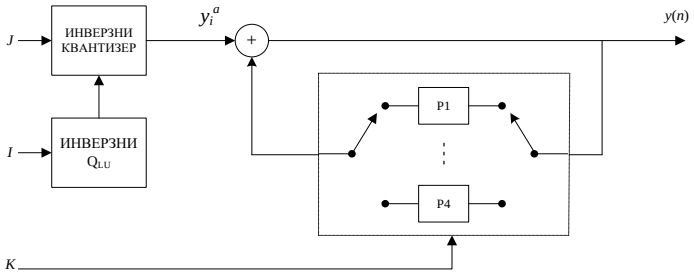
Сл. 5.18 Перформансе у широком опсегу варијансе улазног сигнала за неадаптивни и адаптивни тернарни квантизер (Q_{LU} са L = 32 нивоа): а) SQNR и б) R.

Опис тернарне ADM. На слици 5.19 приказана је блок шема кодера и декодера за предложени тернарни ADM (користи се адаптивни тернарни квантизер и прекидачки предиктор са четири фиксна LP-а другог реда [26] а избор између њих се врши на основу коефицијента корелације ρ). Начин функционисања се може описати кроз следеће кораке:

1. ***Баферовање фрејма.*** Исто као у кораку 1 алгоритма из одељка 4.1.



а)



б)

Сл. 5.19 Предложена тернарна ADM: а) кодер и б) декодер.

2. ***Процена и квантизација варијансе фрејма.*** Исто као у кораку 2 алгоритма из одељка 4.1.
3. ***Процена коефицијената корелације и избор LP-а другог реда.*** $\rho_{1,j}$ се за тренутни j-ти фрејм израчунава помоћу (5.21). Затим се испитује опсег у коме $\rho_{1,j}$ припада да би се извршио избор одговарајућег LP-а другог реда (означени су са P_1, …, P_4 на слици 5.19) у оквиру прекидачког предиктора а информација о изабраном LP-у се кодује са два бита, односно:

$$\begin{cases} \rho_{1,j} \in (0.9,\ 1) \rightarrow P_1 \rightarrow 00 \\ \rho_{1,j} \in (0.5,\ 0.9) \rightarrow P_2 \rightarrow 01 \\ \rho_{1,j} \in (0,\ 0.5) \rightarrow P_3 \rightarrow 11 \\ \rho_{1,j} \in (-1,\ 0) \rightarrow P_4 \rightarrow 10 \end{cases}, \tag{5.58}$$

и преноси до пријемника индексом K.

4. ***Одређивање сигнала грешке предикције.*** За *j*-ти фрејм сигнал грешке предикције се одређује као: $e_j(n) = x_j(n) - y_j(n)$, где је $n = 1,\ldots, M$, а $y_j(n)$ је реконструисана вредност одмерка (излаз из локалног декодера) дата са:

$$y_j(n) = a_{1,(k)} y_j(n-1) + a_{2,(k)} y_j(n-2) + y_i^a \operatorname{sgn}\left(e_j(n)\right), \quad k \in \{1,\ldots,4\}, \tag{5.59}$$

где су $a_{1,(k)}$ и $a_{2,(k)}$ коефицијенти *k*-тог LP-а другог реда а y_i^a је ниво адаптивног тернарног квантизера (квантована вредност одмерка грешке предикције).

5. ***Адаптивна квантизација.*** Адаптивни тернарни квантизер се за *j*-ти фрејм пројектује као у кораку 3 алгоритма из одељка 4.1, тј. множењем кодне књиге неадаптивног квантизера са фактором скалирања који је у овом случају дат са (израз (5.9)):

$$g_j = 10^{V_i/20} \sqrt{1 + a_{1,(k)}^2 + a_{2,(k)}^2 - 2\left(a_{1,(k)}\rho_1 + a_{2,(k)}\rho_2 - a_{1,(k)}a_{2,(k)}\rho_1\right)}. \tag{5.60}$$

Затим се одмерци грешке предикције $e_j(n)$ пропуштају кроз адаптивни квантизер, а кодовани одмерци се преносе до пријемника (индекс *I*).

6. ***Понављајти претходне кораке док сви фрејмови не буду обрађени.***

За предложену тернарну ADM битска брзина се рачуна као:

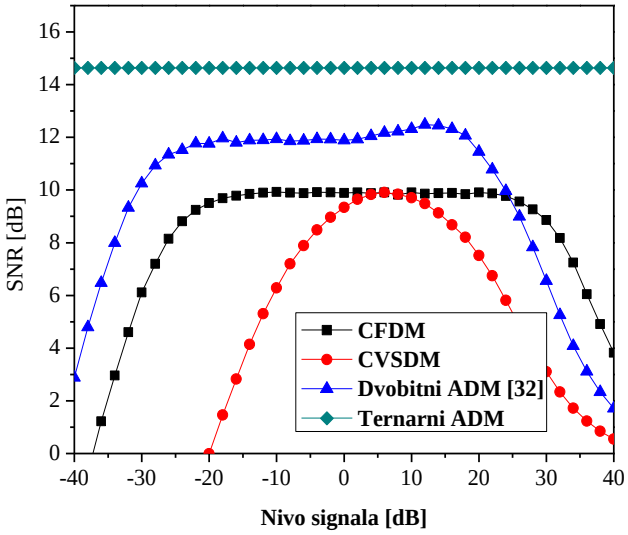
$$R_{\mathrm{DM}} = R + \frac{R_{\mathrm{LU}} + 2}{M} \ [\mathrm{bps}], \tag{5.61}$$

где *R* битска брзина тернарног квантизера чија је вредност дата у табели 3.1.

Примена на говорни сигнал. Тренинг секвенца у трајању од једног минута (фреквенција одмеравања је 22050 Hz) користи се за одређивање параметара прекидачког предиктора. Одговарајуће вредности су дате у табели 5.5, а добијене су за величину фрејма од 10 ms.

Табела 5.5 Коефицијенти за четири LP-а другог реда који се користе у оквиру прекидачког предиктора

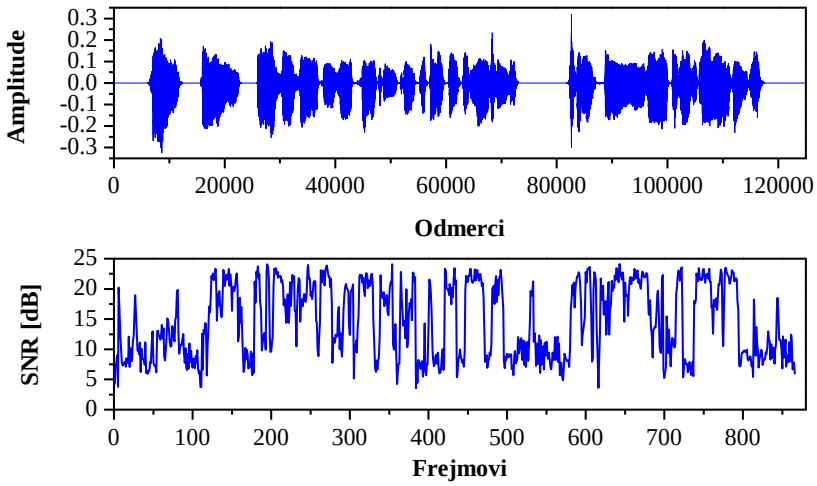
Предиктор	**Опсег ρ_1**	ρ_1	ρ_2	a_1	a_1
P_1	0.9-1	0.97	0.91	1.48	-0.52
P_2	0.5-0.9	0.78	0.52	0.96	-0.23
P_3	0-0.5	0.25	-0.09	0.29	-0.16
P_4	-1-0	-0.26	-0.22	-0.34	-0.31



Сл. 5.20 SNR у функцији амплитуда улазног говора за предложени тернарни ADM (Q_{LU} са L = 32 нивоа, дужина фрејма је 10 ms,), CFDM (α = 1.1), CVSDM (β = 0.9) и двобитни ADM (α = 1.1 β = 1.8, γ = 1.2) [32] при излазној битској брзини од 22050 bps.

Као тест сигнал користи се говор од 8 s одмераван на 22.05 kHz, који није био укључен у тренинг секвенцу.

Са слике 5.20 је очигледно да је тернарна ADM надмоћнија у односу на двобитни ADM [32], CFDM и CVSDM [1, 8]. Постигнут је за 2.5 dB већи SNR од двобитног ADM-а [32] и 4 dB у поређењу са CFDM и CVSDM. Слика 5.21 показује SNR по свим фрејмовима говора дужине 10 ms за разматрани ADM.

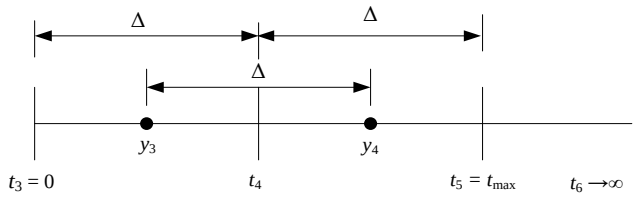


Сл. 5.21 Оригинални говор и SNR по фрејмовима дужине 10 ms за предложену тернарну ADM (Q_{LU} са L = 32 нивоа).

5.4.2 Двобитни ADM са фракционом линеарним предиктором

У овом пододељку описан је ADM алгоритам предложен у раду [74] а заснива на примени двобитног униформног квантизера и фракционог линеарног предиктора (FLP). Иначе, FLP је нови тип предиктора који до сада није био имплементиран у склопу ADM-а.

Опис и пројектовање квантизера. Симетрични униформни квантизер са $N = 4$ нивоа је приказан на слици 5.22. Кодовање се врши фиксном дужином кодних речи, што значи да је $R = \log_2 4 = 2$ bps.



Сл. 5.22 Модел симетричног двобитног униформног квантизера.

Дисторзија овог квантизера, за претпостављену Гаусову PDF и $\sigma^2 = \sigma_{ref}^2 = 1$, је дата са:

$$D = 2\int_0^{\Delta}(x-\Delta/2)^2 p(x)dx + 2\int_{\Delta}^{\infty}(x-3\Delta/2)^2 p(x)dx$$

$$= 1 - \sqrt{\frac{2}{\pi}}\Delta\left(1+2\exp\left\{-\frac{\Delta^2}{2}\right\}\right) + \Delta^2\left(\frac{1}{4}+4Q(\Delta)\right) \ . \qquad (5.62)$$

D искључиво зависи од Δ. Оптимално Δ за коју D постиже минимум дефинисана је следећом лемом.

Лема 5.1. За симетрични двобитни униформни Гаусов квантизер оптимално Δ се добија итеративно као:

$$\Delta^{(i+1)} = \frac{1}{\sqrt{2\pi}}\left(\frac{1+2\exp\left\{-\frac{\left(\Delta^{(i)}\right)^2}{2}\right\}}{\frac{1}{4}+4Q\left(\Delta^{(i)}\right)}\right), i = 0,1,\ldots . \qquad (5.63)$$

Доказ. Из услова $\partial D/\partial\Delta = 0$ добија се следећа интегрална једначина:

$$\frac{1}{\sqrt{2\pi}}\left(1+2\exp\left(-\frac{\Delta^2}{2}\right)\right)=\Delta\left(\frac{1}{4}+4Q(\Delta)\right), \tag{5.64}$$

из које се Δ може изразити као:

$$\Delta=\frac{1}{\sqrt{2\pi}}\left(\frac{1+2\exp\left\{-\frac{\Delta^2}{2}\right\}}{\frac{1}{4}+4Q(\Delta)}\right), \tag{5.65}$$

чиме је доказ завршен.

У изразу (5.63) се јавља Q-функција која захтева нумерички прорачун. Да би се рачунска комплексност смањила и тиме поједноставио процес пројектовања користи се апроксимација Q-функције. За дати проблем користи се параметарска апроксимација дата изразом (3.24), али за разлику од одељка 3.4 овде се параметар a оптимизује за опсег вредности аргумената Δ. Доња и горња граница опсега означене са Δ^{low} Δ^{up}, респективно, добијају се из израза (5.63) када се уместо $Q(x)$ користи апроксимација (3.31) и апроксимација из рада [97] (означена је са $F^{[97]}(x)$) која је дата са:

$$F^{[97]}(x)=\frac{1}{\sqrt{2\pi}}\frac{x}{x^2+1}\exp\left(-\frac{x^2}{2}\right). \tag{5.66}$$

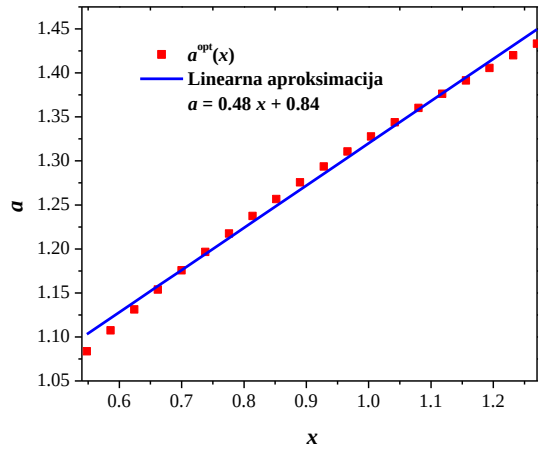
За почетну вредност се усваја $\Delta^{\text{low}(0)}=\Delta^{\text{up}(0)}=2\sqrt{\log N}$, $N = 4$ [80].

Оптимизација параметра a из (3.24) се изводи на следећи начин. Опсег $[\Delta^{\text{low}}, \Delta^{\text{up}}]$ се унифромно подели на I тачака, и онда се за сваку вредност аргумента $x = \Delta(i)$, $i = 1,\ldots, I$, одреди вредност $a^{\text{opt}}(\Delta(i))$ за коју је релативна грешка при апроксимацији Q-функције најмања, односно:

$$a^{\text{opt}}\big(\Delta(i)\big)=\underset{a}{\arg\min}\left\{\frac{\left|F(\Delta(i),a)-Q(\Delta(i))\right|}{Q(\Delta(i))}\right\}, i=1,\ldots,I, \tag{5.67}$$

где је $\Delta^{\text{low}} = \Delta(1)$ и $\Delta^{\text{up}} = \Delta(I)$. Затим се добијене вредности $a^{\text{opt}}(\Delta(i))$ апроксимирају линеарном правом $a(x) = k{\cdot}x + m$, при чему се k и m бирају тако да средња квадратна грешка буде најмања (слика 5.23). Дакле, апроксимација (3.24) се за проблем двобитне униформне квантизације своди на облик:

$$F(x)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{x^2+0.48x+0.84}}\exp\left\{-\frac{x^2}{2}\right\}. \tag{5.68}$$



Сл. 5.23 Одређивање параметра a за апроксимацију (3.24) за проблем двобитне униформне квантизације.

