Univerza v Mariboru
Fakulteta za elektrotehniko, računalništvo in informatiko

DOKTORSKA DISERTACIJA

VREDNOTENJE KAKOVOSTI VEČMODALNIH STORITEV V SODOBNIH TELEKOMUNIKACIJSKIH SISTEMIH

Maribor, november 2014 Tomaž Lovrenčič

Univerza v Mariboru
Fakulteta za elektrotehniko, računalništvo in informatiko

DOKTORSKA DISERTACIJA

VREDNOTENJE KAKOVOSTI VEČMODALNIH STORITEV V SODOBNIH TELEKOMUNIKACIJSKIH SISTEMIH

Maribor, november 2014

Avtor: Tomaž Lovrenčič
Mentor: izr. prof. dr. Andrej Žgank

Avtor:	Tomaž Lovrenčič, univ. dipl. inž. telekomunikacij
Naslov disertacije:	Vrednotenje kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih
Naslov v angleščini:	Quality assessment of multimodal services in contemporary telecommunication systems
UDK:	621.391:[004.932:004.934](043.3)
Ključne besede:	kakovost storitev, večmodalne vsebine, kakovost videa, kakovost govora, procesiranje in analiza slik, analiza avdia, klasifikacija
Število strani:	227
Število izvodov:	10
Ilustriranje:	Vlasta Lovrenčič
Obdelava besedila in slik:	Tomaž Lovrenčič
Lektoriranje (slovenščina):	doc. dr. Darinka Verdonik, prof. slovenščine
Lektoriranje (angleščina):	mag. phil. Boštjan Šauperl
Kraj in datum:	Maribor, november 2014

Univerza v Mariboru

Slomškov trg 15
2000 Maribor, Slovenija

Maribor, 19. 6. 2013
Številka: DR 59/2013/428-DM

Na osnovi 287., 140., 142. in 144. člena Statuta Univerze v Mariboru (Statut UM-UPB10, Ur. l. RS, št. 46/2012) ter sklepa 22. redne seje Senata Univerze v Mariboru, ki je potekala 18. 6. 2013 v zvezi z vlogo doktorskega kandidata Tomaža Lovrenčiča za sprejem odločitve o predlagani temi doktorske disertacije in mentorja

izdajam naslednji

SKLEP

Odobri se tema doktorske disertacije Tomaža Lovrenčiča s Fakultete za elektrotehniko, računalništvo in informatiko z naslovom »Vrednotenje kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih«. Za mentorja se imenuje izr. prof. dr. Andrej Žgank. Kandidat mora članici predložiti izdelano doktorsko disertacijo v zadostnih izvodih najpozneje do 17. 6. 2017.

Obrazložitev:

Kandidat Tomaž Lovrenčič je 14. 1. 2013 na Fakulteti za elektrotehniko, računalništvo in informatiko vložil vlogo za potrditev teme doktorske disertacije z naslovom »Vrednotenje kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih«. Za mentorja je bil predlagan izr. prof. dr. Andrej Žgank.

Senat Fakultete za elektrotehniko, računalništvo in informatiko je na osnovi pozitivnega mnenja komisije za oceno teme doktorske disertacije, ki je ugotovila, da kandidat izpolnjuje pogoje za pridobitev doktorata znanosti, in ocenila, da je predlagana tema ustrezna, sprejel pozitivno mnenje in poslal predlog teme doktorske disertacije s predlogom mentorja v odobritev Senatu univerze.

Senat Univerze v Mariboru je po proučitvi vloge in na osnovi določil Statuta Univerze v Mariboru sprejel svojo odločitev o predlagani temi doktorske disertacije in imenoval mentorja, kot izhaja iz izreka.

V skladu s 144. členom Statuta Univerze v Mariboru mora kandidat za pridobitev doktorata znanosti najpozneje v štirih letih od dneva izdaje tega sklepa, članici predložiti izdelano doktorsko disertacijo v zadostnih izvodih. Kandidatu je bil določen rok za oddajo izdelane doktorske disertacije glede na datum sprejetja teme na pristojnem organu.

Pouk o pravnem sredstvu:
Zoper ta sklep je možna pritožba na Senat Univerze v Mariboru v roku 8 dni od prejema tega sklepa.

Rektor:
Prof. dr. Danijel Rebolj

Obvestiti:
1. Kandidata.
2. Fakulteto.
3. Arhiv.

UNIVERZA V MARIBORU

Zahvala

Iskreno se zahvaljujem svojemu mentorju izr. prof. dr. Andreju Žganku za vso pomoč, strokovno usmerjanje in nasvete v času podiplomskega študija.

Zahvaljujem se tudi kolektivu raziskovalne skupine Mobitel za njihove nasvete in konstruktivne debate, ki so mi pomagale pri premagovanju raziskovalnih problemov.

Na tem mestu se zahvaljujem podjetjema Mobitel d.d. in Telekom Slovenije d.d. ter Tehnološki agenciji Slovenije, ki so v okviru razpisa »Mladi raziskovalci iz gospodarstva – generacija 2009« sofinancirali doktorski študij in raziskovalne aktivnosti, ki so omogočile nastanek te doktorske disertacije.

Zahvaljujem se vsem sodelavcem Laboratorija za digitalno procesiranje signalov za prijetno delo v času študija in nastajanja doktorske naloge.

Posebej bi se rad zahvalil svojim najbližjim, predvsem staršema, ki sta me podpirala in mi omogočila brezskrbno doseganje zastavljenih ciljev.

Posebej bi se rad zahvalil tudi Melisi za razumevanje in podporo v najtežjih trenutkih.

Zahvala gre tudi prijateljem, ki so me nesebično podpirali in mi stali ob strani.

UDK: 621.391:[004.932:004.934](043.3)

Ključne besede: kakovost storitev, večmodalne vsebine, kakovost videa, kakovost govora, procesiranje in analiza slik, analiza avdia, klasifikacija

Povzetek:

V doktorski disertaciji obravnavamo problematiko vrednotenja kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih. Pri tem smo izpostavili degradacije, ki vplivajo na uporabniško kakovost in jih glede na izvor razdelimo v izvorne in omrežne. Njihov vpliv lahko izmerimo s subjektivnimi ali z objektivnimi metodami. Ker so večmodalne storitve lahko obojesmerni sistemi, je potreben nadzor degradacij na vhodnih in izhodnih modalnostih sistema. Pri tem prihaja do medmodalnega učinka kot posledice karakteristik človeške zaznave. Osredotočenost uporabnika na polja interesa (ROI) daje degradacijam v teh območjih večji vpliv, kar lahko izkoristimo za porazdeljeno vrednotenje. Cilj disertacije je predlagati model za vrednotenje kakovosti večmodalnih storitev in izdelati vzorčen koncept evalvatorja, ki bo upošteval omenjena dejstva.

Za dosego cilja smo nalogo razdelili na tri področja: v prvem smo določili vpliv degradacij na vhodno modalnost, v drugem smo zgradili primerno večmodalno bazo HD-posnetkov in naredili subjektivno in objektivno vrednotenje izhodne modalnosti, v tretjem pa predlagali nov model večmodalnega porazdeljenega vrednotenja kakovosti. Pri vrednotenju kakovosti vhodne modalnosti sistema smo analizirali storitev IVR s funkcijo razpoznavanja govora, kjer smo na podlagi meritev povprečne objektivne ocene kakovosti (objMOS) iz govorne baze SpeechDat(II) ovrednotili vpliv degradacije transkodiranja in izgube paketov (PL). Govorni kodeki so pri tem pokazali precejšnja odstopanja, tudi med različnimi konfiguracijami istih govornih kodekov. Govorna izguba je degradirala signal do te mere, da je bila potrebna uporaba robustnejše modalnosti v obliki DTMF-izbiranja. Na podlagi analize smo predlagali klasifikator vhodne modalnosti na osnovi Gaussovih modelov (GMM). V učni fazi smo analizirali različne konfiguracije klasifikatorja. Testna faza je pokazala uspešno delovanje klasifikatorja za izbiro vhodne modalnosti v različnih scenarijih izgube paketov.

Pri raziskavi vpliva degradacij na izhodno modalnost smo izdelali večmodalno bazo posnetkov s štirimi vrstami vsebine. Baza je vsebovala posnetke z

avdiom (A, kodek AAC, 48kbps), videom (V, kodek H.264/AVC, 1920x1080) in avdio-videom (AV) pri različnih scenarijih izgube paketov. Izvedli smo subjektivno testiranje z 20 osebami na 240 posnetkih, pri katerih smo dobili povprečne subjektivne ocene kakovosti (subMOS), kar je služilo za referenco objektivnemu vrednotenju. Objektivno vrednotenje je potekalo s standardom PESQ, pri video modalnosti pa smo iz nabora 26 slikovnih metrik izbrali tisto z najboljšo korelacijo s subjektivno oceno: slikovno metriko NQM. Na podlagi rezultatov smo predlagali model vrednotenja kakovosti večmodalne storitve, ki je upošteval tip modalnosti, tip scene, količino degradacij in enomodalne ocene objMOS. Korelacija na testnem naboru je bila 0,892.

Pri analizi osredotočenosti uporabnika storitve na ROI in možnosti porazdeljenega vrednotenja smo uporabili detektor vizualne razpoznave strukture obraza, ki temelji na algoritmu Viola-Jones s kaskadnimi klasifikatorji s šibkimi Haarovim podobnimi značilkami, ki smo ga ustrezno modificirali, da smo dosegli čim boljšo detekcijo obraza. Z analizo smo določili pristop porazdeljenega vrednotenja vizualne informacije z enostavnim vrednotenjem ozadja (ne-ROI) z metriko PSNR in kompleksnejšim vrednotenjem obraza (ROI) z metriko NQM. Pomembnost porazdeljenega vrednotenja kakovosti storitev smo potrdili s subjektivnimi testi.

UDK: 621.391:[004.932:004.934](043.3)

Keywords: quality of service, multimodal content, video quality, speech quality, image processing and analysis, audio analysis, classification

QUALITY ASSESSMENT OF MULTIMODAL SERVICES IN CONTEMPORARY TELECOMMUNICATION SYSTEMS

Abstract:

This thesis focuses on quality assessment of multimodal services in contemporary telecommunication systems. It addresses quality degradations which affect user experience. Depending on their origin, they can be categorized as source or network impairments. Their impact can be measured with subjective or objective methods. Since multimodal services can be bi-directional systems, it is necessary to have control over input and output modalities of the system. This leads to intermodal influences between the modalities as a consequence of human perception. Furthermore, the users' focus on Regions-of-Interest (ROI) gives degradations in those particular regions greater impact on the overall quality, which we can use for differentiated quality assessment. The aim of this thesis is to propose a model for quality assessment of multimodal services and develop the concept of the quality evaluator, which takes the above mentioned facts into account. Therefore, the thesis is divided into three sections. In the first section, the impact of quality degradations on the input modality is determined. In the second, a suitable multimodal database comprising HD recordings is established. This section also presents subjective and objective assessment of output modality, where subjective mean opinion score (subMOS) and objective mean opinion score (objMOS) were conducted. Based on the results, a new model of multimodal quality assessment is proposed. The last section addresses differential quality evaluation based on ROI.

As part of the evaluation of the effect of quality degradations on the input modality, a voice-driven IVR service with a built-in speech recognition module (ASR) is analyzed. Assessment begins by measuring objMOS values of the samples from the SpeechDat(II) database. Samples were degraded by transcoding and packet

loss. There were substantial differences between the speech codecs used, even when the exact same codec was used with different configurations. Generally, deterioration was greater for codecs with lower bandwidth. The voice signal degraded to such an extent that it was necessary to use a more robust modality, i.e. DTMF dialing. After an analysis of the results, a classifier of input modality based on the Gaussian Mixture Models (GMM) was proposed. When training the classifier, different classification parameters were conducted. Test phase confirmed the successful operation of the classifier regarding the input modality with various packet loss scenarios.

For the purpose of assessing the impact of degradations on the quality of output modality, a specifically designed multimodal database was established. It comprised audio (AAC at 48 kbps), video (H.264/AVC at a resolution of 1920x1080 pixels) and combined audio and video clips for a total of 240 samples, used in various packet loss scenarios. After that, subjective tests with 20 subjects were conducted, which gave reference data for objective quality assessment. Objective quality was measured separately for audio and video modalities. To assess the audio modality, standardized PESQ speech quality metric was used, and to assess the video modality NQM video metric was applied. Then, using the regression method, a linear model for evaluating the quality of multimodal services was proposed, which takes into account the type of modality, type of scene, amount of degradation and unimodal objMOS scores. Correlation yields 0.892.

The differential quality evaluation consists of two stages. First, a ROI face detector was used, based on the Viola-Jones object detection algorithm with weak Haar-like feature-based cascade classifiers. Then, using good detection results, an analysis of the optimization possibilities due to differential quality assessment of visual modality is presented. This investigation proposed evaluating the quality of ROI regions with a more complex algorithm (NQM) since those regions have higher visual attention, and using a simpler quality metric (PSNR) for the background, i.e. non-ROI regions. The importance of differential quality assessment was confirmed with subjective image quality evaluation.

Izvirni znanstveni prispevki

V disertaciji predlagamo naslednje izvirne znanstvene prispevke:

1. **Model za izbiro vnosne modalnosti.** Model sestavlja klasifikator izbire vhodne modalnosti (govor/DTMF) glede na karakteristike govornih podatkov (transkodiranje, omrežne degradacije). Tak pristop je aktualen v sodobnih TK-storitvah, kjer je kakovost uporabniškega vmesnika še posebej pomembna in so uporabniške zahteve stroge. Raziskava vključuje komercialno primerljiv razpoznavalnik govora kot eno izmed možnosti pri izbiri vnosne modalnosti pri storitvi IVR.
2. **Zasnova s podporo sodobnim TK-storitvam.** Zasnova sistema za delovanje v omrežjih naslednje generacije z vsebinami visoke kakovosti se od tradicionalnih pristopov loči predvsem v novih uporabniških zahtevah, npr. ločljivosti vsebin (HD), uporabljenih prenosnih parametrih (heterogena omrežja) in novih funkcijah. Spremembe pogojev, tj. večanje uporabniških zahtev in posledično tudi prilagoditev objektivnih meritev kakovosti TK-storitev, pomenijo tudi potrebo po analizi v spremenjenih razmerah ter oceni korelacije subjektivnih in objektivnih postopkov.
3. **Večmodalna evalvacija in kombinacija različnih metrik kakovosti za posamezno modalnost.** Trenutni sistemi za določanje kakovosti storitev po večini temeljijo na enomodalnih pristopih. Vendar nam kognitivna spoznanja o HVS/HAS kažejo, da prihaja do medmodalnih učinkov, ki jih je potrebno pri končni subjektivni in posledično objektivni evalvaciji upoštevati. Z integracijo različnih enomodalnih metrik kakovosti, kjer vsaka deluje najboljše na svojem področju delovanja, lahko z ustrezno kombinacijo in upoštevanjem medmodalnega vpliva dosežemo optimalnejšo delovanje objektivnega evalvatorja. Dodatno delovanje s HD-vsebinami odpira nove raziskovalne pristope.
4. **Porazdeljeno vrednotenje kakovosti storitev z uporabo ROI.** Detektor ROI je model za prepoznavo vzorcev slike z višjo pomembnostjo, ki za razliko od objektivnih metrik, delujočih na nivoju slikovnih pik ali nižjih kognitivnih plasteh možganskega korteksa, deluje na nivoju detekcije struktur, ki so prepoznane na nivoju osmisljenja, npr. detekcija obraza napovedovalca. Takšen

pristop je primeren tudi pri porazdeljenem vrednotenju, saj se uporabniki bolj osredotočajo na ROI in je potrebno močneje obtežiti morebitne napake v teh poljih.

5. **Zasnova in način izgradnje večmodalne baze.** V znanstveni literaturi je opaziti precejšnje število referenčnih baz, ki raziskovalcem omogočajo učenje, testiranje in verifikacijo njihovih raziskav. Pri tem je vidno pomanjkanje večmodalnih baz, če posebej takšnih, ki bi bile primerne za vsebine visoke ločljivosti. Zasnova takšne referenčne baze zahteva izbiro primernih testnih scenarijev, upoštevanje standardov pri postavitvi in izvedbi subjektivnih testov, izbiro reprezentativnih testnih uporabnikov, primerno izbrane testne podatke in pravilno izbiro statističnih metod za obdelavo. Referenčna baza vsebuje posnetke s slovensko govorno vsebino, kar poveča uporabnost takšne baze za slovensko raziskovalno sfero.

Kazalo

Kazalo slik

Kazalo tabel

Kazalo algoritmov

Seznam uporabljenih kratic

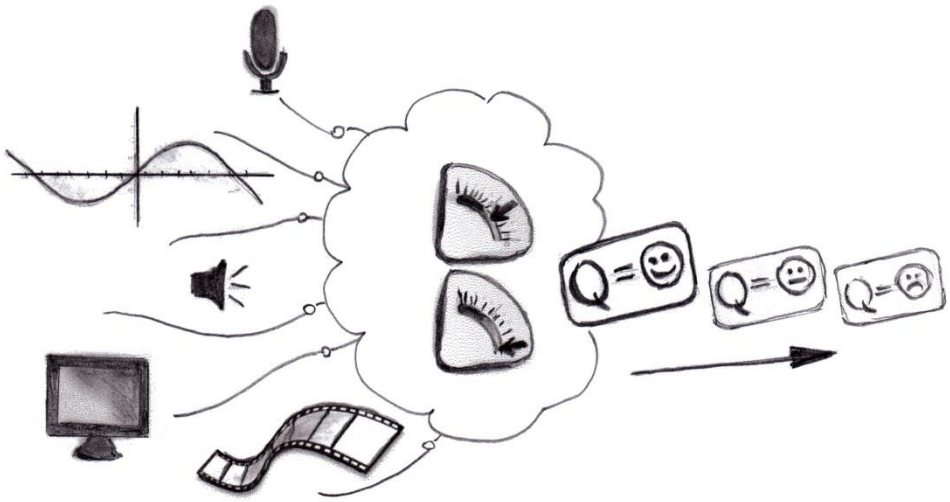
AAC	advanced audio coding	napredno avdio kodiranje
ABR	average bitrate	povprečna bitna hitrost
ACELP	algebraic code-excited linear predictive coding	algebraično kodno vzbujano linearno napovedno kodiranje
AMR-NB	adaptive multi-rate narrow band	adaptivni večhitrostni kodek za ozek frekvenčni pas
AMR-WB	adaptive multi-rate wide band	adaptivni večhitrostni kodek za širok frekvenčni pas
ASR	automatic speech recognizer	avtomatski razpoznavalnik govora
AVR	average bitrate	povprečna bitna hitrost
BSD	bark spectral distortion	spektralno popačenje Barka
CAVLC	context-adaptive variable-length coding	kontekstno-adaptivno dolžinsko-variabilno kodiranje
CBR	constant bitrate	konstantna bitna hitrost
CELP	code excited linear prediction	kodno vzbujana linearna predikcija
CFF	critical flicker frequency	kritična frekvenca utripanja
CIF	common interchange format	splošni izmenjalni format
CMYK	cyan, magenta, yellow, black	cijan, magenta, rumena, črna
CNG	comfort noise generator	generator komfortnega šuma
CSF	contrast sensitivity function	funkcija senzitivnosti kontrasta
CW-SSIM	Complex-wavelet structural similarity index	kompleksna valjčna funkcija strukturne podobnosti
DCT	discrete cosine transform	diskretna kosinusna transformacija
DCR	degradation category rating	kategorijska lestvica degradacije
DFT	discrete Fourier transformation	Diskretna Fourierjeva transformacija
DMOS	differencial mean opinion score	diferencialna povprečna ocena
DMOS	degradation mean opinion score	povprečna ocena degradacije
DS	double stimulus method	metoda z dvojnim dražljajem
DSIS	double stimulus impairment scale	degradacijska lestvica z dvojnim dražljajem
DTMF	Dual tone multi-frequency	dvotonska večfrekvenčna

		signalizacija
DWT	discrete wavelet transform	diskretna valjčna transformacija
ETSI	European telecommunications standards institute	Evropski inštitut za telekomunikacijske Standarde
FFT	flicker fusion threshold	prag fuzije utripanja
FFT	fast Fourier transformation	hitra Fourierjeva transformacija
FLAC	free lossless audio codec	prosti brezizgubni avdio kodek
FR	full reference (quality metric)	metrika kakovosti s polno referenco
FRExt	fidelity range extensions	razširitve razpona verodostojnosti reprodukcije
GAST	gradient ascent subjective quality testing	gradientno dvigajoče subjektivno testiranje
GOP	group of pictures	skupina slik
HAVS	human audio-visual system	človekov avdiovizualni sistem
HAS	human auditory system	človekov avditorni sistem
HCI	human-computer interaction	interakcija človek – računalnik
HILN	harmonic and individual lines noise	šum harmoničnih in posamičnih linij
HSV	hue, saturation, value	odtenek, nasičenost, svetlost
HVS	human visual system	človekov vizualni sistem
HVXC	harmonic vector excitation coding	kodiranje z vzbujanim harmoničnim vektorjem
IEC	International electrotechnical commission	Mednarodna elektrotehniška komisija
IKT	information and communication technology	informacijska in komunikacijska tehnologija
IQA	image quality assessment	vrednotenje kakovosti slik
ISO	International organization for standardization	Mednarodna organizacija za standardizacijo
ITU	International telecommunication union	Mednarodna telekomunikacijska zveza
IVR	Interactive Voice Response	interaktivni govorni odzivnik
JND	just noticeable difference	komaj zaznavna razlika

KRCC	Kendall rank correlation coefficient	Kendallov koeficient korelacije
LMSE	Laplacian mean square errror	Laplaceova meritev kvadrata povprečne napake
MAE	mean absolute error	povprečna absolutna napaka
MD	maximum difference	maksimalna razlika
MDCT	modified discrete cosine transform	modificirana diskretna kosinusna transformacija
MNB	measuring normalizing blocks	meritev normalizacijskih blokov
MOS	mean opinion score	povprečna ocena kakovosti
MPEG	Moving picture experts group	Ekspertna skupina za gibljive slike
MSE	mean squared error	srednja kvadratna napaka
MS-SSIM	multi scale SSIM	meritev strukturne podobnosti na več obsegih
NQM	noise quality measure	meritev kakovosti šuma
NR	no reference (quality metric)	metrika kakovosti brez reference
objMOS	objective MOS	objektivna MOS
objMOSa	objective MOS of audio modality	objektivna MOS avdio modalnosti
objMOSa_av	objective MOS of audio modality for audio-video content	objektivna MOS avdio modalnosti za avdiovideo vsebine
objMOSav	objective MOS of audio-video modality	objektivna MOS avdiovideo modalnosti
objMOSv	objective MOS of video modality	objektivna MOS video modalnosti
objMOSv_av	objective MOS of video modality for audio-video content	objektivna MOS video modalnosti za avdiovideo vsebine
PAMS	perceptual analysis measurement system	merilni sistem za zaznavno analizo
PCM	pulse code modulation	pulznokodna modulacija
PEAQ	perceptual evaluation of audio quality	zaznavna evalvacija kakovosti avdia
PESQ	perceptual evaluation of speech quality	zaznavna evalvacija kakovosti govora
PLCC	Pearson linear correlation coefficient	Pearsonov linearni koeficient korelacije

PSNR	peak signal to noise ratio	razmerje vršnih vrednosti signal-šum
PSQM	Perceptual speech quality measure	meritev zaznavne kakovosti govora
PVD	prefered viewing distance	zaželena oddaljenost opazovanja
QCIF	quarter common intermediate format	četrtina splošnega izmenjalnega formata
QoE	quality of experience	kakovost izkušnje
QoS	quality of service	kakovost storitve
RASTA	relative spectra filtering of log domain coefficients	relativno spektralno filtriranje koeficientov v log domeni
RFSIM	Riesz-transform based feature similarity	kakovostna metrika podobnosti značilk Rieszove transformacije
RGB	red, green, blue	rdeča, zelena, modra
ROI	region of interest	polje (vizualnega) interesa
RR	reduced reference (quality metric)	metrika kakovosti z delno referenco
SAMVIQ	subjective assessment method for video quality evaluation	subjektivna ocenjevalna metoda za evalvacijo kakovosti videa
SC	structural content	strukturna vsebina
SD TV	standard definition TV	TV standardne ločljivosti
SI	spatial information	prostorska informacija
SLS	scalable lossless (audio coding)	stopenjsko brezizgubno (avdio kodiranje)
SNR	signal to noise ratio	razmerje signal – šum
SRCC	Spearman’s rank correlation coefficient	Spearmanov koeficient korelacije
SS	single stimulus (method)	metoda z enojnim dražljajem
SS-SSIM	single scale structural similarity index	meritev strukturne podobnosti na enem obsegu
SSIM	structural similarity index	meritev strukturne podobnosti
subMOS	subjective MOS	subjektivna MOS
subMOSa	subjective MOS for audio modality	subjektivna MOS avdio modalnosti
subMOSav	subjective MOS for audio-video modality	subjektivna MOS avdiovideo modalnosti
subMOSv	subjective MOS for video modality	subjektivna MOS video modalnosti

UE	user equipment	uporabniška oprema
UX	user experience	uporabniška izkušnja
VCEG	Video coding experts group	Skupina za razvoj standardov za kodiranje videa
VoIP	voice over IP	govor preko IP
W-CWSSIM	weighted complex wavelet SSIM	utežena CW-SSIM
WSNR	weighted signal to noise ratio	obtežena SNR



1. Uvod

Vrednotenje kakovosti storitve ima pomembno vlogo pri zagotavljanju ustrezne kakovosti uporabniške izkušnje. Z večanjem števila večmodalnih vsebin in dodatnih funkcij so tudi zahteve končnih uporabnikov postale višje. Ponudniki storitev tako na eni strani poskušajo zadostiti uporabniškim zahtevam, po drugi strani pa so omejeni iz tehnoloških in ekonomskih vidikov. Ker večmodalne storitve zahtevajo velike prenosne zmogljivosti sodobnih heterogenih omrežij, prihaja do degradacij tako na izvorni strani distributerja večmodalnih vsebin kot tudi v samem prenosnem omrežju, kar odločilno vpliva na nivo kakovosti. Če ta ni zagotovljen, pomeni to izgubo naročnikov in je pomemben tehnološki pokazatelj ustreznosti in sprejemljivosti storitve za končnega uporabnika. Zato je nujen nadzor kakovosti, predvsem iz stališča obravnave zaznavne kakovosti in uporabniške izkušnje. Za to obstajajo modeli subjektivnega in objektivnega vrednotenja. Lastnost prvih je zelo natančna ocena kakovosti, a so časovno in stroškovno potratni, lastnost drugih pa zmožnost sprotnega spremljanja v realnem času, vendar je njihova uspešnost odvisna od upoštevanja odziva sistemskih parametrov pri zasnovi objektivnega evalvatorja.

1.1. Opis raziskovalnega problema

Potreba po primerjavi in ovrednotenju *kakovosti storitve* (angl. Quality of Service – QoS) je že v preteklosti narekovala klasifikacijo parametrov kakovosti. Na primer v analogni telefoniji so uporabniki zahtevali *čas razpoložljivosti storitve*, *število pravilno zvezanih linij*, *čas dostopa* in *usmerjanja*, *kontinuiteto povezave* ter *razmerje signal – šum* [1]. Z razvojem sodobnih storitev so telekomunikacijski trg preplavile nove uporabniške storitve, ki ne samo dopolnjujejo tradicionalne, temveč nudijo tudi dodano vrednost z namenom prijaznejše uporabe, kot je večmodalna izkušnja, enostavnost uporabe, visoka kakovost vsebin, stalna povezljivost itd. Primer takšne napredne TK-storitve je *interaktivni govorni odzivnik* (angl. Interactive Voice Response, IVR). Pri IVR je kakovost uporabniškega vmesnika definirana z zanesljivostjo upravljanja in nadzora nad storitvijo ter odzivom sistema na uporabniške zahteve. Vhodna modalnost sistema je lahko izbrana v obliki vnosa preko številčnice ali govornega vnosa, s tem da je slednji za uporabnika primernejši, saj predstavlja najenostavnejši način komunikacije [2], [3]. Govorni vnos poteka preko modula *avtomatskega razpoznavalnika govora* (angl. Automatic Speech Recognizer, ASR), katerega naloga je pretvorba govora v stroju razumljivo obliko, primerno za nadaljnje procesiranje. Četudi je uporaba ASR zaželena, je tehnična zasnova pogosto zahtevna naloga [4] zaradi velike kompleksnosti analize in obdelave govora ter zaradi različne izgovorjave (naglasi), spreminjanja tonsko-spektralnih lastnosti, glasnosti, hitrosti govora in drugo [5], [6]. Uspešnost ASR je omejena z naborom dovoljenih besed, tj. števila možnih izhodnih stanj, naborom uporabljenih značilk, nivojem prisotnega šuma in degradacijami [7]. S povečevanjem števila omenjenih parametrov se povečuje tudi kompleksnost razpoznavanja in posledično občutljivost sistema, kar pomeni manjšo robustnost ASR na morebitne degradacije v govornem signalu [8]. Degradacije glede na izvor delimo na izvorne in tiste, ki se zgodijo na prenosni poti (omrežne).

Primer prvega je transkodiranje, katerega namen je prilagoditev uporabniških podatkov na prenosni sistem (interoperabilnost) ali napravo, po večini z uporabo kompresijskih metod [9]. Vpliv poslabšanja kakovosti je še posebej velik pri večkratnem transkodiranju, npr. v primeru prehoda uporabniških podatkov čez heterogena omrežja z omejeno pasovno širino (G.729, AMR).

Na drugi strani pa omrežna okvara predstavlja kvazistohastičen pojav, ki je težko nadzorljiv. Posledica tega so izguba, latenca, trepetanje in podvajanje paketov,

omejevanje pasovne širine in bitne napake, do katerih prihaja zaradi implementiranih brezpovezavnih protokolov (UDP/RTP), kjer se prioritizira pretočnost podatkov in delovanje v realnem času. Tako iz perspektive uporabnika kot DSP-modulov (ASR) je najbolj destruktivna izguba paketov, ki je posredno prav tako lahko posledica drugih okvar, npr. latence, ki postane izguba paketov, ko je pomnilnik proti trepetanju (angl. dejitter buffer) premajhen ali prezaseden [10], [11]. Takšne napake imajo pomemben vpliv tudi na delovanje ASR [12].

Vpliv omrežnih degradacij je lahko tako močan, da izgubljeni in zavrženi RTP-paketi degradirajo govorni signal do te meje, da uspešnost razpoznavalnika pade pod definiran prag kakovosti in je potrebna uporaba metode, ki izkazuje bolj robustno delovanje: DTMF-izbiranje. S tem namenom predlagamo vključitev odločitvenega modula kakovosti vnosa v IVR, kar določa izbiro primerne modalnosti z namenom zagotavljanja boljše uporabniške izkušnje [13]. Zato je potrebna analiza, v kateri se s pomočjo objektivnih metrik kakovosti govora izmerjene vrednosti kakovosti govora preslikajo v objektivne ocene kakovosti govora [14], [15], [16], [17], [18], [19], [20]. Metoda je kompleksna in delno odvisna tudi od uporabljenega ASR in njegovih lastnosti, uporabljenega jezika, modelov, stopnje in tipa degradacij, distribucije in frekvence napak, uporabljenega kompresijskega algoritma, stopnje kompresije itd. Uporabe takšnega pristopa v IVR-sistemu v obstoječi literaturi ni zaslediti.

S ciljem obogatitve uporabniške izkušnje dostavljene vsebine v sodobnih omrežjih vključujejo tudi večmodalnosti. Zaradi storitvenih zmogljivosti se vključujejo predvsem vizualne komunikacije visoke ločljivosti (HD), ki pa zahtevajo zmogljivejši prenosni sistem. Fuzija modalnosti v multimedijske tokove, npr. video na zahtevo (angl. Video on Demand, VoD), IPTV, spletne vsebine ter P2P (angl. Peer-to-Peer), zahteva sprotni nadzor kakovosti, ki ga lahko izvajamo z objektivnimi metrikami kakovosti [21]. Ocena objektivne kakovosti mora zato čim bolje korelirati s subjektivno oceno (MOS, angl. Mean Opinion Score), ki bi jo podal uporabnik, da sistem primerno ukrepa samo v primeru, kadar se uporabniško zaznavna kakovost zmanjša [22]. Glede na prisotnost referenčnega signala zato ločimo 3 kategorije metrik kakovosti: a) takšne *brez reference* (angl. No Reference – NR), b) z *delno referenco* (angl. Reduced Reference – RR) in c) s *polno referenco* (angl. Full Reference – FR). Slednje štejemo med najbolj natančne in robustne, saj imajo dostop do izvornih podatkov in so posledično tudi bolj primerne za generične meritve vseh vrst degradacij in distorzij na področju videa [23]. Večmodalna FR-metrika je v splošnem sestavljena iz napovedi posamičnih ocen kakovosti

modalnosti ter združevanja teh v končno objektivno oceno kakovosti. Objektivno vrednotenje videa v splošnem poteka z določanjem kakovosti prostorskih delov (po navadi video okvirjev) s slikovnimi metrikami, ki se nato združijo v skupno objektivno oceno (prostorsko-časovni pristop). Med slikovnimi metrikami so najpogostejše in najpreprostejše tiste, ki delujejo na nivoju slikovnih pik. Primer takšnih slikovnih metrik so *razmerje signal – šum* (angl. Signal to Noise Ratio, SNR), *PSNR* (angl. Peak SNR), *MSE* (angl. Mean Square Error) in *MAE* (angl. Mean Absolute Error). Prednost teh je matematična učinkovitost in preprostost, slabost pa slabša koreliranost s človeškim dojemanjem v določenih primerih uporabe [24].

Percepcijsko bližje so metrike, ki izkoriščajo znanje o človeškem vizualnem sistemu (angl. Human Visual System, HVS) [25], kot je strukturna podobnost [26], pasovna/časovna/barvna dekompozicija slike [27], [28], [29], efekt prostorskega in časovnega maskiranja [30], [31], [32], [33], kontrastna senzitivna funkcija (angl. Contrast Sensitivity Function, CSF) [34] in adaptacija na svetilnost [35]. Njihova prednost je boljša metodologija v evalvaciji in obdelavi vizualne informacije, slabost pa potreba po dobrem razumevanju HVS, tj. fizioloških in psiholoških spoznanj [36]. Model percepcije človeka na vizualne dražljaje se lahko simulira po pravilu od zgoraj navzdol [37] ali od spodaj navzgor [38]. Prvi mehanizem obravnava vizualno pozornost (angl. visual attention) iz biološkega stališča obdelave vizualne informacije, tj. kognitivnega procesiranja v možganih, ko je informacija zbrana na podlagi cilja, namena in prepričanj opazovalca [39], [40], [41]. Pristop od spodaj navzgor pa deluje na podlagi psihofizičnih karakteristik modalnih kanalov, na katere vplivajo zunanji, fizični dražljaji. Primer tega je združevanje posameznih atributov kakovosti (ostrina, svetlost, barva) v bolj splošne, združene parametre, dokler ne pridemo do zadnjega nivoja združitve.