Заменом (5.68) у (5.63) добија се апроксимативна вредност за Δ (означена је са Δ^a) која се такође одређује итеративно, али се избегава прорачун Q-функције и тиме смањује рачунска комплексност. Такође, применом (5.68) у (5.62) добија се апроксимативни израз за дисторзију:

$$D^a = 1 - \sqrt{\frac{2}{\pi}}\Delta^a\left(1 + 2\exp\left\{-\frac{(\Delta^a)^2}{2}\right\}\right) + +(\Delta^a)^2\left(\frac{1}{4} + 4F(\Delta^a)\right), \tag{5.69}$$

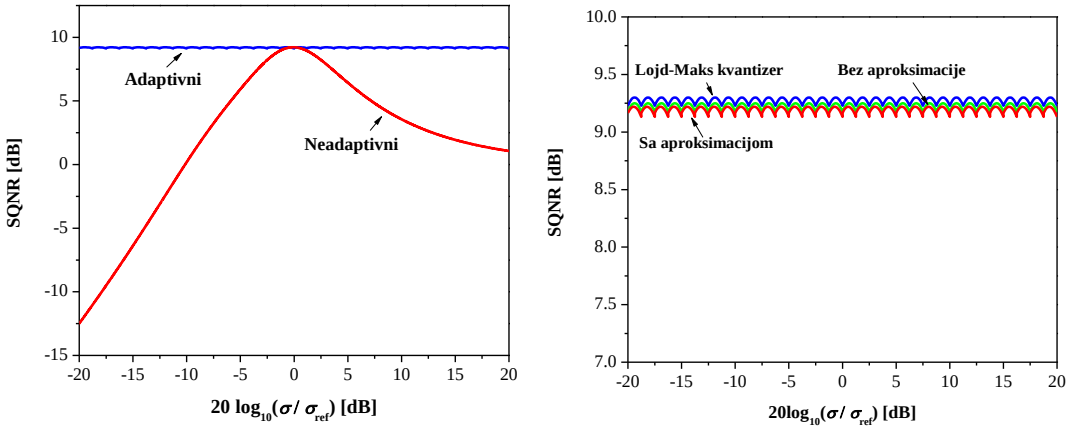
док је апроксимативни израз за SQNR дат са:

$$\mathrm{SQNR}^a = 10\log_{10}\left(\frac{1}{1-\sqrt{\frac{2}{\pi}}\Delta^a\left(1+2\exp\left\{-\frac{(\Delta^a)^2}{2}\right\}\right)+(\Delta^a)^2\left(\frac{1}{4}+4F(\Delta^a)\right)}\right). \tag{5.70}$$

Анализа теоријских резултата. Табела 5.6 приказује апроксимативне вредности за параметре двобитног униформног квантизера када се користе различите апроксимације и RE добијене у том случају. RE је најмања (испод 1 %) када се користи предложена апроксимација (5.68), што указује на изузетно високу тачност изведених израза.

Табела 5.6 Апроксимативне вредности за параметре двобитног униформног квантизера и одговарајуће релативне грешке, за различите апроксимације Q-функције

Апроксимација	Δ^a	δ_Δ **(%)**	D^a	δ_D**(%)**	**SQNR^a [dB]**	δ_{SQNR}**(%)**
$F^{[100]}(x)$	0.9759	1.9886	0.1270	6.9024	8.9620	3.1135
$F^{[98]}(x)$	0.8742	12.2025	0.1753	47.5589	7.5622	18.2465
$F^{[96]}(x)$	1.2704	27.5886	0.0409	65.5724	13.8828	50.0843
$F(x)$	0.9936	**0.2109**	0.1197	**0.7576**	9.2191	**0.3341**



Сл. 5.24 Поређење перформанси адаптивног двобитног квантизера (са апроксимацијом Q-функције) са: а) неадаптивним двобитним униформним квантизером (са апроксимацијом Q-функције) и б) адаптивним двобитним униформним и адаптивним двобитним Лојд-Макс квантизером.

На слици 5.24-а) се могу видети значајно веће перформансе (SQNR) адаптивног (Q_{LU} са $L = 32$ нивоа) у односу на неадаптивни двобитни униформни квантизер (са апроксимацијом Q-функције), а слика 5.24-б) указује на компарабилне перформансе овог квантизера у односу на адаптивни двобитни униформни и адаптивни двобитни неуниформни (Лојд-Макс) квантизер [1, 2, 8].

Опис предиктора. Користи се фракциони линеарни предиктор (FLP) код кога се предикована вредност сигнала добија као [126−128]:

$$x_p(n) = b/h^{\lambda}\big(x(n-1) - \lambda x(n-2)\big), \tag{5.71}$$

где је b коефицијент FLP-а, h је периода одмеравања а λ је реалан број. Дефинисани FLP модел користи меморију од два одмерка, исто као и LP другог реда (одељак 5.1), а за фиксно λ одређен је само са једним коефицијентом исто као и LP првог реда (одељак 5.1). За $\lambda = 0$ дати FLP модел је еквивалентан LP-у првог реда, па се λ бира тако да се постигну што је могуће веће перформансе у односу на LP првог реда.

Сигнал грешке предикције $e(n)$ је дат са:

$$e(n) = x(n) - x_p(n) = x(n) - b/h^{\lambda}\big(x(n-1) - \lambda x(n-2)\big), \tag{5.72}$$

а σ_e^2 је дата са:

$$\sigma_e^2 = E\left[\left(x(n) - x_p(n)\right)^2\right] = \sigma_x^2\left(1 - \frac{2b}{h^\lambda}(\rho_1 - \lambda\rho_2) + \left(\frac{b}{h^\lambda}\right)^2(1 - 2\lambda\rho_1 + \lambda^2)\right). \quad (5.73)$$

Решавањем једначине $\partial\sigma_e^2/\partial b = 0$ добија се оптимално b [126]:

$$b^{opt} = h^\lambda\left(\frac{\rho_1 - \lambda\rho_2}{1 - 2\lambda\rho_1 + \lambda^2}\right). \quad (5.74)$$

Заменом (5.74) у (5.73) добија се:

$$\sigma_{e,min}^2 = \sigma_x^2\left(1 - \frac{(\rho_1 - \lambda\rho_2)^2}{(1 - 2\lambda\rho_1 + \lambda^2)}\right), \quad (5.75)$$

а добитак предикције за разматрани FLP модел је дат са:

$$G = 10\log_{10}\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10\log_{10}\left(\frac{1}{1 - \frac{(\rho_1 - \lambda\rho_2)^2}{(1 - 2\lambda\rho_1 + \lambda^2)}}\right). \quad (5.76)$$

Опис двобитне ADM са FLP-ом. Блок дијаграм ADM система са адаптивним двобитним униформним квантизером и адаптивним FLP-ом је дат на слици 5.25. У зависности од тога да ли се процесира фрејм чије су вредности све нуле или не дати систем ради у једном од два мода. Принцип рада у првом моду је следећи:

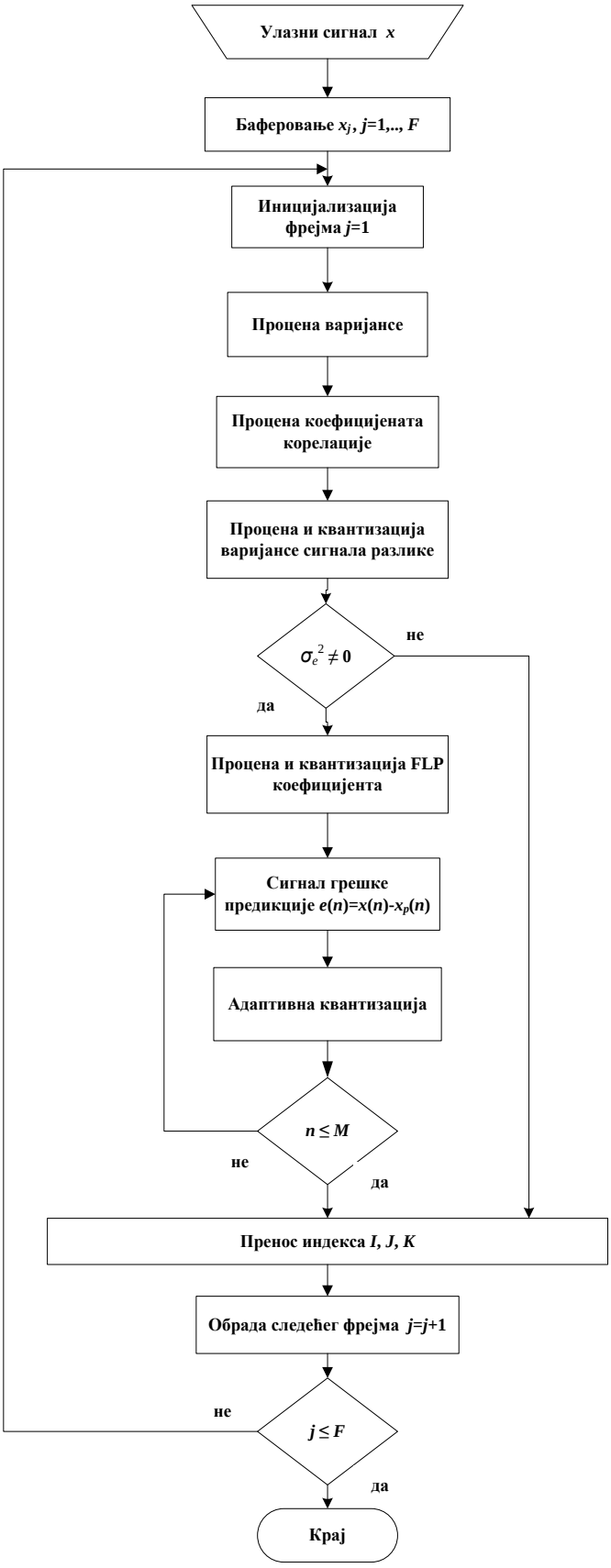
1. ***Бафероване фрејма.*** Исто као у кораку 1 алгоритма из одељка 4.1.
2. ***Процена варијансе фрејма.*** Исто као у кораку 2 алгоритма из одељка 4.1.
3. ***Процена коефицијената корелације.*** Коефицијенти корелације ρ_1 и ρ_2 се за j-ти фрејм естимирају као [1, 3]:

$$\rho_{k,j} = \frac{\frac{1}{M-k}\sum_{n=1}^{M-k} x_j(n)x_j(n+k)}{1/M\sum_{n=1}^{M} x_j^2(n)}, j = 1, \ldots, F, k = 1,2. \quad (5.77)$$

4. ***Процена и квантизација варијансе грешке предикције.*** За j-ти фрејм, $\sigma_{e,j}^2$ се процењује према (5.75) а квантовање се врши на начин како је описано у кораку 2 алгоритма из одељка 4.1.
5. ***Процена и квантизација FLP коефицијента.*** За j-ти фрејм b_j се процењује на основу израза (5.74) и затим се помоћу квантизера Q_b унифомно квантује на један од B репрезентационих нивоа:

$$Q_b: \quad b_j \in b_B \mid b_B = b_{min} + \frac{2t-1}{2}\Delta_b, t = 1, \ldots, B, \quad (5.78)$$

где је $\Delta_b = (b_{\max} - b_{\min}) / B$ величина корака квантизера Q_b, а $b_{\min}$ и $b_{\max}$ означавају минималну и максималну процењену вредност коефицијента FLP-а, респективно. За пренос информације о коефицијенту FLP-а (једном по фрејму) користи се $R_b = \log_2 B$ битова (индекс K).



Сл. 5.25 Предложени двобитни ADM са FLP-ом.

6. ***Одређивање сигнала грешке предикције.*** За j-ти фрејм, сигнал грешке предикције се одређује као $e_j(n) = x_j(n) - y_j(n)$, где је $y_j(n)$ реконструисана вредност одмерка и дата је са:

$$y_j(n) = b_j/h^{\lambda}\left(y_j(n-1) - \lambda y_j(n-2)\right) + y_i^a \operatorname{sgn}\left(e_j(n)\right), \qquad (5.79)$$

где је y_i^a квантована вредност одмерка $e_j(n)$.

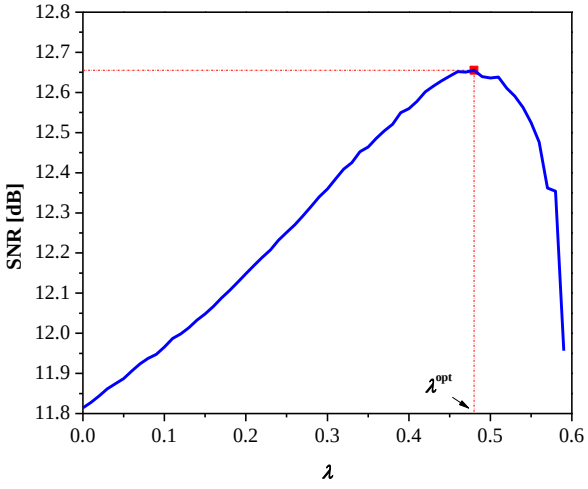
7. ***Адаптивна квантизација.*** Адаптивни двобитни униформни квантизер се за j-ти фрејм добија као у кораку 3 алгоритма из одељка 4.1. Затим се $e_j(n)$ квантује и та информација се преноси до декодера (индекс I).
8. ***Понављајти претходне кораке док сви фрејмови не буду обрађени.***

У другом режиму рада (вредности фрејма су све нуле), варијанса фрејма и варијанса грешке предикције су једнаке нули па су нивои квантизера подешени на нулу, што има за последицу да су улаз и излаз ADM алгоритма идентични. Из тог разлога се не ради кодовање већ се до декодера преноси само информација о нултој вредности варијансе (једна кодна реч квантизера Q_{LU} је резервисана за ту сврху), и тада декодер на свом излазу генерише фрејм који садржи све нуле.

Битска брзина за описани ADM је:

$$R_{\mathrm{DM}} = 2 + \frac{\log_2 L + \log_2 B}{M} \text{ [bps]}. \qquad (5.80)$$

Примена на говорни сигнал. За процену параметра λ користи се тренинг секвенца од приближно 30 минута говора екстрахованог из TIMIT базе [113]. На основу слике 5.26 усваја се $\lambda = 0.48$.

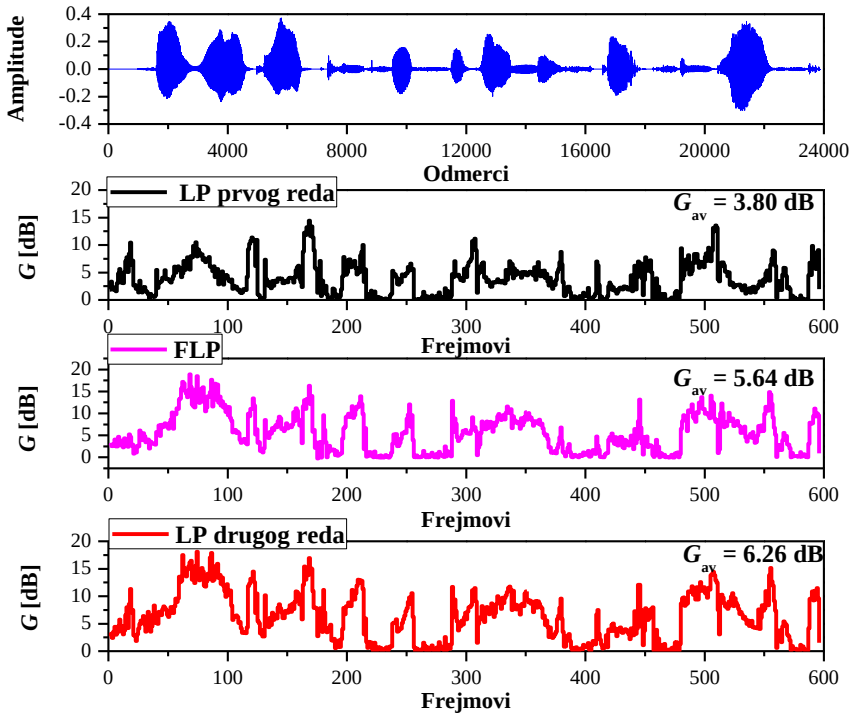


Сл. 5.26 Избор оптималне вредности за λ на тренинг секвенци.

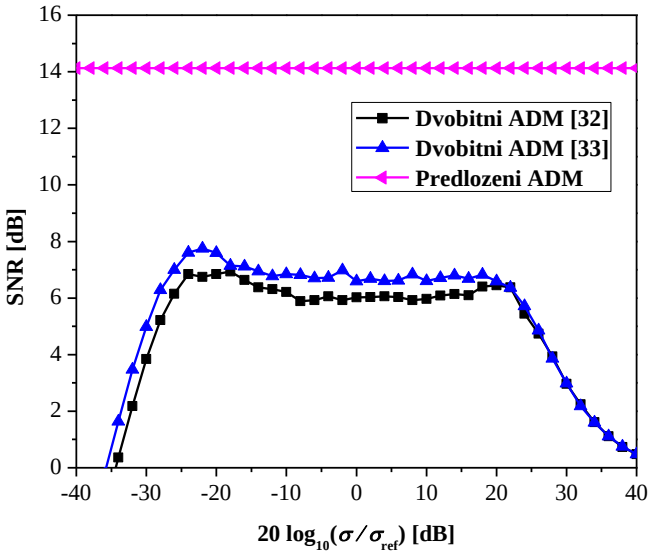
Табела 5.7 Перформансе предложеног двобитног ADM-а (Q_{LU} са L = 32 нивоа и Q_b са N_b = 32 нивоа, величина фрејма је 5 ms) на тест сигналу за неколико типова предиктора

	двобитни ADM (LP првог реда)	**двобитни ADM (FLP)**	**двобитни ADM (LP другог реда)**
SNR [dB]	12.67	14.13	14.80
***R* [bps]**	2.25	2.25	2.375

За тестирање се користи говор (24 000 одмерака) из TIMIT базе који није био укључен у тренинг секвенцу. Из табеле 5.7 се види да је FLP модел (λ = 0.48) боље решење од LP-а првог реда (λ = 0), јер тада предложени двобитни ADM при истој битској брзини од 2.25 bps остварује већи SNR за око 1.5 dB. Са друге стране, при мањој битској брзини, добијен је компетитивни SNR у односу на случај када се користи LP другог реда. На слици 5.27 су приказане перформансе (добитак предикције G) наведених предиктора на нивоу фрејма, како би се додатно истакла предност FLP модела.



Сл. 5.27 Добитак предикције на нивоу фрејма за разматрани FLP и LP првог и другог реда.



Сл. 5.28 Перформансе предложеног двобитног ADM-а у односу на постојећа двобитна ADM решења при излазној битској брзини од 16 kbs.

Слика 5.28 показује да предложени двобитни ADM (λ = 0.48, Q_{LU} са L = 32 нивоа и Q_b са N_b = 32 нивоа, величина фрејма је 5 ms) при излазној битској брзини од 16 kbs има доста боље перформансе у поређењу са постојећим напреднијим двобитним ADM решењима [32] и [33].

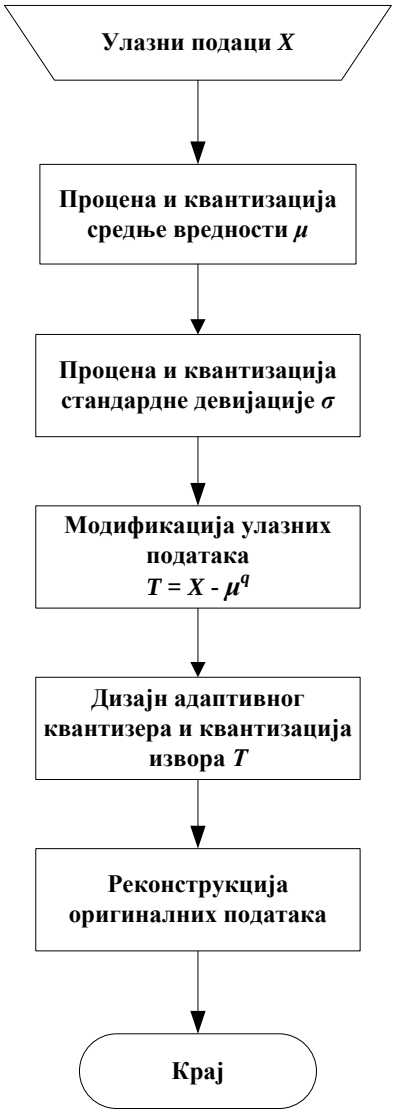
6. Решења на бази скаларне квантизације за компресију неуронских мрежа

У овом поглављу дисертације биће разматрано неколико нових модела склараних квантизера за компресију NN-а. Акценат ће бити на решењима за компресију тежина (обучене) NN, с обзиром на њихов доминантан утицај на величину мреже. Биће представљени квантизери са малим бројем нивоа јер су они значајнији са становишта компресије, а ако се правилно пројектују могу да обезбеде добар компромис између величине и перформанси NN-а. Принцип пројектовања квантизера је другачији у односу на постојеће моделе. Теоријско пројектовање се базира на Лапласовом извору, а биће анализиран и ефекат неприлагођења на варијансу и предложиће се адекватене мере у циљу потискивања овог ефекта. Најпре ће бити разматран један тип бинарног квантизера који се интензивно примењује у компресији NN-а и извршиће се унапређење модела у смислу оптимизације перформанси и адаптације (користи се слична техника као код говора), а биће показано да у том случају квантована NN постиже веће перформансе. Затим ће бити предложено двобитно решење на бази униформне квантизације, за који ће бити показано да, тиме што даје већи SQNR од постојећих двобитних модела, значајно поправља перформансе квантоване NN. На крају овог поглавља биће разматрано још једно двобитно решење на бази логаритамских компандинг квантизера које због својих добрих особина (пре свега робусност) може бити погодна алтернатива за адаптивне квантизере.

6.1 Метод за адаптацију квантизера

У стварној ситуацији као што је квантизација параметара NN-а конвергенција модела NN-а зависи од неколико аспеката, укључујући величину скупа података, архитектуру, број епоха итд., тако да се могу јавити разлике између варијансе података који се квантују и варијансе за коју је квантизер пројектован. Стога је од нарочитог интереса да се ефекат неприлагођења на варијансу или избегне или

сведе на што мању меру, имајући у виду његов штетан утицај на SQNR а који даље може утицати на перформансе квантоване NN. За ту сврху користи се адаптација. Метод адаптације који се овде предлаже се углавном ослања на метод описан у одељку 4.1, а осим тога повезан је и са процесом нормализације који се широко користи код NN-а [44, 50, 115]. Нека x_i представљају податке улазног извора X, где је $i = 1, \ldots, M$, а M је дужина података.



Сл. 6.1 Адаптација квантизера за примену у компресији NN-а.

Дијаграм тока за предложени метод је дат на слици 6.1 и се може описати помоћу следећих корака:

1. ***Процена и квантизација средње вредности података.*** Средња вредност улазних података се процењује као [1−3]:

$$\mu = \frac{1}{M}\sum_{i=1}^{M} x_i \ . \tag{6.1}$$

Овај параметар се квантује коришћењем *floating-point* квантизера [114], а квантована вредност μ^q се чува са 32 бита.

2. ***Процена и квантизација стандардне девијације.*** Стандардна девијација улазних података се процењује као [1−3]:

$$\sigma = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(x_i - \mu\right)^2} \ . \tag{6.2}$$

Овај параметар се такође квантује помоћу *floating-point* квантизера користећи 32 бита [114] на чијем излазу се добија квантована вредност σ^q.

3. ***Модификација улазних података.*** Од сваког елемента извора X одузима се μ^q и добија се нови скуп података T:

$$T = X - \mu^q \ . \tag{6.3}$$

Ово заправо има за циљ правилну примену квантизера с обзиром на то да се квантизери пројектују за нулту средњу вредност.

4. ***Дизајн адаптивног квантизера и квантизација извора T.*** Квантизер прилагођен улазним подацима добија се скалирањем параметара квантизера (обично се пројектује за $\sigma_{\text{ref}}^2 = 1$) са σ^q:

$$t_i\left(\sigma\right) = \left(1+\varepsilon\right)\sigma^q t_i\left(\sigma_{\text{ref}}\right),\ y_i\left(\sigma\right) = \left(1+\varepsilon\right)\sigma^q y_i\left(\sigma_{\text{ref}}\right), \tag{6.4}$$

где је ε константа која се користи за компензацију несавршености између теоријског модела и расподеле података (тежина). Затим се подаци $t_i \in T$, $i = 1, \ldots, M$, квантују адаптивним квантизером и добијају се квантовани подаци t_i^q.

5. ***Реконструкција оригиналних података***. Пошто се средња вредност одузима од оригиналних података и даље квантује (користећи 32 бита), мора се извршити инверзни процес да би се оригинални подаци реконструисали:

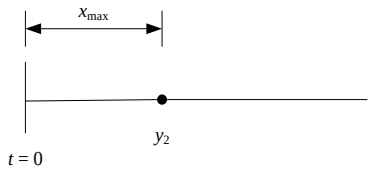
$$x_i^Q = t_i^q + \mu^q,\ i = 1,\ldots,M \ , \tag{6.5}$$

где x_i^Q означава реконструисане податке.

6.2 Бинарни квантизер тип 2

У овом одељку разматран је квантизер предложен у раду [65], који заправо представља побољшање модела из рада [115, 116] а који има важну улогу у компресији NN-а.

Опис и пројектовање квантизера. Симетрични бинарни квантизер [115, 116] је приказан на слици 6.2, а ниво у позитивном делу је подешен на максималну вредност улазних података ($y_2 = x_{max}$).

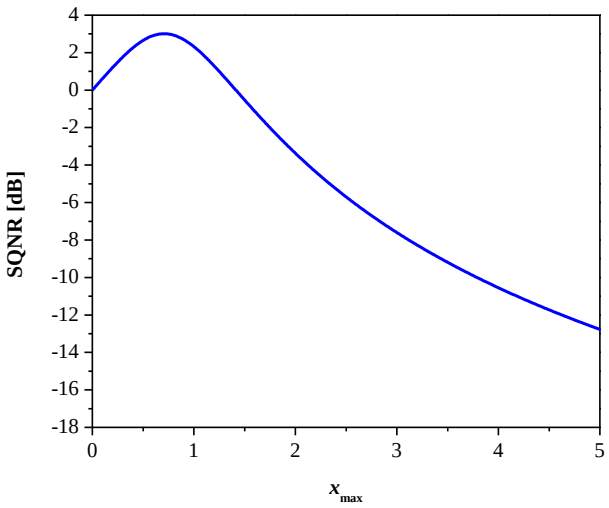


Сл. 6.2 Модел бинарног квантизера.

Дисторзија овог квантизера је за Лапласов извор ($\sigma^2 = \sigma_{ref}^2 = 1$) дата са:

$$D = 2\int_0^{\infty} \left(x - x_{max}\right)^2 p\left(x\right) dx = 1 - \sqrt{2}x_{max} + x_{max}^2 . \tag{6.6}$$

Дакле, на D утиче само x_{max} (односно y_2). Правилним избором овог параметра могуће је остварити максимални SQNR ($y_2 = x_{max} = 0.7071$) као што то показује показује слика 6.3, а добијена је иста вредност као код модела из одељка 4.2.

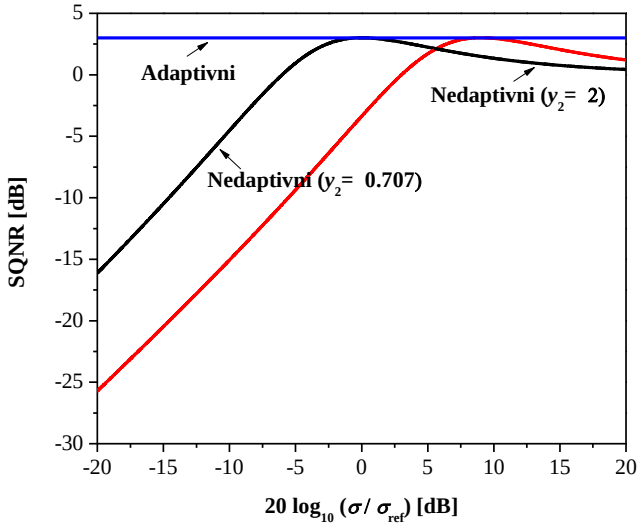


Сл. 6.3 SQNR у функцији од x_{max} за бинарни квантизер тип 2.