Podobno pri evalvaciji zvočnega dela multimedijskih vsebin v scenarijih s polno referenco ločujemo metrike kakovosti glede na upoštevanje zgradbe človeškega slušnega sistema (angl. Human Auditory System, HAS) in kognitivnih procesov ter tiste, ki delujejo na enostavnejših matematičnih modelih, npr. z določanjem kakovosti glede na količino šuma. V preteklosti so različni avtorji predstavili evalvacijske algoritme, temelječe na zaznavnih modelih za ocenjevanje nelinearnih in degradiranih zvočnih komunikacijskih sistemov. Potencialno zanimive, z dobro uspešnostjo na velikem naboru raznolikih scenarijev s parametri, kot so tip degradacije, vsebine in omrežne lastnosti, so sprejete v priporočilih standardizacijskih teles. Beerends in

Stemerdinkov model PSQM (angl. Perceptual Speech Quality Measure) je bil sprejet kot ITU-T-priporočilo P.861 [42] in kasneje nadgrajen v PSQM99. Kasneje se priporočilu priključi tudi alternativni sistem, temelječ na meritvi normalizacijskih blokov (angl. Measuring Normalizing Blocks, MNB) [43]. MNB zagotavlja dobro korelacijo ocene pri omrežnih degradacijah, kot so bitne napake in izguba zvočnih okvirjev.

Drug model, imenovan PAQM (angl. Perceptual Audio Quality Measure), je bil združen z nekaterimi drugimi avdio modeli v metodo PEAQ (angl. Perceptual Evaluation of Audio Quality) in sprejet v standardu BS.1387. PEAQ je primeren za generične zvočne posnetke visoke kakovosti (frekvenca vzorčenja 44,1-48 kHz). Izdelan je bil z namenom evalvacije kodirnih algoritmov, zato ni primeren za ocenjevanje večjih degradacij, kot so izgube IP-paketov, ki nosijo zvočni signal.

Iz nadgradnje Holierjevega modela BSD (angl. Bark Spectral Distortion) [44] se je razvil PAMS (angl. Perceptual Analysis Measurement System), ki prvi upošteva tudi vpliv filtriranja (degradacija izvora) in variabilne latence od točke do točke (degradacija na prenosni poti) [45]. PAMS parametrizira degradacije in jih preslika v objektivno oceno preko linearne kombinacije posameznih kvadratnih funkcij. Ker so posamezni parametri med seboj (vsaj delno) odvisni, je definiranje nabora parametrov ter pripadajoče funkcije preslikav netrivialna naloga.

Vpliv določenih tipov degradacij (izguba paketov, kompresijska degradacija, šum v signalu) pa kljub temu povzroča nenatančne rezultate omenjenih modelov [46]. Objektivna metrika, ki izkazuje najboljše rezultate pri evalvaciji predvsem govornih posnetkov, je tako združek omenjenih algoritmov: standard P.862 ali PESQ (angl. Perceptual Evaluation of Speech Quality) [47]. PESQ in njegova širokopasovna razširitev (PESQ-WB, P.862.2) [48] je pogosto uporabljena metoda za analizo vpliva izgubljenih govornih IP-paketov v fiksnih [49], [50], [51], brezžičnih [52], [53] in mobilnih (3G, 4G) omrežjih [54], [55]. Avtorji v [56] omenjajo korelacijo PESQ s subjektivnimi testi v povprečju 0,942 pri prenosu v fiksnem omrežju, 0,921 pri internetni telefoniji (VoIP) in 0,962 v mobilnem omrežju. Uspešnost algoritma PESQ temelji na zvočni spektralni razdalji (angl. Auditory Spectrum Distance, ASD) [57], ki s pomočjo naučenega kognitivnega modela na veliki množici posnetega materiala in psiho-akustičnih podatkov interpretira diferenčno funkcijo med originalnim in degradiranim signalom. Zasnovan kognitivni model, podobno kot pri vizualni modalnosti, obravnava percepcijo in procesiranje zvočnih signalov v primarnem

avditornem korteksu. Psiho-akustično modeliranje pa določa omejitve človeškega HAS, tj. časovno maskiranje, frekvenčni in amplitudni razpon ter zvočno lokalizacijo.

Večina avtorjev se pri evalvaciji kakovosti osredotoča na uporabo ene metrike kakovosti, ki jo optimirajo za delovanje v določenih pogojih za vizualne [58], [59], [60], [61], [62], zvočne [63], [64], [65], [20], [66] ali multimedijske vsebine [67], [68]. Nekateri uporabljajo združen sistem enomodalnih metrik kakovosti, kar daje boljše rezultate in korelacijo z MOS v raznolikih pogojih in pri širokem spektru uporabe [69], [70], [71]. Zato je pri konceptu večmodalnih metrik kakovosti, z razumevanjem kognitivnih procesov, potrebno obtežiti in združiti signale zaznave na različne načine [72], odvisno od karakteristik multimedije in namena uporabe. Zaradi kompleksnosti tega je v literaturi zaslediti le malo temeljnih raziskav, ki obravnavajo modele vrednotenja, ki združujejo različne metrike kakovosti večmodalnih vsebin HD-kakovosti [73], ki so združene v skupno objektivno oceno kakovosti ter pri tem upoštevajo primernost in odvisnost metrik za dano TK-storitev.

Čeprav so človeški senzorni organi ločeni, pa človeški možgani povežejo dražljaje v za njih smiselne pojme. Vpliv, ki ga ima posamezen dražljaj na HVS/HAS, pa je že v osnovi odvisen od več parametrov, npr. konteksta, v katerem se nahajamo [74], vsebine [75], [76], [77], tipov modalnosti in njihovih medmodalnih učinkov [78], [79] ter delno tudi drugih dejavnikov [80]. Raziskave potrjujejo, da se ocena MOS posamezne modalnosti lahko precej spremeni v prisotnosti oz. odsotnosti druge, s tem pa tudi korelacija z MOS na celotnem večmodalnem toku uporabniških podatkov [81], [82]. Dokazano je tudi, da so testne osebe drugače ocenile nekoherentne tokove v testih avdiovizualne predstavitve govora, in sicer glede na to, ali je bila prisotna asinhronost vizualnega ali govornega dela [83] ter odvisno od količine degradacije v signalu [84]. Pri vplivu degradacij je potrebno upoštevati tudi njihovo pomembnost. Polje interesa vzorcev na sliki in videu imenujemo polje ROI (angl. Region Of Interest). To je pojem, ki predstavlja skupino vizualnih gradnikov, npr. slikovnih pik z iskanimi lastnostmi. Primer uporabe je zakoreninjen že v samih kodirnih standardih, npr. uporablja se z namenom adaptacije kodirne matrike JPEG pri kodiranju slik [85] ali intranapovedi za algoritem odpornosti na napake v videu H.264/AVC [86]. Za detekcijo polj ROI so bile predstavljene različne metode. Nekatere temeljijo na nižjenivojskem modelu senzornega sistema statičnih [87], [88] ali dinamičnih podatkov [89], [90], [91], spet druge na modelu osredotočenosti z uporabo verjetnostnih meritev [92] in opazovanjem dejanskih uporabnikov, npr. z inštrumenti za sledenje oči [93], [94], [95], [96], [97]. Nekateri so

poiskušali tudi z določanjem položaja izvora zvoka/govora s predpostavko o medmodalnih povezavah [98]. Znanje o ROI lahko zato uporabimo za nadgradnjo objektivnih video metrik s »pametnimi funkcijami« [99]. Te funkcije nizkonivojskim kognitivnim procesom pripišejo še dojemanje sveta preko zavestnih miselnih procesov. Dodatno je potrebno upoštevati tudi fizikalne lastnosti, tj. zgradbo očesa. Raziskave namreč potrjujejo, da so lastnosti vizualnih polj ROI od tega močno odvisne, npr. upadanje senzitivnosti očesa na vizualne anomalije z oddaljenostjo od centralne osi vidnega fokusa [100], [101]. Degradacije perifernih območij (ne-ROI polja) zato nosijo manjše posledice na skupno MOS [102]. To dejstvo se lahko uporabi tudi pri evalvaciji kakovosti večmodalnih vsebin in izdelavi naprednih metrik kakovosti [103], [104]. Pri pregledu odvisnosti med zaznavno kakovostjo videa in različnimi atributi izgube paketov (dolžina, vpliv, lokacija, število in vzorec izgubljenih paketov) namreč opazimo precejšnjo variacijo med perifernimi in ROI-območji, zato je uporaba v namene evalvacije kakovosti smiselna [105].

1.2. Cilji in teza doktorske disertacije

Glavni cilj disertacije je predlagati model za vrednotenje kakovosti sodobnih večmodalnih TK-storitev in izdelati vzorčni koncept evalvatorja. Pri tem je potrebno analizirati delovanje vhodne in izhodne modalnosti večmodalnega sistema pod vplivom različnih degradacij, ki obstajajo v heterogenih omrežjih. Na podlagi tipa in karakteristik transkodiranja (izvorna degradacija) in omrežnih napak (prenosna degradacija) govorne modalnosti na naboru testnih scenarijev predlagamo klasifikator, ki je zmožen določanja vhodne modalnosti uporabniškega vmesnika storitve IVR glede na zaznano kakovost uporabniških podatkov.

Izhodna modalnost sistema bo vključevala izdelavo govorno-vizualne baze posnetkov v HD-ločljivosti, kjer se bodo upoštevali temu primerni parametri, ki ločujejo te vsebine od tistih z nižjo kakovostjo. Cilj je izdelava takšne baza posnetkov in njihovo vrednotenje s subjektivnim in objektivnim ocenjevanjem. Predlog pristopa vrednotenja kakovosti storitve za večmodalni primer izhoda bo takšen, da bo zajemal objektivno vrednotenje kakovosti posameznih modalnosti in združevanje objektivnih ocen v skupno oceno kakovosti storitve za multimedijske vsebine. Cilj je ovrednotiti delovanje objektivnih algoritmov vrednotenja ter primerjati in analizirati te s subjektivnimi

rezultati. Vključena bo metoda porazdeljenega vrednotenja kakovosti storitev na podlagi polj ROI na izhodu vizualne modalnosti. Dodaten cilj je upoštevati dejstvo, da bo model evalvatorja deloval v realnem ali skoraj realnem času na sodobni strojni opremi, saj je ta vidik pomemben s stališča ostalih gradnikov v telekomunikacijskem omrežju.

Na osnovi opredeljenega raziskovalnega problema smo oblikovali tezo, da je možno izbiro vhodne modalnosti sistema, ki podpira govorni in DTMF-vnos, določiti na podlagi objektivne ocene kakovosti, ki upošteva degradacijo uporabniških podatkov. V skladu s tem postavimo naslednjo hipotezo:

Hipoteza 1:

D*efiniramo lahko klasifikator vhodne modalnosti, ki na podlagi objektivne ocene kakovosti uporabniškega vnosa določa tip vhodne modalnosti*.

Za uporabnika storitve je pomembna tudi sprejeta večmodalna vsebina. Ker je ta v IP-omrežjih pogosto degradirana, bomo predlagali model za oceno kakovosti takšnih vsebin. Ker so zahteve uporabnikov visoke, se bomo osredotočili na vsebine visoke ločljivosti, kar bo razširilo uporabnost glede na metrike kakovosti, ki trenutno obstajajo. Da zadostimo uporabi modela na širokem spektru pogojev, bomo upoštevali karakteristike vsebine in s tem izkoristili tehnološki in biološki faktor dojemanja uporabniške kakovosti, kar bo dalo boljšo oceno v primerjavi z obstoječimi metrikami, ki tega ne upoštevajo. Pri integraciji ocen kakovosti posameznih modalnosti bo potrebno tudi upoštevati medmodalni učinek. Na osnovi razumevanja kakovosti na kognitivnem nivoju tudi vemo, da je osredotočenost uporabnika večja na polja interesa. Ta niso odvisna od prostorskih lastnosti okvirja, temveč definirajo logično povezana stanja, kot jih razume človek. Takšen primer je človeški obraz. To dejstvo se s pridom uporablja pri detekciji ROI v sliki, mi pa bomo tak pristop uporabili pri evalvaciji kakovosti. Na tak način je možno definirati vizualna polja, kjer je potrebna precizna evalvacija objektivne kakovosti, na ozadju pa z uporabo enostavnejših metrik zmanjšamo čas izvedbe skupne evalvacije video okvirja in posledično tudi celotnega okvirja. Glede na omenjeno smo postavili naslednjo hipotezo:

Hipoteza 2:

Objektivno oceno kakovosti večmodalnih vsebin lahko napovemo s primerno združitvijo enomodalnih ocen kakovosti ter z upoštevanjem medmodalnega učinka. Takšen model kakovosti bo izkazoval dobro korelacijo za različne tipe posnetkov in degradacij. Pri tem lahko definiramo dodatne funkcije, ki zmanjšajo časovno kompleksnost evalvacije z upoštevanjem vpliva polja ROI.

1.3. Struktura doktorske disertacije

V uvodnem poglavju smo podrobneje predstavili področje obravnave problematike v tej doktorski disertaciji. Predstavili smo stanje raziskav, kjer se ukvarjajo z vrednotenjem kakovosti večmodalnih vsebin. Naslednje poglavje opisuje poglavitne pojme, ki jih bomo obravnavali v naši nalogi. Pri tem se seznanimo s pojmi kakovosti večmodalnih storitev tako iz perspektive ponudnika storitve kot tudi iz stališča uporabnika storitve. Poglavje 3 izpostavi nelinearnost človeškega senzornega sistema. Ti vidiki so pomembni pri razumevanju zaznave degradacije pri interakciji uporabnika z večmodalno storitvijo. Poglavje 4 predstavi vzroke in posledice degradacij, ki vplivajo na zaznavo uporabniške kakovosti takšnih storitev. Obravnava degradacij je ločena glede na tip modalnosti, v kateri se pojavlja, kar daje boljši pregled za nadaljnje razumevanje. Poglavje 5 predstavi metodologijo pri izvedbi subjektivnih testov. Predstavi prednosti in slabosti različnih testnih metod in ocenjevalnih lestvic vrednotenja subjektivne kakovosti, pravilne postavitve eksperimentalnega okolja in pazljivosti pri izvedbi eksperimentov. Poglavje 6 seznani z objektivno evalvacijo večmodalnih storitev. Najprej so predstavljene prednosti teh metod pred subjektivnimi ter prednosti in težave znanih metrik kakovosti. To poglavje podrobneje obravnava metrike s polno referenco, ki so osnova objektivne evalvacije v tej disertaciji. V poglavju 7 je predstavljena zasnova eksperimentalnega sistema za vrednotenje večmodalnih storitev. Zasnova je razdeljena na tri področja: prvo obravnava vrednotenje vhodne modalnosti, drugo vrednotenje izhodne modalnosti ter tretje pristop porazdeljenega vrednotenja večmodalnega sistema. Poglavje 8 zajema rezultate analize vpliva degradacij na vhodno in izhodno modalnost sistema. Nato je predlagan klasifikator vhodne modalnosti, ki na podlagi objektivne meritve kakovosti uporabniškega govora odloči o tipu vhodne modalnosti. Nadalje je za določanje vpliva

kakovosti izhodne modalnosti predlagan model vrednotenja kakovosti večmodalnega sistema, ki združuje ocene posameznih modalnosti ter parametre izgube paketov in tipa scene testnega posnetka. Nazadnje je predstavljena analiza koncepta porazdeljenega vrednotenja kakovosti. Ta obsega analizo rezultatov detekcije obraza v posnetkih s sceno tipa intervju. Na podlagi uspešne detekcije je nato predstavljena performančna analiza z uporabo polj ROI s predlogom porazdeljenega vrednotenja kakovosti večmodalne storitve. Smiselnost porazdeljenega vrednotenja je potrjena s subjektivnimi testi. Podrobnejši podatki in rezultati, ki se navezujejo na zasnovo eksperimentalnega modela in rezultate, so dodani v prilogah. V zaključku so ovrednoteni glavni prispevki disertacije ter potrditve hipotez.

2. Osnovni koncepti kakovosti

2.1. Pojem kakovosti, kakovosti storitve in kakovosti izkušnje

Pojem *kakovost* v splošnem določa stopnjo odličnosti nečesa, merjeno v primerjavi z referenčno točko podobnih lastnosti. Terminologija ISO razlaga definicijo kot:

- *skupek lastnosti in značilnosti proizvoda ali storitve, ki se nanašajo na njegovo zmožnost, da izpolnijo posredno ali neposredno izražene potrebe* (ISO 8402:1986, 3.1) [106].

V telekomunikacijskem (TK) svetu kakovost predstavlja zmožnost TK-sistema, da zagotavlja zahtevano stopnjo funkcionalnosti in uporabnosti IKT-storitve, definirano s parametri kakovosti. Nadzor kakovosti poteka na nivoju prenosnega sistema, npr. v lokalnem (LAN), širokopasovnem fiksnem (xDSL, kabelski) ali mobilnem omrežju (2G, 3G, 4G, WiMAX), ter zajema tudi neomrežne vidike. Pri tem je stopnja kakovosti

primerjava dojemanja uspešnosti delovanja in zadovoljstva s storitvijo v odvisnosti od pričakovanj uporabnika, ki so za različno publiko in različne kontekste uporabe lahko različni [107], [108]. Standardizirana definicija *kakovosti storitve* (angl. Quality of Service – QoS) je pojmovana kot:

- *skupek značilnosti telekomunikacijske storitve, ki se nanaša na njeno sposobnost, da izpolni izražene in implicitne potrebe uporabnika te storitve. »Značilnosti« lahko opazujemo in izmerimo. Če so značilnosti definirane, postanejo parametri, ki jih je možno izraziti z metrikami kakovosti* (ITU-T E.800) [109].

Podrobneje lahko kakovost storitve razdelimo glede na perspektivo obravnave. Ločimo dve perspektivi: a) iz stališča distribucije storitve, tj. *distribucijska perspektiva*, in b) iz stališča uporabnika, tj. *uporabniška perspektiva*. Pri a) se kakovost določi na podlagi izmerljivih vrednosti, tj. karakteristik sistema in/ali storitve ter prenosnega okolja (omrežja). Te vrednosti sestavljajo [110]:

- *tehnološki faktorji*, povezani s prenosno infrastrukturo od točke distribucije do uporabniške opreme (angl. User Equipment – UE). V to kategorijo sodijo tipične meritve kakovosti vira ter degradacije na omrežnem nivoju,
- *aplikacijski faktorji*, ki opredeljujejo učinke na kakovost na sejnem, predstavitvenem in aplikacijskem nivoju,
- *storitveni in kontekstni faktorji*, ki definirajo vpliv komunikacijske semantike, prioritete, vrednosti in sistema zaračunavanja na doseženo kakovost.

Omenjeni faktorji vplivajo na učinkovitost, razširljivost, fleksibilnost, možnost vzdrževanja, pravilnost zaračunavanja, zmogljivost ter zanesljivost prenosnega omrežja. Slednje je ključnega pomena, ker največje IP-omrežje, tj. internet, deluje po principu *najboljše možne storitve* (angl. best-effort service), kar pomeni, da uporabniku ni zagotovljeno, kdaj in če sploh se njegovi podatki, natančneje IP-paketi, dostavijo ali pošljejo sprejemniku. Uspešnost sprejema je neposredno odvisna od količine degradacij, ki jih tak prenosni sistem vnese v komunikacijsko pot. Tipe omrežnih degradacij in vpliv na kakovost storitve opisuje poglavje 4.2. Potrebno je poudariti, da nekateri pod pojmom *QoS* obravnavajo samo vidike degradacij, ki izhajajo izključno iz omrežnih virov in pojavov. V našem delu razširjamo ta pogled z dejstvom, da lahko imajo podatki

že pred distribucijo *nepopolno* kakovost. Glede na izvor degradacije ločimo dva tipa, in sicer:

- *vpliv poslabšanja kakovosti na izvoru* in
- *vpliv poslabšanja kakovosti v prenosnem omrežju*.

Po drugi strani *uporabniška perspektiva* upošteva uporabniški vidik in človeške dejavnike vpliva na doseženo kakovost. Ta h kvantitativno merljivi tehnološki komponenti dodaja komponento *človeške* oz. *uporabniške izkušnje* (angl. User Experience – UX). Različni tehnološki parametri različno vplivajo na UX, pri tem pa se posamezni vplivi razlikujejo v odvisnosti od tipa storitve, konteksta uporabe, lastnosti uporabnika itd. Ker je zadovoljstvo uporabnikov s storitvijo dejansko merilo uspešnosti, to predstavlja ključni diferenciator na tržišču [111], [112]. Sodobni pristopi tako v uporabniški vidik zajamejo celoten »vmesnik« in upoštevajo tudi lastnosti ter prostorsko-časovno stanje uporabnika [113], kar lahko vključuje:

- človekov senzorni sistem,
- uporabnikove potrebe,
- pričakovanja,
- subjektivno interpretacijo,
- zadovoljstvo s ceno,
- kontekst uporabe,
- prepričanje,
- način uporabe,
- motivacijo itd.

Tak vidik imenujemo *kakovost uporabniške izkušnje* (angl. Quality of Experience – QoE). QoE je več-dimenzionalni konstrukt, ki je posredno in neposredno odvisen od tradicionalne QoS, pri tem pa njegovi atributi neposredno izhajajo iz opažanj končnega uporabnika [114]. Nekateri avtorji obravnavajo QoE na višjem, abstraktnem nivoju, poimenovanem tudi *nivo psevdodojemanja* [115], ali ga skušajo horizontalno umestiti v aplikacijski nivo modela OSI [116]. Predlaganih definicij QoE je več, bolj uveljavljene in relevantne za to disertacijo razlagajo, da je QoE:

- *splošna sprejemljivost aplikacije ali storitve, kot jo subjektivno zazna končni uporabnik; QoE zajema celoten učinek na storitveni/aplikacijski sistem od konca do konca in je lahko pod vplivom uporabnikovih pričakovanj in konteksta uporabe* (ITU-T P.10) [117],
- *stopnja ugodja oz. neugodja uporabnika pri uporabi sistema oz. storitve; to je posledica njegovih pričakovanj glede na uporabnost in/ali prijetnost uporabe storitve z vidika osebnosti in trenutnega stanja uporabnika; v kontekstu komunikacijskih storitev na QoE vplivajo storitev, vsebina, naprava, aplikacija in kontekst uporabe* (Qualinet White paper) [118],
- *QoE izhaja iz sodbe o zaznani kompoziciji subjekta glede na želeno oz. pričakovano kompozicijo* (Jekosch) [119],
- *je merilo učinkovitost uporabe IKT-storitev oz. produktov, ki temelji na objektivnih in subjektivnih, tj. psiholoških meritvah* (ETSI TR 102 643) [120].

Iz stališča uporabnika tako »boljša« vrednost tradicionalnih QoS-parametrov posledično ne pomeni tudi boljše ocene QoE [121]. Pri določanju zadovoljstva uporabnikov s storitvijo je zato potrebno obravnavati tudi vpliv omenjenih atributov.

2.2. Večmodalna izkušnja in storitve

Sodobni IKT-sistemi ponujajo storitve z bogato uporabniško izkušnjo, tako da uporabniku omogočajo večmodalno interakcijo, ki je človeku bolj intuitivna in daje več svobode, dostopnosti in univerzalnosti v primerjavi z interakcijo samo preko enega komunikacijskega kanala [122], [123], [124]. Večmodalnost je definirana s pojmoma »več«, kar dobesedno pomeni »več kot ena«, in »modalnost«, ki opisuje človeške kanale percepcije [125]. *Percepcija* je proces pridobivanja, interpretacije, zbiranja in organiziranja čutnih informacij preko človeških čutov (čutil). Tipi in število prvinskih človeških čutov, ki se v literaturi pojavljajo, niso enotni, tradicionalno je veljalo grobo pravilo, da je teh pet: vid, sluh, dotik, okus in vonj. V resnici je čutov *vsaj* devet, pri tem literatura omenja še: termopercepcijo (vročina), nocicepcijo (bolečina), čut za ravnotežje in čut zavedanja. Nekateri avtorji so naredili delitev na podlagi oddaljenosti zaznavanja, tako je skupina *čutov za oddaljeno zaznavanje*, kot npr. vid in sluh, zmožna »oddaljene« percepcije, *čuti za bližinsko zaznavanje* pa »bližinske« percepcije, kot je

dotik [126]. Čeprav zelo pomembni s stališča fizioloških lastnosti ljudi, so v računalniškem svetu nekateri čuti neizkoriščeni (vonj, okus) in težko kakovostno in kvantitativno merljivi glede na sedanje komunikacijske paradigme. V literaturi, v sklopu HCI, term *večmodalnosti* uporabljajo po večini kot primer *vhodne modalnosti* interakcije, npr. uporaba tipkovnice in govornega vnosa. Vendar bomo v tej disertaciji *večmodalnost* obravnavali za vsak konceptualni model medsebojne komunikacije med človekom (njegovimi čutili) in strojem (storitev).

V literaturi tako obstaja več definicij večmodalne interakcije. Za to disertacijo relevantne pojem opisujejo kot:

- *večmodalnost je proces interakcije preko več modalnosti; modalnost pojmuje človeške čute, ki omogočajo sprejem in obdelavo sprejete informacije* (Wasinger) [127],
- *večmodalna interakcija pomeni, da obstaja več kot en način medsebojne komunikacije/interakcije uporabnika s sistemom, na primer ponuditi uporabniku izbiro med govorjenjem in tipkanjem ali, v določenih primerih, dovoliti uporabniku kompozitni vnos* (W3C-EMMA) [128],
- *večmodalna HCI-interakcija je interakcija z virtualnim in fizičnim okoljem skozi človeku naravne modalnosti za komunikacijo, to je modalnosti, ki vključujejo pet človeških čutov* [129].

Večmodalni sistemi zajemajo večsignalno integracijo, tj. dva ali več vhodnih in/ali izhodnih načinov interakcije, kjer so komplementarne modalnosti sinergijsko združene, da omogočajo potencialno boljšo robustnost sistema ter dajejo večjo možnost izražanja. Primer takšnega sistema kaže slika 2.1.

Slika 2.1: Primer večmodalnega sistema.

Taksonomija nekaterih avtorjev razlikuje celo tri razrede, in sicer: *enomodalna*, *bimodalna* in *večmodalna* interakcija [130]. Če je tok informacije v obe smeri, je to *obojestranska interakcija*, sicer pa *enosmerna*.

Pri tem vhodna modalnost takšnega sistema zajema veličine, ki jih lahko razdelimo na:

- avditorno interakcijo (govor),
- taktilno interakcijo (dotik, pero),
- otipno interakcijo (haptična naprava),
- gestikularno interakcijo (kretnje oči, glave, telesa),
- EEG-interakcijo (kontrola z možganskimi valovi) in
- interakcijo preko standardnega vmesnika (tipkovnica, miška), imenovano tudi WIMP (angl. Windows-Icons-Menus-Pointers).

Potrebno je poudariti, da človek za dojemanje sveta neprestano uporablja kombinacijo svojih čutov. V vsakdanji HCI je najpogostejša predstavitev informacij uporabniku *vizualna* in *slišna* modalnost, kjer tradicionalno enomodalne enosmerne storitve konvergirajo v integrirano obliko, tj. *multimedijo*. Med pojmoma *multimedija* in *večmodalnost* obstaja sorodnost, v literaturi jo zasledimo tako:

- multimedija predstavlja primer večmodalnega sistema (uporabe več komunikacijskih kanalov), a opisuje tehnologijo oz. medij za interakcijo, npr. zvok, video, grafika, večmodalnost pa je način percepcije čutov in proces posredovanja informacij skozi tak medij; osnovna razlika je tako v dejstvu, da večmodalni sistem razume semantiko informacij, ki jih posreduje, multimedijski pa ne (Wasinger) [127],
- iz stališča sistemske obravnave je multimedijski sistem tudi večmodalen, ker ponuja preko različnih medijev uporabniku večmodalni izhod, tj. zvočno in vizualno informacijo; dodatno pri uporabi večmodalnega sistema ta uporabniku ponuja večmodalni vhod, na katerega se uporabnik odzove z uporabo (želene) modalnosti, ki je priročno sredstvo za interakcijo (Anastopoulou et al.) [131].

Glede na to, kako večmodalni sistem obravnava podatke iz različnih vhodov, ločimo tri načine [132]:

- *sekvenčni vnos*, kjer je možna uporaba samo 1 modalnosti sočasno, npr. uporaba tipkovnice,
- *simultani vnos*, kjer sistem pridobiva informacije iz več modalnosti sočasno, vendar jih obdela ločeno glede na čas zajema informacije; vhodna modalnost se skozi čas lahko spreminja, isti element vnosa pa je možno predstaviti v različnih modalnostih; primer tega je sočasna uporaba izbiranja DTMF in ASR,
- *kompozitni vnos*, kjer se večmodalne informacije obravnavajo kot skupen integriran, kompozitni vhod; primer tega je sočasna uporaba kretenj in govora, kjer uporabnik gestikularno nadomesti pomen besed, npr. »želim pridi do *tja*«.

Hkratnost vhodnih signalov je tako odvisna od pogojev uporabe, načina izražanja in zasnove celotnega sistema.

Prednosti večmodalnega sistema pred enomodalnim so predvsem v:

- *robustnosti*, saj redundanca iste ali podobne informacije skozi različne kanale povečuje kakovost komunikacije med uporabnikom in sistemom ter verjetnost razpoznave [133]; dodatno se lahko doseže večja stopnja in fina nastavitev uporabniškega vnosa z uporabo fuzijskih algoritmov več modalnosti [134],

- *naravnosti*, saj se večmodalni pristop približuje intuitivni komunikaciji, podobni tisti med ljudmi; v raziskavi je 95 %-100 % uporabnikov preferiralo večmodalno interakcijo pred enomodalno [135],
- *fleksibilnosti in prilagodljivosti*, saj večmodalni vmesnik omogoča uporabniku, da zaznava in strukturira njegovo komunikacijo na različne načine za različne kontekste uporabe, npr. simultano govor (v mirnem okolju) in pisanje (v hrupnem okolju); tako se izkoristijo semantične, časovne in sintaktične prednosti pred enomodalnim pristopom [127], [136],
- *učinkovitosti*, saj paralelni vnos doprinese k hitrosti in preprostosti uporabe [137], [138]; dodatno se uporabnik nagiba k izbiri modalnosti, ki se mu zdi manj nagnjena k napakam in je zato manj samo-popravljanja, spontanih ponovitev in napačne sinhronizacije med uporabnikom in sistemom [137]; uporabniki naredijo do 36 % manj napak ob uporabi večmodalnega vmesnika [139],
- *dostopnosti*, saj uporaba komplementarnih modalnosti daje boljše pogoje za interakcijo ljudem s posebnimi potrebami; primer tega so slabovidni ljudje, ki si pomagajo z uporabo razpoznavalnika govora (namesto vizualne interakcije) [140], ali delavci, ki nosijo rokavice in so zato omejeni pri uporabi tipkovnice [141],
- *komplementarnosti*, kajti hkratna uporaba več modalnosti kot prirojena shema uporabnikovega izražanja pomeni komplementarnost semantične informacije pri doseganju cilja, in ne redundantnosti [142].

Vendar imajo večmodalni sistemi tudi slabosti. Iz tehnološkega vidika so zbiranje podatkov, gradnja in testiranje takšnih sistemov kompleksnejši in dražji. Razvoj takega sistema zahteva tudi interdisciplinarno znanje. Pri tem je treba paziti, da ne pride do kognitivne preobremenitve, pri kateri je uporabnik izpostavljen stimulaciji s preveč medijskimi tokovi in vmesniki [143]. Potrebno je tudi razumevanje vpliva prenosnih degradacij na posameznih kanalih modalnosti [74].

2.3. Izbira modalnosti

Večmodalni vmesniki imajo vrsto prednosti (podpoglavje 2.2), pri tem pa izbira uporabljene modalnosti vpliva na uporabniško izkušnjo in zadovoljstvo ter hitrost in

kompleksnost opravila [144]. Primer sodobnega komercialnega sistema, ki uporablja večmodalni vhodni vmesnik, je storitev IVR [145]. Ta omogoča govorno vodenje z uporabo funkcije avtomatskega razpoznavalnika govora (angl. Automatic Speech Recognition – ASR), ki je še posebej primerno takrat, ko uporabnik nima prostih rok [146] ali kadar je potreben daljši dialog ali pridobivanje informacij od uporabnika. Govorni vnos je primeren tudi za slepe in slabovidne [147]. Vendar uporaba funkcije ASR ni trivialna, kadar prihaja do degradacij v govornem signalu [148]. Takrat je potrebno uporabiti robustnejšo metodo, kot je npr. DTMF-izbiranje. Kdaj je govorni vnos nezadovoljiv za razpoznavanje govora in je potreben prehod na DTMF, lahko določamo z izmerjeno kakovostjo vhodnih govornih podatkov. To je tipično narejeno s klasifikatorjem, ki se na podlagi karakteristik govornega signala odloča o izbiri vhodne modalnosti. Obstaja množica različnih algoritmov za izločanje značilk govora, ki so potrebne za klasifikacijo. Med njimi so pogosto uporabljeni algoritmi:

- *mel kepstralni koeficienti* (angl. Mel-frequency Cepstral Coefficients – MFCC) [149],
- *kodiranje z linearno predikcijo* (angl. Linear Predictive Coding – LPC) [150],
- *zaznavna linearna predikcija* (angl. Perceptual Linear Predicton – PLP) [151] in
- *relativno spektralno filtriranje koeficientov v domeni log* (angl. Relative Spectra Filtering of Log Domain Coefficients – RASTA) [152].

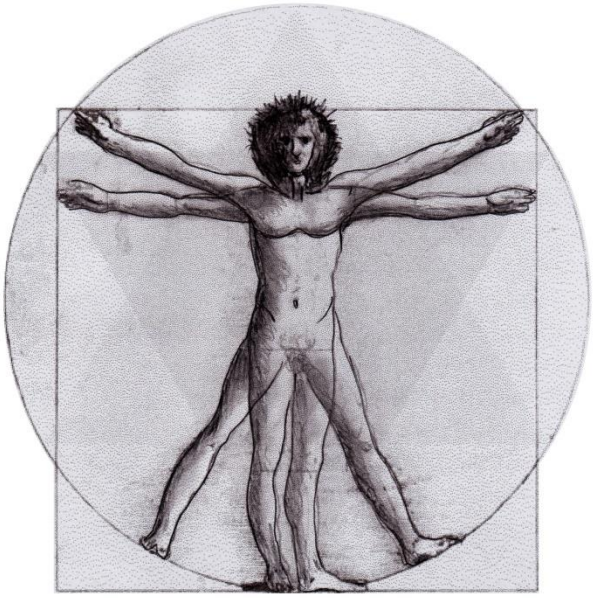
Med temi je metoda z MFCC-koeficienti, ki deluje na podlagi hitre Fourierjeve transformacije magnitude spektra, med bolj popularnimi zaradi enostavne uporabe in zanesljivosti. Značilke MFCC so predstavljene z vektorji, kjer dimenzija vektorja predstavlja število skalarnih komponent različnih značilk. Izločanje značilk je zato pomembna naloga, da dobimo dobre rezultate klasifikacije. Tipična vrednost dimenzije za uporabo govornih značilk v ASR-modulu je 13 značilk vsakih 10 ms, pri tem pa se dodajo še prvi in drugi odvodi, kar daje skupno 39-dimenzionalen vektor značilk [153].