Оптимална вредност за $x_{\max}$ се може одредити и аналитички, на следећи начин:

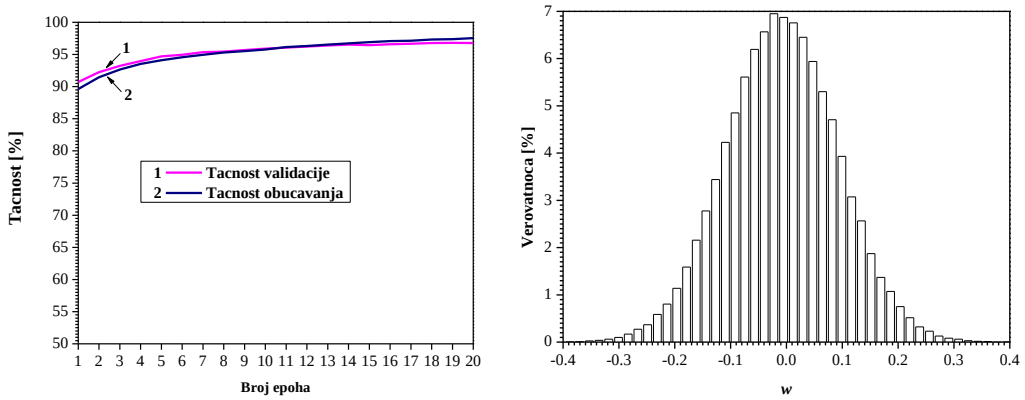
$$\frac{\partial D}{\partial x_{\max}} = 0 \Rightarrow x_{\max} = x_{\max}^{\text{opt}} = \frac{1}{\sqrt{2}}. \tag{6.7}$$

Анализа теоријских резултата за адаптивни бинарни модел. Као што је речено, адаптација има за циљ да омогући ефикасну обраду параметара NN-а. Слика 6.3 показује да адаптивни модел (y_2 = 0.7071, ε = 0) остварује изузетно добре перформансе у широком опсегу варијансе и да значајно поправља перформансе неадаптивног модела са оптимално (y_2 = 0.7071) и произвољно одабраним нивоом (y_2 = 2) [115, 116].



Сл 6.4 Зависност SQNR-а од варијансе сигнала за неадаптивни (са оптимално и произвољно одабраним нивоом) и адаптивни бинарни квантизер.

Експериментални резултати. Као тест подаци користе се тежине обучене MLP NN [117] развијене за потребе класификације слика. Подаци за тренирање и тестирање MLP мреже узимају се из MNIST базе (60 000 слика се користе за тренирање а 10 000 слика за тестирање, а димензије слика су 28 × 28 пиксела) [118]. MLP NN се састоји из улазног, скривеног и излазног слоја са по 784, 128 и 10 неурона, респективно. Хиперпараметри имају следеће вредности: *regularization rate* = 0.01, *learning rate* = 0.0005 и *batch size* = 128.



а) б)

Сл. 6.5 а) криве учења за MLP NN и б) хистограм тежина за обучену MLP NN.

Слика 6.5-а) приказује криве валидације и обуке за MLP NN у првих 20 епоха. На слици 6.5-б) дат је хистограм за тежине између улазног и скривеног слоја (укупно их има 100 352) обучене MLP NN, одакле се види да Лапласова PDF добра апроксимира расподелу тежина.

У табели 6.1 су приказане перформансе (тачност класификације) за MLP NN пре и након примене бинарне квантизације. Види се да предложени модел квантизера остварује већи SQNR (за око 1 dB) од модела из [115, 116], па је тиме повећана тачност квантоване MLP NN за око 1.3 %. У поређењу са MLP моделом пуне прецизности, квантована MLP има за око 5 % мању тачност али и значајно мањи капацитет (степен компресије износи 32).

Табела 6.1 Перформансе квантоване MLP NN добијене применом два модела бинарног квантизера

	Предложени бинарни квантизер	**Бинарни квантизер [115,116]**	**Без компресије**
Тачност [%]	91.28	89.96	96.70
SQNR [dB]	4.287	3.205	-

6.2 Двобитни униформни квантизер

Сада ће бити анализирано решење предложено у раду [75], које се базира на униформној квантизацији а која представља први избор када је једноставност имплементације од примарног интереса.

Опис и пројектовање квантизера. Модел двобитног (N = 4) униформног квантизера је већ представљен у одељку 5.4.2, где је пројектовање рађено за Гаусов извор а имплементиран је у ADM алгоритму за потребе кодовања говора. Међутим, овде се пројектовање ради за други тип извора (Лапласов извор) а другачија је и намена (компресија тежина NN-а).

Дисторзија D двобитног квантизера за Лапласову PDF ($\sigma^2 = \sigma_{ref}^2 = 1$) је дата са:

$$\begin{aligned} D &= 2\left(\int_0^{t_4} (x-y_1)^2 p(x)dx + \int_{t_4}^{\infty} (x-y_2)^2 p(x)dx \right) \\ &= 1 - \frac{\Delta}{\sqrt{2}} + \frac{\Delta^2}{4} - \sqrt{2}\Delta \exp\left\{-\sqrt{2}\Delta\right\} \end{aligned} \quad . \tag{6.8}$$

D зависи од Δ. Следећа лема дефинише оптималну вредност за Δ.

Лема 6.1. За двобитни униформни Лапласов квантизер оптимално Δ се може одредити итеративно, као:

$$\Delta^{(i+1)} = \sqrt{2} - \frac{\sqrt{2}}{2 + \frac{1}{2}\exp\left\{\sqrt{2}\Delta^{(i)}\right\}} \quad , \; i = 0,1,... \quad . \tag{6.9}$$

Доказ. Диференцирањем дисторзије (6.8) по Δ добија се:

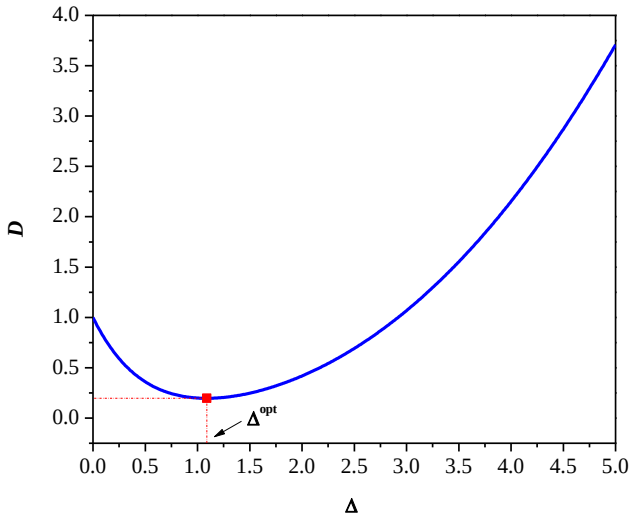
$$\frac{\partial D}{\partial \Delta} = \frac{\Delta}{2} - \frac{1}{\sqrt{2}} + (2\Delta - \sqrt{2})\exp\left\{-\sqrt{2}\Delta\right\} \quad . \tag{6.10}$$

Изједначавањем последњег израза са нулом и решавањем по Δ добија се:

$$\Delta = \sqrt{2} - \frac{\sqrt{2}}{2 + \frac{1}{2}\exp\left\{\sqrt{2}\Delta\right\}} \quad , \tag{6.11}$$

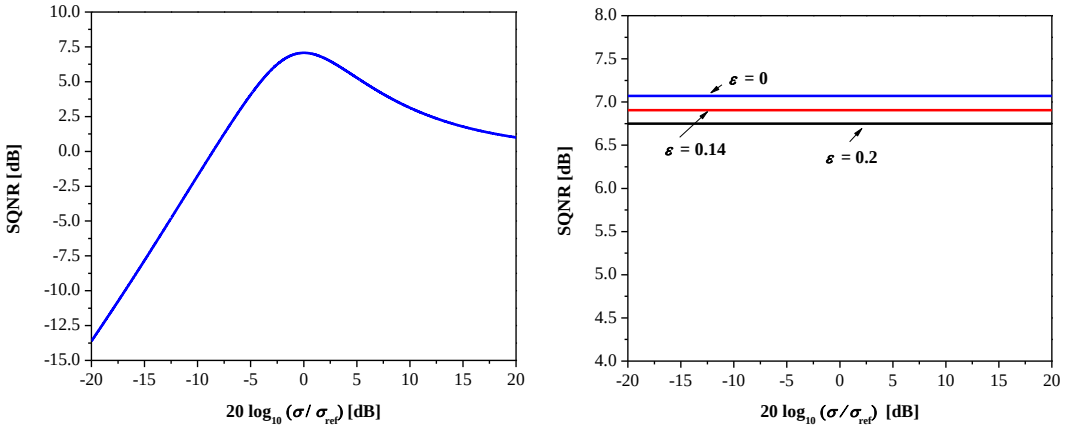
што указује на то да се Δ може одредити итеративно, што доказује лему.

На основу леме 6.1 добија се $\Delta = \Delta^{opt} = 1.089$, а ова вредност се у потпуности поклапа са вредношћу Δ^{opt} са слике 6.6 која је добијена нумерички.



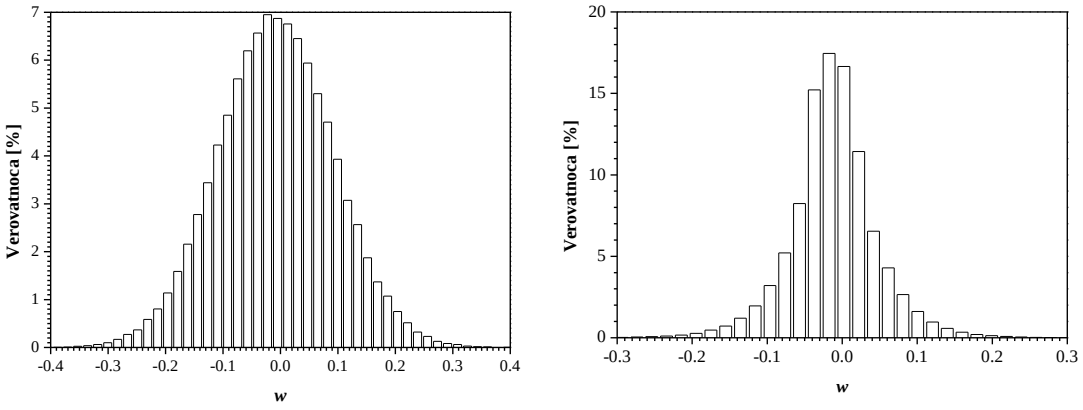
Сл. 6.6 Дисторзија у зависности од Δ за двобитни униформни квантизер.

Анализа теоријских резултата за адаптивни двобитни униформни квантизер. Слика 6.6 даје приказ перформанси за неадаптивни и адаптивни двобитни униформни Лапласов квантизер (за неколико вредности ε). Адаптивни модел има боље перформансе јер постиже константан SQNR у целом посматраном опсегу варијансе, а нешто ниже вредности за SQNR се добијају са порастом ε јер адаптација на варијансу тада није савршена (израз (6.4)).



Сл. 6.7 Перформансе двобитног квантизера у широком опсегу варијансе: а) неадаптивни и б) адаптивни.

Експериментални резултати. Користе се класична MLP и конволуциона (CNN) NN [117] развијене за класификацију слика, где се подаци узимају из MNIST базе [118]. MLP NN има исту структуру и вредности хиперпараметара као у одељку 6.1. CNN се састоји из конволуционог, *max-pooling*, потпуно повезаног (енг. *fully connected*) и излазног слоја. Број филтера у конволуционом слоју је 32, док је величина кернела 3×3. Величина *pooling* прозора је подешена на 2×2. *fully connected* слој има 100 а излазни слој 10 неурона. CNN мрежа се обучава у 10 епоха, при чему је *batch size* = 128.

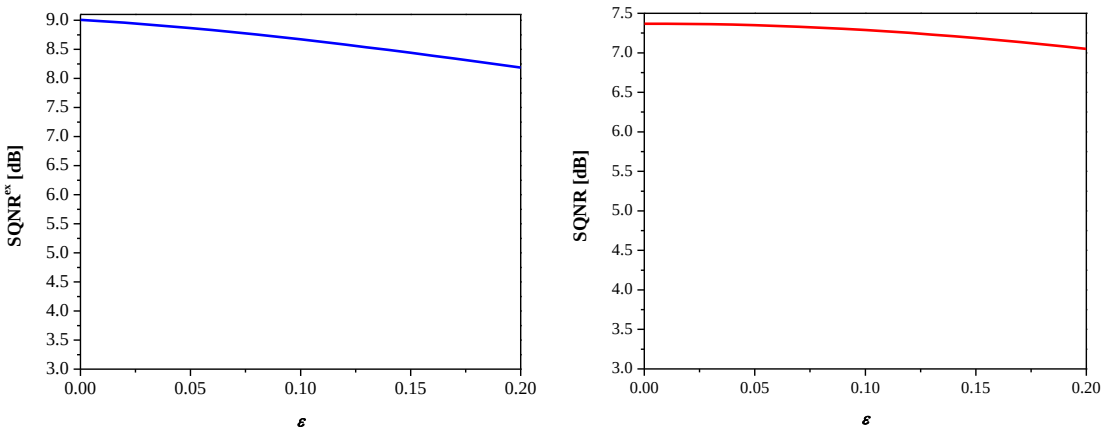


а) б)

Сл. 6.8 Хистограми тежина обучене NN: а) MLP и б) CNN.

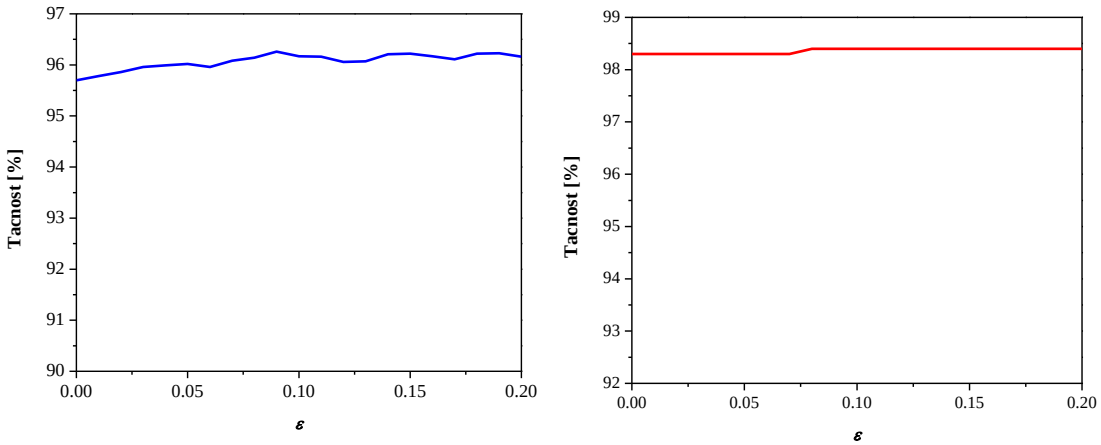
Слика 6.8 приказује хистограме тежина за обучену MLP и обучену CNN мрежу. Слике 6.9 и 6.10 показују како ε утиче на перформансе (SQNR) квантизера

односно перформансе квантоване NN. Са порастом ε перформансе квантизера се смањују што је у складу са теоријским резултатима на слици 6.6-б), а перформансе квантоване NN благо расту. За релевантну вредност параметра ε усваја се она за коју квантована NN постиже максимум у перформансама и то $\varepsilon = 0.09$ у случају MLP и $\varepsilon = 0.08$ у случају CNN.