Metoda za klasifikacijo, ki jo lahko uporabimo z dobljenimi vektorji značilk, je odvisna od lastnosti uporabljenega primera, v našem eksperimentu bo to klasifikacija vhodne modalnosti. Med bolj znane metode štejejo tiste, ki so zasnovane na:

- *modelih GMM* (angl. Gaussian Mixture Models),

- *skritih modelih Markova* (angl. hidden Markov models),
- *nevronskih mrežah* (angl. neural networks),
- *metodi podpornih vektorjev* (angl. Support Vector Machines) ali
- kombinaciji omenjenih modelov.

Pri tem je metoda z modeli GMM zasnovana na semiparametrični funkciji, ki določi maksimalno posteriori oceno disperzije Gaussovih funkcij in predstavlja zelo splošen model gostote porazdelitve. Zaradi doseganja boljšega rezultata modeliranja je funkcija gostote verjetnosti izražena kot linearna kombinacija osnovnih funkcij. Metoda GMM predpostavlja, da obravnavan signal vsebuje različne komponente, ki so med seboj neodvisne. Zato je zmožnost GMM-modelov ta, da tvorijo dobro aproksimacijo tudi funkcijam s poljubno obliko. V kombinaciji s procesom razpoznavanja govora, kot je npr. HMM, vodi to do hibridnih modelov. Posledično lahko določimo različne razrede klasifikacije v različnih domenah, npr. izbiro vhodne modalnosti, meritve kakovosti govora, identifikacijo govorca in oceno izgube paketov [154], [155], [156]. Dodatna prednost GMM je tudi obvladljivost učnih scenarijev za primer velikega in raznolikega nabora učnih podatkov.



3. Človeški senzorni sistem

Naše znanje o zunanjem svetu je odvisno od načina dojemanja, tj. zaznavanja informacij iz okolja na primitivni senzorni ravni. Informacije iz zunanjega okolja, ki so najpomembnejše za obstoj človeka, lahko organizem sprejme preko receptorjev, jih kodira v živčne signale ter obdeluje in shranjuje v mrežah osrednjega živčevja. Mreže so zmožne procesov, ki jih opredeljuje psihološki pojem **percepcije.** Ta zajema analizo informacij preko usmerjanja pozornosti, identifikacije, kategorizacije, interpretacije in konstrukcije smiselnih podob. Teh procesov še ne zmoremo preučevati neposredno; lahko uporabimo psihološke metode in raziskujemo nevrobiološke pojave, ki jih spremljajo. Predstavljamo si, da v nevronskih mrežah nastajajo notranje predstavitve okolja kot živčni modeli zunanjega sveta, ki pri človeku omogočajo tudi predvidevanje dogodkov in s tem ustreznejše vedenje. V seriji živčnih procesov (recepcija, kodiranje, prenos, zaznava, prepoznava) se dražljaj kot posledica zaznave informacije obdeluje v specializiranih senzoričnih kanalih, pri tem pa psihofizična analiza obravnava nekaj glavnih atributov dražljaja, npr. modalnost, jakost, trajanje, lokacijo. Te značilnosti se

uporabljajo za določanje pragov senzitivnosti, prostorske in časovne ločljivosti ter za razumevanje izkušnje [157].

V svetu telekomunikacijskih naprav so najpomembnejše modalnosti človeka vizualna (vid), avditorna (govor in sluh) in taktilna (dotik). V tej disertaciji bomo obravnavali prvi dve, ki zajemata pretežni del interakcije med uporabnikom in storitvijo. V sledečih podpoglavjih bomo zato predstavili osnovne koncepte človeškega vizualnega in avditornega sistema, ki bodo pomagali pri razumevanju vrednotenja uporabniške izkušnje, ki jo dojemamo po teh modalnostnih kanalih.

3.1. Človeški vizualni sistem

Človeški vizualni sistem (angl. Human Visual System – HVS) nelinearno zaznava, obdeluje in se odziva na vizualne dražljaje in jih posreduje možganskemu korteksu za nadaljnjo obdelavo in osmišljenje. Čeprav ni natančne meje, kje se določene funkcije procesiranja vizualnih podatkov obdelujejo, lahko hierarhično funkcije delimo na tri nivoje, odvisne od mehanizmov zaznave, in sicer na percepcijo na:

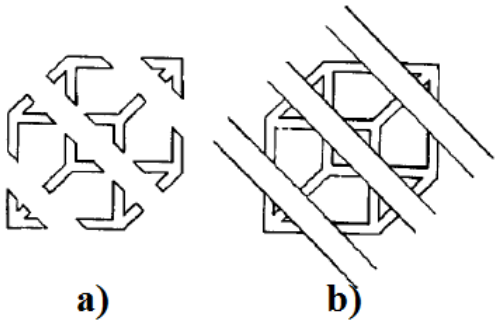
- *nižjem nivoju,*
- *srednjem nivoju in*
- *višjem nivoju.*

Percepcija na nižjem nivoju HVS obravnava analizo fizičnih dražljajev na nivoju mrežnice in je posledica elektromagnetnih lastnosti svetlobe (valovne dolžine 400 do 700 nm). Tukaj vpadni svetlobni žarki sprožijo fototransdukcijo, tj. pretvorbo svetlobne energije v spremembo membranskega potenciala fotoreceptorjev v mrežnici. Pragovne funkcije senzitivnosti veličin (kontrast, adaptacija, filtriranje frekvenc) se uporabljajo pri modeliranju nizkonivojskega modela vizualne kakovosti [158]. Percepcija obravnava:

- kontrast,
- barvo,
- filtriranje barvnih frekvenc,
- intenziteto/svetlost,

- adaptacijo na dražljaje (svetloba, kontrast itd.).

Percepcija na srednjem nivoju HVS obravnava elementarne vizualne primitive, kot so obrisi, strukture in površine, vendar so le-ti brez pomena. Primer mehanizma, ki loči ta nivo od zaznave na nižjem nivoju, je dojemanje navidezno izoliranih fragmentov (slika 3.1a), ki jih zaznamo kot *strukturo*, v primeru, da jih prekriva monotona površina, npr. trije beli pasovi (slika 3.1b) [159].



Slika 3.1: Mehanizem percepcije na srednjem nivoju: a) izolirani fragmenti, b) navidezno združeni fragmenti slike v strukturo.

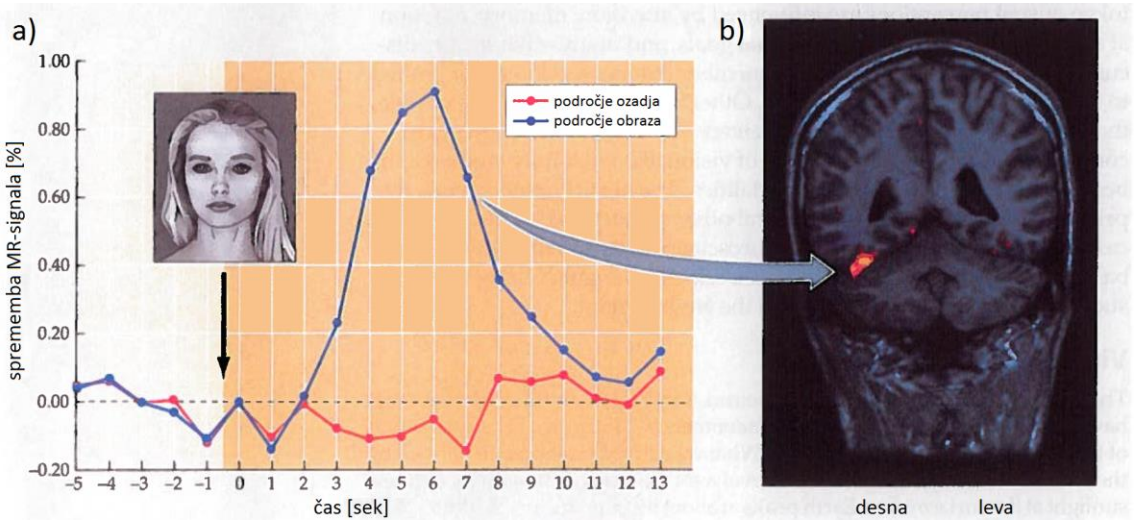
Ta nivo je tudi prvi, pri katerem lahko govorimo o pojavu osredotočanja in izločanja relevantnejše vizualne informacije [160]. Pojav je prisoten ob neregularnosti in izstopanju določenih vizualnih področij (barve, kontrasta, gibanja itd.) ter se detektira v nižjih delih vizualnega predela cerebralnega korteksa (V1) in v talamusu (LGN) [161]. Na tem nivoju HVS zaznava časovno-vizualne lastnosti:

- gibanje,
- obrise in sence,
- velikost,
- strukture,
- polja interesa,
- orientacijo,
- vzorce in
- teksture.

Razumevanje vizualne informacije na višjem nivoju HVS je pod vplivom spomina in zavestnega dojemanja, kjer je glavni pristop razumevanje človeških zmožnosti *za prepoznavo* in *kategoriziranje* podobnih objektov ter razumevanje podobnosti. Teorij razlage je več, nekateri trdijo, da so objekti analizirani kot konkavni deli povezanih robov in v spominu »shranjeni« v obliki abstraktnih komponent (cilindrov), ki jih iz slike lahko izločimo ne glede na spremembe v orientaciji [162]. Tukaj namreč posebni mehanizmi binokularnega in monokularnega vida tvorijo prostorsko informacijo (3D). Drugi predlagajo, da so različni pogledi objektov predstavljeni kot nabor značilnih potez oz. značilk in da je to orientacijsko odvisna funkcija. Po tej teoriji je v spominu shranjena množica pogledov objekta [163]. Na tem nivoju človek tudi prepoznava obraze. Vedenjski dokazi potrjujejo, da sicer delujejo drugi mehanizmi za objekte in obraze, pri tem pa so slednji bolj kritični in dovzetni za spremembe [164], [165]. Če povzamemo, na tem nivoju HVS dojema mehanizme:

- prepoznave obrazov,
- pomena vizualnih objektov in dogodkov,
- lokacije opazovanega predmeta,
- globine,
- integracije instanc objekta v dan kontekst in sceno,
- fokusa, namena in motivacije opazovalca,
- semantičnega in prostorskega konteksta.

Različni nivoji HVS so biološko, fiziološko in sociološko odvisni ter različno vplivajo na zaznavo kakovosti vizualnih karakteristik storitve in aplikacij. Primer tega je zaznava in prepoznava obraza. Dokazi iz področja nevrološke znanosti so pokazali veliko razliko v MR-odzivu pri prepoznavi znanih obrazov med zdravim človekom in tistim, ki trpi za prozopagnozijo, tj. redko obliko napake kognitivnega centra, kjer poškodovani del onemogoča prepoznavo obrazov, pri tem pa so ostale intelektualne in vizualne funkcije nedotaknjene (slika 3.2). To nakazuje na povečan fokus že iz biološkega stališča.

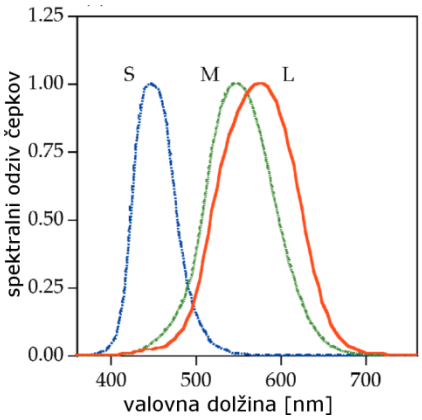


Slika 3.2: Detekcija obraza: a) MR-aktivnost pri detekciji obraza (modro) in ozadja (vijolično), b) področje zaznave strukture obraza (vir: [166]).

Če želi vizualni model kakovosti posnemati nevronski odziv HVS, je potrebno razumeti osnovna načela, po katerih deluje. V nadaljevanju so predstavljeni temeljni pojmi, ki vplivajo na dojemanje kakovosti vizualne informacije.

3.1.1. Kontrastna senzitivnost

Vizualne signale tipično prikazujemo v *prostorski* domeni (zaslon). Osnovni gradnik, *slikovna pika*, ima prostorske koordinate *(x, y* v 2D-prostoru) ter nosi sivinsko (črno-bele slike) oz. barvno vrednost (barvne slike). Svetlost pike izrazimo z *luminanco Y*, kromatično informacijo pa z vrednostjo v barvnem prostoru. Vendar je to samo eden od možnih pristopov, kako predstaviti vizualni signal. Biološki fotoreceptorji v človeškem očesu, natančneje na očesni mrežnici, so zgrajeni iz paličic in čepkov. Pri tem so prvi akromatični in zadolženi za vid v slabših svetlobnih pogojih, drugi pa sodelujejo pri vizualni percepciji ob večji svetlosti. Obstajajo trije razredi čepkov, ki prenašajo informacijo po neodvisnih nevronskih kanalih: L-, M- in S-čepki, ki so dovzetni za *kratke*, *srednje* in *dolge* valovne dolžine svetlobe (slika 3.3). Rečemo tudi, da je človeški vid *trikromatičen*.

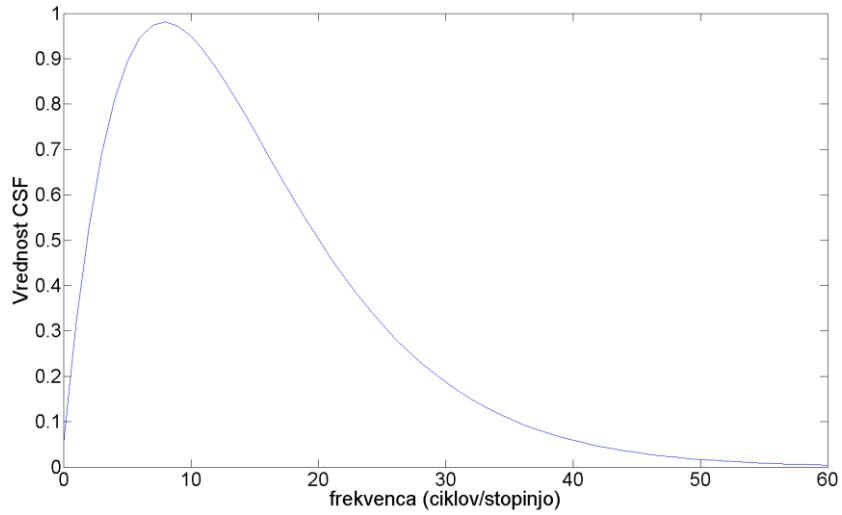


Slika 3.3: Spektralni odziv čepkov v človeškem očesu (vir: [167]).

Ugotovljeno je bilo, da je najmanjši potreben kontrast, da človek zazna vizualno degradacijo, določen s prostorsko frekvenco te spremembe [168]. Tak minimalni kontrast imenujemo *prag detekcije kontrasta*, njemu obratno funkcijo pa *kontrastna senzitivnost*. Ta nam pove, kako senzitivni smo za različne frekvence vizualnega dražljaja, in deluje kot pasovnoprepustni filter (slika 3.4). Avtorji so v [169] definirali *funkcijo senzitivnosti kontrasta* (angl. Contrast Sensitivity Function – CSF) na črno-belih slikah kot funkcijo ciklov, tj. število sinusoidnih luminančnih ciklov na stopinjo vidnega kota *f*:

$$\boldsymbol{CSF(f) = 2,6(0,0192 + 0,114f)e^{-(0,114f)^{1.1}}} \qquad (3.1)$$

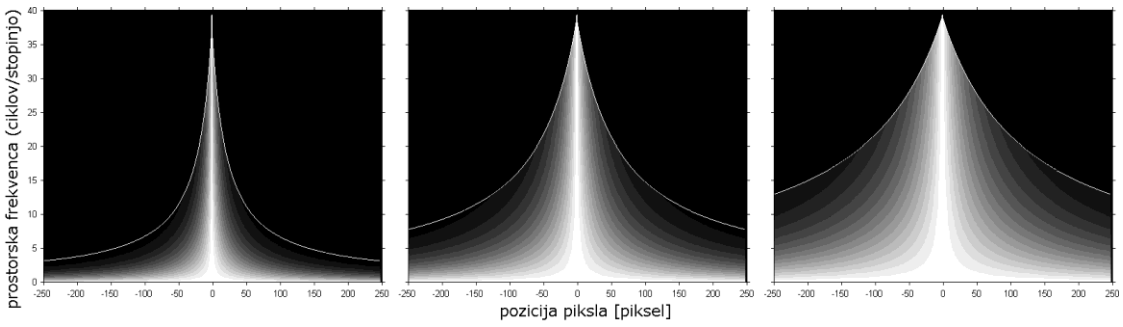
Funkcija ima maksimum pri *f ~ 8 ciklov/stopinjo* in je neznatna za *f > 60 ciklov/stopinjo*. Zgornja meja je posledica optičnih lastnosti očesa, razdalje med fotoreceptorji in kvantnega šuma [170], zmanjšanje senzitivnosti pri nizkih prostorskih frekvencah pa zaradi omejenosti sprejemnega vizualnega kota (optični fokus) in efekta maskiranja DC komponente. Enačba 3.1 predstavlja zgolj poenostavljeno aproksimacijo realnega optičnega modela očesa. V realni situaciji pa opazimo dodatna odstopanja, ki so posledica statičnih ali dinamičnih sprememb pri individualnih opazovalcih. Sem lahko prištevamo spremembe svetlosti ozadja in opazovanega objekta (velikost zenice), geometrijske oblike opazovanega vzorca (senzitivnost diagonalno orientiranega vzorca je manjša kot horizontalnega in vertikalnega), geometrijo očesa (dioptrija, sprememba fokalne razdalje), očesno adaptacijo na luminanco in ostale biološke dejavnike (starost, astigmatizem, barvna slepota, ambliopija) [171], [172] ter časovne vidike [173], [174], [175].



Slika 3.4: Funkcija CSF deluje kot pasovnoprepustni filter.

3.1.2. Prostorska senzitivnost

Fotoreceptorji so neenakomerno porazdeljeni na očesni mrežnici, zato je zaznana prostorska resolucija višja za stvari, ki so bliže fiksacijski točki našega pogleda, tj. na območju rumene pege (slika 3.5) [176]. Pristop neposredno uporabljajo IQA-metrike, ki temeljijo na modelu pozornosti (angl. attention model) (slika 3.6).



Slika 3.5: Model vizualne prostorske senzitivnosti za sliko velikosti $W_{š} \times W_{d} = 512 \times 512$ slikovnih pik iz razdalje H_{w} = 1- (levo), 3- (sredina) in 6-kratna *višina slike*. Svetlost predstavlja normalizirano moč senzitivnosti vizualne napake.

Slika 3.6: Gradientna ekscentričnost vidnega polja na izvorni sliki (levo) in njena grafična predstavitev (desno).

3.1.3. Orientacijska senzitivnost

V IQA-algoritmih je CSF upoštevan tako, da se slika filtrira z 2D-prostorskim filtrom, zasnovanim tako, da ustreza psihofizičnim rezultatom. Avtor je v [177] definiral njegov frekvenčni odziv kot:

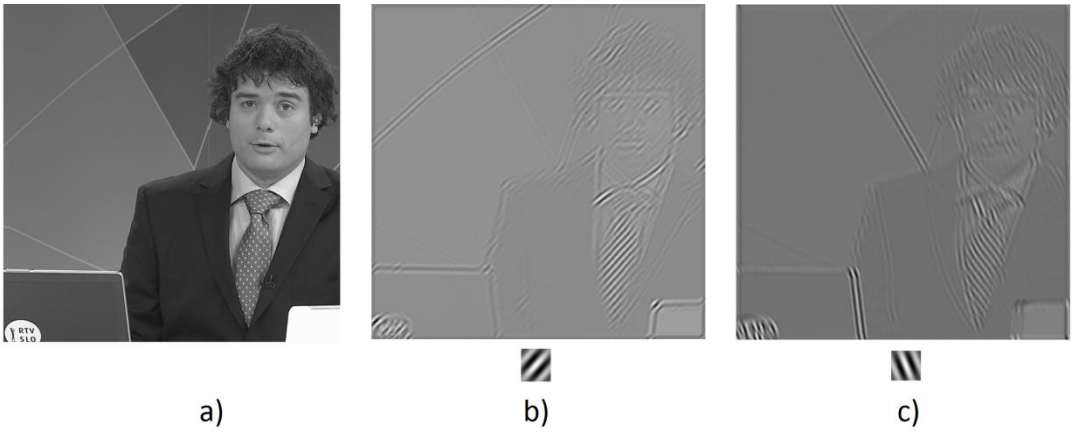
$$\boldsymbol{H(f,\theta)} = \begin{cases} \boldsymbol{2,6(0,019 + \lambda f_\theta)e^{-(\lambda f_\theta)^{1,1}}}, & \boldsymbol{\check{c}e\ f \geq f_{mejni}} \\ \boldsymbol{0,981}, & \boldsymbol{ostalo} \end{cases} \tag{3.2}$$

kjer je f radialna prostorska frekvenca, λ valovna dolžina in $\theta \in [-\pi, +\pi]$ orientacija. *Orientacija* vpliva na percepcijo vizualnega signala že v V1, kjer je nevronski odziv na kardinalni dražljaj (vertikalni ali horizontalni) močnejši kot za poševni, npr. 45° ali 135° dražljaj [178], kar pomeni, da človek precej bolje zaznava vertikalne in horizontalne vzorce (za 2- do 4-krat) kot poševne. *Efekt poševnosti* (angl. oblique effect) je definiran z:

$$\boldsymbol{f_\theta} = \frac{\boldsymbol{f}}{\boldsymbol{[0,15\cos(4\theta) + 0,85]}} \tag{3.3}$$

Pri evalvaciji kakovosti slik je pogosta uporaba Gaborjevega filtra pri vrednotenju teksturne degradacije (slika 3.7). To je orientacijski, linearni, pasovno

prepustni filter za zaznavo robov, katerega frekvenca in orientacija sta podobni, kot ju daje kortikalni odziv t. i. *enostavnih celic* HVS.



Slika 3.7: Odziv Gaborjevih filtrov na sceno *Intervju_napovedovalec*: a) izvorna slika, b) z orientacijo filtra 22° in c) z orientacijo filtra -15°.

3.1.4. Luminančna senzitivnost

Zaznana luminanca opazovanega objekta je odvisna od luminance ozadja, pred katerim se objekt nahaja. Odvisnost je eksponentna za luminanco ozadja manjšo od 70 (na 8-bitni lestvici) in linearna za vrednosti nad 70 [158], [179]:

$$\boldsymbol{J(x,y)} = \begin{cases} \mathbf{17}\left(\mathbf{1} - \sqrt{\dfrac{\bar{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y})}{\mathbf{70}}}\right) & \text{če } \bar{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y}) \leq \mathbf{70} \\ \dfrac{\mathbf{3}}{\mathbf{185}}(\bar{\boldsymbol{A}}(\boldsymbol{x},\boldsymbol{y}) - \mathbf{70}) + \mathbf{3} & \textbf{\textit{ostalo}} \end{cases} \tag{3.4}$$

kjer je *J(x,y)* vidni prag za točko *(x,y)* in $\bar{A}(x,y)$ lokalno povprečje luminance izračunano z utežnim nizko-prepustnim filtrom (slika 3.8).

1	1	1	1	1
1	2	2	2	1
1	2	0	2	1
1	2	2	2	1
1	1	1	1	1

Slika 3.8: Utežni filter luminance.

Mehanizem adaptacije na temo s povečanjem zenice in leče poveča količino fotonov, ki padejo na mrežnico. Pri zelo nizkih svetlobnih pogojih (pod 0,03 cd/m^2) smo nezmožni zaznavanja barv (nočni vid), pri velikih intenzitetah pa nevronski inhibitorji poskrbijo za omejevanja signala na ganglijskih celicah. Tak pojav je opazen pri preklopih scen na ekranu, kjer alfa adaptacija, sprožena z živčnimi mehanizmi na nižjem nivoju, poskrbi za hitro zmanjšanje vzdraženosti nevronov še pred skrčenjem zenice. Ker je skupna adaptacija očesa odvisna od združenih dražljajev vidnega polja, oko fizično enake signale (glede amplitude in frekvence svetlobe) dojema drugače: primer tega je enak kvadrat, ki ga obkroža temno ali svetlo ozadje (slika 3.9).

(a) (b)

Slika 3.9: Razlika dojemanja iste barve v odvisnosti od luminance ozadja.

3.1.5. Kritična frekvenca utripanja

Časovna resolucija človeškega očesa je odvisna od *kritične frekvence utripanja* (angl. Critical Flicker Frequency – CFF). V primeru, da je prag CFF presežen, zaznamo

nezveznost gibanja zaporednih slik, kar daje občutek zmanjšanja kakovosti vsebin. CFF je odvisna od luminance ozadja *L* in jo opisuje Ferry-Porterjev zakon [180]:

$$\boldsymbol{CFF = a\,logL + b} \tag{3.5}$$

kjer sta *a* in *b* konstanti. Dodatno je CFF odvisna tudi od lastnosti dražljaja, tj. njegove intenzitete in velikosti [181]. Pri vrednotenju kakovosti videa mora metrika kakovosti zajemati tudi časovno komponento signala, če želimo ovrednotiti stopnjo utripanja. Ta je pogosto pogojena z lastnostmi strojne opreme in premajhno hitrostjo okvirjev. *Prag fuzije utripanja* (angl. Flicker Fusion Threshold – FFT) je za premikajoče se slike običajno definiran z vrednostjo 16 Hz.

3.1.6. Vizualno maskiranje

Lastnost HVS, da izpostavljena območja slike zaznava drugače, imenujemo *vizualno maskiranje* (angl. visual masking). V splošnem pojav nastane takrat, ko se vidnost/zaznavnost opazovanega objekta (tarča) zmanjša ob prisotnosti drugega objekta (maska). Vizualno maskiranje obsega široko področje zaznavanja vizualnih dražljajev, in sicer:

- **Luminančno maskiranje**: matematična aproksimacija dojemanja fizičnih dražljajev je določena z Weberjevim zakonom, ki definira percepcijo fizičnega dražljaja kot funkcijo, kjer je minimalna zaznana razlika (angl. just-noticeable difference) odvisna od magnitude samega dražljaja. Če na ozadju z luminanco L postavimo objekt z luminanco L + ΔL, potem je prag detekcije, da človeško oko zazna spremembo ΔL, odvisen od intenziteta ozadja in je [182]:

$$\frac{\boldsymbol{\Delta L}}{\boldsymbol{(L + L_0)}} = \boldsymbol{konst.} \tag{3.6}$$

  kjer je konstanta L_0 ekvivalentna nevronskemu šumu in predstavlja t. i. »temno svetlobo«. Ker je le-ta majhna, je pomembna samo za nizke vrednosti *L*.

  Fechner je dopolnil Weberjev zakon z interpretacijo subjektivne zaznavnosti svetilnosti. Predpostavil je, da je senzorična magnituda Δr, ki je

najmanjša še zaznana diferenca ΔL/L, konstantna. Weber-Fechnerjev zakon definiramo kot [183]:

$$\Delta \boldsymbol{r} = \boldsymbol{c}\frac{\Delta \boldsymbol{L}}{\boldsymbol{L}} \tag{3.7}$$

Za velike vrednosti L lahko L_0 izpustimo iz generalizirane enačbe. Če predpostavimo integrabilnost Δr in ΔL, lahko določimo senzorični prag h končni magnitudi dražljaja R:

$$\boldsymbol{R} = \int \boldsymbol{dr} = \int \frac{\boldsymbol{dL}}{\boldsymbol{L}} \tag{3.8}$$

$$\boldsymbol{R} = \boldsymbol{c}\,\boldsymbol{ln}\,\boldsymbol{L} + \boldsymbol{A} \tag{3.9}$$

Izpeljemo lahko določen integral med vrednostima L_t in L, kjer je L_t najmanjša zaznana vrednost luminance L. Ker velja, da je $R=0$ za $L=L_t$, enačba dobi obliko:

$$\boldsymbol{R} = \boldsymbol{c}\,\boldsymbol{log}\frac{\boldsymbol{L}}{\boldsymbol{L_t}} \tag{3.10}$$

$$\boldsymbol{R} \approx \boldsymbol{logI} - \boldsymbol{logL_t} \tag{3.11}$$

R predstavlja dejansko *zaznano intenziteto* dražljaja v razmerju z *magnitudo fizičnih dražljajev I*. Weber-Fechnerjeva definicija je primerna tudi za druge tipe dražljajev [184], [185].

Kasneje je idejo razširil Stanley Stevens, ki je predpostavil, da človekova senzitivnostna skala sledi eksponentni funkciji namesto logaritemski. Stevensov zakon je definiran kot:

$$\boldsymbol{R} = \boldsymbol{kI^a} \tag{3.12}$$

kjer je *k* konstanta proporcionalnosti (odvisna od tipa stimulacije in uporabljenih merskih enot), *I* magnituda fizičnega dražljaja in *a* eksponent, odvisen od tipa stimulacije. Za $a \sim 0{,}3$ je oblika Stevens-ove krivulje podobna Weber-Fechnerjevi, kar sovpada z eksperimentalno izmerjenimi vrednostmi za zaznano intenziteto svetlosti.

- **Maskiranje kontrasta**: pomeni prisotnost maskirnega signala s podobno prostorsko lokacijo in frekvenčno vsebino kot jo ima vizualna napaka ter posledično višji prag detekcije takšne napake. Vzrok je v delovanju možganskega korteksa, ki deluje kot pasovno prepustni niz filtrov s pasovno širino 1 oktave in 30-kotnim stopinjskim razmikom med frekvenčnimi pasovi. Prisotnost enega signala povzroči detekcijo tudi drugih komponent signala v istem podpasu ali v sosednjih podpasovih. Tako enačba 3.4 ne deluje za poljubne podatke, prav zaradi nelinearnosti HVS. Je dobro raziskana oblika maskiranja v IQA-metodah evalvacije vizualne kakovosti zaradi neposredne merljivosti veličin [186].

- **Maskiranje obrisov:** HVS je izredno prilagojen na izločanje strukturne informacije, še posebej na robove. To izkoriščajo tudi metrike za vrednotenje kakovosti slik, kot npr. ESSIM [187].

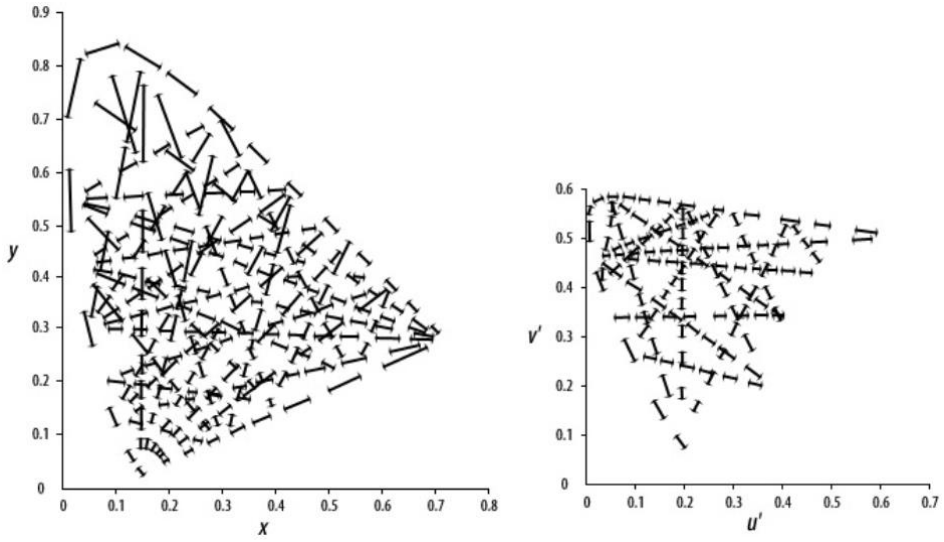
- **Maskiranje svetlosti:** kljub temu da ima oko veliko sposobnost prilagoditve v širokem dinamičnem razponu svetlosti (do 10^9) [188], se senzitivnostna krivulja spreminja tudi z navzočnostjo okoljske svetlobe, ki obkroža opazovani zaslon (adaptacija),

- **Maskiranje tekstur, vzorcev in gradienta:** teksture in vzorci so primeri maskiranja na srednjem nivoju HVS, kjer zaradi prisotnosti vizualnih vzorcev HVS drugače reagira na testni dražljaj [189]. Na tak način je mogoče določiti različne moči prostorske kvantizacije pri kodiranju slik za enak učinek zaznave kakovosti ali upoštevati manjšo pomembnost degradacij (šuma) v poljih slike z več teksturne informacije [190], [182].

3.1.7. Digitalno procesiranje vizualnih signalov

Kako je potrebno vizualno informacijo digitalno zakodirati, predpisuje *barvni prostor*. Ta specificira karakteristike temeljnih barv, iz katerih je mogoče sestaviti kompletni sestav vidnih frekvenc (400 nm ~ 700 nm). Glede na kolorimetrijo, ki definira človeško percepcijo barv, so poznani sledeči barvni prostori:

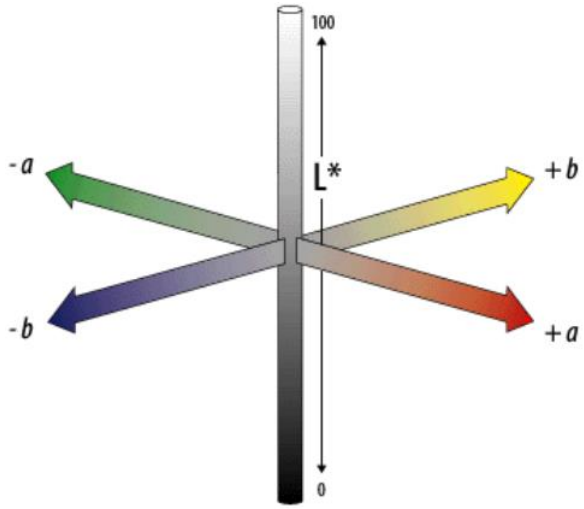
- **CIE:** *Commission Internationale de l'Eclairage* (CIE) je prva matematično definirala in standardizirala barvni prostor na podlagi meritev človeške vizualne percepcije (CIE 1931 XYZ). Problem XYZ je bila nelinearnost oddaljenosti med vrednostmi v kromatičnem diagramu. Uniformnost v smislu zaznane »oddaljenosti« med vrednostmi so nato popravili s prostorom CIELUV (slika 3.10), ki je še posebej uporaben pri aditivnem mešanju barv zaradi svojih linearnih lastnosti translacije vrednosti *XYZ* → *LUV* in uporabo skale svetlosti *L**, ki izvira iz Munsellovega sistema in povezave med *luminanco ali intenziteto* (*Y*) in *svetlostjo ali svetilnostjo* (V) [191] (slika 3.11).



Slika 3.10: Kromatična diagrama CIEXYZ (levo) in CIELUV (desno). Dolžina črt predstavlja enako zaznavno oddaljenost med barvami (vir: [192]).

Slika 3.11: Enaki koraki luminance (levo) in enaki koraki svetlosti (desno).

Ločeno luminančno komponento *L* je prevzel tudi model *nasprotujočih si barv*, imenovan CIELAB. Ta razlaga, da ima opazovana barva določeno vrednost, ki ne more biti seštevek več barv, saj si različne barve, tj. zelena in rdeča ter vijolična in rumena, med seboj nasprotujejo (slika 3.12).

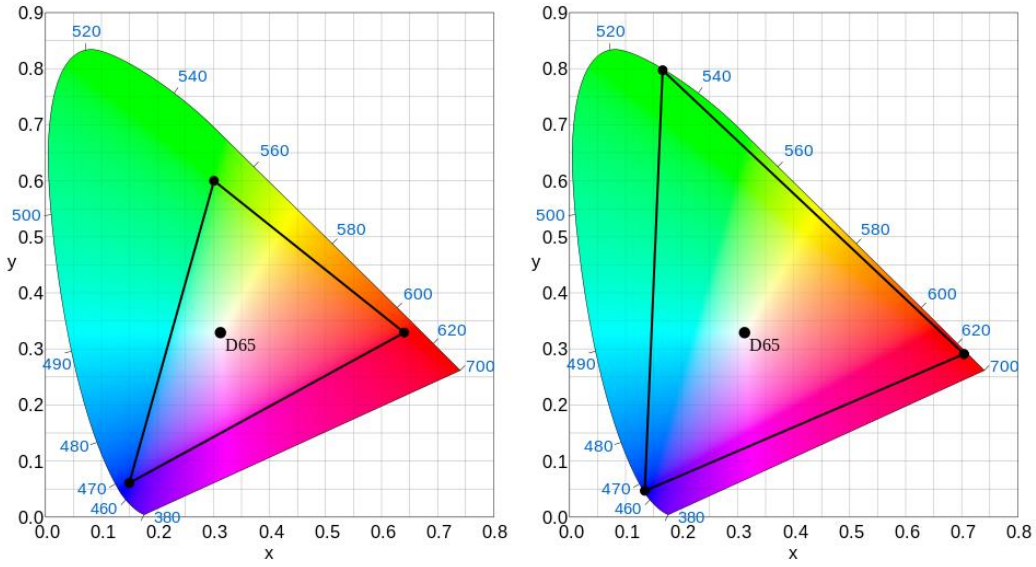


Slika 3.12: Barvni prostor nasprotujočih si barv CIELAB. Vsaka izmed osi *a, b* določa vrednosti temeljnih barv, *L** pa svetlost (vir: [192]).

Prednost CIELAB je njegova neodvisnost od naprave, kar daje boljšo referenčno točko pri obravnavi slik iz različnih naprav.

- **RGB: j**e aditivni barvni model, primeren predvsem za reprodukcijo na zaslonu, TV ekranih in digitalni fotografiji. Barvni spekter se dobi z mešanjem različnih intenzitet RGB-barv, pri tem daje *intenziteta 0* črno, *polna intenziteta* vseh komponent pa belo barvo. Izbira temeljnih barv, tj. valovnih dolžin, ki jih poznamo kot »rdeča«, »zelena« in »modra«, je povezana s fiziologijo očesa, saj te predstavljajo maksimalno oddaljenost med frekvenčnimi odzivi čepkov in posledično določajo velik barvni trikotnik [193]. Popularni barvni prostori, ki se uporabljajo in temeljijo na RGB-modelu, so *sRGB* (ITU-R BT.709), *NTSC*, *PAL*, *UHDTV* in *Apple RGB*. Čeprav ima različica *sRGB* enega izmed ožjih razponov vrednosti, je še vedno pogosto uporabljana (HD-zasloni). Definirana je bila kot

generičen prostor, ki ustreza povprečnemu monitorju, popravljeno z nelinearno gama korekcijo. Najnovejši UHDTV uporablja novi standard ITU-R BT.2020 (slika 3.13).



Slika 3.13: RGB, ki ga uporabljata BT.709 (levo) in BT.2020 (desno) v barvnem prostoru *CIEXYZ* (vir: [194], [195]).

- **CMYK (cijan, magenta, rumena, črna):** je odštevalni barvni prostor in uporablja štiri primarne komponente. Uveljavil se je predvsem v tisku slik. Četrta komponenta predstavlja »črno« barvo, ki se zaradi tehnične narave tiskalnikov ne meša, kot npr. aditivno pri RGB, ampak je uporabljena ločeno. »Odsotnost« barvne informacije predstavlja »belo«.

- **HSV/HSB (odtenek, nasičenost, svetlost):** je različica RGB, kar pomeni, da so njene komponente in kolorimetrija izpeljane iz RGB-sistema. HSV-barve so predstavljene v cilindričnem koordinatnem prostoru (namesto v kartezijskem), kjer vrednosti zastopajo barve v smislu barvnega odtenka in nasičenosti namesto v smislu odštevanja temeljnih komponent, kar daje bolj intuitivno in percepcijsko boljšo razlago.

- **HSL/HSI/HSD (odtenek, nasičenost, luminanca):** je podoben sistemu HSV, vendar »svetlost« zamenjuje »luminanca« ali »svetilnost«. Razlika je v tem, da je

svetlost »čiste« barve enaka svetlosti bele barvne (tj. seštevek vseh barv), medtem ko je luminanca »čiste« barve enaka svetilnosti povprečne vrednosti sive.

IQA-degradacije, kot je npr. *uhajanje barvne informacije,* težko ovrednotimo s sivinskimi metrikami kakovosti, zato je v literaturi zaslediti razširitve teh metrik na barvni prostor [196], [197]. Ovrednotenje se naredi na posameznih barvnih kanalih, ki se združijo v skupno oceno z določeno (uteženo) funkcijo [198]. Nekateri podatke prenesejo v druge barvne prostore, ki so zaznavno bolj uniformni (CIELAB) in je barvna degradacija lažje merljiva [199].

3.2. Človeški avditorni sistem

Človeško uho je organ slušnega sistema, ki skrbi za zaznavo zvočnih informacij. Govor kot tip takšne informacije predstavlja poglavitni tip komunikacije med ljudmi ter komunikacije s stroji (angl. Human to Computer interface – HCI), zato so tudi fiziološke karakteristike najbolj prilagojene na dojemanje govornih signalov. Splošno nam fiziološke in nevrološke karakteristike slušnega sistema postavljajo določene omejitve, ki vplivajo tudi na tehnološke vidike pri uporabi avditorne modalnosti. V naslednjih podpoglavjih izpostavljamo dejstva, ki služijo pri razumevanju zaznave in adaptacije avditorne modalnosti.

3.2.1. Frekvenčna senzitivnost

Slišna frekvenca je med 15 Hz in 20 kHz, vendar se s starostjo frekvenčni razpon nekoliko zoža (20 Hz do 16 kHz za starejše). To vpliva na izbiro frekvenčnega razpona pri omejitvi avdio kanalov, zato je tipično »perfektno« kakovost zvoka možno zakodirati z dvakratnikom te frekvence, kot definira Nyquistov teorem. K temu je dodano še majhno nadvzorčenje (angl. oversampling) z upoštevanjem:

- karakteristik zaznave,

- potrebnega tranzicijskega frekvenčnega pasu pri implementaciji avdio nizko-pasovno prepustnega filtra, ki v realnem okolju ne zmore narediti čistega prehoda na mejni frekvenci,
- tehnoloških dejstev: zgodovinska omejitev kodiranja avdio kanala je zahtevala, da avdio kanal ne sme presegati več kot 3 kodirne vzorce na video vrstico, to pa za televizijski sistem PAL pomeni 294 aktivnih vrstic in 50 pol-okvirjev na sekundo.

Iz omenjenih predpostavk izhaja tudi frekvenca vzorčenja CD-zvoka, ki je 44,1 kHz. Senzitivnost ušesa za frekvence je največja za nižja frekvenčna območja v bližini govornih frekvenc. Človeški glas ima osnovno frekvenco 85 Hz–180 Hz pri moških in 165–255 Hz pri ženskah. To je pomembno pri določanju frekvenčnega razpona govornih storitev, kot sta PSTN in GSM.

3.2.2. Amplitudna senzitivnost

Amplitudna senzitivnost definira najmanjši zaznan signal. Odvisna je od frekvence in je največja med 1 kHz in 3 kHz, kar je pogojeno z resonančnimi lastnostmi ušesnega sistema (velikost slušnega kanal). Najmanjša amplituda, tj. variacija pritiska zraka oz. vibracij, ki jo zaznavamo, je 0.2×10^{-9} del pritiska ozračja. Lastnosti sluha tudi omogočajo zelo velik dinamičen razpon (merjeno v dB), pri tem pa zaščitni mehanizmi ušesa inhibirajo nevronski signal za glasne zvoke. Razlika amplitude, ki jo zaznamo, je približno *1,4x trenutna intenziteta*, in iz tega sledi naslednje pravilo:

$$\frac{\boldsymbol{zaznana\ razlika\ amplitude\ intenzitete}}{\boldsymbol{intenziteta}} = \mathbf{0,4} \qquad (3.13)$$

kar ustreza Weberjevemu zakonu, podobno kot velja tudi za percepcijo vizualnih signalov (enačba 3.6).

Največja amplituda zvoka, ki jo uho zazna brez bolečine, je 120 dB.

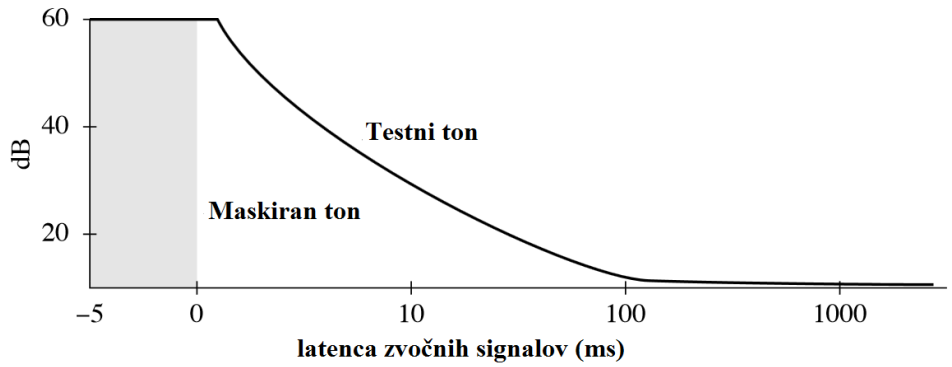
3.2.3. Avditorno maskiranje

Poznamo več vrst avditornega maskiranja. Glede na mehanizem in domeno delovanja ločujemo:

- simultano maskiranje,
- časovno ali nesimultano maskiranje in
- frekvenčno ali spektralno maskiranje.

Simultano maskiranje se pojavi, ko originalen zvok postane nerazumljiv zaradi prisotnosti drugega, npr. ton z veliko amplitudo in frekvenco 1 kHz zamaskira ton z manjšo amplitudo pri frekvenci 1.1 kHz. Lastnosti takšnega maskiranja lahko opišemo s Fletcherjevim spektralnim modelom [200].

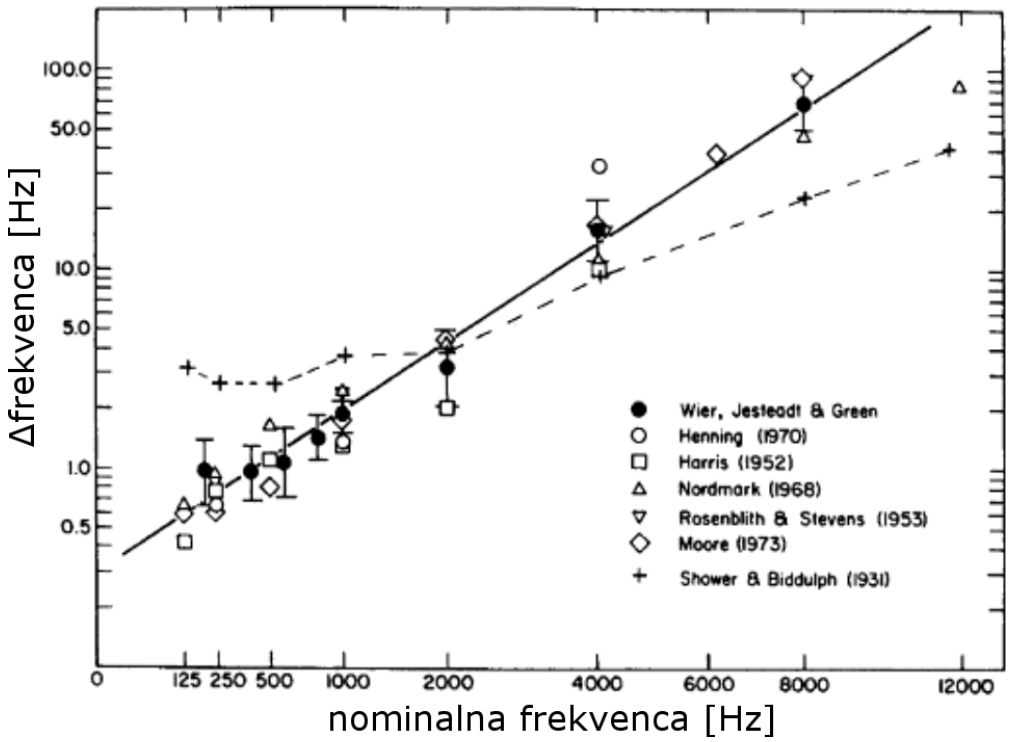
Časovno maskiranje nastane, ko neželen signal (šum) v časovni domeni sledi originalnemu ali je slišen pred tem. Dolžina lateralne inhibicije je pri tem približno 100 ms in 20 ms [201]. Dolžina časovno maskiranega signala je odvisna tudi od amplitude maskiranega in testnega tona (slika 3.14). Glasnejši kot je testni signal, manjša je latenca med zvočnima signaloma, da razločimo signala.



Slika 3.14: Časovno maskiranje v odvisnosti od amplitude signalov [202].

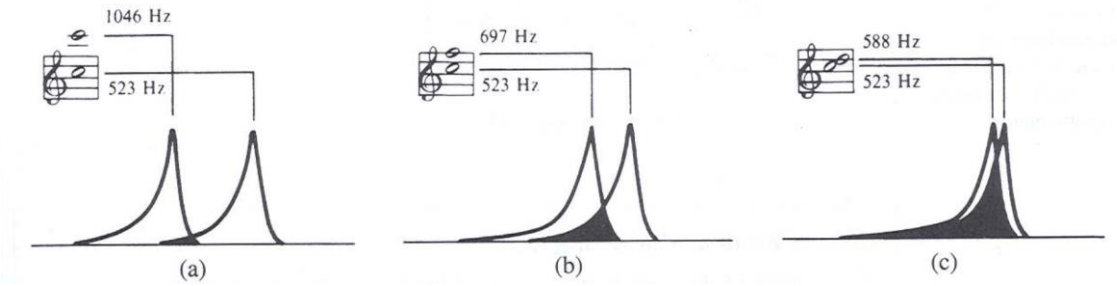
Prag frekvenčne diskriminacije za slišne zvoke s povprečno amplitudo je približno 0,3 % za frekvence okrog 100 Hz in se približno linearno povečuje z večanjem nominalne frekvence slišnega tona (slika 3.15). Na sliki vidimo izsledke več raziskav, kjer je zaradi pogojev testiranja prihajalo do majhnih variacij, v splošnem pa opazimo linearno odvisnost *nominalne frekvence* (abscisna os) od *frekvenčne razlike tokov* (ordinatna os). S tem dobimo približno 1300 različnih frekvenc zaznave, 600 za pas

med 15 Hz in 2 kHz ter 720 za frekvence med 2 kHz in 16 kHz. Ta načela upoštevajo tudi avdio kodeki.



Slika 3.15: Rezultati več raziskav meritev praga frekvenčne diskriminacije (Vir: [203]).

Pojav frekvenčnega maskiranja je pogojen v stimulaciji slušnih zaznavnih mehanizmov. V primeru, da sta dve frekvenci v frekvenčnem prostoru narazen, prihaja do le malo prekrivanja v smislu živčnega dražljaja iste frekvence, saj le malo slišnih dlačic reagira na obe frekvenci. Vendar z zmanjševanjem frekvenčnega intervala med njima prihaja do prekrivanja, ki v primeru frekvenčne razdalje, ki je manjša od kritične, daje občutek zaznave iste frekvence (slika 3.16).



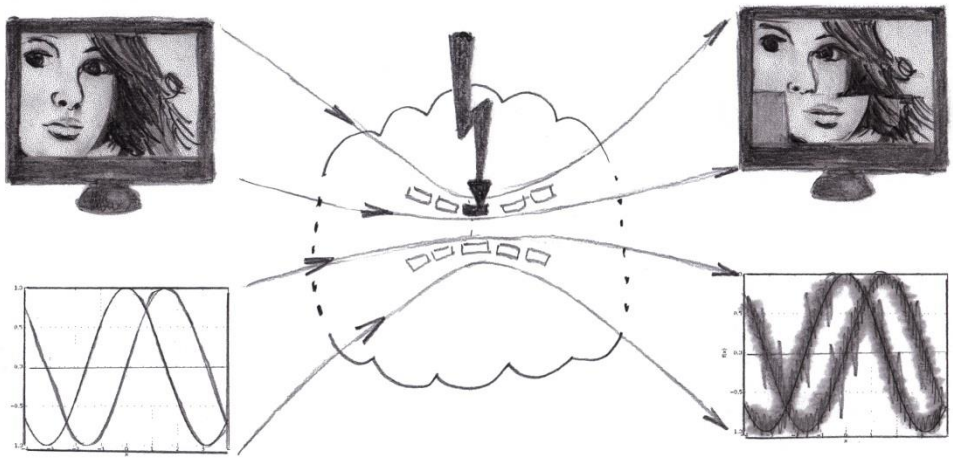
Slika 3.16: Frekvenčni odziv para čistih tonov.

Omenjene predpostavke izkoriščajo mehanizmi izločanja redundantne informacije v govornih kodekih na račun privarčevanja na pasovni širini kodeka.

3.3. Medmodalni vpliv vizualne in avditorne modalnosti

Dojemanje zunanjih dražljajev je v realnem okolju le v redkih primerih enomodalne narave. Pogosto namreč prihaja do medmodalnih učinkov in si predstavo zunanjih dražljajev dokončno predstavljamo šele v možganih. Dokazano je bilo, da izvor govora ob prisotnosti vizualne informacije, npr. videa z govorcem, dojemamo, da prihaja iz govorčevih ust kot pa iz dejanske lokacije zvočnega izvora [204]. To imenujemo pojav ventrilokvista. Podobno tudi različno razumemo pomen besede ob prisotnosti različne vizualne (premiki ustnic), a iste zvočne informacije (McGurkov efekt) [205]. Spet drugi so ugotovili, da uporabniki en svetlobni utrip luči zaznajo kot množico svetlobnih utripov v primeru, da jim sočasno predvajajo več zaporednih zvočnih piskov [206]. Omenjeno daje indikacije o večji medmodalni interakciji, kar so nekateri upoštevali tudi pri vrednotenju telekomunikacijskih sistemov [207], [208]. Tak primer je tudi model vrednotenja kakovosti avdio-video vsebin pri IP-prenosu, ki upošteva prostorsko in časovno kakovost, s paketno izgubo degradiranih video okvirjev in medmodalni učinek vizualne in avditorne modalnosti [209].

Kvantitativno predstavitev in statistično primerjavo odziva različnih modalnosti ter medmodalnega učinka v različnih testnih scenarijih je potrebno ovrednotiti. Najbolj relevantna metoda za to je subjektivno testiranje, ki nam daje povprečen odziv uporabnika storitve v danih testnih pogojih.



4. Degradacije multimedije in vpliv na kakovost storitve

Degradacije v IP-omrežjih lahko na grobo razdelimo v dve kategoriji:

- *izvorne degradacije* in
- *omrežne degradacije*.

Tipičen primer izvorne degradacije je *izgubna kompresija podatkov*, ki predstavlja poslabšanje originalnega signala že pri distributerju storitve. Kompresijski dejavniki degradirajo surov format videa in/ali avdia z namenom zmanjšanja potrebne pasovne širine in prilagoditve podatkov na prenosni/uporabniški sistem. V primeru prehoda skozi heterogen medij so uporabniški podatki podvrženi transkodiranju, da se zagotovi interoperabilnost in adaptacija TK-sistemov [210], npr. prenos multimedije iz interneta do klienta v 3G-omrežju, klic med VoIP- in GSM-naročnikom itd. Ob predpostavki, da so algoritmi transkodiranja izgubni, se kakovost podatkov v vsakem koraku transkodiranja zmanjša. Čeprav kompresijske metode upoštevajo dejstvo, da je v signalu

redundantna informacija, ki je HAVS ne more ali pa jo komajda zazna, je gostota informacij in njihova pomembnost pri stisnjenih podatkih večja na enoto potrebne pasovne širine signala. To pomeni, da so podatki tudi bolj dovzetni za morebitne napake pri prenosu. Prenos skozi omrežje lahko vnese neželene okvare, tipično *izgubo IP-paketov*, ki se pojavijo zaradi ozkih grl, nasičenih pomnilniških vrst na omrežni opremi, nepravilnega usmerjanja in podobno.

4.1. Izvorne degradacije

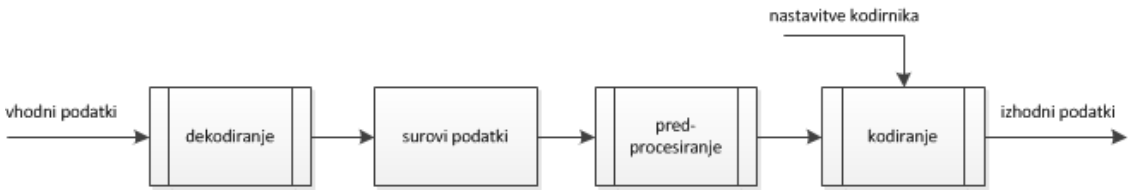
4.1.1. Transkodiranje

Transkodiranje je proces digitalne pretvorbe vsebine, npr. slike, videa ali avdio datoteke, multimedijskega toka ali tekstovne datoteke, iz ene v drugo kodirno konfiguracijo. Če se pri tem podatki zakodirajo v drug kodirni standard, imenovan kodek, govorimo o *heterogenem*, sicer pa o *homogenem* transkodiranju. Primer homogenega transkodiranja je transkodiranje video izvora visoke kakovosti v tarčni video tok nižje kakovosti z drugačno prostorsko-časovno resolucijo in bitno hitrostjo, vendar isto kodirno shemo [211]. Transkodiranje multimedijskih vsebin tako predstavlja eno ali več funkcionalnosti:

- spremembo bitne hitrosti,
- pretvorbo formata,
- spremembo prostorske razločljivosti (video),
- spremembo časovne resolucije (zvok in video) ter
- spremembo odpornosti na napake.

Trenutno obstaja več transkodirnih standardov za zvok in video za različne večmodalne vsebine in storitve. Standard se lahko uporabi in je optimiran za določen spekter uporabe, večjo generičnost transkodirnika pa dosežemo s kaskado dekodirnika, sistema za obdelavo (opcijsko) in kodirnika [212]. Pri tem dekoder ustvari sekvenco surovih podatkov (PCM za zvok, YUV za video), ki so lahko podvrženi želeni obdelavi,

npr. filtraciji šuma, prilagoditvi frekvenčnega razpona itd., in nato zakodirani z novim naborom kodirnih parametrov (slika 4.1).

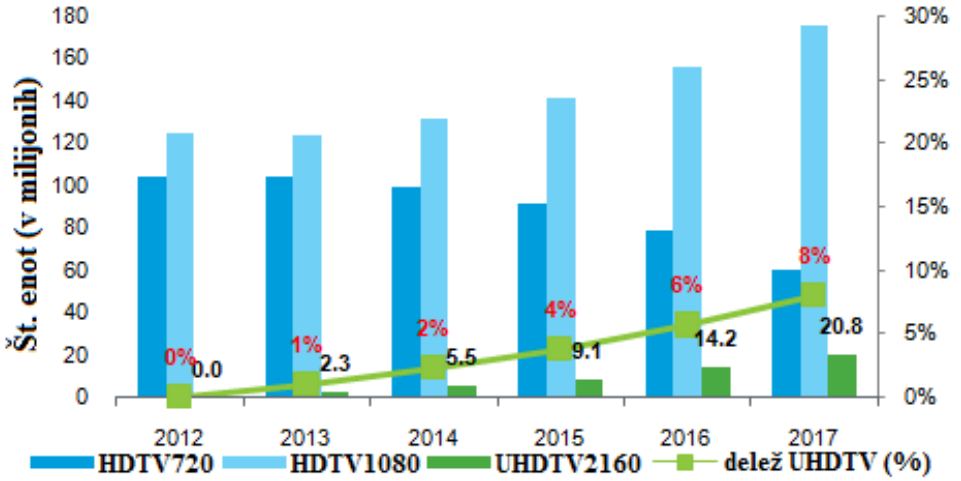


Slika 4.1: Kaskadni transkodirnik.

V preteklosti je bil osrednji razlog transkodiranja obvladovanje omejitev hrambe podatkov in pasovne širine omrežja, še posebej za video [213]. Sprememba bitne hitrosti skrbi za ujemanja statičnih, tj. *konstantne bitne hitrosti* (angl. Constant Bit Rate – CBR), ali dinamičnih omejitev omrežja, tj. *spremenljive bitne hitrosti* (angl. Variable Bit Rate – VBR). Primer takšne uporabe je združevanje več video tokov v en transportni tok [214]. CBR je tako enostavnejši iz vidika sinhronizacije in kontrole pomnilnika (primerno za pretočne vsebine), VBR pa daje boljše rezultate v smislu porabe pasovne širine in prilagodljivosti. Različica, imenovana *aditivni VBR*, kot ga npr. uporablja MP3, določi tarčno bitno hitrost, vendar v primeru kompleksnega avdia le-to poveča. S tem se zagotovi minimalna pasovna širina kodeka. V nasprotju s tem pa je *odštevalni VBR* omejen navzgor, vendar lahko zmanjša pasovno širino za manj kompleksni avdio, kar je primerno za spletne vsebine, saj zagotovimo minimalno kakovost in privarčujemo bitno širino za »enostavni« avdio. Konsistentno kakovost dosežemo s t. i. *kodeki fiksne kakovosti* in uporabo *povprečne bitne hitrosti* (angl. Average Bitrate – ABR). ABR spreminja podatkovno hitrost v odvisnosti od pričakovane kakovosti, kar je bolj primerno za aplikacije, kjer uporabnike bolj zanimata kakovost in velikost avdio datotek kot kontrola pasovne širine.

V omrežjih naslednje generacije (NGN) je potreba po transkodiranju posledica vse večje raznolikosti komunikacijskih tehnologij skupaj s spremembami v načinu uporabe večmodalnih vsebin s strani uporabnikov. Transkodiranje je tako postalo proces adaptacije na sodobna, heterogena komunikacijska okolja, tako na nivoju omrežij kot UE-opreme, UE-aplikacij in storitev, npr. v primeru, ko sprememba formata omogoča tarčnim napravam kompatibilnost s formatom originalnih podatkov. Evolucija multimedijskih storitev je tako pripeljala do povečanega števila zaslonov vse do ultra

visokoresolucijskega formata (UHDTV). Najpogostejša TV-resolucija je *HDTV1080*, pri tem pa se njen delež na tržišču še povečuje (slika 4.2).



Slika 4.2: Predviden delež ekranov tipične resolucije na tržišču (vir: [215]).

Manjše resolucije ostajajo in so primerne za prenosne naprave in prenos po omrežju, standardne prikazuje tabela 4.1. Da se izognemo hrambi več kodiranih različic istega izvornega materiala, se uporabijo transkoderji, ki prilagodijo material visoke kakovosti/resolucije na specifične terminalske zahteve, npr. HDTV na mobilni video [216].

Tabela 4.1: Video formati.

Format/ akronim	**x[slikovne pike]**	**y [vrstice]**	**Akronim (razmerje stranic)**	**x[slikovne pike]**	**y [vrstice]**
SQCIF	128	96	VGA (4:3)	640	480
QCIF	176	144	SVGA (4:3)	800	600
SCIF	256	192	WSVGA (~17:10)	1024	600
SIF(525)	352	240	XGA (4:3)	1024	768
CIF/SIF(625)	352	288	XGA+ (4:3)	1152	864
DCIF	528	384	WXGA (16:9)	1280	720
4SIF(525)	704	480	WXGA (5:3)	1280	768
4CIF/4SIF(625)	704	576	WXGA (16:10)	1280	800
16CIF	1408	1152	SXGA (5:4)	1280	1024
SDTV NTSC[*1]	720[*2]	525/483[*3]	HD (~16:9)	1360	768
SDTV PAL[*1]	720[*2]	625/576[*3]	SXGA+ (4:3)	1400	1050
HDTV 720p	1280	720	WXGA+ (16:10)	1440	900
HDTV 1080i[*1]	1920	1080	FHD (16:9)	1920	1080
HDTV 1080p	1920	1080	WUXGA (16:10)	1920	1200
UHDTV 2160p	3840	2160	QWXGA (16:9)	2048	1152
UHDTV 4320p	7680	4320	WQHD (16:9)	2560	1440
			WQXGA (16:10)	2560	1600

**1 Prepleten video.*
**2 Horizontalna resolucija 720 slikovnih pik je najpogostejša v digitalnih sistemih, vendar ni fiksno določena. Pri konfiguraciji 4:3 so vidne 704 slikovne pike . SDTV uporablja način skeniranja s prepletanjem.*
**3 Dejanske vrstice/ vidne vrstice.*

Naslednja prednost uporabe transkodiranja je na omrežnem nivoju s prilagoditvijo mehanizmov za odpornost na napake z namenom povečanja robustnosti na komunikacijski kanal z drugačnimi lastnostmi, kot je bilo prvotno mišljeno. Primer tega je kvalitetna optična povezava, kjer je zadnji nivo prenosa občutljiva brezžična povezava, ki zahteva višjo stopnjo odpornosti na napake.

Prav zaradi karakteristik prenosnega medija se v praksi pri transkodiranju uporablja *kompresija*, povečini z namenom zmanjševanja pasovne širine. Glede na zmožnost reproduciranja izvornega signala v transkodiranem toku ločimo *neizgubne* in *izgubne* standarde. Prvi so v teoriji informacij omejeni s Shannonovim teoremom, kjer je optimalna dolžina kod določena z $-log_b P$, pri čemer je b število simbolov, iz katerih so sestavljeni izhodni kodi, in P verjetnost vhodnega simbola. Tak način je pogostejši pri kodiranju zvoka, manj pa za video kodirnike, ki brezizgubno kompresijo uporabljajo samo za entropijsko kodiranje sintaksičnih elementov in vrednosti transformacijskih koeficientov, ne pa signala. Za H.264 se za to uporablja *kontekstnoadaptivno binarno aritmetično kodiranje* (angl. Context-Adaptive Binary Arithmetic Coding – CABAC), *kontekstnoadaptivno dolžinskovariabilno kodiranje* (angl. Context-Adaptive Variable-

Length Coding – CAVLC) in druge strukturne tehnike kodiranja, npr. eksponentno Golombovo kodiranje.

Izguba vizualne informacije je lahko zelo velika ob isti ali podobni kakovosti, saj je količina psihomodalne redundance velika. Takšen primer je prostorsko-časovna vizualna redundanca HD-videa [217].

Različne A- in V-formate pokrivata organizaciji ITU in ISO, med njimi so trenutno najpogosteje zastopani predstavljeni v tabeli 4.2 in tabeli 4.3. Krovna skupina ITU se bolj posveča TK-storitvam, npr. video telefoniji (H.26x), ISO pa fokus usmerja na potrošnika (JPEG za slikovne in MPEG za video vsebine). Njuni standardi so namenjeni specifični uporabi, npr. standard za video, standard za slike, standard za zvok itd. Zaradi podobnosti obravnavane tematike, so bile ustanovljene skupine strokovnjakov iz obeh organizacij, kot npr. iniciativa *združene video skupine* (angl. Joint Video Team – JVT), kjer si skupen standard lasti doprinos obeh organizacij, v tem primeru standard *H.264/MPEG-4 AVC*.

Tabela 4.2: Video kodirni in kompresijski standardi.

Oznaka standarda	Leto izdaje	Bitna hitrost	Format	Barvni model	Izguben format	Uporaba
MPEG-1 del 2	1993	< 100Mbit/s	Do 4095×4095	4:2:0	DA	DVD, VCD, CD-ROM
MPEG-2 del 2/ H.262	1996	Omejeno s profilom in nivojem kodirnika, priporočeno: 2 Mbit/s do 30 Mbit/s [218]	Brez omejitev	4:2:0 4:2:2	DA	DVD Digitalna TV HDTV, DVB-T
MPEG-4	1999	Priporočeno: < 64 kbit/s za mobilne in < 2 Mbit/s za TV [219]	Omejeno s profilom in nivojem kodirnika, do 4096×2304	4:2:0 4:2:2 4:4:4	DA	Mobilni internet fiksni internet HDTV okolja Flash player blue-ray video v realnem času
H.261	1988	40kbit/s do 1.92Mbit/s [220]	QCIF	4:2:0	DA	Videotelefonija, videokonferenca, v sistemih ISDN
H.263	1996	Ni omejitve, VBR	Sub-QCIF QCIF CIF 4CIF 16CIF	4:2:0	DA	Internetni video, videokonferenca, telemedicina
MPEG-4 del 10/H.264	2003	Omejeno s profilom in nivojem kodirnika (do 960 Mbit/s za profil »High 4:4:4« in nivo 5.2)	Omejeno s profilom in nivojem kodirnika, do 4096×2304	4:2:0 4:2:2 4:4:4	DA	Mobilni internet, fiksni internet, HDTV okolja, Flash player, Blu-ray, video v realnem času

Avditorna informacija je dodana s postopkom multipleksiranja v skupen multimedijski tok. Kako so posamezni podatkovni elementi, tj. video, zvok, podnapisi, slike, metapodatki, združeni in predstavljeni, je zapisano v multimedijskem vsebniku (angl. multimedia container). Čeprav formalno format vsebnika opredeljuje samo ovoj digitalnih tokov, novejši modularni dizajn z namenom kompatibilnosti določa še kodirnik/dekodirnik (kodek), npr. PNG za slike, ogg-vorbis za zvok. Multimedijski vsebnik tako skrbi za prilagoditev na način, da ustreza sistemu prenosa ter sistemu obdelave. Omrežja enostavnih izvedb z naročniškimi linijami, npr. ISDN, zato

zahtevajo strogo CBR-shemo in nizko kompleksnost obdelave, tista z dragoceno in omejeno pasovno širino (GSM) pa omejujejo prenosno količino podatkov (kljub degradaciji prenosnega kanala) z uporabo kodekov, ki izkoriščajo psiho-avditorno redundanco. Tako zaradi lastnosti linearnosti, frekvenčnega razpona in frekvenčne diskriminacije HAS ter maskiranja zvočnih signalov dosegamo majhen ali transparenten učinek na kakovost zvočnega signala. Dobro uveljavljene avdio formate prikazuje tabela 4.3.

Tabela 4.3: Avdio kodirni in kompresijski formati.