а) б)

Сл. 6.9 SQNR у функцији од ε: а) MLP и б) CNN.



а) б)

Сл. 6.10 Тачност класификације за различите вредности ε: а) квантована MLP и б) квантована CNN.

У табелама 6.2 и 6.3 дате су перформансе за квантовану MLP и квантовану CNN, респективо, када се користи предложени али и други постојећи модели

двобитних квантизера [115, 119−121]. У односу на остале примењене моделе предложени квантизер је знатно ефикаснији са становишта SQNR-а, па су и перформансе квантоване MLP и CNN доста веће него у осталим случајевима. Квантована MLP постиже за 0.6 % мању тачност од MLP модела пуне прецизности, док деградација у случају CNN модела износи око 0.2 %. Такође, оба квантована NN модела су за око 16 пута мања од NN модела пуне прецизности.

Табела 6.2 Перформансе (тачност класификације и SQNR) квантоване MLP за различите примењене моделе двобитних квантизера

	Двобитни униформни [119]	**Двобитни униформни [115]**	**Двобитни униформни [120]**	**Двобитни неуниформни [121]**	**Двобитни предложени**	**Без компресије**
Тачност [%]	94.70	94.49	92.38	92.73	96.26	96.86
SQNR [dB]	1.63	1.19	-8.89	-2.41	8.71	-

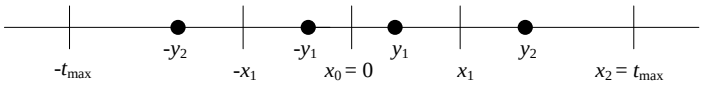
Табела 6.3 Перформансе (тачност класификације и SQNR) квантоване CNN за различите примењене моделе двобитних квантизера

	Двобитни униформни [119]	**Двобитни униформни [115]**	**Двобитни униформни [120]**	**Двобитни неуниформни [121]**	**Двобитни предложени**	**Без компресије**
Тачност[%]	96.1	96.6	97.8	97.2	98.4	98.6
SQNR [dB]	-14.12	2.36	7.00	5.56	7.32	-

6.3 Двобитни логаритамски компандинг квантизер

Овај одељак даје опис једног неуниформног двобитног решења за компресију неуронских мрежа које је предложено у раду [76], а засновано је логаритамској компандинг квантизацији.

Опис и пројектовање квантизера. На слици 6.11 приказан је симетрични неуниформни двобитни ($N = 4$) квантизер.



Сл. 6.11 Илустрација двобитног неуниформног квантизера.

Иначе, реализација разматраног неуниформног квантизера се изводи применом компандинг технике. Користи се логаритамска компресорска функција са μ-законом компресије $c(x)$ дата изразом (2.19). Инверзна компресорска функција $c^{-1}(x)$ која испуњава услов $c^{-1}(c(x)) = x$ је дефинисана са [1, 2]:

$$c^{-1}(x) = \frac{t_{\max}}{\mu}\left((1+\mu)^{x/t_{\max}} - 1\right)\operatorname{sgn}(x) . \tag{6.12}$$

Прагови одлуке и нивои двобитног логаритамског компандинг квантизера (LCSQ) се рачунају помоћу инверзне компресорске функције и дати су са:

$$t_0 = 0, \quad t_1 = \frac{t_{\max}}{\mu}\left((1+\mu)^{\frac{1}{2}} - 1\right), \quad t_2 = t_{\max} , \tag{6.13}$$

$$y_1 = \frac{t_{\max}}{\mu}\left((1+\mu)^{\frac{1}{4}} - 1\right), \quad y_2 = \frac{t_{\max}}{\mu}\left((1+\mu)^{\frac{3}{4}} - 1\right). \tag{6.14}$$

Компоненте дисторзије се рачунају на следећи начин:

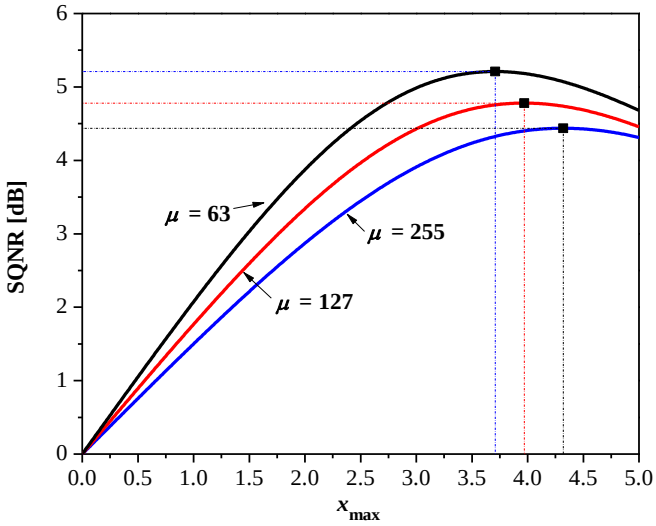
$$D_g = 2\int_{x_0=0}^{x_1} (x-y_1)^2 p(x)dx + 2\int_{t_1}^{t_{\max}} (x-y_2)^2 p(x)dx , \tag{6.15}$$

$$D_o = 2\int_{t_{\max}}^{+\infty} (x-y_2)^2 p(x)dx . \tag{6.16}$$

На основу (6.13)−(6.16) за дисторзију се добија следећи израз у затвореном облику:

$$\begin{aligned} D = 1 &- \frac{\sqrt{2}t_{\max}}{\mu}\left((1+\mu)^{1/4} - 1\right) + \frac{t_{\max}^2}{\mu^2}\left((1+\mu)^{1/4} - 1\right)^2 \\ &+ \exp\left(-\frac{\sqrt{2}t_{\max}\left(-1+\sqrt{1+\mu}\right)}{\mu}\right)\frac{t_{\max}}{\mu}(1+\mu)^{1/4} \\ &\times\left(\frac{t_{\max}}{\mu}\left(2\sqrt{1+\mu} - 2 + \mu\left((1+\mu)^{1/4} - 2\right)\right) - \sqrt{2}\left(\sqrt{1+\mu} - 1\right)\right) \end{aligned} . \tag{6.17}$$

На D утичу $t_{\max}$ и μ. За дато μ, на D утиче само $t_{\max}$. Оптимално $t_{\max}$ се одређује тако да D има минималну вредност односно SQNR максималну.



Сл. 6.12 SQNR као функција од t_{max} за двобитни LCSQ, за различито μ.

На слици 6.12 је дат је SQNR у функцији од t_{max} за различито μ. За свако μ може се идентификовати оптимално t_{max} за које SQNR постиже свој максимум, и ове вредности су дате у табели 6.4. Из табеле 6.4 се види да се за мање μ добија већи SQNR.

Табела 6.4 Оптималне вредности за t_{max} и SQNR постигнут у том случају, за неколико вредности μ

	t_{max}^{opt}	SQNR [dB]
μ = 63	3.707	5.21
μ = 127	3.965	4.78
μ = 255	4.318	4.44

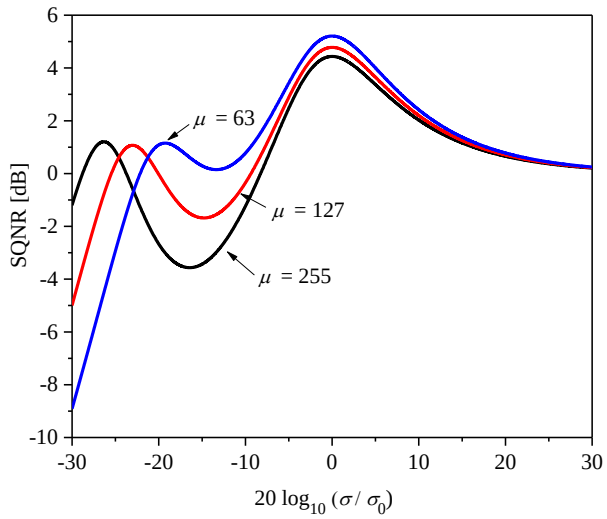
Анализа робусности двобитног LCSQ-а. За процену перформанси двобитног LCSQ-а у случају неприлагођења на варијансу користи се се следећи израз:

$$D(\sigma) = 2\int_{0}^{t_1} (x-y_1)^2 p(x,\sigma)dx + 2\int_{t_1}^{+\infty} (x-y_2)^2 p(x,\sigma)dx\,, \tag{6.18}$$

где је $p(x,\sigma)$ дефинисано изразом (2.6). Решавањем горњих интеграла добија се следећи израз у затвореном облику:

$$D(\sigma)=\sigma^2\left[1-\frac{\sqrt{2}t_{\max}\left((1+\mu)^{1/4}-1\right)}{\mu\cdot\sigma}+\frac{t_{\max}^2\left((1+\mu)^{1/4}-1\right)^2}{\mu^2\sigma^2}\right.$$
$$+\exp\left\{-\frac{\sqrt{2}t_{\max}\left(-1+\sqrt{1+\mu}\right)}{\mu\sigma}\right\}\frac{t_{\max}}{\mu\cdot\sigma}(1+\mu)^{1/4}\cdot \tag{6.19}$$
$$\left.\cdot\left(\frac{t_{\max}}{\mu\cdot\sigma}\left(2\sqrt{1+\mu}-2+\mu\left((1+\mu)^{1/4}-2\right)\right)-\sqrt{2}\left(\sqrt{1+\mu}-1\right)\right)\right]$$

SQNR за двобитни LCSQ у широком опсегу варијансе дат је на слици 6.13. Свака SQNR крива (добијена за различито μ) поред глобалног (за $\sigma^2 = \sigma_{\text{ref}}^2$) постиже и локални максимум, због тога што за неку специфичну вредност варијансе $-y_1$ и y_1 имају већи утицај на дисторзију од $-y_2$ и y_2, па се двобитни LCSQ тада понаша као бинарни LCSQ. Глобални SQNR максимуми су идентични са вредностима из табеле 6.4.

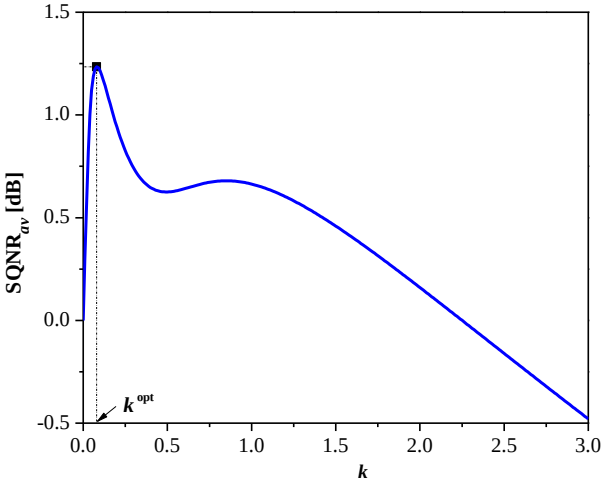


Сл. 6.13 SQNR у функцији варијансе сигнала за двобитни LCSQ и различито μ.

Како би се перформансе иницијалног LCSQ модела поправиле врши се додатна адаптација параметра $t_{\max}$ као:

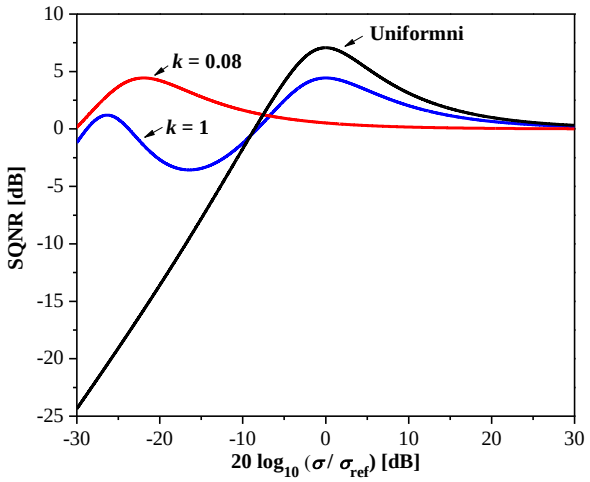
$$t_{\max}^{new}=k\cdot t_{\max}, k\in R, \tag{6.20}$$

где је k реална константа која се бира тако да средњи SQNR (израз (2.12)) у разматраном опсегу варијансе буде максималан.



Сл. 6.14 SQNR$_{av}$ у односу на k за разматрани двобитни LCSQ (μ = 255).

За μ = 255 бира се $k = k^{opt}$ = 0.08, као што је дато на слици 6.14. На слици 6.15 приказане су перформансе за двобитни LCSQ (μ = 255) пре (k = 1) и након скалирања параметра t_{max} (k = 0.08). У поређењу са двобитним униформним квантизером (Δ = 1.087 [1, 3, 8]) који постиже већи максимални SQNR, LCSQ (k = 0.08) је робуснији јер остварује већи SQNR$_{av}$ за више од 3.8 dB.



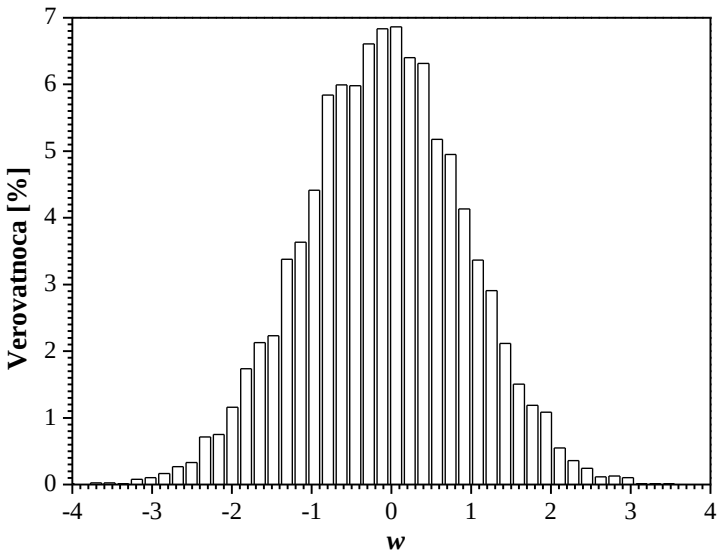
Сл. 6.15 Поређење перформанси двобитног LCSQ-а (μ = 255) и двобитног униформног квантизера (Δ = 1.087) у широком динамичком опсегу варијансе.

Табела 6.5 сумира оптималне вредности за k и SQNR$_{av}$ у том случају, за све посматране вредности μ. У поређењу са вредностима за SQNR$_{av}$ када је k = 1 остварен је добитак од око 0.6 dB.