Oznaka standarda	Leto prve izdaje	Vhodna hitrost vzorčenja [Hz]	Izhodna bitna hitrost [kbit/s]	Organizacija	Kompresija / izgubnost	Uporaba
MPEG-1/ MPEG-2 avdio nivo I (MP1)	1993	32k, 44,1k, 48k (MPEG-1) 16k, 22,05k, 24k (MPEG-2)	32 – 448	ISO/ IEC	DA/ DA	Medijski predvajalniki DCC
MPEG-1/ MPEG-2 avdio nivo II (MP2)	1993	32k, 44,1k, 48k (MPEG-1) 16k, 22,05k, 24k (MPEG-2)	32 – 384	ISO/ IEC	DA/ DA	Digitalna avdio radiofuzija, internetni radio zvok v DVD, VCD, SVCD
MPEG-1/ MPEG-2 avdio nivo III (MP3)	1995	32k, 44,1k, 48k (MPEG-1) 16k, 22,05k, 24k (MPEG-2)	32 – 320 (MPEG-1) 8 – 160 (MPEG-2)	ISO/ IEC	DA/ DA	Internetne vsebine, digitalni prenosni avdio predvajalniki
WMA	1999	Do 48k	64 do 192	Microsoft	DA/ DA	Medijske vsebine na PC
MPEG-2/ MPEG-4 Audio (AAC)	1997	8k do 96k	do 96 za 5.1	ISO/ IEC	DA/ DA	Avdio v MPEG multimedijskih vsebnikih
Dolby Digital (AC3)	1987	do 96k	32k do 640k	Dolby Laboratories	DA/ DA	Digitalna TV, DVD-video, igralne konzole, internetne vsebine, DVB, Blu-ray
Ogg-Vorbis	2000	8k do 192k	45 do 500 (pri 44,1 kHz)	Xiph.org	DA/ DA	Internetni radio, video igre, prenos avdia preko spleta
FLAC	2001	1 do 655,350	Odvisno od vhodnega signala	Xiph.org	DA/ NE	Predvajalniki vsebin visoke kakovosti

AMR-NB	1998	8k filtrirano v govorni spekter: 200 - 3400	4,75 FR/HR 5,15 FR/HR 5,90 FR/HR 6,70 FR/HR 7,40 FR/HR 7,95 FR/HR 10,20 FR 12,20 FR	3GPP	DA/ DA	Prenos govora v mobilnih omrežjih (GSM/UMTS)
AMR-WB/ G.722.2	2002	16k filtrirano na frekvenčni pas: 50 - 7000	6,60 8,85 12,65 14,25 15,85 18,25 19,85 23,05 23,85	3GPP/ ITU	DA/ DA	Širokopasovna telefonija v GSM- in 3G-omrežjih, medijski strežniki, govorni predal
G.711/ PCM	1988	8k	64	ITU	DA/ DA	Digitalna telefonija v PSTN in ISDN omrežjih, VoIP
G.722	1988	16k	48 56 64	ITU	DA/ DA	VoIP Radio preko kanala ISDN-B
G.723.1	1996	8k	5,3 6,3	ITU	DA/ DA	VoIP
G.729A	1996	8k	8[1]	ITU	DA/ DA	VoIP, konferenčni klici v internetnih omrežjih

[1] Kasnejši aneksi so omogočali dodatne funkcije in hitrosti, npr. 6.4kbit/s, 11.8kbit/s.

Kodeki za zvok glede na ciljno uporabo sodijo v eno od dveh skupin:

- splošni kodeki,
- govorni kodeki.

Slednji so prirejeni govoru in tipično vsebujejo model s filtrom časovno spremenljive sinteze, ki ustreza modelu vokalnega trakta ter upošteva akustične lastnosti človeških zvokovno reprodukcijskih (vibracija glasilk, vibracija ustnic itd.) in slušnih organov (frekvenčni in amplitudni obseg zvočnega signala, kot ga sprejema uho). Zato ti načeloma slabše reproducirajo negovorno vsebino, vendar je po drugi strani potrebna pasovna širina precej manjša v primerjavi s splošnimi kodeki, kar izhaja iz omenjenih predpostavk in tipa uporabe, npr. enokanalni VoIP, GSM, DECT. Na drugi strani pa so splošni kodeki zasnovani za glasbo, zvočne efekte, govor in ostale zvoke. Ker je uporaba vezana pretežno na naprave z večjo zmogljivostjo in pričakovano višjo kakovostjo, so večji tudi kompleksnost, število avdio kanalov, frekvenčni razpon ter posledično pasovna širina. Stereo kanal tako potrebuje približno 20 do 50 % večjo pasovno širino glede na mono, odvisno od separacije med kanaloma: večja kot je separacija, več bitov je potrebnih za dosego iste zaznavne kakovosti [221].

Kodeki za zvok temeljijo na standardiziranih kodirnih/kompresijskih algoritmih (tabela 4.5). Kodirni algoritmi določajo, kako je zvočna informacija shranjena na podlagi atributov zvočnega signala: bitne širine, pasovne širine signala, zakasnitve, števila kanalov, kompleksnosti in zahtev shranjevanja (velikost medpomnilnika), predvidene ocene kakovosti čistega signala in senzitivnosti degradacije v signalu. Kodeki kot ADPCM kljub dvakratni kompresiji originala ohranjajo subMOS na visokem nivoju (tabela 4.4). Po drugi strani pa CELP ne ohranja takšne kakovosti, vendar na račun velikega kompresijskega razmerja.

Tabela 4.4: Kakovost govora, kompresiranega z ADPCM in CELP (vir: [222]).

kompresija signala	**bitna hitrost [kbps]**	**razmerje kompresije**	**povprečna ocena MOS**
original	64	1,0	4,3
ADPCM	32	2,0	4,1
CELP	4,8	6,5	3,2

Tabela 4.5: Kodirni algoritmi za zvok.

Ime algoritma	Tip	Mehanizem	Variacije	Vidiki kakovosti
CELP (angl. Code-Exited Linear Prediction)	Govor / zvok	Nabor pulzov določa kodno-definirano *linearno predikcijo* (LP), kjer se uporablja fiksna in/ali adaptivna kodna tabela preko izračuna in kvantizacije linearnih predikcijskih koeficientov (LPC). Uporablja kvantizacijo vektorjev.	*Algebraični CELP* (angl. Algebraic CELP). Kodni nabori imajo specifično algebraično strukturo. *Sproščen CELP* (angl. Relaxed CELP). Uporabi samo približno predikcijo (1 parameter za višino na okvir), ki opisuje silhueto govornega signala. *CELP z majhno zakasnitvijo* (angl. low-delay CELP). Uporablja povratno adaptacijo prediktorjev z linearnim filtrom 50. reda (G.728). Zakasnitev je 5 vzorcev, tj. 0,625ms. *CELP z vektorsko vsoto* (angl. Vector Sum Excited Linear Prediction). Implementira hitro in efektivno proceduro iskanja kodnih vektorjev pri nizki kodirni hitrosti (GSM Half Rate).	Zelo visoko razmerje med izvorno kakovostjo in pasovno širino. Tip in umestitev degradacije nelinearno vplivata na izhodno kakovost. Visoka kompleksnost originalne implementacije (izdelava kodne knjige), vendar z učinkovitimi implementacijami možna uporaba tudi v vgrajenih napravah (mobilni telefoni). Je standard za MPEG-4 Audio.
PCM (angl. Pulse-Code Modulation)	Govor / zvok	Zelo nizko-kompleksna digitalizacija analognih signalov s diskretnimi pulzi. Iz omrežnega vidika predstavlja zelo potratno porabo omrežnih virov.	*Linearni PCM* (angl. Linear PCM). Predstavlja vzorce na linearni skali (WAV, Audio CD, DVD). *μ/A-zakon*. Standard ITU G.711, ki definira pretvorbo 13/14-bitnih linearnih vzorcev na 8-bitno logaritemsko skalo. *Adaptivni diferencialni PCM* (angl. Adaptive Differential PCM). Implementira variabilno velikost kvantizacijskega koraka ter diferenčno pasovno kodiranje, npr. kompresija 16-	Visoka izhodna kakovost izvora s pomanjkljivostjo potratne porabe pasovne širine. Vpliv degradacij je linearen glede na tip izgubljenih paketov.

			bitnega vzorca v 4-bitnega.	
HVXC (angl. Harmonic Vector Exication Coding)	Govor	Uporablja parametrično govorno kodiranje. Standardiziran v MPEG-4 Part 3 (MPEG-4 Audio) in namenjen zelo nizkim podatkovnim hitrostim, npr. 1.2 kbit/s.	Originalen algoritem, specificiran v MPEG-4 Audio. *HVXC extended*, definiran v MPEG-4 Audio Version 2.	Namenjen kodiranju pri zelo nizkih pasovnih širinah, tj. 2 in 4 kbps, z dobro izhodno kakovostjo. Namenjen mobilni, satelitski komunikaciji in paketiziranim medijskim vsebinam. Kompleksnost je podobna CELP.
AAC (angl. Advanced Audio Coding)	Zvok	MPEG standard MPEG-2 Part 7 in kasneje MPEG-4 Part 3. Že v osnovi uporablja širokopasovno kodirno shemo s strategijo izločanja navidezno nepomembnih izvornih signalov in redundantnosti v kodiranem signalu.	AAC uporablja modularni pristop, kjer so profili algoritma odvisni od dovoljene kompleksnosti bitnega toka, želenih performanc in sprejemljive kakovosti na izhodu. Popularnejši v okviru uporabe v MPEG-vsebniku so: *AAC-LC* (angl. Low Complexity AAC), *AAC-HE* (angl. High Efficienty AAC), *AAC-SSR* (angl. AAC with Scalable Sampling Rate) in *AAC-LD* (angl. AAC with Low Delay).	Primeren za internetne AV-vsebine z razmeroma visoko pasovno širino, tipično 128 kbps, 256 kbps ali 320 kbps, polnim obsegom frekvenc (44,1 ali 48 kHz) ter več-kanalnim kodiranjem. Kompleksnost AAC je visoka (HE-AAC) ali zmerna (LC-AAC), sorazmerna je tudi kakovost (zmerno dobra/zelo dobra).
HILN (angl. Harmonic and Individual Lines Plus Noise)	Govor / zvok	Vhodni signal je razdeljen na komponente, za katere so dobljeni parametri modela za oceno vhodnih podatkov in združeni v hibridni parametrični model, tj. *individualna*	/	Potencialne degradacije, povezane z *izvornimi modeli* (slaba dekompozicija, slaba estimacija parametrov), *percepcijskimi modeli*

		sinusoida, definirana kot funkcija frekvence in amplitude, *harmonični ton* kot funkcija frekvence, amplitude in spektralne ovojnice njenih delov, in *šum* kot funkcija amplitude in spektralne ovojnice. Parametri komponent so kvantizirani, kodirani, multipleksirani v končni podatkovni tok. Je tip parametričnega kodiranja standardiziran v drugi verziji MPEG-4 standarda.		(kvantizacija, pomanjkanje fazne informacije, izbira relevantnosti posameznih komponent). Govorni model neprimeren za glasbo zaradi drugačne sestave spektralne ovojnice (pop glasba-premočna komponenta za *šum*, orkestralna glasba – premočna komponenta za *harmonično mrežo*).
Vorbis	Zvok	Naprejadaptiven monolitni kodek, ki temelji na *modificirani diskretni kosinusni transformaciji* (angl. Modified Discrete Cosine Transform – MDCT). Rezultat transformiranih podatkov v frekvenčni domeni je razdeljen na *nivo šuma* (angl. noise floor) in ostale komponente, ki so kvantizirane in entropijsko kodirane (kodna knjiga) z vektorsko kvantizacijskim algoritmom.	Zaradi odprto-kodne licence množična popularnost predvsem v domeni interneta (podpora v HTML5) in PC-aplikacij (VLC, MPlayer).	Pred-odmev in reverberacija zaradi metode pri določanju nivoja šuma v izvornem signalu. Dobra robustnost na omrežne napake ter zelo dobra izhodna kakovost zaradi prilagodljive pasovne širine (31 kbps – 500 kbps).

4.1.2. Degradacije transkodiranja slik in videa

Surovi format videa zahteva veliko dragocene pasovne širine, zato je transkodiranje nujna operacija pri prenosu le-tega skozi omrežje, še posebej za vsebine visoke ločljivosti (tabela 4.6). Cilj kodiranja videa je dobiti najboljšo kakovost slike ob najmanjši zaznani degradaciji. Transparentnost H.264-kodiranega videa, definirano s prostorsko in časovno resolucijo video materiala, kjer človeško oko ne razločuje degradacij, so avtorji dosegli že s kompresijo vira ~ 1 : 400 (tabela 4.7). Video lahko obravnavamo kot množico slik v časovnem sosledju in tako ga obravnavamo tudi pri evalvaciji kakovosti. To imenujemo evalvacija *prostorske degradacije*. Ker transkodiranje uvede tudi *časovno* degradacijo, je potreben pristop združenja ocen posameznih slik v skupno oceno kakovosti videa [223] (angl. temporal pooling).

Tabela 4.6: Surovi video format.

Velikost okvirja [slikovne pike]	**Barvni model**	**Barvna globina [bitov/ barvo]**	**Pre-pletenost videa**	**Podatk. velikost slike [Mbit]**	**Hitrost okvirjev [fps]**	**Hitrost videa [Mbit/s]**
1920×1080	YUV 4:2:0	8	ne	49,77	30	746,50
1920×1080	YUV 4:2:2	8	ne	49,77	30	995,33
1920×1080	YUV 4:4:4	8	ne	49,77	30	1492,99
1920×1080	YUV 4:2:0	8	da	49,77	30	373,25
1920×1080	YUV 4:2:2	8	da	49,77	30	497,66
1920×1080	YUV 4:4:4	8	da	49,77	30	746,50

Tabela 4.7: Razmerje kompresije izgubnega videa H.264 pri vizualno transparentnem kodiranju.

	[224]		[225]	
	H.264/AVC	**Surov video**	**H.264/AVC FRExt**	**Surov video**
Velikost okvirja [slikovne pike]	768×432	768×432	1920×1080	1920×1080
Hitrost okvirjev [fps]	25p	25p	23,98p	23,98p
Pasovna širina [Mbit/s]	0.5 – 1.46	199	16	597
Razmerje kompresije	**1 : (136 – 399)**		**1 : 37**	

Tipični potek od surove do kompresirane oblike podatkov je:

- **predprocesiranje slike:** sem sodijo particioniranje, segmentiranje slikovnih pik (npr. bloki 8×8 pri DCT), filtriranje (pod-, nad-vzorčenje) in ostale operacije,
- **kompresija:** je dejanska izgubna operacija pretvorbe izvornega signala v drugi prostor (frekvenčni, valjčni, itd.) na podlagi določenih pravil translacije,
- **kvantizacija:** dodatna izgubna operacija, ki kvantizira transformacijske elemente na podlagi določenih parametrov,
- **post-procesiranje:** določa način pretvorbe podatkov v kodirano obliko.

Čeprav bi bilo to najnaravnejše s stališča človeške percepcije, se kompresija slik ne izvaja v prostorski domeni. Pomanjkljivosti prostorske metode, tj. velika redundantnost in majhna zgoščenost informacij, dajeta prednost metodam v drugih domenah. Pogoste operacije transformacije vizualnih informacij temeljijo na:

- **Fourierjevi transformaciji,**
- **kosinusni transformaciji** ali
- **valjčni transformaciji.**

Uporaba teh metod temelji na dejstvu, da je mogoče vsak vizualni signal predstaviti kot vsoto osnovnih operacij oz. funkcij (recimo kosinusnih pri DCT). Pri tem velja, da je aproksimacija lažja za gladke signale, tj. za te je potrebno manjše število osnovnih

funkcij, kar je osnovni princip kompresije takšnega tipa. Ker se s povečevanjem kompresijskega razmerja sčasoma začne vidno degradirati tudi rekonstruirana slika, transformacije prinašajo določeno stopnjo vizualnih napak.

4.1.3. Fourierjeva transformacija

Fourierjeva transformacija je linearna operacija pretvorbe signala v Fourierjevo frekvenčno domeno, tako, da je ta predstavljen z neskončno množico sinusnih in kosinusnih funkcij. Pri tem je definicijsko območje neskončno število sinusoid iz množice kompleksnih števil. V digitalnem svetu so te funkcije predstavljene diskretno, kar pomeni, da imajo časovne in frekvenčne spremenljivke končno vrednost, in to imenujemo *diskretna Fourierjeva transformacija* (angl. Discrete Fourier Transformation – DFT). Če je *f(m,n)* diskretna funkcija, ki opisuje sliko in ima neničelno vrednost za dano, končno območje $0 \leq m \leq M\text{-}1$, $0 \leq n \leq N\text{-}1$ z enakomernim prostorskim vzorčenjem, potem so koeficienti DFT definirani kot:

$$\boldsymbol{DFT_{pq}} = \sum_{\boldsymbol{m=0}}^{\boldsymbol{M-1}} \sum_{\boldsymbol{n=0}}^{\boldsymbol{N-1}} \boldsymbol{f(m,n)} \boldsymbol{e}^{\frac{\boldsymbol{-j2\pi pm}}{\boldsymbol{M}}} \boldsymbol{e}^{\frac{\boldsymbol{-j2\pi qn}}{\boldsymbol{N}}} \qquad (4.14)$$

kjer sta *M* in *N* horizontalna in vertikalna dimenzija DFT in $0 \leq p \leq M\text{-}1$, $0 \leq q \leq N\text{-}1$.

Velikost v Fourierjevi domeni je enaka velikosti v prostorski domeni (slikovne pike) in sega od *–(M/2-1)* do *(M/2 -1)* ter od *–(N/2-1)* do *(N/2 -1)*. Kompleksnost DFT je $O(n^2)$, vendar lahko število izračunov zmanjšamo z uporabo algoritmov *hitre Fourierjeve transformacije* (angl. Fast Fourier Transformation – FFT). Algoritmov FFT je več, osrednji problem pa predstavlja dekompozicija celotne DFT v množico manjših blokov (tipično velikosti korena 2), na katerih lahko izvedemo matematične operacije s kompleksnostjo *O(n)*. Kompleksnost FFT je tako $O(n \cdot \log n)$.

4.1.4. Kosinusna transformacija

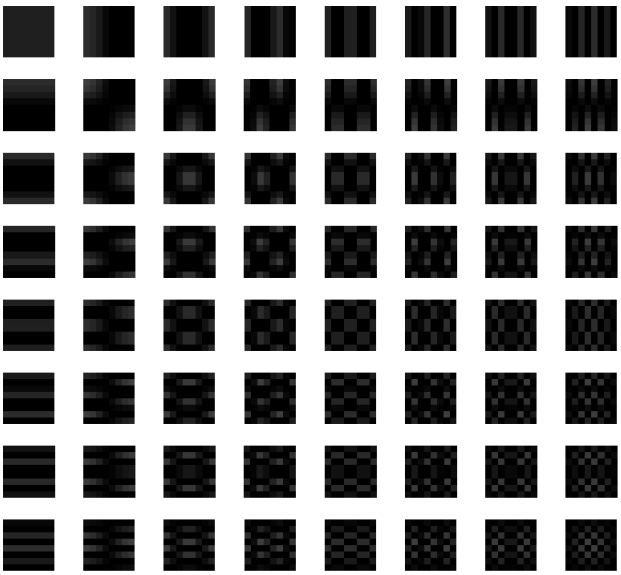
Diskretna kosinusna transformacija (angl. Discrete Cosine Transformation – DCT) je primer FT, le da uporablja zgolj realna števila in kosinusne komponente. DCT je

pogosta in priljubljena metoda pri kompresiji slik, saj je računsko manj zahtevna, jo je relativno enostavno implementirati in ima boljšo energijsko kompaktnost, tj. več informacije je vsebovane v začetnih DCT-koeficientih kot pri DFT [226]. Različne frekvence dekompozicije dajejo različno magnitudo k pomembnosti vizualne informacije, zato se pri postopku obdelave podatkov višje frekvenčne komponente izločijo, saj nosijo »manj informacije« za razliko od enosmernega koeficienta, imenovanega tudi *DC-komponenta* (angl. Direct Current – DC). V tem pogledu izločanje komponent DFT določa stopnjo izgube, kar se s pridom izkorišča pri kompresiji s standardi JPEG, MPEG in drugih [227]. V teoriji slik se uporablja tehnika *večdimenzionalne DCT-II*, katere koeficienti so definirani kot:

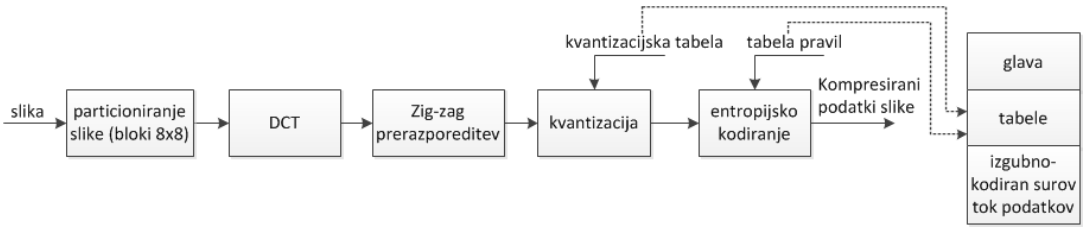
$$DCT_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N} \quad (4.15)$$

pri tem je A matrika, tj. slika velikosti $M \times N$ ter $0 \leq p \leq M-1$, $0 \leq q \leq N-1$.

Tipično se operacija izvede na bloku slikovnih pik velikosti 8 × 8, tj. $m_1 = 8$ in $n_2 = 8$, kjer »pomembnost« informacije (komponente nižjega reda imajo večjo »pomembnost«) pada od leve proti desni in od zgoraj navzdol (zig-zag). Matematično torej blok predstavlja utežno funkcijo. Primer vertikalno-horizontalne dekompozicije kaže slika 4.3, postopek kodiranja DCT pa slika 4.4.



Slika 4.3: Dekompozicija DCT na bloku velikosti $m_1 = 8$ in $n_2 = 8$.



Slika 4.4: Blokovna shema DCT-kodirnika.

4.1.5. Valjčna transformacija

Tudi valjčna transformacija sodi v tip frekvenčnih transformacij, katere prednost je, da lahko signale analiziramo pri različno skaliranih resolucijah [228]. Ta večresolucijska lastnost dobro sovpada s senzitivno funkcijo HVS. Prednost pred Fourierjevimi metodami je tudi pri obdelavi signalov, ki vsebujejo nezveznosti in ostre vrhove, kot je to značilno za večino naravnih slik. *Diskretna valjčna transformacija* (angl. Discrete Wavelet Transformation – DWT) je tip valjčne transformacije, ki za razliko od *kontinuirane valjčne transformacije* (Continuous Wavelet Transform – CWT) diskretizira signal v nabor ortogonalnih komponent, osnovnih funkcij. Osnovna valjčna transformacija, imenovana transformacija Haar, sestavlja nabor nizko- (angl. Low Pass – LP) in visokoprepustnih filtrov (angl. High Pass – HP). LP-filter izvede povprečenje, HP pa določanje razlik, kar lahko izrazimo kot:

$$LP = \frac{1}{\sqrt{2}}(1,1) \tag{4.16}$$

$$HP = \frac{1}{\sqrt{2}}(1,-1) \tag{4.17}$$

Polno transformacijo Haar nato zapišemo kot:

$$DWT_{Haar} = \ W_N A\, W_N^T \tag{4.18}$$

kjer je A 2D-matrika, ki predstavlja sliko, W_N transformacijska matrika vrstic slike (prva operacija) in $W_N{}^T$ transformacija stolpcev iz prve operacije. W_N je definirana kot matrika:

$$WN = \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & & & & \\ & & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & & \\ & & & & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & \\ & & & & & & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \hline -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & & & & \\ & & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & & & \\ & & & & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & & \\ & & & & & & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} \quad (4.19)$$

$$LP / HP =$$

v primeru transformacije slike velikosti *4×4 slikovnih pik*.

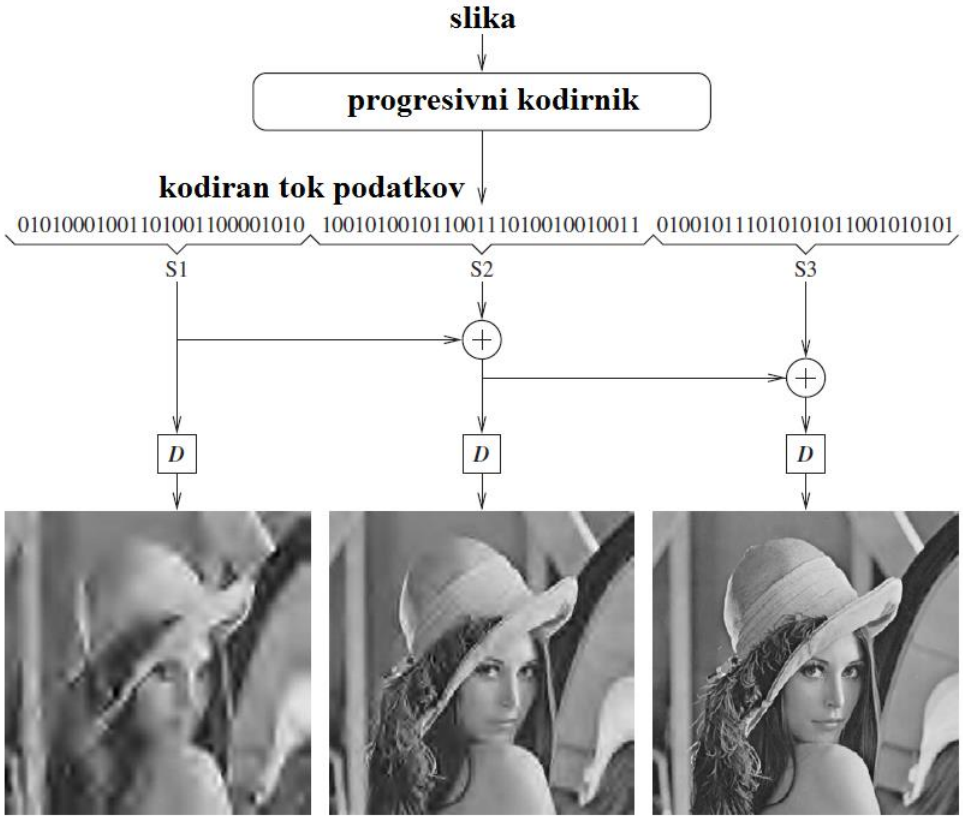
Z operacijama povprečenja in določanja razlik lahko določimo poljubni 2 števili, pri tem pa dejstvo, da sosednje slikovne pike po večini vsebujejo le majhno amplitudno diferenco, omogoča primerno kompresijo, v kolikor majhno razliko enačimo z 0.

Podobna je tudi različica Daubechiesove valjčne transformacije, ki se od Haarove razlikuje v definiciji normiranih operacij in povprečenja, ki zajamejo nekoliko večje število vrednosti signala.

DWT-transformacijo uporablja tudi standard JPEG2000 [229]. Kompresijske napake se razlikujejo od tistih pri DCT (slika 4.5); DWT daje boljšo kompresijsko razmerje pri isti kakovosti slike (večja učinkovitost zaradi optimalnejše bitne alokacije), a ima večjo računsko kompleksnost [230].

Slika 4.5: Kompresijske napake transformacije DCT (levo) in DWT (desno).

DWT je atraktivna zaradi enostavne implementacije prenosne podatkovne sheme, kjer se progresivno s sprejemom podatkov izboljšuje kakovost prejete slike (slika 4.6). Prednost tega je, da je kodiran tok podatkov možno prekiniti na določeni točki, ko je želena kakovost dosežena.



Slika 4.6: Progresivni valjčni kodirnik.

4.1.6. Drugi tipi video transformacij

V preteklosti je bila pogosta uporaba transformacije Walsh-Hadamard. Koeficienti te so samo +1 ali -1, zato je implementacija zelo enostavna, zgolj z operacijo seštevanja in odštevanja. Uporaba sodobnih procesorskih enot z enocikličnim multiplikatorjem je izničila to prednost.

Različne transformacije delujejo na izločanju značilk iz originalne slike. Radonova transformacija predstavlja sliko kot zbirko projekcij slikovne matrike v določeni smeri, kar matematično predstavlja 2D-usmerjeni linijski integral. Pri kompresiji slik se uporablja v kombinaciji z DCT, kjer so Radonove točke, ki predstavljajo linije v prostorski domeni in pomenijo translacijo singularnosti *črta → točka*, kodirane z DCT-koeficienti [231].

Hilbertova dekompozicija služi kot za zamik invariantna valjčna transformacija, pri tem pa računska metoda temelji na FFT, DCT ali podobno [232]. Hilbertov operator kot primer singularnega integralnega operatorja pri harmonični analizi daje analitično predstavitev signala. Zaradi redundantnosti se metode, ki temeljijo na Hilbertovi transformaciji, po večini ne uporabljajo v kompresijskih algoritmih, ampak najdejo mesto pri npr. aplikacijah z vodnim žigom, računalniški tomografiji in podobno [233], [234].

Zanimive so tudi Rieszove transformacije, ki so invariantne na zamik, skaliranje, rotacije in translacije. Transformacije različnega reda učinkovito in efektivno izločijo ključne značilke slike, kot so npr. robovi, faza in orientacija na *ključnih točkah*. Te točke so sidrne pri vizualni zaznavi in pomembnosti teh področij (ROI), kar ji daje prednost pri percepcijskem kodiranju.

4.1.7. Degradacije transkodiranja avdia

Izgubna kompresija govora tipično deluje na psiho-akustičnem modelu HAS. Priljubljene tehnike kompresije izkoriščajo sledeče lastnosti:

- **Maskiranje**: efekt nastane v primeru, ko glasnejši ton zakrije tišjega in poslušalec ni zmožen slišati dveh različnih tonov v določenem časovnem trenutku (*frekvenčno ali spektralno maskiranje*). Če glasnejši ton sledi tišjemu z

majhno latenco, je to *časovno maskiranje*, kar je posledica nasičenosti slušnih receptorjev in je funkcija razlike med testnim in maskirnim signalom (slika 3.14).

- **Določanje praga HAS**: človeško uho je dojemljivo za frekvence med 20Hz in 20 kHz (odvisno od starosti poslušalca), pri tem je prag slišnosti odvisen od frekvence *f* [202]:

$$\boldsymbol{prag(f) = 3,64\left(\frac{f}{1000}\right)^{-0,8} - 6,5e^{-0,6\left(\frac{f}{1000}-3,3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4} \quad \text{[dB]} \tag{4.20}$$

  Pri tem je maksimalen dinamični razpon, tj. razmerje med maksimalno in minimalno zvočno amplitudo, ki jo HAS še zazna, 120 dB.

Izpopolnjenost in posledično kakovost zvočnih posnetkov, kodiranih z omenjenimi avdio kodirnimi algoritmi, je precej boljša, kot to velja za video, iz stališča *zaznane uporabniške izkušnje* v odvisnosti od *sistemskih zmogljivosti*. Degradacije so predvsem posledica omejene pasovne širine sistema, kar ima za posledico:

- **Slabše karakteristike nizko-prepustnega filtra:** izguba pasovne širine nad 16 kHz poslabša »polnost« in »naravnost« zvočnega signala, zato novejši standardi predpisujejo frekvenčni spekter do 20 kHz ali več.

- **Pred- in poodmevi**: nastanejo pri ostrih zvočnih tranzicijah, npr. kjer vsebino vhodnega zvočnega okvirja zakodira linearni predikcijski filter (ACELP). Raztegnjen in mehkejši odziv ima drugačne lastnosti, kar privede do slišnih sprememb.

- **Kvantizacija:** zmanjšanje resolucije privede do pojava izgube globine, okvirjanja (angl clipping), ne-naravnosti glasu in podobno.

- **Detekcija šuma**: sistem detekcije šuma, kot npr. *detekcija aktivnosti govora* (angl. Voice Activity Detection – VAD) določa nivo šuma, tj. nekoristni signal, ki ga sistem kodira z majhno pasovno širino ter dekodira/simulira s *komfortnim*

generatorjem šuma (angl. Comfort Noise Generator – CNG) na sprejemni strani pri digitalnih govornih sistemih, npr. VoIP. Previsoko nastavljen nivo privede do nepravilnega ali prekinjajočega zvoka ozadja.

Glede na omenjeno, transkodiranje govornega signala z različnimi standardnimi kodeki degradira referenčni signal za tudi več kot 1 oceno MOS (tabela 4.8).

Tabela 4.8: Primerjava degradacije transkodiranja govornega signala.

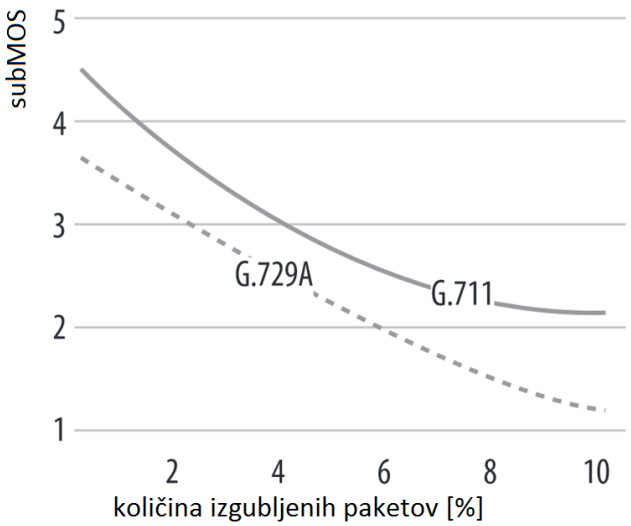
Kodek	**Povprečna vrednost MOS**	**ΔMOS od ref. signala (PCM)**
AMR 4,75 k	2,59	1,00
AMR 5,90 k	2,90	0,69
AMR 6,70 k	2,90	0,69
AMR 7,40 k	3,00	0,59
AMR 7,95 k	3,10	0,49
AMR 10,2 k	3,08	0,51
AMR 12,2 k	3,20	0,39
G.726 32 k	3,35	0,70
G.726 40 k	3,58	0,41
G.729 8 k	3,74	0,31
G.723.1 5,3 k	3,45	0,54
G.723.1 6,3 k	3,46	0,53
G.722 64 k	3,54	0,94
G.722 56 k	3,56	0,92
G.722 48 k	3,33	1,15

4.2. Omrežne degradacije

V omrežjih IP prihaja do degradacij uporabniških podatkov na omrežnem nivoju, ki so posledica vsote naključnih nepravilnosti na poti med storitvijo in uporabnikom. Vzroki nepravilnosti so v stanju pomnilniških vrst omrežnih elementov, različno uteženih poti IP-paketov med omrežnimi vozlišči (večpotni presih), izločanju paketov s predolgo življensko dobo (angl. Time-To-Live – TTL), napak na fizičnem nivoju (WiFi, 3G), omrežnih zastojih in preobremenjenosti, zavrnitvi poškodovanih paketov med tranzitom, napak na omrežni opremi, gonilnikih in pri usmerjanju. Dodatno so lahko dovzetni za napake zaradi uporabljenih protokolov, kot je to v primeru brezpovezavne in nepotrjevane omrežne komunikacije (UDP/RTP). Tipične degradacije, ki jih srečamo v omrežju, so:

- izguba IP-paketov,
- bitne napake,
- latenca,
- potresavanje,
- pomanjkanje prepustnosti omrežja, nepravilno poravnane vrste paketov in
- podvajanje paketov.

Naštete napake vplivajo na omrežni promet na različne nedeterministične načine. Najbolj destruktivna je pri tem izguba IP-paketov (angl. packet loss – PL), ki pomeni popolno izgubo uporabniške informacije. To še posebej velja za časovno senzitivne storitve, kjer ni možnosti ponovnega pošiljanja, npr. pri realnočasovni obojesmerni komunikaciji, kot je VoIP. Na uporabniško ter sistemsko kakovost storitve zraven količine degradacij vpliva tudi distribucija napake ter tip izgubljenega paketa oz. natančneje količina ali tip informacije, ki jo paket nosi [235]. Koliko je degradacija zaznavna, je odvisno tudi od tipa modalnosti, in iz tehnološke perspektive lahko rečemo, da ima enaka amplituda degradacije različne učinke na kakovost storitve. Vpliv na avditorno modalnost je zato precej manjši kot na vizualno modalnost. K temu vpliva biološka tolerantnost na napake. Za avditorno modalnost so zato povprečne sprejemljive vrednosti PL v rangu nekaj % PL, recimo 3 % PL (G.729A) in 6 % PL (G.711) za *MOS = 2,8* (slika 4.7), za vizualno modalnost pa so te vrednosti precej nižje, saj v določenih primerih subjektivna ocena MOS pade pod 2.8 že pri *PL = 0,1 %* [236].



Slika 4.7: Subjektivna ocena MOS za govorna kodeka G.711 in G.729A (vir: [237]).

Naslednja podpoglavja opisujejo tipe degradacij, ki jih opazi uporabnik in so posledica tako vpliva transkodiranja kot tudi omrežnih napak.

4.3. Tipi video degradacij

Degradacija pretočnega videa, kot npr. pri sistemih IPTV, je posledica transkodiranja, omrežnih nepravilnosti in sistemskih omejitev. Tipi degradacij, ki pri tem nastanejo, so:

- **Bločna degradacija (angl. blockiness)**: vidna kot kockast vzorec, ki nastane zaradi neodvisne kvantizacije individualnih blokov slike pri DCT. Degradacija je velikosti večkratnika osnovnega bloka. Ob uporabi barvnega sistema *YUV 4:2:*0 je osnovni blok tipično matrika *8×8 slikovnih pik* za kromatične podatke in 16×16 za luminančne podatke. Degradacija je nezvezni učinek na robovih bloka (slika 4.8). Pri tem je vidnost robov sorazmerna z bitno hitrostjo kodeka. Napaka je bolj zaznavna na zveznih površinah z majhnim gradientnim odstopanjem, tj. majhno razliko kontrastov in na svetlejših površinah [238]. Delna rešitev problema je uporaba filtra bločne degradacije (angl. deblocking filter).

Slika 4.8: Bločna degradacija: izvorni (levo) in močno kvantiziran video *Intervju_narator* s kompresijo JPEG z 0,12 bita/slikovno piko (desno).

- **Zamegljenost (angl. blur):** pomeni izgubo prostorskih podrobnosti in zmanjšanje ostrine robov (slika 4.9). Pojav nastane zaradi izločanja visoko-frekvenčnih koeficientov in uporabe nizko-prepustnega filtra bločne degradacije ali zaradi pomanjkanja višje-spektralnih komponent pri DWT. Pojav zamegljenosti ni toliko viden v domeni DCT, je pa osrednja degradacija pri

uporabi metode DWT. Distribucija takšne zamegljenosti je tipično Gaussova ali semi-Gaussova in jo je v času evalvacije kakovosti potrebno razločevati od izvorne zamegljenosti, ki je posledica napake ostrine. Merilo zamegljenosti je povprečna debelina robov [239].

Slika 4.9: Zamegljenost: izvorni (levo) in degradiran video *Intervju_napovedovalec* (desno), zamegljenost = 10,76 [240].

- **Uhajanje barvne informacije (angl. color bleeding)**: lastnosti barv, tj. gostota oz. intenzivnost, prehajajo med področji z veliko kontrastno razliko (slika 4.10). Nastane z zatiranjem visokofrekvenčnih koeficientov kromatičnih komponent z operacijo *kroma pod-vzorčenja* in *kvantizacije* [241]. V primeru kromatičnega pod-vzorčenja, kot npr. v barvni shemi *4:2:0*, se degradacija razširi po celotnem bloku. Degradacija je bolj vidna v nenaravnih scenah (risani film), kjer je vsebnost sosednjih monotonih, med seboj močno kontrastnih površin, večja [242].

Slika 4.10: Uhajanje barvne informacije: izvorni (levo) in degradiran video (desno).

- **Stopničavost poševnih linij**: DCT-algoritem je naklonjen prikazu horizontalnih in vertikalnih linij (mreža blokov je horizontalno-vertikalno orientirana), pri čemer pa prikaz drugače orientiranih linij zahteva visoko-frekvenčne koeficiente

DCT za natančno rekonstrukcijo. Kvantizacija teh koeficientov tako povzroči efekt stopničavosti (slika 4.11).

Slika 4.11: Stopničavost poševnih linij.

- **Zvonjenje (angl. ringing)**: pojem je fundamentalno povezan z Gibbsovim fenomenom [243] in je viden ob visoko-kontrastnih robovih na sicer enakomerno-kontrastnih površinah (slika 4.12). Je rezultat kvantizacije in posledic nepravilnosti pri rekonstrukciji signala pri DCT in valjčnih operacijah. Zvonjenje se pojavi tako na luminančni kot tudi kromatični komponenti video signala.

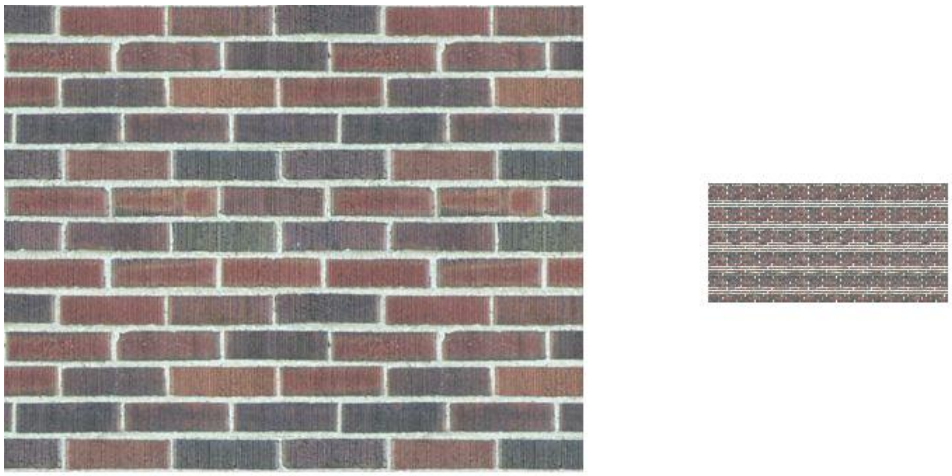
Slika 4.12: Efekt zvonjenja na robovih velikih kontrastov. Na slikah z velikimi in ostrimi kontrastnimi prehodi (levo) se pojavi viden efekt, kjer je degradacija približno srednja vrednost obeh barv v barvnem prostoru (desno).

- **Utripanje (angl. flicker)**: je časovna degradacija, ki nastane kot rezultat spreminjajoče se kvantizacije. Nastane na scenah z večjo stopnjo detajlov. Utripanje je možno odpraviti s tehniko povečanja časovne resolucije, kot npr. prepleten video. Prepletanje zmanjša čas osvežitve okvirjev pri isti pasovni širini, vendar uvede napako »razdeljenih« polokvirjev, kar je še posebej razvidno pri scenah z veliko stopnjo gibanja (slika 4.13).

Slika 4.13: Hitro gibanje v načinu prepletanja. Prizor prehoda pešca v videu *Intervju_Robert*.

- **Mosquito noise**. je časovna degradacija, podobna degradaciji zvonjenja, le da je dodatno vidna v časovnem prostoru. Je posledica razlike kodiranja istih predelov slike v zaporednih okvirjih videa in spominja na prisotnost letečih »komarjev« na sliki.

- **Stopničenje (angl. aliasing)**: označuje efekt popačenja poševnih črt in krivulj v sliki zaradi premajhne ločljivosti zaslona. Opazen je na območjih scene, ki so nad Nyquistovo mejo v časovni ali prostorski domeni, npr. pri zmanjšanju prostorske ali časovne resolucije vzorčnega signala, kjer ponavljajoči se vzorci interferirajo in tvorijo novo frekvenčno ovojnico (slika 4.14). Čeprav je rekonstrukcija takšnih struktur kompleksna, je efekt delno obvladljiv (in popravljiv) z uporabo metode mehčanja robov (angl. anti-aliasing).



Slika 4.14: Stopničenje: zmanjšanje resolucije vzorca (levo) za faktor 12 privede do strukturnih degradacij (desno).

- **Zatikanje (angl. jerkiness)**: se nanaša na motnjo zveznosti gibanja objekta v videu. Nastane zaradi pomanjkanja časovne resolucije ali nezadostne kompenzacije gibalnih vektorjev (angl. motion vectors).

- **Zamegljenost gibanja (angl. motion blur):** področja gibanja so prikazana megleno.

- **Ostalo:** ostale degradacije nastanejo večinoma kot posledica prej omenjenih, npr. izgubljena slika, zamrznjena slika.

Izvor omenjenih degradacij je opisan v tabeli 4.9.

Tabela 4.9: Tipične degradacije video kakovosti ter njihovi izvori.

Degradacija	Izvor				
	kodiranje	prenos	prostorska	časovna	prostorsko-časovna
Bločna degradacija	X	X	X		
Zamegljenost	X		X		
prehajanje barve	X		X		
Stopničavost linij	X	X	X		
Zvonjenje	X		X		
Utripanje	X			X	
mosquito noise					X
Stopničenje			X		
Zatikanje		X		X	
Zamegljenost gibanja				X	
Izguba slike		X			X
Zamrznjena slika		X		X	

4.4. Tipi avdio degradacij

Posledica izvornih in omrežnih degradacij v avdiu so slišne napake, ki jih zaznavamo kot distorzije avdio signala, izgubo frekvenc, odmev, pomanjkanje zvoka, nepravilno poravnavo stereo zvoka in podobno [244], [245]. Tipične degradacije, ki so posledica transkodiranja avdio signala so:

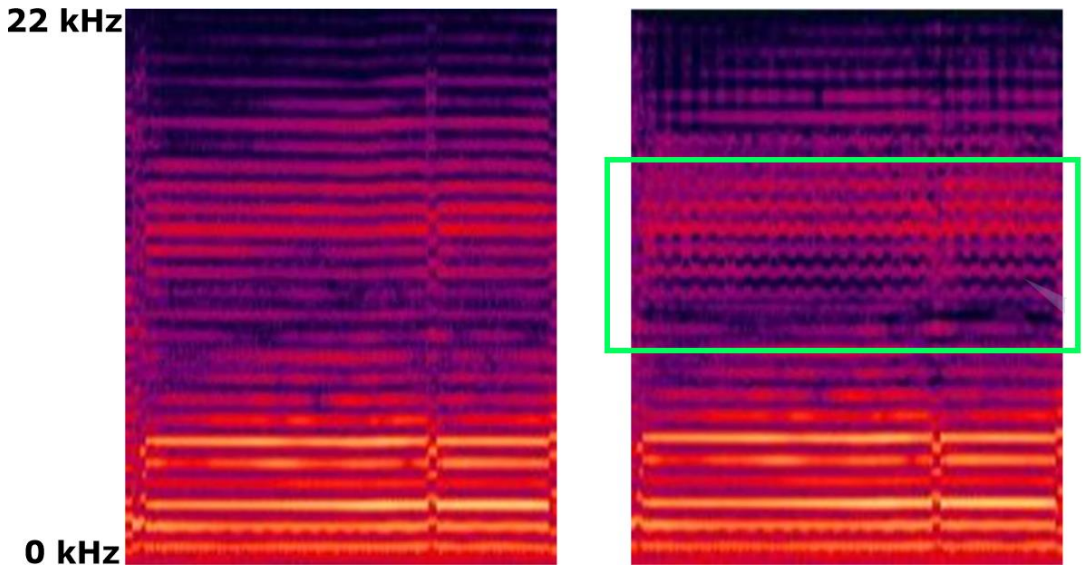
- **Šum:** najpogostejša sta beli in koloriziran šum. Posledica prvega so napake na avdiu signalu, ki se pojavljajo enakomerno razpršene po celotnem frekvenčnem pasu signala, vzrok drugega pa je kvantizacija bank filtrov avdio signala, kjer pa spekter šuma sledi silhueti spektra signala.

- **Stopničenje in kvantizacijska nelinearnost**: napake aliasinga so posledica nepravilnosti pri rekonstrukciji signala iz nabora filtrov (angl. filter banks), zaradi sprememb kvantizacije med frekvenčnimi pasovi in časovnimi okvirji. Ker mora kvantizacija zajeti čim širši razpon frekvenc in amplitud zvoka z omejeno širino, prihaja do nelinearnosti kvantizacije, še posebej v mejnih območjih dinamičnega razpona.

- **Izguba visokih frekvenc in omejitve pasovne širine**: nastanejo kot posledica uporabe pasovno-prepustnih filtrov in krčenja signala. Rekonstrukcija visokih frekvenc ni nikoli čisto enaka originalu, lahko pa se umetno obnovi, npr. z metodo ekstrapolacije.

- **Predodmev:** nastane na prehodu avdio signalov in ga je možno zaznati, če so uporabljena dolga časovna okna okvirjev. Ta so sicer zaželena zaradi višje stopnje kodirne učinkovitosti, vendar če se avdio degradacija pojavi znotraj takšnega okvirja, se ta ob dekodiranju razširi po celotnem okvirju.

- **Tonsko drhtenje** (angl. tone trembling): pojav ponavljajočih se frekvenc v višjih frekvenčnih pasovih (slika 4.15).

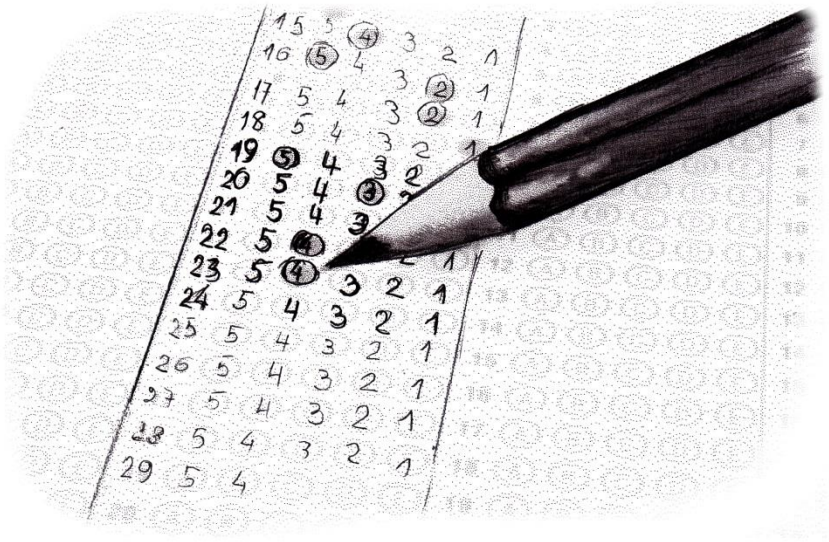


Slika 4.15: Tonsko drhtenje: na levi je spektrogram izvornega signala, na desni pa spektrogram istega signala, ki vsebuje tonsko drhtenje [246].

- **Sprememba sterea**: v primeru, da se izgubi diferenčni signal med avdio signaloma, poslušalec dobi občutek eno-kanalnega avdia.

Pri prenosu skozi omrežja IP izguba omrežnih paketov, v katerih je avdio informacija, običajno pomeni pomanjkanje slišne informacije za dolžino avdio okna, tipično med 20 in 30 ms. Moč slišne degradacije je odvisna od stopnje odvisnosti med različnimi avdio okvirji. Metoda *modificirane diskretne kosinusne transformacije* (angl.

Modified Discrete Cosine Transform – MDCT), ki jo uporablja kodek AAC, tako uporablja prekrivanje okvirjev, kar daje večjo robustnost v primeru izgube avdio okvirja. Izgubljeni paketi in posledično izgubljeni avdio okvirji se zakrivajo s tehnikami *zakrivanja izgubljenih paketov* (angl. Packet Loss Concealment – PLC). Najpogosteje se uporablja metoda ponovitve istega okvirja, kjer se pravilno sprejet okvir kopira na izgubljeno mesto. Z drsečim oknom dekodirnika se konci signala v kopiranem okvirju nato zlijejo s silhueto avdio signala, kar daje majhno slišno napako. Za krajše izgube se okvirji lahko preskočijo, ampak s tem prekinemo sinhronizacijo v primeru večmodalnih vsebin. Za daljše, rafalne izgube okvirjev se tipično vstavi tišina ali pa dekodirnik generira zvok ozadja (angl. Comfort Noise Generation). Naprednejše metode zakrivanja izgubljenih paketov manjkajoč signal rekonstruirajo z interpolacijo signala iz informacije prejšnjih in naslednjih pravilno sprejetih okvirjev.



5. Merjenje kakovosti storitve s subjektivnimi metodami

Subjektivno ocenjevanje predstavlja najnatančnejšo in zanesljivo metodo določanja kakovosti storitve, saj daje ocenjevalcu informacijo o tem, kako končni uporabnik storitev dejansko dojema in zaznava. Govorimo o zaznani kakovosti storitve. Ocenjevanje zajema vrednotenje, primerjavo ali oceno testnih posnetkov, ki so podobni tistim v obravnavanem sistemu oz. storitvi. Ločimo dva razreda subjektivnega ocenjevanja, tj. *ocenjevanje uspešnosti/zmožnosti sistema v idealnih pogojih* in *ocenjevanje kakovosti sistema v degradiranih pogojih*. Slednji so primerni za evalvacijo delovanja sistemov v času distribucije storitve, ko prihaja do neoptimalnih pogojev, ki se nanašajo na prenos ali emisijo. Primer tega je subjektivna ocena vpliva omrežnih napak na storitev ali sistemsko opremo.

Pri izvedbi subjektivnega testiranja je najprej potrebno izbrati značilnosti, ki najbolje ustrezajo ciljem in okoliščinam ocenjevanja. To zajema:

- pripravo testnega okolja,
- pripravo testnega gradiva,
- izbiro ocenjevalcev,
- izbiro ocenjevalnih metod,
- izvedbo in potek subjektivnega testiranja in
- obdelavo subjektivnih ocen.

Standardna metoda evalvacije poteka tako, da testne osebe izpolnijo vprašalnik za vsak testni scenarij v testnem naboru, ki ga ocenjevalec predloži. Rezultat takšne evalvacije je *povprečna subjektivna ocena kakovosti* (angl. Subjective Mean Opinion Score – subMOS).

5.1. Priprava testnega okolja in gradiva

Vsako subjektivno testiranje kakovosti se začne na nivoju izbire primernih testnih vzorcev in scenarijev. Referenčne testne sekvence, ki se uporabljajo za namene testiranja, morajo biti vedno ponovljive in enake, da se ohrani transparentnost rezultatov med različnimi evalvacijami. Testno gradivo, scenariji, postopki izvedbe in izbira testnih oseb so izbrani tako, da ustrezajo predpisanim standardom. Pri enomodalni evalvaciji so eksperimentalna zasnova in testne metode izbrane tako, da ustrezajo namenu testiranja ter istočasno izločijo oz. omejijo vpliv drugih modalnosti na percepcijo testnega sistema.

Izdaja tehničnih priporočil in standardov je pod okriljem standardizacijskih teles. Pomembnejše standardizacijske ustanove z mednarodnim vplivom, ki (vsaj delno) zajemajo IT-področja, opisana v tej disertaciji, so:

- ITU,
- ISO,
- IEC,
- W3C in
- ETSI.

ITU je specializirana agencija Združenih narodov, odgovorna posebej za razvoj in nadzor v IKT-segmentu. Razdeljena je na 3 sektorje: ITU-R, ITU-D in ITU-T, pri tem pa zadnji skrbi za izdajo telekomunikacijskih standardov. ITU-T tako določa dobro uveljavljena pravila pri izmenjavi zvočnih, video in podatkovnih sporočil, ITU-R pa izdaja radiokomunikacijska pravila, med katera sodi tudi evalvacija radijsko distribuiranih vsebin. Za podobno vlogo ISO skupaj z IEC skrbi v okviru komiteja ISO/IEC JTC 1. Konzorcij W3C je sestavljen iz skupine organizacij, ki so specializirane posebej za spletne vsebine. Na drugi strani je bil ETSI ustanovljen pretežno z namenom, da podpre evropske industrijske in regulativne zahteve na področju tehnologij IKT, vendar je z globalnim povezovanjem uspel doseči mednaroden vpliv. Omenjeni so z namenom združenja strokovnega znanja iz industrije, raziskovalnih skupin in univerz ustanovili posebne delovne skupine strokovnjakov. Takšni sinergijski vzgibi so privedli do ustanovitve sledečih skupin:

- ***VQEG* (angl. Video Quality Experts Group)**: opravlja subjektivne teste kakovosti videa, validira objektivne modele vrednotenja kakovosti videa in skupinsko razvija nove tehnike.

- ***MPEG* (angl. Moving Picture Experts Group):** je delovna skupina ISO/IEC, zadolžena za standardizacijo avdio in video kompresije in prenosa.

- ***JPEG* (angl. Joint Photographic Experts Group):** Komite združenja ISO/IEC in ITU-T, ki je ustvaril standarde kodiranja slik, znane pod istoimensko oznako JPEG.

- ***VCEG* (angl. Video Coding Experts Group)**: Neformalno ime študijske skupine *ITU-T Q.6/SG 16*, odgovorne za standardizacijo video kodekov H.26x.

- ***JVT* (angl. Joint Video Team):** Združenje skupin VCEG in MPEG z namenom izdelave standardov za napredno video kodiranje, npr. H.264.

- ***Delovna skupina MMI* (angl. Multi-Modal Interaction Working Group):** Vizija delovne skupine MMI je omogočiti dinamično izbiro najprimernejšega načina interakcije za uporabnika, glede na njegove potrebe in zmožnosti.

V Sloveniji za pripravo tehničnih standardov in regulativ deluje *Slovenski inštitut za standardizacijo* (SIST), ki deluje v sodelovanju z naštetimi organizacijami.

5.2. Standardi ITU

Standardi in priporočila ITU se uporabljajo pri subjektivni evalvaciji večmodalnega gradiva. Testno metodologijo, pripravo okolja in izvedbo subjektivnega ocenjevanja, ki so ključni za razumevanje te disertacije, definirajo standardi, predstavljeni v tabeli 5.1. Na podlagi teh je možno zasnovati okolje in določiti pogoje za subjektivno evalvacijo videa visoke kakovosti (tabela 5.2).

Tabela 5.1: Obseg področja standardov ITU-T in ITU-R.

Oznaka ITU	Leto zadnje verzije	Obseg področja	Modal-nost
P.910 [247]	2008	Neinteraktivne subjektivne metode in pogoji za evalvacijo kakovosti digitalnega videa različnih video razredov (tabela 5.2). Predvidene testne metode so primerne za izbiro algoritmov, testiranje karakteristik video sistema in evalvacijo video povezave.	V
P.911 [248]	1998	Neinteraktivne subjektivne metode in pogoji za evalvacijo skupne enosmerne AV-kakovosti za multimedijske aplikacije, kot so npr. videokonference, prenos AV-vsebin itd. Predlagane so karakteristike testnega gradiva (dolžina posnetkov, tip vsebine, število sekvenc itd.), relacije med A-, V- in AV-modalnostmi ter vpliv kakovosti.	AV
P.913 [249]	2014	Definira neinteraktivne metode subjektivnega vrednotenja kakovosti enosmerne avdio, video in/ali avdio-video kakovosti za storitve, kot so internetni video in distribucija kakovostnega videa. Definirane metode so primerne za različne namene subjektivne evalvacije, predvsem za primerjavo kakovosti naprav ter vrednotenje zmogljivosti naprav v različnih okoljih.	AV
BT.500 [250]	2012	Definira dva razreda subjektivnih testiranj za televizijske sisteme: evalvacijo sistemov v optimalnih pogojih in evalvacijo sistemov v neoptimalnih pogojih, tj. zmožnost zagotavljanja kakovosti storitve pod vplivom degradacij.	V
BT.710 [251]	1998	Izpeljanka standarda BT.500, primerna za evalvacijo v kontekstu HDTV.	V
P.800 [252]	1996	Procedure in metode za izvedbo subjektivnega ocenjevanja kakovosti prenosa.	A
P.830 [253]	1996	Subjektivno ocenjevanje zmogljivosti digitalnih govornih kodekov. Namen metodologije je meritev degradacij na prenosni poti.	A
BT.1129	1998	Subjektivno ocenjevanje SD TV-sistemov pod vplivom degradacij, npr. BER.	V
BT.1788 [254]	2007	Metodologija za subjektivno ocenjevanje kakovosti videa v neinteraktivnih multimedijskih aplikacijah.	V

Tabela 5.2: Parametri subjektivne evalvacije videa.

Parameter	**Vrednost**
Video signal – video razred	TV0: ITU-R BT.601: maks.: < 500 ms(videokonferenca :< 150 ms); 270*
	TV1: ITU-R BT.601: maks.: < 500 ms(videokonferenca :< 150 ms); 18 - 50
	TV2: ITU-R BT.601: maks.: < 500 ms(videokonferenca :< 150 ms); 10 - 25
	TV3: ITU-R BT.601: maks., občasna ponovitev okvirja: < 500 ms; 1.5 - 8
	MM 4a: ITU-R BT.601: 25 ali 30 fps: <~150 ms; ~ 1.5
	MM 4b: CIF: 25 ali 30 fps: <~150 ms; ~ 0.7
	MM 5a: CIF: 10 - 30 fps: <~1000ms; ~ 0.2
	MM 5b: CIF: 1 - 15 fps: <~1000ms; ~ 0.05
	MM 6: CIF: od 0 fps: brez restrikcij; < 0.05 ~ 0 (0 fps)
Testne metode **	ACR, SSCQE, DSCQS, DCR, CCR
Tip degradacije	Omrežni nivo: količina izgube paketov,
Vsebina ***	A ena oseba, glava in ramena, omejeni detajli in premikanje.
	B ena oseba, glava in ramena z grafiko in/ali detajli.
	C Več oseb.
	D Grafika.
	E Premikanje objektov in kamere.
Število sekvenc in dolžina testiranja	Število sekvenc je definirano glede na eksperimentalno zasnovo. Dolžina testiranja < 30 min.
Oddaljenost opazovanja ****	1 – 9
Maksimalna luminanca zaslona	100 – 200 cd/m
Osvetljenost prostora	<= 20 lux
Testne degradacije	Vsaj 4 različne scene ter porazdeljen spekter degradacij s percepcijsko različnimi pogoji dražljajev.
Testni zaslon	CRT, LCD, plasma, projektor ali drugo.
Osvetlitev prostora	navadne ali fluorescentne žarnice, med 2700 °K in 6500 °K barvne toplote

** - Označeno kot »razred:resolucija:hitrost osveževanja:latenca:nominalna bitna hitrost(v Mbit/s)«*

*** Večina metod ocenjevanja je občutljiva na spremembe obsega in distribucijo pogojev, zato je potrebno vključiti polni rang vplivajočih degradacijskih faktorjev.*

**** Izbira scenskih karakteristik ter njene pripadajoče prostorske in časovne informacije igrajo pomembno vlogo pri stopnji in lastnostih digitalnih posnetkov ter posledično vplivu nepravilnosti pri prenosu skozi (oddajni, distribucijski) podatkovni kanal. Set sekvenc mora upoštevati kompleten nabor možnih napak za obravnavan tip multimedije.*

***** Oddaljenost opazovanja je odvisna od velikosti opazovanca, tj. ekrana. Preferirano oddaljenost opazovanja (angl. Prefered Viewing Distance - PVD) prikazuje tabela 5.3.*

Tabela 5.3: Priporočena oddaljenost opazovanja.

Diagonala zaslona		Velikost zaslona	PVD
Razmerje 4/3	**Razmerje 16/9**	**(m)**	**(H)**
12	15	0,18	9
15	18	0,23	8
20	24	0,30	7
29	36	0,45	6
60	73	0,91	5
> 100	> 120	> 1,53	3 - 4

5.3. Testne metode in osnove točkovanja

Edini zanesljiv način, da ocenimo kakovost storitve je z izvedbo subjektivnih testov. To zagotavlja referenčne podatke pri raziskavah uporabniške izkušnje (QoE) in pri oceni zmogljivosti objektivnih metod. Da se zagotovi relevantnost rezultatov testiranja, je potrebna pravilna zasnova in dober nadzor eksperimentalnega okolja. Uporaba testne metode vpliva na postopek stimulacije testnih oseb ter na dobljene rezultate evalvacije. Testne metode lahko razdelimo v kategorije glede na vrsto stimulacije:

- **z enojnim dražljajem** (angl. Single Stimulus Method – SS),
- **z dvojnim dražljajem** (angl. Double Stimulus Method – DS),
- **s trojnim dražljajem** (angl. Triple Stimulus Method – TS).

Pri metodah SS je ocenjevalcu predložen en posnetek naenkrat, ki ga mora oceniti. Pri tem ocenjevalec ne ve, ali gre za *referenčni* ali *testni posnetek* (metoda s skrito referenco). Pri DS sta prikazana dva posnetka, lahko drug ob drugem (vzporedna evalvacija) ali zaporedno, pri tem pa se ocenjuje en posnetek v odnosu do drugega [255]. Pri metodi TS obstajajo trije posnetki v eni testni sekvenci (»A«, »B« in »C«). Pri tem je »A« *znan referenčni posnetek* in se predvaja najprej, nato pa »B« in »C«, ki naključno zajemata vlogo *skrite reference* in *testnega posnetka*. Ocenjevalec ocenjuje a) kakovost posnetka »B« v primerjavi z »A« ter b) »C« v primerjavi z »A« (ITU-R BS.1116-1). Glede na potek ocenjevanja lahko to poteka *sprotno* ali *ob koncu*.

Prednost sprotnega načina je informacija o porazdelitvi kakovosti znotraj enega testnega posnetka, slabost pa potreba po združevanju parcialnih ocen kakovosti v končno, skupno oceno testnega posnetka. Dodatno je potrebno posvetiti pozornost

vplivu kontinuirane interakcije ocenjevalca z ocenjevalno opremo, namesto da bi se osredotočil na testni posnetek. Na drugi strani pa je *ocenjevanje ob koncu* omejeno z maksimalno dolžino testnega posnetka (ta ne sme biti prevelika) in s časom pojavitve degradacije znotraj enega testnega scenarija.

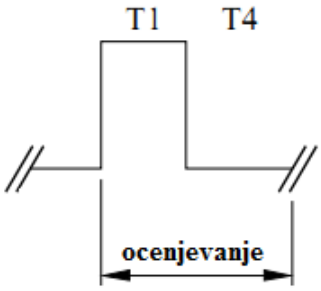
Večina metod je občutljivih na spremembe obsega in distribucijo pogojev, zato je potrebno vključiti čim popolnejši rang vplivnih faktorjev, npr. kombinacijo vseh možnih scenarijev. Če se to ne da, je potrebno testne posnetke naključno razporediti, da ne pride do pojava zaporednih vzorcev. Raziskovalec mora poskrbeti za to, da imajo vključeni testni nabori tudi scenarije, ki predstavljajo robne pogoje.

Pri ocenjevanju ocenjevalci uporabljajo lestvico kakovosti. Ta ima lahko skalo z *neposredno sidranimi vrednostmi*, npr. standardna diskretna lestvica MOS z vrednostmi od 1 do 5, ali pa je skala lahko *zvezna* (slika 5.2d). Prednosti prve so natančno izražene ocene, druga pa daje večjo svobodo ocenjevalcem in predstavlja manjši šum in kvantizacijsko napako.

Ocenjevalnih metod je veliko, pri tem pa je pomembno izbrati tisto, ki najbolje kvantitativno in kvalitativno zajame problem raziskave. V nadaljevanju so opisane standardizirane in najpogosteje uporabljene metode pri subjektivnem ocenjevanju večmodalnih storitev.

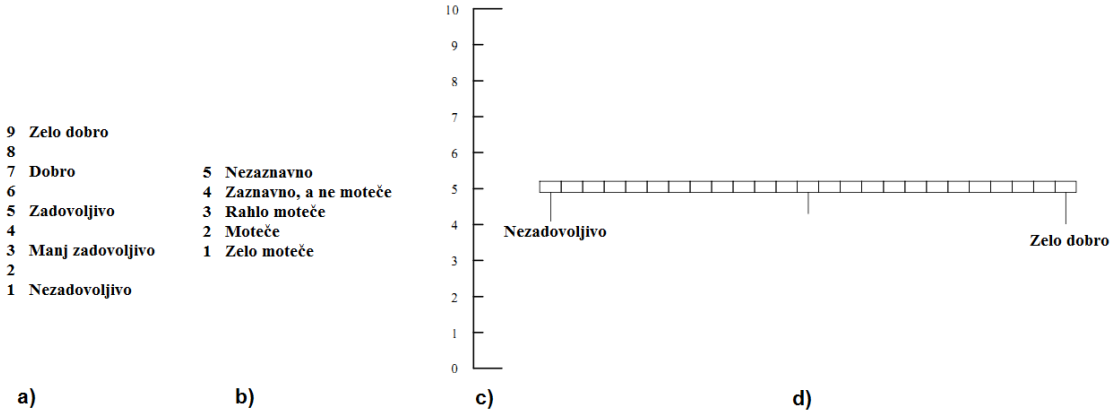
5.3.1. Absolutna kategorijska ocena

Absolutna kategorijska ocena (angl. Absolute Category Rating – ACR) sodi v kategorijo metod SS, kjer so testni posnetki predstavljeni posamično ter nato ocenjeni neodvisno na kategorijski lestvici. V zaporedju prikazanih posnetkov je prikazan tudi referenčni posnetek. Čas za ocenjevanje je možen ob koncu predvajanja testnega posnetka (*T4*), nato se cikel ponovi (slika 5.1).



Slika 5.1: Potek ocenjevanja po metodi ACR.

Pri ocenjevanju se uporabi eden izmed načinov točkovanja, pri čemer je najpogostejša *absolutna 5-stopenjska semantična lestvica kakovosti* (slika 5.2a). V določenih primerih je numerična lestvica upravičena, saj se pri tem izognemo napačni presoji pomena semantične lestvice. Posledično je možna tudi večstopenska lestvica ali celo takšna, ki ni kategorizirana (slika 5.2b, c). Prednost teh je manjša napaka zaradi diskretiziranosti ocenjevalnih vrednosti.

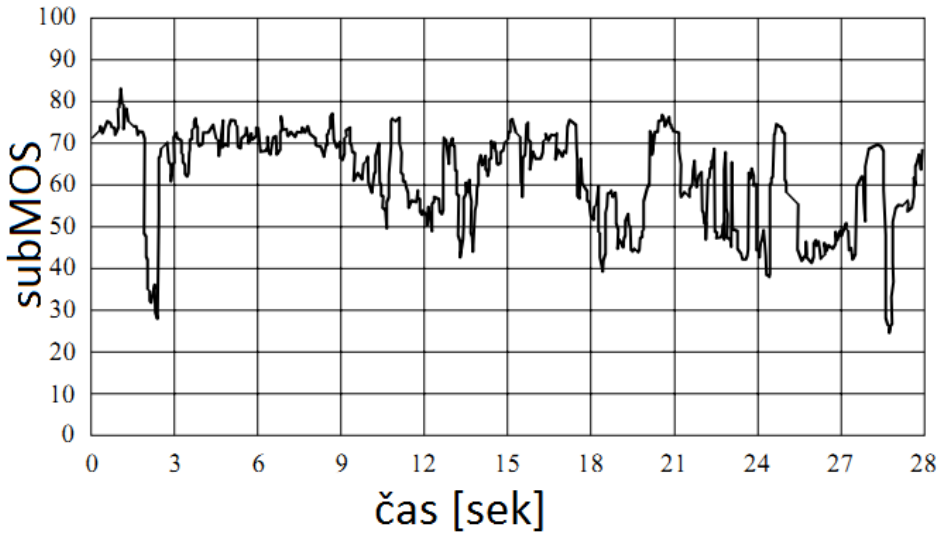


Slika 5.2: Lestvice kakovosti: a) 5/9-stopenska semantična, b) 5-stopenjska semantična, c) 11-stopenjska numerična, d) zvezna grafična.