Табела 6.5 k^{opt} и одговарајући $SQNR_{av}$ за разматрани двобитни LCSQ

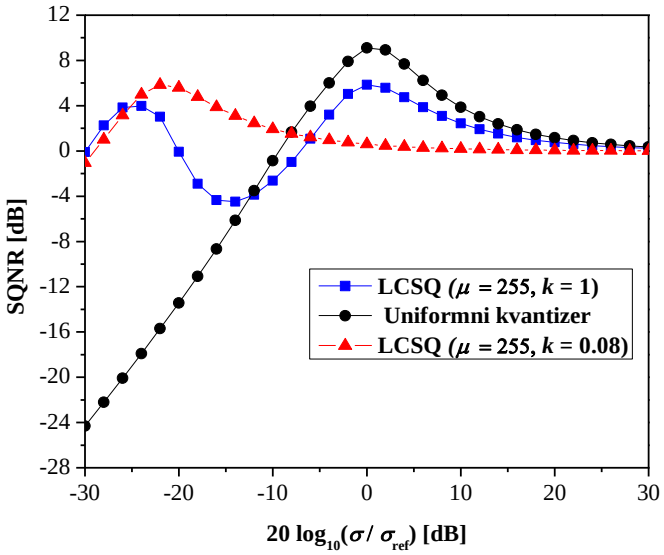
$N = 4$	$\mu = 63$	$\mu = 127$	$\mu = 255$
k^{opt}	0.4	0.09	0.08
$SQNR_{av}^{k=kopt}$ [dB]	1.67	1.37	1.23
$SQNR_{av}^{k=1}$ [dB]	1.09	1.03	0.66

Експериментални резултати. Користи се MLP NN за класификацију слика [117], а подаци се узимају из MNIST базе [118]. Архитектура MLP-а је једноставна и садржи улазни слој са 784 неурона и излазни слој са 10 неурона, а користе се следеће вредности хиперпараметра: *learning rate* = 0.5 и *batch size* = 250. Обучавање је рађено у 20 епоха, а хистограм тежина обучене MLP NN је дат на слици 6.16.



Сл. 6.16 Хистограм тежина обучене MLP NN.

На слици 6.17 су приказане SQNR вредности за предложени двобитни LCSQ-а (k = 1 и k = 0.08) али и за двобитни униформни квантизер, добијене при квантовању тежина чије се варијансе мењају у опсегу [-30 dB, 30 dB] у односу на референтну σ_w^2 (σ_w^2 је варијанса тежина обучене MLP мреже). Добијени резултати за SQNR се добро слажу са теоријским резултатима на слици 6.15; са друге стране, предност у односу на доста примењивани униформни квантизер је потврђена и на реалним подацима.



Сл. 6.17 Перформансе предложеног двобитног LCSQ-а (μ = 255) и униформног квантизера при квантовању тежина са различитим варијансама.

7. Закључак

У овом завршном поглављу ће бити сумирани најзначајнији резултати остварени у овој дисертацији:

1. За кодовање Гаусовог извора разматрани су квантизери са малим бројем нивоа и ентропијским кодом, за разлику од досадашњих истраживања која су се махом спроводила за квантизере са великим бројем нивоа и фиксном дужином кодних речи. Првенствена намена ових решења јесте компресија извора, с обзиром на то да су вредности SQNR-а ниске. Изложени су и нови начини пројектовања квантизера који су рачунски мање захтевни од стандардно коришћених метода (Лојд-Макс алгоритам). Такође, коришћене су и апроксимације *Q*-функције, имајући у виду да се при анализи Гаусових квантизера *Q*-функција неизоставно јавља, чиме је процес пројектовања додатно упрошћен.
2. За потребе PCM кодовања говора развијено је више нових решења која користе нове моделе квантизера и примењују нов начин процесирања. Поред квантизера за неограничен Лапласов извор, извршено је и пројектовање квантизера за ограничен Лапласов извор што није тако често рађено у литератури. Анализирани су и двомодни квантизери који се заснивају на наизменичном коришћењу квантизера за ограничен и неограничен извор односно квантизера за ограничен извор. Основна карактеристика двомодних квантизера је та да ако се правилно пројектују могу да остваре значајно веће перформансе у односу на квантизер пројектован за неограничен извор. На пример, анализом теоријских резултата показано је да PCM кодек са двомодним квантизером може да пружи исти квалитет сигнала мерен SQNR-ом као и стандардизовани G.711 квантизер при битској брзини мањој за 0.8125 bps. Такође, резулати за SQNR добијени на основу теоријских израза и на реалном говорном сигналу покалапају се доста добро.
3. Предложене су нове ADM конфигурације које заправо представљају надградњу базичне ADM шеме, где су имплементирани нови модели

квантизера и предиктора. Развијена ADM решења користе нов начин адаптације за квантизер/предиктор, а показано је да се на овај начин постиже већи максимални SNR и шири динамички опсег у односу на постојећа ADM решења. На пример, за двобитну и дводигитну ADM експериментално (на реалном говору) је утврђено да су значајно ефикаснији од својих еквивалената из класе тренутно адаптивне ADM али и од неких популарних алгоритама из ове класе попут CFDM-а и CVSDM-а. Осим тога, за тернарну ADM је показано да пружа за 4.5 dB већи SNR у односу на основну ADM шему. Такође, изузетно добре перформансе су остварене и код двобитног ADM-а са FLP-ом који је по први пут примењен у обради говора.

4. Анализирана је и област компресије неуронских мрежа и за ту сврху предложено је неколико модела скаларних квантизера. Квантизери (униформни и неуниформни) су намењени за *post-trening* квантизацију тежина, а приликом дизајна узета је у обзир расподела тежина и то је битна разлика у односу на већину доступних решења. Развијен је и ефикасан метод за адаптацију квантизера (у циљу ефикасног процесирања тежина), а на реалним примерима показано је да се овакавим приступом поправљају перформансе квантоване NN. Са предложеним бинарним решењем, које је добијено редизајнирањем иницијалног модела бинарног квантизера (чиме је повећан SQNR) и даљом адаптацијом модела, квантована MLP NN остварује за 1.3% већу тачност у односу на случај када се користи иницијални модел. Као доста добро решење у односу на постојеће двобитне квантизере показао се и адаптивни двобитни униформни квантизер (остварује већи SQNR), а код квантоване MLP и квантоване CNN примећена је мала деградација перформанси у односу на моделе пуне прецизности (0.6% у случају MLP и 0.3% у случају CNN). Такође, за двобитни LCSQ је показано да је боље решење за квантовање тежине од широко коришћеног униформног квантизера, а због тога што је робустан може се користити и као алтернатива адаптивним квантизерима.

Литература

[1] N. S. Jayant, P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, New Jersey, 1984.

[2] A. Gersho, R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Masachusetts, 1992.

[3] W. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*, John Wiley & Sons, 2003.

[4] R. Goldberg, L. Riek, *A Practical Handbook of Speech Coders*, CRC Press, 2000.

[5] L. Hanzo, C. Somerville, J. Woodard, *Voice and Audio Compression for Wireless Communications*, John Wiley & Sons - IEEE Press, 2007.

[6] I. V. McLoughlin, *Speech and Audio Processing − a Matlab based Approach*, Cambridge University Press, 2016.

[7] A. Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, John Wiley & Sons, 2004.

[8] K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann, 2018.

[9] D. Salomon, *A Concise Introduction to Data Compression*, Springer, New York, 2008.

[10] D. Salomon, *Variable-Length Codes for Data Compression*, Springer, London, 2007.

[11] R.M. Gray, D.L. Neuhoff, "Quantization", *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2384, 1998.

[12] S. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129-136, 1982.

[13] Y. Hiwasaki, H. Ohmuro, T. Mori, S. Kurihara, A. Kataoka, "A G.711 embedded wideband speech coding for VoIP conferences", *IEICE Trans. Information and Systems*, vol. E89-D, no. 9, pp. 2542– 2552, 2006.

[14] Y. Hiwasaki, T. Mori, S. Sasaki, H. Ohmuro, “A wideband speech and audio coding candidate for ITU-T G.711 WBE standardization”, *Proc. IEEE ICASSP*, Las Vegas, USA, pp. 4017–4020, 2008.

[15] A. Ortega, M. Vetterly, “Adaptive scalar quantization without side information”, *IEEE Trans. on Image Processing,* vol. 6, no. 5, pp. 665–676, 1997.

[16] J. Nikolić, Z. Perić, A. Jovanović, “Two forward adaptive dual-mode companding scalar quantizers for Gaussian source”, *Signal Processing*, vol. 120, pp.129−140, 2016.

[17] Z. Perić, J. Nikolić, “High-quality Laplacian source quantisation using a combination of restricted and unrestricted logarithmic quantisers”, *IET SignalProcessing,* vol. 6, no.7, pp. 633–640, 2012.

[18] Z. Perić, J. Nikolić, “An adaptive waveform coding algorithm and its application in speech coding”, *Digital Signal Processing*, vol. 22, no. 1, pp. 199−209, 2012.

[19] J. Nikolić, Z. Perić, “Lloyd-Max's algorithm implementation in speech coding algorithm based on forward adaptive technique”, *Informatica,* vol. 19, no. 2, pp. 255−270, 2008.

[20] M. Dinčić, Z. Perić, A. Jovanović, “New coding algorithm based on variable-length codewords for piecewise uniform quantizers”, *Informatica,* vol. 27, no. 3, pp. 527–548, 2016.

[21] Z. Perić, M. Dinčić, D. Denić, A. Jocić, “Forward adaptive logarithmic quantizer with new lossless coding method for Laplacian source”, *Wireless Personal Communications,* vol. 59, pp. 625–641, 2010.

[22] Z. Perić, M. Savić, M. Dinčić, D. Denić, M. Praščević, “Forward adaptation of novel semilogarithmic quantizer and lossless coder for speech signals compression”, *Informatica,* vol. 21, no. 3, pp. 375–391, 2010.

[23] D. G. Zrilić, *Circuits and Systems Based on Delta Modulation,* Springer, Berlin, 2005.

[24] S. Sarade, “Speech compression by using adaptive differential pulse code modulation (ADPCM) technique with microcontroller”, *Journal of Electronics and Communication Systems*, vol.2, no.3, pp. 1–9, 2017.

[25] S. Uddin, I.R. Ansari, S. Naaz, “Low bit rate speech coding using differential pulse code modulation”, *Advances in Research*, vol. 8, no. 3, pp. 1–6, 2016.

[26] V. Despotović, Z. Perić, L. Velimirović, V. Delić, “DPCM with forward gain-adaptive quantizer and simple switched predictor for high quality speech signals”, *Advances in Electrical and Computer Engineering*, vol. 10, no. 4, pp. 95–98, 2010.

[27] Z. Perić, B. Denić, V. Despotović, “Delta modulation system with a limited error propagation”, *Proceedings of XIII International Conference SAUM*, Niš, 2016.

[28] P.W.M. Tsang, W.K. Cheung, T.C. Poon, “Near computation-free compression of Fresnel holograms based on adaptive delta modulation”, *Opt. Eng*. vol. 50, no. 8, Article ID: 085802, 2011.

[29] S.H. Dandach, S. Dasgupta, B.D.O Anderson, “Stability of adaptive delta modulators with forgetting factor and constant inputs”, *Int. J. Adapt. Contr. Signal Process.*, vol. 25, pp. 723–739, 2011.

[30] F. Gomez-Estern, C. Canudas-de-Wit, F.R. Rubio, “Adaptive delta modulation in networked controlled systems with bounded disturbances”, *IEEE Transactions on Automatic Control*, vol. 56, no. 1, pp. 129–134, 2011.

[31] H. Zheng, Z. Lu, “Research and design of a 2-bit delta modulator encoder/decoder,” *Proceedings of the 24th Chinese Control and Decision Conf*. (CCDC), Taiyuan, 2012.

[32] E. A. Prosalentis, G. S. Tombras, “A 2-bit adaptive delta modulation system with improved performance”, *EURASIP Journal on Advances in Signal Processing*, Article ID 16286, 2007.

[33] M. A. Aldajani, A. H. Sayed, “Stability and performance analysis of an adaptive sigma-delta modulator”, *IEEE Transactions on Circuits and Systems II*, vol. 48, issue 3, pp. 233–244, 2001.

[34] G. S. Tombras, “New adaptation algorithm for a two-digit adaptive delta modulation system”, *International Journal of Electronics*, vol. 68, no. 3, pp. 343–349, 1990.

[35] Z. Perić, B. Denić, V. Despotović, N. Simić, “Delta modulation with improved prediction and quasilogarithmic quantizer for Laplacian source”, *Proceedings of 14th International Conference SAUM*, Niš, 2018.

[36] N. Simić, Z. Perić, B. Denić, M. Tančić, “Speech signal coding scheme based on multibit delta modulation and LMS algorithm”, *Proceedings of 14th International Conference SAUM* , Niš, 2018.

[37] A. Krizhevsky, I. Sutskever, G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, *Proceedings of the International Conference on Neural Information Processing Systems*, Harrahs and Harveys, Lake Tahoe, NV, USA, pp. 1097–1105, 2012.

[38] S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks”, *Proceedings of the Conference on Advances in neural information processing systems (NeurIPS)*, Montreal, Canada, pp. 91–99, 2015.

[39] A. Conneau, H. Schwenk, L. Barrault, Y. Lecun, “Very deep convolutional networks for text classification”, *arXiv preprint arXiv: 1606.01781*, 2016.

[40] M. Dinčić, Z. Perić, M. Tančić, D. Denić, Z. Stamenković, B. Denić, “Support region of μ-law logarithmic quantizers for Laplacian source applied in neural networks”, *Microelectronics Reliability,* vol. 124, article ID: 114269, 2021.

[41] Z. Perić, B. Denić, M. Savić, M. Dincić, D. Mihajlov, “Quantization of weights of neural networks with negligible decreasing of prediction accuracy”, *Information Technology and Control*, vol. 50, no. 3, pp. 558−569, 2021.

[42] N. Wang, J. Choi, D. Brand, et al. “Training deep neural networks with 8-bit floating point numbers”, *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada, pp. 1–10, 2018.

[43] R. Banner, I. Hubara, E. Hoffer, et al. “Scalable methods for 8-bit training of neural networks”, *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montréal, Canada, pp. 1–9, 2018.

[44] R. Banner, Y. Nahshan, D. Soudry, “Post training 4-bit quantization of convolutional networks for rapid-deployment”, *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 2019.

[45] J. Choi, “Accurate and efficient 2-bit quantized neural networks”, *Proceedings of the 2nd Systems and Machine Learning Conference*, Palo Alto, 2019.

[46] J. Choi, S. Venkataramani, V. Srinivasan, K. Gopalakrishnan, Z. Wang, P. Chuang, “Accurate and efficient 2-bit quantized neural networks”, *Proceedings of the 2nd SysML Conference*, Stanford, CA, USA, 2019.

[47] J., Choi, et al.: “Bridging the accuracy gap for 2-bit quantized neural networks (QNN)”, *arXiv 1807.06964v1* [cs.CV], 2018.

[48] C. Zhu, S. Han, H. Mao, et al., “Trained ternary quantization”, *Proceedings of the International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.

[49] H. Qina, R. Gonga, X. Liu, X. Baie, J. Songc, N. Sebed, “Binary neural networks: A survey”, *Pattern Recognition*, vol. 105, Article ID: 107281, 2020.

[50] T. Simons, D. J. Lee, “A review of binarized neural networks”, *Electronics*, vol. 8, pp. 1−25, 2019.

[51] Y. Wang, J. Lin, Z. Wang, “An energy-efficient architecture for binary weight convolutional neural networks”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, pp. 280 – 293, 2018.

[52] Y. Li, Y. Bao, W. Chen, “Fixed-sign binary neural network: An efficient design of neural network for Internet-of-Things devices”, *IEEE Access*, vol. 8, pp. 164858–164863, 2018.

[53] Z. Perić, B. Denić, Milan Savić, Nikola Vučić, Nikola Simić, V. Despotović, “Binary quantization analysis of neural networks weights on MNIST dataset”, *Elektronika IR Elektrotechnika*, vol. 27, no. 4, pp. 55−61, 2021.

[54] P. Pham, J. Abraham, J. Chung, “Training multi-bit quantized and binarized networks with a learnable symmetric quantizer”, *IEEE Access*, vol. 9, pp. 47194–47203, 2021.

[55] K. Paliwal, K. Wojcicki, “Effect of analysis window duration on speech intelligibility”, *IEEE Signal Processing Letters*, vol. 15, pp. 785–788, 2008.

[56] S. Gazor, W. Zhang, “Speech probability distribution”, *IEEE Signal Process Letters*, vol. 10, issue 7, pp. 204–207, 2003.

[57] R. Banner, Y. Nahshan, E. Hoffer, et al. “ACIQ: Analytical clipping for integer quantization of neural networks”, *arXiv preprint, arXiv:1810.05723*, 2018.

[58] R. Arshad, A. Saleem, D. K. Demirhan, “Performance comparison of Huffman coding and double Huffman coding”, *Proc. Int. Conf. on Innovative Computing Technology (INTECH)*, Dublin, Ireland, 2016.

[59] M.E. Demirhan, U. Arıöz, “Huffman coding of a PCM-quantized speech signal”, *Proc. Int. Conf. Signal Processing and Communications Applications Conf. (SIU)*, Malatya, Turkey, 2015.

[60] L. Velimirović, Z. Perić, B. Denić, “Application of extended Huffman coding on three-level scalar quantizer for Gaussian source”, *The Second National Conference Information Theory and Complex Systems*, Niš, 2014.

[61] M. Dinčić, Z. Perić, “Design of quantizers with Huffman coding for Laplacian source”, *Electronics and Electrical Engineering*, vol. 10, pp. 129−132, 2010.

[62] K. Fredriksson, J. Tarhio, “Efficient string matching in Huffman compressed texts”, *Fundamenta Informaticae*, vol. 62, pp. 1−16, 2004.

[63] L. Velimirović, Z. Perić, B. Denić, “Design and analysis of the two-level scalar quantizer with extended Huffman coding”, *The First National Conference Information Theory and Complex Systems*, Belgrade, 2013.

[64] Z. Perić, B. Denić, V. Despotović, “Gaussian source coding based on variance-mismatched three-level scalar quantization using *Q*-function approximations”, *IET Communications*, vol. 14, pp. 594–602, 2020.

[65] Z. Perić, B. Denić, M. Savić, V. Despotović, “Design and analysis of binary scalar quantizer of Laplacian source with applications”, *Information*, vol. 11, 18 pages, 2020.

[66] B. Denić, Z. Perić, V. Despotović, “Forward adaptive speech coding with low bit rates and variable word length”, *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 15, pp. 125–136, 2016.

[67] Z. Perić, J. Nikolić, B. Denić, V. Despotović, “Forward adaptive dual-mode quantizer based on the first-degree spline approximation and embedded G.711 codec”, *Radioengineering*, vol. 28, no. 4, pp. 729−739, 2019.

[68] B. Denić, Z. Perić, N. Vučić, V. Despotović, “Forward adaptive Laplacian source coding based on restricted quantization”, *Information Technology and Control*, vol. 47, pp. 209−219, 2018.

[69] B. Denić, Z. Perić, V. Despotović, “Three-level delta modulation for Laplacian source coding”, *Advances in Electrical and Computer Engineering*, vol. 17, pp. 95−102, 2017.

[70] Z. Perić, B. Denić, V. Despotović, “Novel two-bit adaptive delta modulation algorithms”, *Informatica*, vol. 30, pp. 117−134, 2019.

[71] Z. Perić, B. Denić, V. Despotović, “An efficient two-digit adaptive delta modulation for Laplacian source coding”, *International Journal of Electronics*, vol. 106, pp. 1085−1100, 2019.

[72] Z. Perić, B. Denić, V. Despotović, “Multilevel delta modulation with switched first-order prediction for wideband speech coding”, *Elektronika IR Elektrotechnika*, vol. 24, pp. 46−51, 2018.

[73] Z. Perić, B. Denić, V. Despotović, “Three-level delta modulation with second-order prediction for Gaussian source coding“, *Advances in Electrical and Computer Engineering*, vol. 18, pp. 95−102, 2018.

[74] Z. Perić, B. Denić, V. Despotović, “Algorithm based on 2-bit adaptive delta modulation and fractional linear prediction for Gaussian source coding”, *IET Signal Processing*, vol. 15, issue 6, pp. 410–423, 2021.

[75] Z. Perić, M. Savić, N. Simić, B. Denić, V. Despotović, “Design of a 2-bit neural network quantizer for Laplacian source”, *Entropy*, vol. 14, pp. 594–602, 2020.

[76] Z. Perić, B. Denić, M. Dinčić, J. Nikolić, “Robust 2-bit quantization of weights in neural network modeled by Laplacian distribution”, *Advances in Electrical and Computer Engineering*, vol. 21, no. 3, pp. 3−10, 2021.

[77] S. Na, D. L. Neuhoff, “Monotonicity of step sizes of MSE optimal symmetric uniform scalar quantizers”, *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1782–1792, 2019.

[78] S. Na, D. L Neuhoff, “On the convexity of the MSE distortion of symmetric uniform scalar quantization”, *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2626–2638, 2018.

[79] S. Na, D.L. Neuhoff, “On the convexity of the MSE distortion of symmetric uniform scalar quantization”, *IEEE Trans. Inf. Theory*, vol. 64, pp. 2626–2638, 2017.

[80] D. Hui, D. L. Neuhoff, “Asymptotic analysis of optimal fixed-rate uniform scalar quantization”, *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 957–977, 2001.

[81] S. Na, D. L. Neuhoff, “On the support of MSE-optimal, fixed-rate, scalar quantizers”, *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2972–2982, 2001.

[82] S. Na, “On the support of fixed-rate minimum mean-squared error scalar quantizers for a Laplacian source”, *IEEE Transactions on Information Theory*, vol. 50, no. 5, pp. 937–944, 2004.

[83] Z. Perić, J. Nikolić, “An effective method for initialization of Lloyd-Max's algorithm of optimal scalar quantization for laplacian source”, *Informatica*, vol. 18, no. 2, pp. 279−288, 2007.

[84] Z. Perić, N. Simić, J. Nikolić, “Design of single and dual-mode companding scalar quantizers based on piecewise linear approximation of the Gaussian PDF”, *J. Frankl. Inst.*, vol. 357, pp. 5663–5679, 2020.

[85] Z. Perić, G. Petković, B. Denić, A. Stanimirović, V. Despotović, L. Stoimenov, "Gaussian source coding using a simple switched quantization algorithm and variable length codewords", *Advances in Electrical and Computer Engineering*, vol. 20, pp. 11−18, 2020.

[86] M. Dinčić, Z. Perić, D. Denić, Z. Stamenković, "Design of robust quantizers for low-bit analog-to-digital converters for Gaussian source", *Journal of Circuits, Systems and Computers*, vol. 28, no. supp01, 1940002, 2019

[87] Z. Perić, S. Suzić, T. Delić, N. Simić, "Support region of semilogarithmic quantizer for Laplacian source", *Elektronika ir Elektrotechnika*, vol. 28, no. 4, pp. 64−67, 2018.

[88] A. D. Lyon, "The μ-law CODEC", *Journal of Object Technology*, vol. 7, no. 8, pp. 17–31, 2008.

[89] ITU-T Recommendation G.711: Pulse code modulation of voice frequencies, 1972.

[90] Z. Perić, J. Nikolić, M. Petković, "A class of tight bounds on the *Q*-function with closed-form upper bound on relative errors", *Math. Methods Appl. Sci.*, vol. 42, pp. 1786–1794, 2019.

[91] A. Marković, Z. Perić, S. Panić, et al.: 'An improved method for ASEP evaluation over fading channels based on *Q*-function approximation', *IETE J. Res.*, vol. 64, no. 6, pp. 777–784, 2018.

[92] J. Nikolić, Z. Perić, A. Jovanović, "Novel approximations for the Q-function with application in SQNR calculation", *Digit. Signal Process.*, vol. 65, pp. 71–80, 2017.

[93] D. Sadhwani, R.N. Yadav, S. Aggarwal, "Tighter bounds on the Gaussian *Q*-function and its application in Nakagami-m fading channel", *IEEE Wireless Commun. Lett.*, vol. 6, no. 5, pp. 574–577, 2017.

[94] Lopez-Benitez, M.: "Average of arbitrary powers of Gaussian *Q*-function over η–μ and κ–μ fading channels", *Electron. Lett.*, vol. 51, no. 11, pp. 869–871, 2015.

[95] V.N.Q. Bao, L. P. Tuyen, H. H. Tue: "A survey on approximations of onedimensional Gaussian *Q*–function", *Rev J. Electron. Commun.*, 5, pp. 13–16, 2015.

[96] A. Gasull, F. Utzet, “Approximating Mills ratio”, J. *Math. Anal. Appl.*, vol. 420, no. 2, pp. 1832–1853, 2014.

[97] P. Fan, “New inequalities of Mill’s ratio and its application to the inverse *Q*-function approximation”, *Aust. J. Math. Anal. Appl.*, vol. 10, no. 1, pp. 1–11, 2013.

[98] W. M. Jang, “A simple upper bound of the Gaussian *Q*-function with closed-form error bound”, *IEEE Commun. Letters,* vol. 15, no. 2, pp. 157–159, 2011.

[99] S. Na, “Asymptotic formulas for variance-mismatched fixed-rate scalar quantization of a Gaussian source”, *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2437−2441, 2011.

[100] G. Karagiannidis, A. Lioumpas, “An improved approximation for the Gaussian *Q*-Function”, *IEEE Commun. Lett*ers, vol. 11, no. 8, pp. 644–646, 2007.

[101] Tellambura, C., Annamalai, A.: “Efficient computation of erfc(x) for large arguments”, *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 529–532, 2000.

[102] S. Na, D. L. Neuhoff, “Asymptotic MSE distortion of mismatched uniform scalar quantization”, *IEEE Transactions on Information Theory,* vol. 58, no. 5, 2012.

[103] S. Na, “Variance-mismatched fixed-rate scalar quantization of Laplacian sources”, *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4561–4572, 2011.

[104] S. Na, “Asymptotic formulas for variance-mismatched fixed-rate scalar quantization of a Gaussian source”, *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2437–2441, 2011.

[105] S. Na, “Asymptotic formulas for mismatched fixed-rate minimum MSE Laplacian quantizers”, *IEEE Signal Processing Letters*, vol. 15, pp. 13–16, 2008.

[106] P. Demonte, “HARVARD speech corpus—Audio recording 2019”, University of Salford Collection, 2019.

[107] L. Schumaker, *Spline Functions: Basic Theory,* New York, USA: Cambridge University Press, 2007.

[108] D. Aleksić, Z. Perić, J. Nikolić, “Support region determination of the quasilogarithmic quantizer for Laplacian source”, *Przeglad Elektrotechniczny*, vol. 88, no. 7A, pp. 130–132, 2012.

[109] Z. Perić, N. Simić, J. Nikolić, “Design of single and dual-mode companding scalar quantizers based on piecewise linear approximation of the Gaussian PDF”, *Journal of the Franklin Institute*, vol. 357, no. 9, pp. 5663–5679, 2020.

[110] B. Denić, Z. Perić, V. Despotović, N. Vučić, P. Petrović, “Dual-mode quasi-logarithmic quantizer with embedded G.711 codec”, *Journal of Electrical Engineering-Elektrotechnicky Casopis*, vol. 69, pp. 46−51, 2018.

[111] ITU-T Recommendation G.712: Transmission performance characteristics of pulse code modulation (PCM), 1992.

[112] L. Velimirović, S. Marić, “New adaptive compandor for LTE signal compression based on spline approximations”, *ETRI Journal*, vol. 38, no. 3, p. 463–468, 2016.

[113] S. Garofolo, et al., *TIMIT Acoustic-phonetic Continuous Speech Corpus*, Linguistic Data Consortium, Philadelphia, USA, 1993.

[114] Z. Perić, M. Savić, M. Dinčić, N. Vučić, D. Đjošić, S. Milosavljević, “Floating point and fixed point 32-bits quantizers for quantization of weights of neural networks”, *Proceedings of the 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, Romania, 2021.

[115] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, Y. Bengio, “Quantized neural networks: Training neural networks with low precision weights and activations”, *J. Mach. Learn. Res.*, vol. 18, 1–30, 2018.

[116] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, D. Kalenichenko, “Quantization and training of neural networks for ecient integer-arithmetic-onlyinference”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018.

[117] A. Zhang, Z.C. Lipton, M. Li, A.J. Smola, *Dive into Deep Learning*, *arXiv 2020, arXiv:2106.11342*, 2020.

[118] Y. LeCun, C. Cortez, C. Burges, *The MNIST Handwritten Digit Database*. (yann.lecun.com/exdb/mnist/)

[119] Y. Bhalgat, J. Lee, M. Nagel, T. Blankevoort, N. Kwak, "LSQ+: Improving low-bit quantization through learnable offsets and better initialization", *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, 2020.

[120] Y. Li, X. Dong, W.Wang, "Additive powers-of-two quantization: An efficient non-uniform discretization for neural networks", *Proceedings of the International Conference on Learning Representations (ICLR)*, Virtual Conference, Formerly Addis Ababa, Ethiopia, 2020.

[121] C. Baskin, N. Liss, E. Schwartz, E. Zheltonozhskii, R. Giryes, M. Bronstein, A. Mendelso, "Uniq: Uniform noise injection for non-uniform quantization of neural networks", *ACM Trans. Comput. Syst*., vol. 37, pp. 1–15, 2021.