Razširitev ACR je metoda *s skrito referenco* (angl. ACR with Hidden Reference – ACR-HR), kjer se lahko pojavi tudi referenčni posnetek. Ta postopek daje možnost odstranitve neželene subjektivne všečnosti konteksta, tj. izkušnje in pričakovanja uporabnika, navideznih popačenj v referenčnem signalu in opremi. Pri podatkovni analizi rezultatov se izračuna t i. *diferencialni MOS* (angl. Differencial mean opinion score – DMOS) med dejansko testno sekvenco in skrito referenco. ACR-HR se po navadi uporablja pri evalvaciji posnetkov z nizkim deležem degradacij.

5.3.2. SS stalna evalvacija

Kontinuirana evalvacija kakovosti z enim dražljajem (angl. Single Stimulus Continuous Quality Evaluation – SSCQE) se uporablja pri ocenjevanju TV-sistemov, kjer testne osebe sprotno ocenjujejo kakovost video sekvence na linearni lestvici kakovosti, po navadi z drsnikom, kolesom, ali gumbom (slika 5.3) [256], [257], [258]. Video sekvence pri metodi SSCQE običajno trajajo dlje, tj. več kot 10 minut, in so predvajanje samo enkrat (degradiran posnetek).



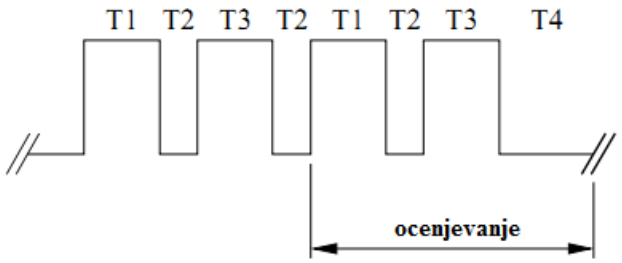
Slika 5.3: Primer sprotnega subjektivnega ocenjevanja kakovosti posnetkov.

5.3.3. Ocena z dvojnim dražljajem

Kontinuirana lestvica kakovosti z dvojnim dražljajem(angl. Double Stimulus Continuous Quality Scale – DSCQS) se uporablja v primerih, ko ni možno zagotoviti celotne palete kombinacij degradacij, še posebej v primerih, ko določamo kakovost zelo natančno z manjšim spektrom degradacij, npr. jamčenje določenega nivoja QoS. Ko je potrebno izbrati omejen nabor testnih sekvenc, se namesto izčrpnega iskanja (angl. Exhaustive Search – ES) uporabi ena od optimizacijskih metod, npr. *gradientno naraščajoče subjektivno testiranje* (angl. Gradient Ascent Subjective Quality Testing – GAST) [259]. Testni posnetki so predstavljeni v parih tako, da:

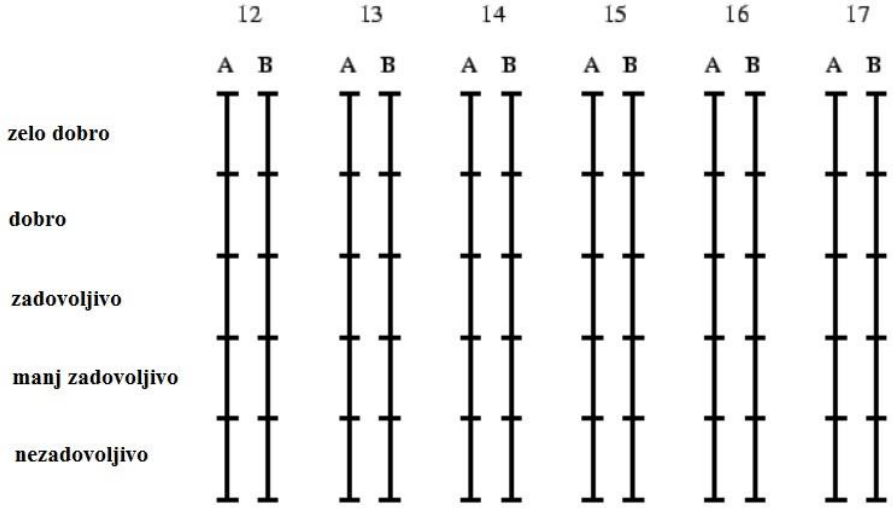
- je zaporedje referenčnega in testnega materiala opazovalcu nepoznano, zato se od njega zahteva, da oceni oba posnetka iz para (slika 5.4),
- se znotraj vsakega testnega scenarija zaporedje naključno menja.

Ocenjevalna lestvica je tipično med vrednostima 0 in 100.



Slika 5.4: Ocenjevanje z metodo z dvojnim dražljajem.

Primer testne pole prikazuje slika 5.5.

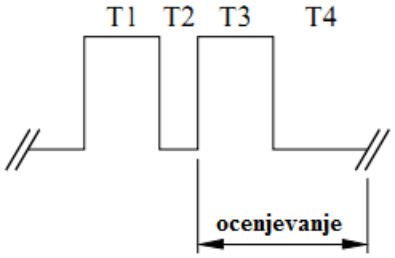


Slika 5.5: Testna pola za ocenjevanje po metodi DSCQS.

5.3.4. Kategorijska ocena degradacije

Kategorijska ocena degradacije (angl. Degradation Category Rating – DCR) je metoda DS, ki jo je standardizirala inštitucija ITU in je imenovana tudi *degradacijska lestvica z dvojnim dražljajem* (angl. Double Stimulus Impairment Scale – DSIS) pri EBU. Pri tej ocenjevalec ocenjuje stopnjo zaznavnosti degradacije. Lestvica kakovosti je lahko enaka

tisti za metodo ACR, le da se za skalo uporabijo degradacijske vrednosti namesto absolutnih (slika 5.2b). Pri tem je zaporedje prikaza posnetkov vedno takšno, da ocenjevalec najprej vidi referenčni posnetek, tj. *T1 = REF*, in nato testno sekvenco, tj. *T3 = DEG* (slika 5.6). Pri tem so določeni časovni parametri, da se ohrani osredotočenost ocenjevalca na testiranje [247], [260] (tabela 5.4).



Slika 5.6: Potek ocenjevanja po metodi DCR.

Tabela 5.4: Priporočeni časovni parametri.

Parameter	**Čas**
T1 in T3	~ 10 sek. (za sliko)
	~ 10, a ne več kot 30 sek. (za video)
T2	≤ 10 sek.
T4	~ 10 sek.
Celotna seja	≤ 30 min.

Senzitivnost te metode je večja od ACR, zato je večja verjetnost, da ocenjevalci izbirajo mejne vrednosti. DCR je primerna za evalvacijo storitev, kjer se ocenjujejo karakteristike degradacij [261], in ne sistemov kot takšnih, ki lahko že v osnovi vsebujejo mejo kakovosti, ki je »perfektna«. Zato je naloga izvajalca, da pripravi sistem, ki je kar najbolj transparenten za dane pogoje. Rezultati evalvacije, ki jih pri tem dobimo, so meritve *subjektivne ocene količine degradacij* (angl. Degradation Mean Opinion Score – DMOS) [262].

Zaradi večje variacije rezultatov DMOS je smiselno upoštevati tudi vsebino testnih posnetkov, npr. pri evalvaciji videa, ki uporablja AVC – ob predpostavki, da ostaja testni parameter (npr. količina izgubljenih paketov) enak – lahko kakovost različnih scen videa iz perspektive opazovalca niha, saj so nižjenivojski parametri prenosa, tj. kodirnega algoritma H.264, drugačni.

5.3.5. Kategorijska primerjalna ocena

Pri kategorijski primerjalni oceni (angl. Comparison Category Rating – CCR) podobno kot pri DCR testni opazovalec primerja pare, vendar brez vednosti, kateri izmed posnetkov je referenčni in kateri testni. Med posnetkoma se primerja kakovost samo relativno na sedemstopenjski lestvici kakovosti (tabela 5.5).

Tabela 5.5: Sedemstopenjska primerjalna lestvica.

Ocena	Vrednotenje
3	Mnogo boljše
2	Boljše
1	Nekoliko boljše
0	Približno enako
-1	Nekoliko slabše
-2	Slabše
-3	Mnogo slabše

Testni sistem načeloma vsebuje *n(n-1)* kombinacij parov testnih režimov, tj. AB, BA, CA, AC itd. [263].

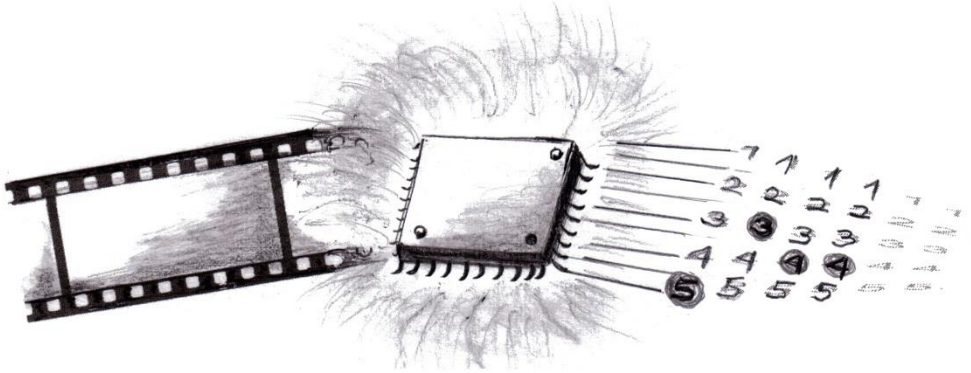
5.3.6. Primerjalna pridevniško kategorijska presoja

Primerjalna presoja s pridevniškimi kategorijami (angl. Stimulus Comparison Adjectival Categorical Judgment – SCACJ) je metoda primerjave dveh video sekvenc, ki se predvajata simultano. Testni opazovalec na koncu oceni primerjavo posnetkov, po navadi s sedemstopenjsko primerjalno lestvico (tabela 5.5).

5.3.7. Subjektivna ocenjevalna metoda evalvacije kakovosti videa

Subjektivna ocenjevalna metoda za evalvacijo kakovosti videa (angl. Subjective Assessment Method for Video Quality evaluation – SAMVIQ) je razvita s strani EBU. Pri tej metodi je med testno proceduro ocenjevalcu dovoljeno predvajati in oceniti poljubno sekvenco. Lahko je dovoljeno v poljubnem trenutku tudi predvajati referenčni posnetek. Če referenčni posnetek ni znan in je pomešan v testni nabor, govorimo o metodi s skrito referenco ali SAMVIQ-HR.

Kategorijska lestvica SAMVIQ je standardna ocenjevalna lestvica MOS z možnostjo vmesnih ocen v razponu vrednosti med 0 in 100.



6. Merjenje kakovosti storitev z objektivnimi metodami

Podobno kot tradicionalne subjektivne metode vrednotenja so objektivne metrike kakovosti potrebne za vrednotenje kakovosti storitve, pri tem pa je zahtevana čim boljša korelacija s subjektivno oceno. Idealna objektivna metoda je tako zmožna natančno in avtomatizirano posnemati odziv povprečnega uporabnika storitve. Glede na prisotnost referenčnega signala, ki velja za posnetek brez degradacij, metrike delimo v 3 skupine:

- **metrike s polno referenco** (angl. Full Reference – FR),
- **metrike z delno referenco** (angl. Reduced Reference – RR) in
- **metrike brez reference** (angl. No Reference – NR).

NR-metode so priročne v praksi, vendar imajo slabšo korelacijo in so kompleksnejše. Pri metodah RR referenčni signal ni popolnoma dosegljiv, namesto tega pa obstaja nabor značilk, pridobljenih iz referenčnega signala, z namenom evalvacije kakovosti testnega posnetka. FR-metrike se izmed vseh izkažejo za najbolj natančne in

robustne ter bolj prilagodljive za generične meritve [23]. Ločimo 2 tipa FR-metrik glede *na pristop detekcije in magnitude degradacij*. Pri tem metrike iz prve kategorije izmerijo napake med originalno in degradirano sekvenco za vsak element posebej (slikovna pika ali časovni okvir), kar privede do degradacijskih območij, kjer se z metodo združevanja izloči skupna ocena kakovosti. Pri metrikah iz druge kategorije pa zaznava posameznih degradacij vodi do skupne ocene sekvence. Glede na *upoštevanje HAVS* je v literaturi zaslediti dve kategoriji: metrike iz prve delujejo na enostavnejšem matematičnem izračunu vrednosti z malo znanja o HAVS ter kasneje upoštevajo a priori znanje o združevanju in percepciji zaznanih degradacij v višjih slojih možganskega procesiranja, metrike iz druge kategorije pa vključujejo informacijo percepcije na nižjih nivojih HAVS, npr. efekt maskiranja, pasovna dekompozicija itd.

Metrike vrednotenja vizualne modalnosti v grobem delimo glede na tip degradacije, ki ga obravnavajo: prostorsko ali časovno komponento oziroma združen, kombiniran pristop. Prostorski pristop deluje na nivoju posamičnih okvirjev, zato govorimo tudi o slikovnih metrikah kakovosti (angl. Image Quality Assessment – IQA). Na drugi strani časovni pristop upošteva še časovno odvisne mehanizme degradacij, ki nastanejo kot posledica distorzij gibanja ali maskiranja teh med zaporednimi okvirji videa.

Naša doktorska naloga obravnava FR-metrike za avdio-video vsebine, pri tem pa vizualno modalnost ovrednotimo s prostorskim pristopom zaznavanja degradacij. Na tej osnovi temeljijo tudi predstavljene slikovne in avdio metrike v naslednjih podpoglavjih.

6.1. Objektivne slikovne metrike s polno referenco

Vrednotenje videa v prostorski domeni zajema vrednotenje v času isto-ležnih gradnikov, npr. video okvirjev oz. slik. Enostavne meritve kakovosti okvirjev delujejo na računsko preprostih meritvah. Primer tega sta *povprečna razlika* (angl. Average Difference – AD), ki izračuna matematično povprečno vrednost razlike med referenčnim in testnim signalom, ter *maksimalna razlika* (angl. Maximum Difference – MD), ki določa maksimum razlike vrednosti slikovne pike na celotnem okvirju. Metriki sta definirani kot:

$$\boldsymbol{AD}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathbf{1}}{\boldsymbol{MN}} \sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}} \sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}} (\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n}) - \boldsymbol{y}(\boldsymbol{m},\boldsymbol{n})) \tag{6.1}$$

$$\boldsymbol{MD}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{MAX}|\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n}) - \boldsymbol{y}(\boldsymbol{m},\boldsymbol{n})| \tag{6.2}$$

kjer sta *M* in *N* število slikovnih pik, *x* referenčna in *y* testna slika velikosti *m* * *n* slikovnih pik.

Rezultati teh metrik ne predstavljajo zadovoljive korelacije pri uporabi IQA, predvsem zaradi ne-konsistentnosti predznaka [264]. Povprečno razliko pogosto prikažemo v absolutnem merilu, saj je pomembnejše dejansko odstopanje med slikama v smislu količine degradacije med njima. Iz tega lahko izpeljemo *povprečno absolutno napako* (angl. Mean Absolute Error – MAE):

$$\boldsymbol{MAE}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{1}}{\boldsymbol{MN}} \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} \sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} |\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n}) - \boldsymbol{y}(\boldsymbol{m},\boldsymbol{n})| \tag{6.3}$$

Kadar primerjamo ocene kakovosti slik z različnimi dinamičnimi razponi vrednosti slikovne pike, je potrebna normalizacija vrednosti, kar lahko izrazimo z *normalizirano absolutno napako* (angl. Normalized Absolute Error – NAE):

$$\boldsymbol{NAE}(\boldsymbol{x},\boldsymbol{y}) = \frac{\sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} \sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} |(\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n}) - \boldsymbol{y}(\boldsymbol{m},\boldsymbol{n})|}{\sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} \sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} |\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n})|} \tag{6.4}$$

Omenjene metode se pogosteje uporabljajo v postprocesiranju, npr. pri izračunu numerične razlike med mapami degradacij posameznih okvirjev.

Med bolj uveljavljenimi in široko uporabljanimi IQA-metrikami sta *kvadrat povprečne napake* (angl. Mean Squared Error – MSE) in povprečna objektivna meritev *maksimalnega razmerja signal-šum* (angl. Peak Signal to Noise Ratio – PSNR). Uporabni sta zaradi nizke računske kompleksnosti in enostavne implementacije. To daje možnost sprotnih meritev v realnem času, pri tem pa predstavlja zadovoljivo korelacijo s subMOS. Definiramo ju kot:

$$\boldsymbol{MSE}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{1}}{\boldsymbol{MN}} \sum_{\boldsymbol{m=1}}^{\boldsymbol{M}} \sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} (\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n}) - \boldsymbol{y}(\boldsymbol{m},\boldsymbol{n}))^{\boldsymbol{2}} \tag{6.5}$$

$$\boldsymbol{PSNR}(\boldsymbol{x},\boldsymbol{y}) = \mathbf{10}\boldsymbol{log}_{\mathbf{10}}\frac{\boldsymbol{L}^{\mathbf{2}}}{\boldsymbol{MSE}} \tag{6.6}$$

kjer *x* predstavlja testno, tj. okvarjeno sliko in *y* originalno sliko, *M* in *N* je število slikovnih pik, *L* pa maksimalna vrednost slikovne pike, definirana kot:

$$\boldsymbol{L} = \mathbf{2}^{\boldsymbol{n}} - \mathbf{1} \tag{6.7}$$

kjer je *n* bitna globina slike.

Za primerjavo odstopanj transkodiranja slik se uporablja tudi *koren kvadrata povprečne napake* (angl. Root Mean Squared Error – RMSE), saj je enota v istem merilu kot merjena veličina. Definiramo jo z:

$$\boldsymbol{RMSE}(\boldsymbol{x},\boldsymbol{y}) = \sqrt{\boldsymbol{MSE}(\boldsymbol{x},\boldsymbol{y})} \tag{6.8}$$

Metrike MSE, RMSE in PSNR predstavljajo kvantitativno numerično oceno degradacije med referenčno in testno sliko, ponavadi izraženo v logaritemski enoti, zaradi velikega dinamičnega razpona vrednosti. Enota PSNR je zato decibel (dB) in zavzema vrednosti od 0 do neskončno: višja kot je vrednost, boljša je izračunana kakovost. Metrike se uporabljajo predvsem pri evalvaciji kakovosti rekonstrukcije slik pri uporabi kodekov s kompresijsko izgubo, npr. *JPEG2000*, kjer so vizualne napake precej enakomerno prostorsko razpršene.

Ker robovi nosijo višjo prostorsko informacijo, gledano s stališča HVS, je pri vrednotenju degradacije teh smiselno uporabiti *Laplace-ovo meritev kvadrata povprečne napake* (angl. Laplacian Mean Square Error – LMSE) [265]. LMSE je občutljiv na visoke prostorske frekvence, tj. robove, in izmeri spremembo razlike slikovnih pik z Laplaceovim operatorjem. Definiran je kot:

$$\boldsymbol{LMSE}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}}\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}}\left[\frac{\boldsymbol{L}(\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n})) - \boldsymbol{L}(\boldsymbol{y}(\boldsymbol{m},\boldsymbol{n}))}{\sum_{\boldsymbol{m}=\mathbf{1}}^{\boldsymbol{M}}\sum_{\boldsymbol{n}=\mathbf{1}}^{\boldsymbol{N}}\boldsymbol{L}(\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n}))}\right]^{\mathbf{2}} \tag{6.9}$$

Drug tip metrik kot vhod uporabi korelacijsko meritev. Primer tega je korelacijska metoda imenovana s*trukturna vsebina* (angl. Structural Content – SC). SC določa korelacijo slikovnih pik, definirano kot:

$$\boldsymbol{SC} = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\boldsymbol{x}(\boldsymbol{m},\boldsymbol{n})^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}\boldsymbol{y}(\boldsymbol{m},\boldsymbol{n})^2} \tag{6.10}$$

Težava matematičnih meritev je, da neposredno ne upoštevajo bioloških faktorjev. Nekateri jim zato pripisujejo slabšo korelacijo s subMOS še posebej v primeru, ko je prisoten širok spekter vizualnih degradacij [266]. Kot rešitev za izboljšanje korelacije predlagajo obtežitev koeficientov pri značilkah določenih degradacij. Primer tega so metrike *obtežene SNR* (angl. Weighted Signal to Noise Ratio – WSNR) [267], *PSNR-HVS* (angl. HVS based Peak Signal to Noise Ratio – PSNRHVS) [268] in *modificiran PSNR-HVS* (angl. modified HSV based PSNR – PSNRHVSM) [269]. Kot primer, WSNR izkorišča CSF z manjšo obtežitvijo napak, ki nastanejo v visokofrekvenčnih področjih, in predstavlja razmerje povprečne utežene signalne moči in povprečne utežene moči šuma.

Še vedno pa so strogo matematično definirani modeli kakovosti atraktivni zaradi dveh razlogov: a) nizke računske kompleksnosti in b) neodvisnosti od karakteristik opazovalca in pogojev opazovanja.

Avtorji so v [59] predlagali nov matematični model, imenovan *univerzalni indeks kakovosti slike* (angl. Universal Image Quality Index – IQI). Zaradi nizke računske kompleksnosti je algoritem primeren za obdelavo visoko-resolucijskih slik v realnem času. Avtorji pa neodvisnost od lastnosti opazovalca in opazovanja opravičujejo z dejstvom, da so (določeni) pogoji uporabe variabilni in zato nekateri podatki načeloma niso na voljo objektivnemu sistemu vrednotenja. IQI je definiran kot:

$$\boldsymbol{Q}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{4\sigma_{xy}\,\overline{x}\,\overline{y}}}{\left(\boldsymbol{\sigma_x^2} + \boldsymbol{\sigma_y^2}\right)[\boldsymbol{\overline{x}^2} + \boldsymbol{\overline{y}^2}]} \tag{6.11}$$

kar lahko zapišemo tudi kot:

$$\boldsymbol{Q}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{\sigma_{xy}}}{\boldsymbol{\sigma_x\sigma_y}} \cdot \frac{\boldsymbol{2\overline{x}\,\overline{y}}}{\boldsymbol{(\overline{x})^2} + \boldsymbol{(\overline{y})^2}} \cdot \frac{\boldsymbol{2\,\sigma_x\sigma_y}}{\boldsymbol{\sigma_x^2} + \boldsymbol{\sigma_y^2}} \tag{6.12}$$

pri tem je:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{in} \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{6.13}$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \quad \text{in} \quad \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{6.14}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \tag{6.15}$$

kjer sta $x = (x_i \mid i = 1, 2, ..., N)$ in $y = (y_i \mid i = 1, 2, ..., N)$ signala referenčne in testne slike. Enačba 6.12 ponazarja dejstvo, da lahko vsako degradacijo modeliramo kot kombinacijo treh komponent: *izgube korelacije*, *degradacije luminančne informacije* in *degradacije kontrastne informacije*. Primerjavo z modelom MSE vidimo na sliki 6.1.

Slika 6.1: Slika "Lena", primerjava MSE in indeksa UIQI za različne degradacije: a) referenčna slika, MSE = 0, Q = 1, b) impulzivni šum sol-in-poper (angl. impulsive noise and pepper noise), MSE = 255, Q = 0,7227, c) aditivni Gaussov šum, MSE=225, Q=0,3891, d) multiplikativni granularni šum (angl. multiplicative speckle noise), MSE = 255, Q = 0,4408, e) povprečni zamik (angl. mean shift), MSE = 225, Q = 0,9894, f) povečan kontrast, MSE = 225, Q = 0,9372, g) meglenost, MSE = 225, Q = 0,3461, h) kompresija JPEG, MSE = 215, Q = 0,2876 (vir: [270]).

Pod predpostavko, da je HVS močno pogojen z izločanjem strukturne informacije, npr. podobnost med strukturami je pomembnejša kot kromatična napaka, je indeks *meritve strukturne podobnosti* (Structural Similarity – SSIM) tudi eno od meril določanja kakovosti zaznanih vizualnih sprememb. Indeks SSIM v časovni domeni obravnava luminančno podobnost *l*, kontrastno podobnost *c* in strukturno podobnost *s*, kar določa lokalno oceno vsakega okvirja in je definirano kot:

$$\boldsymbol{SSIM}(\boldsymbol{x},\boldsymbol{y}) = [\boldsymbol{l}(\boldsymbol{x},\boldsymbol{y})]^{\boldsymbol{\alpha}} \cdot [\boldsymbol{c}(\boldsymbol{x},\boldsymbol{y})]^{\boldsymbol{\beta}} \cdot [\boldsymbol{s}(\boldsymbol{x},\boldsymbol{y})]^{\boldsymbol{\gamma}} \tag{6.16}$$

$$\boldsymbol{l}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{2\mu_x\mu_y + C_1}}{\boldsymbol{\mu_x^2 + \mu_y^2 + C_1}} \tag{6.17}$$

$$\boldsymbol{c}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{2\sigma_x\sigma_y + C_2}}{\boldsymbol{\sigma_x^2 + \sigma_y^2 + C_2}} \tag{6.18}$$

$$\boldsymbol{s}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{\sigma_{xy} + C_3}}{\boldsymbol{\sigma_x\sigma_y + C_3}} \tag{6.19}$$

kjer so α, β in γ težnostni koeficienti za nastavitev pomembnosti vsake izmed komponent, C_1, C_2, C_3 pa majhne konstante za stabilizacijo vrednosti končne ocene, definirane kot:

$$\boldsymbol{C_1 = (K_1 L)^2}\ , \tag{6.20}$$

$$\boldsymbol{C_2 = (K_2 L)^2}\ , \tag{6.21}$$

$$\boldsymbol{C_3 = C_2/2} \tag{6.22}$$

kjer je *L* dinamični razpon slikovne pike (*L* = 255 za 8-bitno sivinsko sliko), $K_1 \ll 1$ in $K_2 \ll 1$ pa skalarni vrednosti. Vrednosti SSIM so med -1 in 1, kjer vrednost 1 pomeni identičnost med referenčno in testno sliko.

Slika 6.2: "Lena", primerjava MSE, PSNR in indeksa SSIM za različne degradacije (vir: [271]).

V enačbah 6.17–6.19 predstavljata μ_x in μ_y povprečno vrednost, σ_x in σ_y standardno deviacijo originalne in testne slike ter σ_{xy} kovarianco obeh slik, definirano kot:

$$\boldsymbol{\mu_x = \frac{1}{N}\sum_{i=1}^{N} w_i x_i \ , \ \mu_y = \frac{1}{N}\sum_{i=1}^{N} w_i y_i} \tag{6.23}$$

$$\begin{aligned} \boldsymbol{\sigma_x} &= \boldsymbol{\sqrt{\left[\frac{1}{N-1}\sum_{i=1}^{N} w_i (x_i - \mu_x)^2\right]}}, \\ \boldsymbol{\sigma_y} &= \boldsymbol{\sqrt{\left[\frac{1}{N-1}\sum_{i=1}^{N} w_i (y_i - \mu_y)^2\right]}} \end{aligned} \tag{6.24}$$

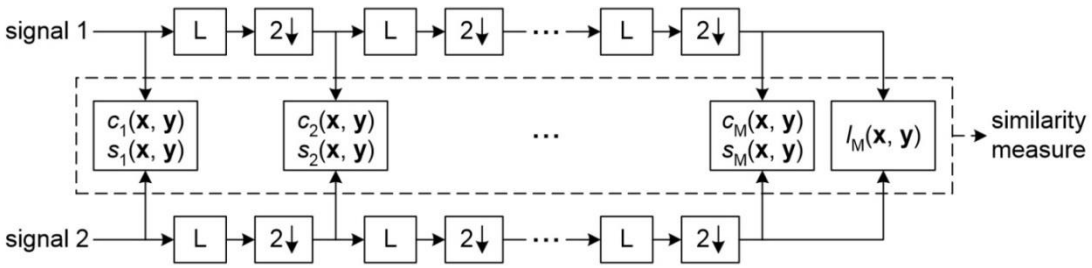
$$\boldsymbol{\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N} w_i (x_i - \mu_x)\,(y_i - \mu_y)} \tag{6.25}$$

pri tem je cirkularna simetrična Gaussova funkcija $w=(w_i|i=1,2,...,N)$ s standardno deviacijo *1,5 vzorca* in normalizirano kot $\sum_{i=1}^{N} w_i = 1$.

Dokazano je bilo, da algoritem SSIM oz. natančneje njegova preprosta izvedba *SS-SSIM* (angl. Single Scale SSIM – SS-SSIM) kljub svoji preprostosti pokaže dobro korelacijo za širok nabor degradacij [26] (slika 6.2). Algoritem SS-SSIM opisuje neko generalizirano rešitev in je primeren za določen obseg opazovalnih pogojev, ker enovito pojmuje stopnjo zaznave (angl. perceivability). Vendar SS-SSIM ne upošteva gostote vzorčenja slikovnega signala, oddaljenosti med opazovalcem in prikazano sliko in posledično sposobnosti opazovalčevega HVS za zaznavanje degradacij pri različnih resolucijah [272]. Razširjena verzija deluje v večkratnih zaporednih iteracijah (praktično je ta med 3 in 5), pri čemer je operacija evalvacije SSIM izvedena na replikah slike v več povečavah, saj s tem zajame dojemanje slikovne podrobnosti, ki je različna za prej omenjene faktorje, npr. oddaljenost opazovalca od zaslona. Metrika kakovosti se imenuje *MS-SSIM* (angl. Multi-Scale Structural Similarity – MS-SSIM) in je definirana kot:

$$\boldsymbol{MS-SSIM}(\boldsymbol{x},\boldsymbol{y}) = [\boldsymbol{l}(\boldsymbol{x},\boldsymbol{y})]^{\alpha_M} \cdot \prod_{\boldsymbol{j=1}}^{\boldsymbol{M}} [\boldsymbol{c_j}(\boldsymbol{x},\boldsymbol{y})]^{\beta_j} [\boldsymbol{s_j}(\boldsymbol{x},\boldsymbol{y})]^{\gamma_j} \qquad (6.26)$$

Podobno kot v osnovni verziji so parametri α_M, β_j in γ_j težnostni koeficienti za nastavitev pomembnosti posamezne komponente. Skalarni parameter *j* definira stopnjo filtriranja in pod-vzorčenja slike (slika 6.3). Na vsaki stopnji se aplicira nizko-prepustni filter ter uporabi pod-vzorčenje za faktor 2. Po *j*-ti operaciji sta kontrastna in strukturna primerjava med slikovnima signaloma *x* in *y* definirana kot $c_j(x,y)$ in $s_j(x,y)$. Avtorji so v [272] eksperimentalno ugotovili težnostne koeficiente na analizi korelacije s subMOS desetih 8-bitnih slik, velikosti *64 x 64* slikovnih pik z *M = 5* in predpostavko, da je $\alpha_j = \beta_j = \gamma_j$ za vse j in $\sum_{j=1}^{M} \gamma_j$: $\beta_1 = \gamma_1 = 0{,}0448$, $\beta_2 = \gamma_2 = 0{,}2856$, $\beta_3 = \gamma_3 = 0{,}3001$, $\beta_4 = \gamma_4 = 0{,}2363$ in $\alpha_5 = \beta_5 = \gamma_5 = 0{,}1333$.



Slika 6.3: Sistem delovanja MS-SSIM.

Izpostavimo lahko lastnost slikovnih metrik, da neposredno merijo vrednosti slikovnih pik na istih prostorskih lokacijah (*x*, *y*). Video kodeki, ki implementirajo časovno kodiranje z uporabo vektorjev premika, kodirajo premike skupkov slikovnih pik skupaj, ter jim dodelijo primerne vrednosti gibalnih vektorjev. Če med prenosom pride do degradacije gibalnih vektorjev, se izgubi informacija o gibanju, zato so predpostavljeni premiki lahko nepravilni. To okvari geometrično informacijo, tj .napako geometrične kletke, kar pomeni, da so istoležne slikovne pike zamaknjene po celotni površini. To za opazovalca percepcijsko ne predstavlja večje napake, iz matematičnega stališča pa dojemanje vizualne informacije pokaže večjo razliko, še posebej za naravne slike, kjer je lokalna deviacija vrednosti sosednjih slikovnih pik velika.

Rešitev za to je translacija slikovne informacije v valjčno domeno. Metrika kakovosti, ki to upošteva je *kompleksna valjčna funkcija strukturne podobnosti* (angl. Complex-Wavelet Structural Similarity - CW-SSIM) [273]. CW-SSIM ohranja strukturno informacijo ter loči amplitudno in fazno napako, kar je smiselno, saj je fazna informacija pomembnejša od amplitudne za naravne slike [274]. Dodatno se izločijo geometrične lastnosti, kot so translacija, skaliranje in rotacija strukturnih elementov s pomočjo dekompozicije slike v *vizualne kanale*, podobne tistim, ki pri ljudeh modelirajo delovanje nevronov v primarnem vizualnem korteksu v obliki lokaliziranih več-dimenzionalnih pasovnoprepustnih orientiranih filtrov [275]. CW-SSIM formuliramo z:

$$\tilde{\boldsymbol{S}}(\boldsymbol{c_x}, \boldsymbol{c_y}) = \frac{\boldsymbol{2}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left|\boldsymbol{c_{x,i}}\right|\left|\boldsymbol{c_{y,i}}\right| + \boldsymbol{K}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left|\boldsymbol{c_{x,i}}\right|^{\boldsymbol{2}} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left|\boldsymbol{c_{y,i}}\right|^{\boldsymbol{2}} + \boldsymbol{K}} \cdot \frac{\boldsymbol{2}\left|\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\boldsymbol{c_{x,i}}\boldsymbol{c_{y,i}^{*}}\right| + \boldsymbol{K}}{\boldsymbol{2}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left|\boldsymbol{c_{x,i}}\boldsymbol{c_{y,i}^{*}}\right| + \boldsymbol{K}} \qquad (6.27)$$

kjer predstavljata $c_x = \{c_{x,i}\ i = 1,2,\dots,N\}$ in $c_y = \{c_{y,i}\ i = 1,2,\dots,N\}$ dva nabora valjčnih koeficientov za isto točko v istem valjčnem pasu za sliki *x* in *y*, c^* je konjugirano kompleksno število od *c* in *K* majhna pozitivna konstanta, potrebna za robustnost lokalizirane meritve, kjer so območja z nizkim razmerjem signal-šum.

Slabost CW-SSIM je občutljivost na strukturno deformacijo, npr. JPEG stiskanje, ki s transkodiranjem v nižjo resolucijo/kakovost onemogoča pravilno detekcijo robov. Pojavi se efekt lažno pozitivnih, tj. ringing, aliasing, ter lažno negativnih vzorcev, tj. meglenost na pregibu podobnih kontrastov (slika 6.4).