[122] P. P. Vaidyanathan, *The theory of Linear Prediction*, ser. Synthesis lectures on signal processing, J. Moura Ed. San Rafael, CA, USA: Morgan & Claypool, 2008.

[123] J.D. Gibson, "On the high rate, independence, and optimal prediction assumptions in predictive coding", *Proceedings of the IEEE Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, 2017.

[124] J. Gibson, "Speech compression", *Information*, vol. 7, no. 32, pp. 1–22, 2016.

[125] J. Lee, S. Na, "A rigorous revisit to the partial distortion theorem in the case of a Laplacian source", *IEEE Communications Letters*, vol. 21, issue 12, pp. 2554–2557, 2017.

[126] V. Despotović, T. Skovranek, Z. Perić, "One-parameter fractional linear prediction", *Comput. Electr. Eng*., vol. 69, pp. 158–170, 2018.

[127] T. Skovranek, V. Despotović, Z. Perić, "Optimal fractional linear prediction with restricted memory", *IEEE Signal Process. Lett*., vol. 26, no. 5, pp. 760–764, 2019.

[128] T. Skovranek, V. Despotović, Z. Perić, "Two-dimensional fractional linear prediction", *Comput. Electr. Eng*., vol. 77, pp. 37–46, 2019.

Списак објављених радова аутора

а) Радови објављени у међународним часописима са IMPACT фактором (M21, M22, M23)

a.1. Z. Perić, **B. Denić**, V. Despotović, "Novel Two-Bit Adaptive Delta Modulation Algorithms", *Informatica*, vol. 30, pp. 117−134, 2019. (M21)
http://dx.doi.org/10.15388/Informatica.2019.200

a.2. Z. Perić, M. Savić, N. Simić, **B. Denić**, V. Despotović, "Design of a 2-Bit Neural Network Quantizer for Laplacian Source", *Entropy*, vol. 14, pp. 594–602, 2020. (M22)
https://doi.org/10.3390/e23080933

a.3. Z. Perić, **B. Denić**, V. Despotović, "Gaussian Source Coding Based on Variance-Mismatched Three-Level Scalar Quantization using Q-function Approximations", *IET Communications*, vol. 14, pp. 594–602, 2020. (M22)
https://doi.org/ 10.1049/iet-com.2019.0431

a.4. Z. Perić, **B. Denić**, A. Jovanović, S. Milosavljevic, M. Savić, "Performance Analysis of a 2-Bit Dual-Mode Uniform Scalar Quantizer for Laplacian Source", рад је прихваћен за штампање у часопису *Information Technology and Control*. (M23)

a.5. Z. Perić, **B. Denić**, V. Despotović, "Algorithm Based on 2-Bit Adaptive Delta Modulation and Fractional Linear Prediction for Gaussian Source Coding", *IET Signal Processing*, vol. 15, Issue 6, pp. 410–423, 2021. (M23)
https://doi.org/ 10.1049/sil2.12040

a.6. M. Dinčić, Z. Perić, M. Tančić, D. Denić, Z. Stamenković, **B. Denić**, "Support Region of μ-Law Logarithmic Quantizers for Laplacian Source Applied in Neural Networks", *Microelectronics Reliability*, vol. 124, article ID: 114269, 2021. (M23)
https://doi.org/10.1016/j.microrel.2021.114269

a.7. Z. Perić, **B. Denić**, M. Savić, M. Dincić, D. Mihajlov, "Quantization of Weights of Neural Networks with Negligible Decreasing of Prediction Accuracy", *Information Technology and Control*, vol. 50, no. 3, pp. 558−569, 2021. (M23)
https://doi.org/10.5755/j01.itc.50.3.28468

a.8. Z. Perić, **B. Denić**, Milan Savić, Nikola Vučić, Nikola Simić, V. Despotović, "Binary Quantization Analysis of Neural Networks Weights on MNIST Dataset", *Elektronika IR Elektrotechnika*, vol. 27, no. 4, pp. 55−61, 2021. (M23)
https://doi.org/10.5755/j02.eie.28881

a.9. Z. Perić, **B. Denić**, M. Dinčić, J. Nikolić, "Robust 2-Bit Quantization of Weights in Neural Network Modeled by Laplacian Distribution", *Advances in Electrical and Computer Engineering*, vol. 21, no. 3, pp. 3−10, 2021. (M23)
http://dx.doi.org/10.4316/AECE.2020.04002

a.10. Z. Perić, G. Petković, **B. Denić**, A. Stanimirović, V. Despotović, L. Stoimenov, "Gaussian Source Coding using a Simple Switched Quantization Algorithm and Variable Length Codewords", *Advances in Electrical and Computer Engineering*, vol. 20, pp. 11−18, 2020. (M23)
http://dx.doi.org/10.4316/AECE.2020.04002

a.11. Z. Perić, **B. Denić,** V. Despotović, "An Efficient Two-Digit Adaptive Delta Modulation for Laplacian Source Coding", *International Journal of Electronics*, vol. 106, pp. 1085−1100, 2019. (M23)
http://dx.doi.org/10.1080/00207217.2019.1582707

a.12. Z. Perić, J. Nikolić, **B. Denić**, V. Despotović, "Forward Adaptive Dual-Mode Quantizer Based on the First-Degree Spline Approximation and Embedded G.711 Codec", *Radioengineering*, vol. 28, no. 4, pp. 729−739, 2019. (M23)
http://dx.doi.org/ 10.13164/re.2019.0729

a.13. B. Denić, Z. Perić, N. Vučić, V. Despotović, "Forward Adaptive Laplacian Source Coding Based on Restricted Quantization", *Information Technology and Control*, vol. 47, pp. 209−219, 2018. (M23)
http://dx.doi.org/10.5755/j01.itc.47.2.16670

a.14. Z. Perić, **B. Denić**, V. Despotović, "Multilevel Delta Modulation with Switched First-Order Prediction for Wideband Speech Coding", *Elektronika IR Elektrotechnika*, vol. 24, pp. 46−51, 2018. (M23)
http://eejournal.ktu.lt/index.php/elt/article/view/20156

а.15. Z. Perić, **B. Denić**, V. Despotović, "Three-Level Delta Modulation with Second-Order Prediction for Gaussian Source Coding", *Advances in Electrical and Computer Engineering*, vol. 18, pp. 95−102, 2018. (M23)
http://dx.doi.org/ 10.4316/AECE.2018.03017

а.16. B. Denić, Z. Perić, V. Despotović, N. Vučić, P. Petrović, "Dual-Mode Quasi-Logarithmic Quantizer with Embedded G.711 Codec", *Journal of Electrical Engineering-Elektrotechnicky Casopis*, vol. 69, pp. 46−51, 2018. (M23)
http://dx.doi.org/10.1515/jee-2018-0006

а.17. B. Denić, Z. Perić, V. Despotović, "Three-Level Delta Modulation for Laplacian Source Coding", *Advances in Electrical and Computer Engineering*, vol. 17, pp. 95−102, 2017. (M23)
http://dx.doi.org/10.4316/AECE.2017.01014

б) Радови објављени у међународним часописима без IMPACT фактора (M24, M51)

б.1. B. Denić, Z. Perić, V. Despotović, "Forward Adaptive Speech Coding With Low Bit Rates and Variable Word Length", *Facta Universitatis, Series: Automatic Control and Robotics*, vol. 15, pp. 125–136, 2016. (M24)
http://casopisi.junis.ni.ac.rs/index.php/FUAutContRob/article/view/1622/1277

б.2. Z. Perić, **B. Denić**, M. Savić, V. Despotović, "Design and Analysis of Binary Scalar Quantizer of Laplacian Source with Applications", *Information*, 2020, vol. 11, 18 pages. (M51)
https://doi.org/10.3390/info11110501

в) Радови саопштени на међународним научним скуповима (M33, M34)

в.1. Z. Perić, N. Vučić, M. Dinčić, D. Ćirić, **B. Denić**, A. Perić, "Design of Uniform Scalar Quantizer for Discrete Input Signals", *Proceedings of 28th Telecommunications Forum (TELFOR)*, Belgrade, 2020. (M33)

в.2. Z. Perić, **B. Denić**, V. Despotović, N. Simić, "Delta Modulation with Embeded G.711 Codec", *Proceedings of 26th Telecommunications Forum Telfor (TELFOR)*, Belgrade, 2018. (M33)

в.3. Z. Perić, **B. Denić**, V. Despotović, N. Simić, "Delta Modulation with Improved Prediction and Quasilogarithmic Quantizer for Laplacian Source", *Proceedings of XIV International Conference SAUM*, Niš, 2018. (M33)

в.4 N. Simić, Z. Perić, **B. Denić**, M. Tančić, "Speech Signal Coding Scheme Based on Multibit Delta Modulation and LMS Algorithm", *Proceedings of XIV International Conference SAUM* , Niš, 2018. (M33)

в.5. N. Vučić, **B. Denić**, Z. Perić, "Design of Switched Quasi-Logarithmic Quantizer of Laplacian Source with Golomb-Rice Codes for Medium Quality of Reconstructed Signal", *Proceedings of 26th International Conference Noise and Vibration*, pp. 161−165, 2018. (M33)

в.6. Z. Perić, **B. Denić**, V. Despotović, "Delta Modulation System with a Limited Error Propagation", *Proceedings of XIII International Conference SAUM*, Niš, 2016. (M33)

в.7. Z. Perić, **B. Denić**, L. Velimirović, "Construction of Three-level Scalar Quantizer with Extended Huffman Coding for Gaussian Source", *Proceedings of XII International Conference SAUM*, Niš, 2014. (M33)

в.8. B. Denić, Z. Perić, N. Vučić, "Forward Adaptive Speech Coding with Variably Words Length and Three Levels Quantizer", *Proceedings of the Third International Conference TAKTONS*, Novi Sad, 2015. (M34)

г) Радови саопштени на националним научним скуповима (M63)

г.1. L. Velimirović, Z. Perić, **B. Denić**, "Application of Extended Huffman Coding on Three-Level Scalar Quantizer for Gaussian Source", *Proceedings of the Second National Conference Information Theory and Complex Systems*, Niš, 2014. (M63)

г.2. L. Velimirović, Z. Perić, **B. Denić**, "Design and Analysis of the Two-Level Scalar Quantizer with Extended Huffman Coding", *Proceedings of the First National Conference Information Theory and Complex Systems*, Belgrade, 2013. (63)

Биографија аутора

Лични подаци

Датум рођења: 03.09.1986.
Место рођења: Врбештица, Штрпце, Република Србија
Место пребивалишта: Београд
e-mail: bojan.denic@elfak.ni.ac.rs

Образовање

- Бојан Денић је завршио основну школу "*Шарски одред*" у Врбештици као носилац Вукове дипломе и средњу школу ЕТШ "*Јован Цвијић*" у Штрпцу са одличним успехом.
- Основне и мастер академске студије завршио је на Факултету техничких наука са привременим седиштем у Косовској Митровици (смер Електроника и телекомуникације), 2010. (просечна оцена 9.58) и 2012. године (просечна оцена 9.89), респективно.
- Докторске студије уписао је школске 2012/13 године на Електронском факултету у Нишу, смер Телекомуникације. Просечна оцена на докторским студијама је 9.67.

Радно ангажовање

Истраживач-сарадник на катедри за Телекомуникације на Електронском факултету у Нишу. Учествовао је на пројектима:

1. Развој и реализација наредне генерације система, уређаја и софтвера на бази софтверског радија за радио и радарске мреже – МПНТР;
2. Обрада сигнала применом модела фракционог реда (Fractional-order signal processing and applications)–Билатерални пројекат са Словачком;
3. Машинско учење применом метода фракционог реда (Fractional calculus approach to machine learning)−Билатерални пројекат са Словачком;
4. Напредне методе квантизације, компресије и машинског учења у вештачкој интелигенцији (Advanced methods of quantization, compression and learning in artificial intelligence)−Фонд за науку Републике Србије.

Научни резултати

Бојан Денић је објавио 29 научних радова, од тога 17 радова у међународним часописима са импакт фактором, 2 рада у домаћим часописима, 8 радова на конференцијама од међународног значаја и 2 рада на конференцијама од националног значаја. Такође, био је рецезент у међународним часописима са импакт фактором.

ИЗЈАВЕ АУТОРА

ИЗЈАВА О АУТОРСТВУ

Изјављујем да је докторска дисертација, под насловом

Пројектовање квантизера за примену у обради сигнала и неуронским мрежама

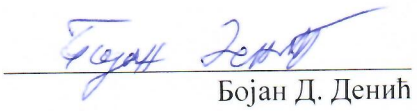
која је одбрањена на Електронском факултету Универзитета у Нишу :

- резултат сопственог истраживачког рада;
- да ову дисертацију, ни у целини, нити у деловима, нисам пријављивао на другим факултетима, нити универзитетима;
- да нисам повредио ауторска права, нити злоупотребио интелектуалну својину других лица.

Дозвољавам да се објаве моји лични подаци, који су у вези са ауторством и добијањем академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада, и то у каталогу Библиотеке, Дигиталном репозиторијуму Универзитета у Нишу, као и у публикацијама Универзитета у Нишу.

У Нишу, _____

Потпис аутора дисертације:



Бојан Д. Денић

ИЗЈАВА О ИСТОВЕТНОСТИ ЕЛЕКТРОНСКОГ И ШТАМПАНОГ ОБЛИКА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Наслов дисертације:

Пројектовање квантизера за примену у обради сигнала и неуронским мрежама

Изјављујем да је електронски облик моје докторске дисертације, коју сам предао за уношење у **Дигитални репозитиријум Универзитета у Нишу**, истоветан штампаном облику.

У Нишу, _____

Потпис аутора дисертације:

Бојан Денић

ИЗЈАВА О КОРИШЋЕЊУ

Овлашћујем Универзитетску библиотеку „Никола Тесла" да у Дигитални репозиторијум Универзитета у Нишу унесе моју докторску дисертацију, под насловом:

Пројектовање квантизера за примену у обради сигнала и неуронским мрежама

Дисертацију са свим прилозима предао сам у електронском облику, погодном за трајно архивирање.

Моју докторску дисертацију, унету у Дигитални репозиторијум Универзитета у Нишу, могу користити сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons), за коју сам се одлучио.

1. Ауторство **(CC BY)**

2. Аутроство – некомерцијално **(CC BY-NC)**

3. Ауторство – некомерцијално – без прераде (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима **(CC BY-NC-SA)**

5. Ауторство – без прераде **(CC BY-ND)**

6. Ауторство – делити под истим условима **(CC BY-SA)**

У Нишу, _____

Потпис аутора дисертације:

Бојан Д. Денић