Občutljivost CW-SSIM na distorzijo bločnega prostorskega zamika (angl. block spatial shift), ki se pojavlja ob prisotnosti napak pri dekodiranju zaradi napak iz omrežja, so avtorji v [276] omejili z definicijo *utežene CW-SSIM* (angl. Weighted Complex Wavelet SSIM – WCWSSIM). Uteži posameznih pod-pasov so definirane kot:

$$W_s = \frac{\int_0^{\alpha} C(f) H_s(f) df}{\int_0^{\alpha} H_s(f) df} \tag{6.28}$$

kjer je $C(f)$ frekvenčni odziv CSF, $H_s(f)$ frekvenčni odziv valčnega pod-pasu s ter frekvenca vzorčenja α za sliko velikosti 512 x 512 slikovnih pik, tako da velja:

$$\alpha = \frac{512/2}{tan^{-1}\frac{1}{12}} \approx 54\ ciklov/stopinjo \tag{6.29}$$

Eksperimentalno ugotovljene uteži imajo vrednosti *[0,254 0,254 0,25 0,18 0,061]* od največje do najmanjše prostorske frekvence. WCWSSIM naj bi izkazoval boljšo korelacijo za omenjeno degradacijo (slika 6.5).

Slika 6.4: Efekt distorzije nadpraga (angl. suprathreshold) na metrike kakovosti MSE, SSIM in CWSSIM (vir: [271]).

Slika 6.5: Prednosti WCWSSIM. MSE je na obeh slikah enak, pri tem pa so objektivne ocene na levi sliki: CWSSIM=0,94, WCWSSIM=0,87. Desna slika: CWSSIM=0,84, WCWSSIM=0,89.

Ob upoštevanju, da je velik del vizualne informacije redundanten in da na percepcijo HVS v večji meri vplivajo samo napake na določenih predelih, ki dajejo bistveno vlogo za subMOS, so avtorji v [277] predlagali IQA metriko *podobnosti značilk Riesz-ove transformacije* (angl. Riesz-transform based Feature SIMilarity – RFSIM). RFSIM za značilke uporabi informacijo o robovih preko Cannyjevega operatorja, ki se ekstrahira s prvim in drugim redom Riesz-ove transformacije. Tako so v končni oceni slike pomembne samo degradacije na teh »ključnih« prostorskih lokacijah (robovih), kar je v skladu s procesiranjem informacije na nizkem nivoju.

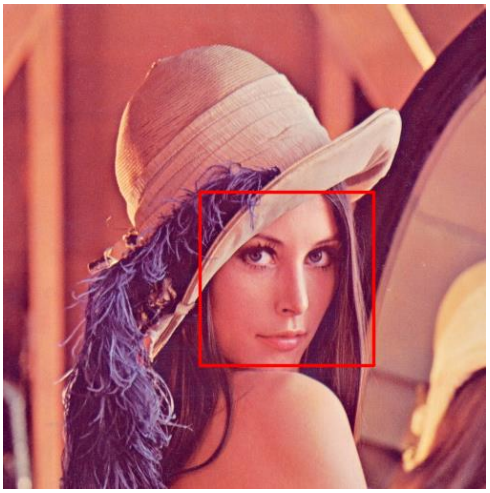
Iz podobnega izhodišča so avtorji v [278] predlagali zasnovo degradacijskega modela. Predlagan model kakovosti *meritve kakovosti šuma* (angl. Noise Quality Measure – NQM) modelira oceno v odvisnosti od prostorsko variabilnega aditivnega šuma. Temelječ na Pelijevi kontrastni piramidi [279], NQM upošteva šum kot lokalne variacije v kontrastni senzitivnosti glede na razdaljo, dimenzijo slike in prostorsko frekvenco, variacije povprečne lokalne luminance, kontrastno interakcijo med prostorskimi frekvencami in efekt maskiranja kontrasta.

6.2. Območja ROI

Na sliki so po navadi območja, ki še posebej privlačijo pogled opazovalca. Osredotočenost uporabnika na ta območja, imenovana *območja interesa* (angl. Region of Interest – ROI), ima vpliv na splošno zaznavo kakovosti. Obstaja množica metod za detekcijo teh področij. Subjektivne metode definirajo lastnoročno izbiro ROI s strani opazovalca [280], nekateri uporabljajo tudi metodo sledenja oči [281], [282]. Rezultati so pokazali, da je v scenah s prikazom ljudi tipičen fokus na obrazu (slika 6.6).

Slika 6.6: Območja ROI na slikah ljudi.

To daje smiselnost porazdeljenemu vrednotenju kakovosti, kjer se polja ROI ovrednotijo na drugačen način kot območja ozadja. Eden izmed postopkov detekcije obraza je s klasifikatorjem na podlagi ekstrakcije značilk. Značilke določijo značilnosti prostorske strukture *obraz* s postopkom statističnega modeliranja. Popularna je uporaba značilk Haar [283], [284], [285], s katerimi je bila detekcija dokazano uspešna (slika 6.7).



Slika 6.7: Detekcija obraza z značilkami Haar.

6.3. Združevanje lokaliziranih vrednosti v končno oceno

Rezultat objektivne FR-metrike kakovosti, ki deluje na nivoju istoležnih slikovnih pik, so območja s prostorsko lokaliziranimi vrednostmi *objMOS* (slika 6.8).

Slika 6.8: Območja degradacij (od leve proti desni): original, kompresija JPEG, mapa z vrednostmi absolutne napake, območja z vrednostmi indeksa SSIM (vir: [286]).

Te vrednosti je potrebno združiti v skupno, tj. skalarno oceno objMOS. Metode *združevanja* (angl. pooling methods) se v večini poslužujejo preprostega povprečenja lokaliziranih vrednosti, kot npr. v [26] in [272]. Takšen pristop je računsko preprost in deluje dobro v primerih, kjer sta povprečna amplituda in lokacija degradacije enakomerni (naključna izguba). V primerih, kjer je degradacija bolj enovito prostorsko porazdeljena, pa je potreben drug mehanizem, saj HVS v takšnih primerih posveti ne-proporcionalno večjo pozornost predelu z nižjo kakovostjo [287]. Lokalizirane ocene objMOS i-tega območja slike m_i se variabilno obtežijo:

$$\boldsymbol{M} = \frac{\sum_{i=1}^{N} \boldsymbol{w_i m_i}}{\sum_{i=1}^{N} \boldsymbol{w_i}} \tag{6.30}$$

kjer je w_i utež za i-to območje slike. Utež je definirana kot funkcija odvisna od lokalne ocene objMOS, in sicer:

$$\boldsymbol{w_i} = \boldsymbol{f}(\boldsymbol{m_i}) \tag{6.31}$$

Nekateri pristopi uporabljajo indirektno metodo po modelu Minkowskega:

$$\boldsymbol{M} = \frac{\mathbf{1}}{\boldsymbol{N_{map}}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{m_i^p} \tag{6.32}$$

kjer je N_{map} število območij kakovosti na celotni sliki, p pa eksponent Minkowskega. Ta določa pomembnost degradiranih območij: z večanjem p se povečuje poudarek degradiranih območij, kar je tudi skladno s subMOS. V primeru, da je $p=1$ enačba 6.32 preide v MAE, ko je $p=2$ pa v vrednost MSE.

Združevanje na podlagi vsebine (angl. information content-weighten pooling) določi uteži na podlagi informacije iz vsebine x_i in y_i:

$$\boldsymbol{w_i} = \boldsymbol{g}(\boldsymbol{x_i}, \boldsymbol{y_i}) \tag{6.33}$$

Funkcija $g(x_i, y_i)$ se nanaša na lastnosti slike, npr. energijo kot so to naredili avtorji v [288], kjer je energijskoobtežena metoda definirana kot:

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\sigma_x^2} + \boldsymbol{\sigma_y^2} + \boldsymbol{C} \tag{6.34}$$

kjer sta σ_x^2 in σ_y^2 standardni deviaciji referenčne in degradirane slike, C pa konstanta, ki predstavlja minimalno utež sistema (šum). Funkcija $g(x_i, y_i)$ je lahko odvisna od različnih parametrov, npr. večje pomembnosti območij [88] in vizualne pozornosti [289], usmerjenosti pogleda opazovalca itd.

Izboljšane tehnike mapiranja objMOS s subMOS so ratificirale tudi različne skupnosti v standardih, kot sta ANSI T1.801.03 [290] in ITU-T J.144 [291].

6.4. Seznam obstoječih podatkovnih baz za določanje kakovosti slik in videa

Poglavitna komponenta vsake objektivne evalvacije je podatkovna baza (PB) z naborom referenčnih vrednosti (*subMOS*). V raziskovalni sferi jih obstaja več, razlikujejo se v:

- **Naboru in tipu testnih degradacij:** nabor testnih degradacij določa obseg evalvacije in specifiko opravljenih meritev. Obstoječe PB se osredotočajo na en ali dva tipa degradacij.

- **Formatu, tipu in količini testnega materiala:** količina testnega materiala daje informacijo o relevantnosti *subMOS*, saj večja količina materiala pomeni, da so tudi rezultati *subMOS* bolj generični za kompleten spekter možnih variacij karakteristik posnetkov. Kljub temu pa obseg raziskave opredeljuje določen format in tip referenčnih ter degradiranih posnetkov, kot npr: prostorska informacija, barvna globina in uporabljen barvni prostor, velikost slike, trajanje video posnetka, itd. *Prostorska informacija* (angl. Spatial Information – SI) kot indikator energije na robovih določa kompleksnost slike in posledično vpliva na percepcijo HVS. SI je definirana kot:

$$\boldsymbol{SI} = \sqrt{\frac{\boldsymbol{L}}{\boldsymbol{1080}}}\sqrt{\sum \frac{(\sqrt{\boldsymbol{s_v} + \boldsymbol{s_h}})^2}{\boldsymbol{MN}}} \tag{6.35}$$

kjer sta s_v in s_r sliki, filtrirani s Sobelovim filtrom, L vertikalna resolucija, MN pa število pik na sliki.

Barvnost (angl. colorfulness) in pomanjkanje le-te je pogost vzrok izvorne degradacije (zmanjšanje barvne globine pri transkodiranju), ki je definirano kot [292]:

$$\boldsymbol{CF} = \sqrt{\boldsymbol{\sigma_{rg}^2} + \boldsymbol{\sigma_{yb}^2}} + \boldsymbol{0.3}\sqrt{\boldsymbol{\mu_{rg}^2} + \boldsymbol{\mu_{yb}^2}} \tag{6.36}$$

kjer je $rg = R - G$, $yb = 0{,}5\,(R + G) - B$ ter R, G, B komponente barvnega prostora RGB.

- **Vsebini testnega materiala:** vsebina testnih posnetkov vpliva na všečnost testnim udeležencem (subjektivna odvisnost) ter posledično na sposobnost dojemanja HVS (objektivna odvisnost).

- **Tipu testnih metod in reprezentaciji meritev***:* uporabljene testne metode (poglavje 3) so izbrane na podlagi ciljev testne evalvacije. Rezultati so po navadi predstavljeni na enoto *posnetek* in *tip ter magnituda degradacije*, lahko pa so tudi združeni ter normirani na kombinacijo le-teh. *SubMOS* je predstavljena na linearni skali, tako da je možno z linearno transformacijo le-te primerjati z drugimi eksperimenti. Območji *subMOS* in *objMOS* sta določeni z:

$$\boldsymbol{R^{subjMOS} = P_{100-n}^{subjMOS} - P_{n}^{subjMOS}} \tag{6.37}$$

$$\boldsymbol{R^{objMOS} = P_{100-n}^{objMOS} - P_{n}^{objMOS}} \tag{6.38}$$

kjer sta $P_n^{subjMOS}$ in P_n^{objMOS} deleža *subMOS* in *objMOS* izmed vseh rezultatov. Avtor v [293] je izbral $n = 5$.

- **Karakteristikah testnih udeležencev:** karakteristike sodelujočih, kot npr. število udeležencev, starost, spol, predznanje, so izbrane tako, da so končni rezultati od njih čim manj odvisni in jih lahko posplošimo na večjo množico.

- **Pogojih subjektivnih meritev:** pogoji kot npr. velikost zaslona in ostale opreme, oddaljenost od zaslona, svetlost prostora testiranja, čas ocenjevanja, so določeni s standardi in načeloma ne-relevantni na končni rezultat.

Obširni seznam podatkovnih baz vodi konzorcij *QUALINET* (»European network on quality of experience in multimedia systems and services«) [294]. Avtor je v [293] naredil analizo teh baz, na podlagi katere je določil kriterije za kvantitativno primerjavo testne vsebine, testnih pogojev in subjektivnega ocenjevanja. V nadaljevanju so predstavljene pogosto uporabljane in prosto-dostopne slikovne/video podatkovne baze.

1) *LIVE – Laboratory for Image and Video Engineering image quality database* [295], [296]. Vsebuje 29 referenčnih in 779 degradiranih 24-bitnih barvnih slik tipa BMP. Resolucija slik sega od 634x438 do 768x512 px. Vsebuje 5 degradacijskih tipov: kompresija JPEG, kompresija JPEG2000, aditivni Gaussov šum, Gaussovo zamegljenost, JPEG2000 slike z bitnimi napakami simuliranimi

z Rayleigh-jevim propagacijskim kanalom. Vsak tip degradacije je generiran s 5-6 različnimi nivoji napak. Pri evalvaciji je sodelovalo 29 udeležencev.

2) *EPFL - PoliMI Video Quality Assessment Database* [297], [298]. Vsebuje 156 10-sekundnih video sekvenc kodiranih v H.264/AVC in degradiranih s simulacijo izgube paketov. Polovica posnetkov ima resolucijo CIF, druga polovica pa QCIF. Pri evalvaciji je sodelovalo 40 testnih oseb iz 2 akademskih inštitucij.

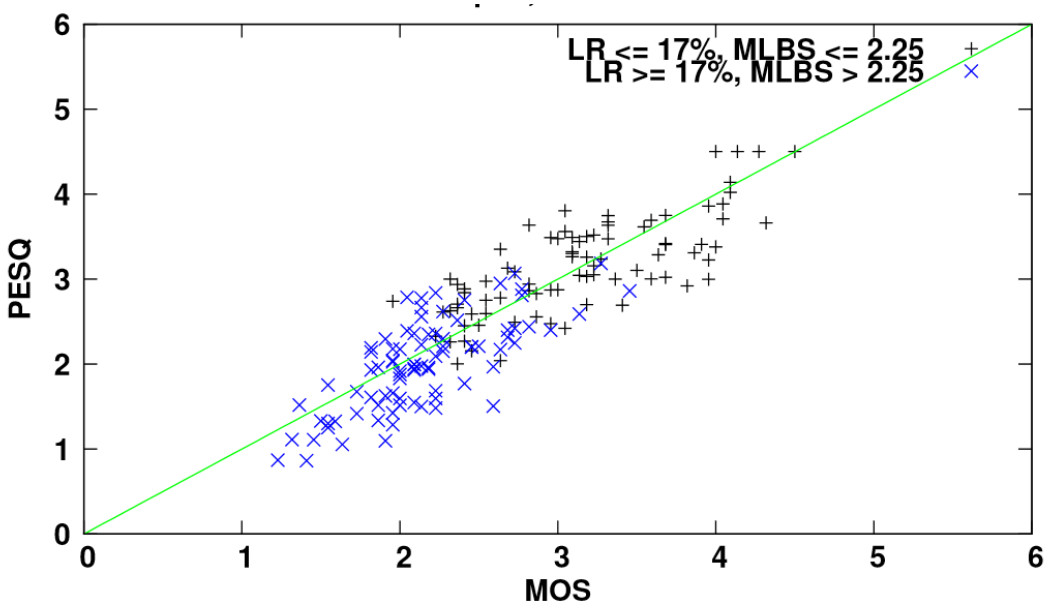
3) *WIQ - Wireless Imaging Quality Database* [299], [300]. Baza vsebuje 7 referenčnih in 80 degradiranih slik. Degradacije so simulirane na brezžičnem prenosnem kanalu v obliki bitnih napak in izgube omrežnih paketov. Gradivo se nahaja v 8-bitnem formatu BMP, tj. črno-bele slike z velikostjo *512x512* slikovnih pik. Subjektivni rezultati so pridobljeni od 30 testnih udeležencev v 2 ločenih testiranjih.

4) *TID2013 – Tampere image database* [301]. TID2013 vsebuje 25 referenčnih slik, vsaka je simulirana s 24 tipi degradacij s 5 nivoji. Od tega se 2 tipa degradacij poslužujeta napak na prenosni poti (»JPEG transmission errors« in »JPEG2000 transmission errors«). Skupno 3000 testnih slik je shranjenih v BMP formatu brez kompresije. Pri evalvaciji je sodelovalo 971 testnih oseb iz petih držav, ki so zagotovile 524.340 primerjalnih MOS ocen in 1.048.680 relativnih ocen (DMOS+σ) parov slik.

5) *Poly@NYU Packet Loss Database* [302]. Vsebuje 17 videov kodiranih s standardom H.264 pod vplivom izgube paketov na prenosni poti. Pri evalvaciji je sodelovalo 32 testnih oseb, ki so ocenjevali posnetke dolžine 2 sekund. Uporabljena testna metoda je SSCQS.

6.5. Objektivne avdio metrike s polno referenco

Obstaja več avdio metrik, pri tem so za to disertacijo pomembne tiste, ki ovrednotijo kakovost govora in delujejo v načinu s polno referenco. Najenostavnejša metoda je razlika signalov oz. izračun razmerja signal-šum. Tak koncept so nekateri razširili z modeli, ki so upoštevali lastnosti HAS, recimo kodirne degradacije, vzorce in distribucije pojave degradacij in podobno [303]. V začetku 1990 je ITU ustanovil komite z namenom razvoja objektivnih avdio metrik tako za zvok (ITU-T BS.1387), kot tudi posebej za govor. Te, ki ovrednotijo govor na modelu HAS sodijo v serijo standardov P.86X. Znane so naslednje: P.861 (PSQM), P.862 (PESQ), P.862.2(PESQ-WB) ter P.863 (POLQA). Najbolj popularna med njimi je metoda PESQ, ki zaradi dobro zasnovanega kognitivnega modela deluje boljše od predhodnika PSQM. Zasnovan kognitivni model, obravnava percepcijo in procesiranje zvočnih signalov v primarnem avditornem korteksu. Zato ocena PESQ daje dobro korelacijo s subjektivnimi testi tudi na širokem razponu degradacijskih vrednosti in dveh tipov porazdelitve na kodeku G.711 (slika 6.9). Slabost algoritma ocene PESQ je, da je preveč optimističen za nižje vrednosti MOS in preveč pesimističen za višje ocene v primerjavi s subjektivnimi ocenami. V ta namen je bila zasnovana tabela mapiranja vrednosti, ki bolje korelira z MOS: PESQ-LQO. Nekateri omenjajo, da je korelacija PESQ s subjektivnimi testi v povprečju 0,942 pri prenosu v fiksnem omrežju, 0,921 pri internetni telefoniji (VoIP) in 0,962 v mobilnem omrežju [56].



Slika 6.9: Korelacija vrednosti PESQ in MOS, *LR* je izguba paketov, *MLBS* pa povprečna velikost izgube rafala paketov (vir: [304]).

Širokopasovna razširitev PESQ, tj. PESQ-WB je primerna za govorne signale v avdiu z večjo pasovno širino, tj. do 16k Hz. Naslednik PESQ je standard POLQA, ki pa, zaradi kratkega časa obstoja v času zasnove disertacije, ni bila aktualna rešitev. POLQA bi naj odpravila omejitve PESQ, zato bi lahko v prihodnjih raziskavah nadomestila metriko PESQ [305].

6.6. Korelacija subjektivnih in objektivnih rezultatov

Različne testne metode in načini evalvacije dajo različne rezultate, ki jih je potrebno preslikati v normalizirane vrednosti primerne za primerjavo statističnih podatkov *objMOS* in *subMOS*. Kako dobro se izmerjeni podatki ujemajo določajo modeli statistične korelacije. Te je možno enostavno izmeriti z meritvami MSE in RMSE. Ko se z raziskovalno hipotezo pričakuje linearna odvisnost kvantitativnih podatkov pa se odvisnost (negativna, pozitivna ali nična) običajno ugotavlja s *Pearsonovim linearnim koeficientom korelacije* (angl. Pearson Linear Correlation Coefficient - PLCC), ki je pogosta oblika ugotavljanja metrične natančnosti pri evalvaciji videa [306]. PLCC je definiran kot:

$$\boldsymbol{PLCC} = \frac{\sum_{i=1}^{N}(\boldsymbol{subjMOS_i} - \overline{\boldsymbol{subjMOS}})(\boldsymbol{objMOS_i} - \overline{\boldsymbol{objMOS}})}{\sqrt{\sum_{i=1}^{N}(\boldsymbol{subMOS_i} - \overline{\boldsymbol{subMOS}})^2}\sqrt{\sum_{i=1}^{N}(\boldsymbol{objMOS_i} - \overline{\boldsymbol{objMOS}})^2}} \tag{6.39}$$

kjer je *N* število primerjanih parov ocen. V primeru, da so podatki *objMOS* in *subMOS* nelinearni, a kljub temu statistično monotono odvisni, relacijo njunih rangiranih vrednosti raje definiramo s *Spearmanovim koeficientom korelacije* (angl. Spearman's Rank Correlation Coefficient - SRCC). SRCC definiramo kot:

$$\boldsymbol{SRCC} = \mathbf{1} - \frac{\mathbf{6}\sum_{i=1}^{N}\boldsymbol{d_i^2}}{\boldsymbol{N}(\boldsymbol{N^2}-\mathbf{1})} \tag{6.40}$$

kjer je d_i^2 razlika med *i*-im rangom objektivne in subjektivne evalvacije ter *N* število razredov.

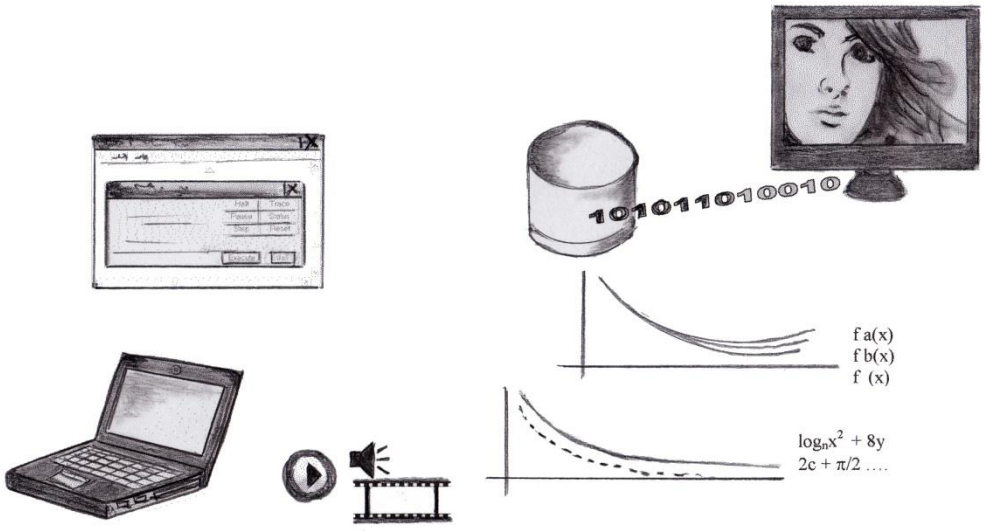
Predikcijo monotonosti ugotavlja tudi *Kendallov koeficient korelacije* (angl. Kendall Rank Correlation Coefficient - KRCC). Ne-parametrični koeficient τ je definiran kot:

$$\boldsymbol{\tau} = \frac{\boldsymbol{N_c} - \boldsymbol{N_d}}{\frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{N}(\boldsymbol{N}-\mathbf{1})} \tag{6.41}$$

kjer je N_c število skladnih in N_d neskladnih parov za katere velja:

$$(\boldsymbol{x_i}, \boldsymbol{y_i}) \;\&\&\; (\boldsymbol{x_j}, \boldsymbol{y_j}) \in \boldsymbol{N_c}, \text{če: } (\boldsymbol{x_i} < \boldsymbol{x_j} \;\&\&\; \boldsymbol{y_i} < \boldsymbol{y_j}) \;||\; (\boldsymbol{x_i} > \boldsymbol{x_j} \;\&\&\; \boldsymbol{y_i} > \boldsymbol{y_j}) \tag{6.42}$$

$$(\boldsymbol{x_i}, \boldsymbol{y_i}) \;\&\&\; (\boldsymbol{x_j}, \boldsymbol{y_j}) \in \boldsymbol{N_d}, \text{če: } (\boldsymbol{x_i} < \boldsymbol{x_j} \;\&\&\; \boldsymbol{y_i} > \boldsymbol{y_j}) \;||\; (\boldsymbol{x_i} > \boldsymbol{x_j} \;\&\&\; \boldsymbol{y_i} < \boldsymbol{y_j}) \tag{6.43}$$



7. Zasnova eksperimentalnega sistema za vrednotenje kakovosti večmodalnih storitev

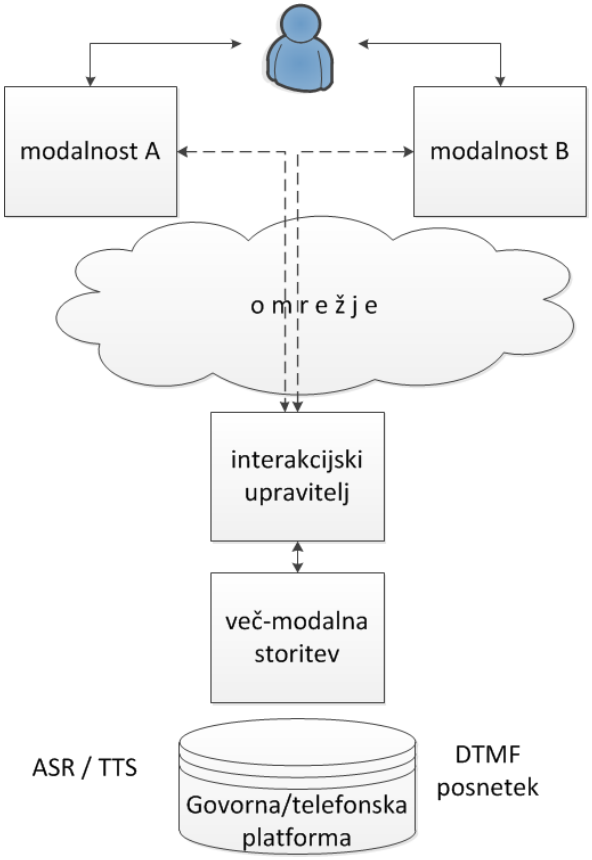
V tem poglavju bomo predstavili zasnovo sistema za vrednotenje kakovosti večmodalnih storitev. Poglavje smo razdelili na tri podpoglavja, kjer se vsako ukvarja z ločeno izpostavljeno problematiko pri vrednotenju večmodalne kakovosti. Prvo podpoglavje opisuje tipičen primer delovanja napredne storitve IVR, kjer poteka interakcija od uporabnika do storitve. Tip modalnosti, ki jo uporabnik v takšnem primeru lahko uporabi, je odvisen tudi od degradacij na prenosni poti ter od uporabljene strojne opreme (mikrofon, uporaba govornega kodeka). Prednostno je to uporaba govora, vendar lahko degradacije povzročijo občutno poslabšanje uspešnosti razpoznavanja govora, zato je v takšnem primeru bolje uporabiti robustnejšo vhodno modalnost. Na podlagi analize govornih podatkov z različnimi scenariji degradacije bomo predlagali klasifikator izbire vhodne modalnosti z Gaussovimi modeli, primeren za takšen sistem IVR.

Drugo podpoglavje obravnava vpliv omrežnih degradacij na kakovost izhodnih modalnosti sodobne TK storitve. Za ta namen bomo zasnovali podatkovno bazo HD posnetkov, kjer bomo uporabili tri tipe posnetkov glede na modalnost: avdio (A), video (V) in avdio-video (AV). Posnetki bodo služili za referenčno točko pri izdelavi modela za vrednotenje objektivne kakovosti. Pri tem bo model objektivnega evalvatorja upošteval tip modalnosti, medmodalni učinek, scenski tip posnetka v videu ter količino omrežnih degradacij.

Tretje podpoglavje obravnava porazdeljeno vrednotenje kakovosti, kjer bomo uporabili rezultate vpliva degradacij na kakovost izhodne vizualne modalnosti sistema. Z boljšo korelacijo objektivne metrike slik, ki je posledica upoštevanja HVS, se eksponentno povečuje tudi kompleksnost algoritmov. Tukaj bomo predstavili koncept porazdeljenega vrednotenja kakovosti, kjer bomo najprej detektirali področja ROI, tj. obraz, nato pa predlagali ločeno evalvacijo ROI in ne-ROI področij z namenom, natančnejšega in računsko učinkovitejšega pristopa.

7.1. Določanje vpliva degradacij na vhodno modalnost sistema

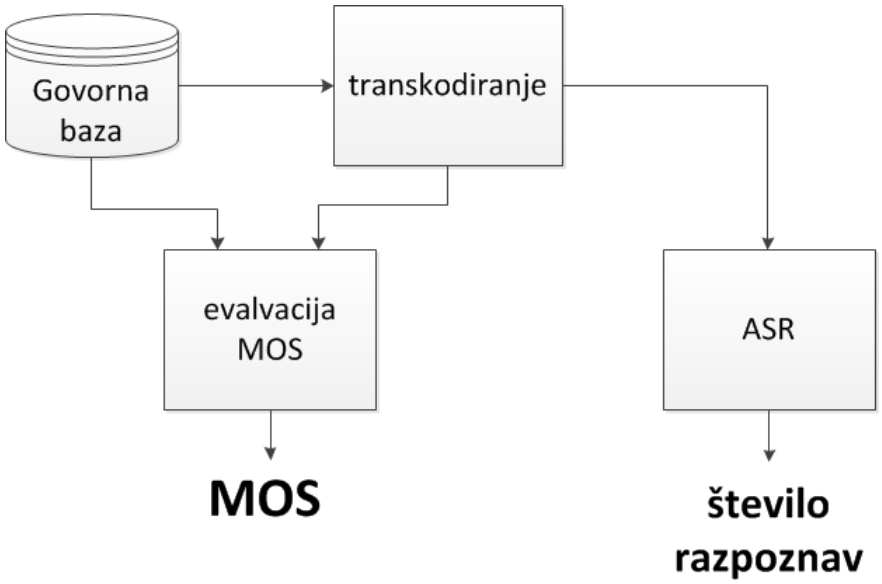
V tem podpoglavju bomo obravnavali vpliv izvornih in omrežnih degradacij na vhodno modalnost večmodalnega sistema. Kot primer bomo uporabili storitev IVR, ki uporabniku omogoča vnos govornih ali/in tonskih (DTMF) signalov (slika 7.1). Privzeta nastavitev vhodnih podatkov je govor, kjer ASR modul skrbi za pretvorbo govornih signalov v digitalno obliko. Interaktivni upravitelj dialoga mora poskrbeti za visok nivo uporabniške kakovosti, kar pomeni, da je v primeru poslabšanja signala potrebno preklopiti na robustnejšo metodo vnosa, tj. DTMF vnos, saj ASR v takšnem primeru ne zagotavlja zadovoljive uspešnosti razpoznavanja govornih ukazov. Do poslabšanja signala pride zaradi transkodiranja ali omrežnih degradacij. Cilj raziskave je zato predlagati model klasifikatorja vhodne modalnosti v odvisnosti od stopnje degradacij vhodnega govornega signala.



Slika 7.1: Sistem IVR kot primer večmodalne storitve.

Simulacija uporabe sistema IVR je najprej zahtevala natančno analizo vpliva degradacij na izbiro vhodne modalnosti. Simulacijo smo razdelili v tri faze.

Prva faza simulacije je bila namenjena evalvaciji izvornih, tj. transkodiranih govornih posnetkov in vplivu izvorne degradacije na storitev IVR. Uporabljena je bila standardizirana govorna baza *1000 FDB SpeechDat(II)* [307], ki je namenjena razvoju ASR za glasovno vodene sisteme IVR (slika 7.2).



Slika 7.2: Simulacijsko okolje za določanje vpliva degradacij na storitev IVR.

Govorna baza je vsebovala nabor ukaznih besed za krmiljenje storitve govorne pošte (angl. voiceMail) z lastnostmi prikazanimi v tabeli 7.1.

Tabela 7.1: Govorna baza.

Lastnost	**Vrednost**
Velikost	1070 posnetkov
Format	PCM
Število kanalov	1 (mono)
Vzorčenje zvočnega signala	8 kHz, 16-bit
Dolžina posnetka	4,096 sek.
Snemalno okolje	Naravno, telefonski kanal
Nabor govorcev	171, mešano M in Ž

Orodja za transkodiranje, ki posnemajo delovanje govornih kodekov v komunikacijskem sistemu, so bila generirana iz ANSI C izvorne kode, pridobljene na uradnih straneh standardizacijskih inštitucij ITU in 3GPP. Uporabili smo znane in

pogosto uporabljene govorne kodeke, ki se uporabljajo v govornih aplikacijah, npr. v VoIP. Testni režim je obsegal standarde in načine kodekov kot so prikazani v tabeli 7.2. Ker smo želeli dobiti informacijo o vplivu različnih mehanizmov kodekov, predvsem različne pasovne širine, je testni režim obsegal različne konfiguracije.

Tabela 7.2: Kodeki in načini delovanja.

Scenarij	Govorni kodek	Konfiguracija
1	G.722	64 kbps
2	G.722	56 kbps
3	G.722	48 kbps
4	G.726	40 kbps
5	G.726	32 kbps
6	G.726	24 kbps
7	G.726	16 kbps
8	G.727	(4, 1)[2]
9	G.727	(4, 0)[2]
10	G.727	(3, 2)[2]
11	G.727	(3, 1)[2]
12	G.727	(3, 0)[2]
13	G.727	(2, 3)[2]
14	G.727	(2, 2)[2]
15	G.727	(2, 1)[2]
16	G.727	(2, 0)[2]
17	G.723.1	6,3 kbps, PF, HPF, VAD/CNG[3]
18	G.723.1	6,3 kbps, PF, HPF
19	G.723.1	6,3 kbps
21	G.723.1	5,3 kbps, PF, HPF, VAD/CNG
22	G.723.1	5,3 kbps, PF, HPF
23	G.723.1	5,3 kbps
24	G.729A	8 kbps
25	AMR	12,2 kbps
26	AMR	10,2 kbps
27	AMR	7,95 kbps
28	AMR	7,4 kbps
29	AMR	6,7 kbps
30	AMR	5,9 kbps
31	AMR	5,15 kbps
32	AMR	4,75 kbps

Za razpoznavanje govora je bila uporabljena konfiguracija slovenskega ASR, ki je nastala v okviru iniciative MASPER [308] in je primerljiva s komercialnimi rešitvami

[2] Označuje (*število_ jedrnih_bitov*, *število_dodatnih_bitov)*. Efektivna hitrost kodeka se izračuna po formuli:
16*(*število_ jedrnih_bitov* + *število_dodatnih_bitov*) kbps.
[3] PF (angl post filter), HPF (angl. High-Pass filter), VAD/CNG (angl. Voice Activity Detection/ Comfort Noise Generation).

za slovenski jezik. Uporabljen ASR je deloval v *od govorca neodvisnem* načinu delovanja, ki deluje na HMM modelih, ki so aplicirani kot kontekstno-odvisni akustični modeli. Uspešnost ASR je bila merjena s pomočjo merjenja pravilnosti razpoznavanja govornih ukazov:

$$\boldsymbol{Acc_{ASR}}(\%) = \frac{\boldsymbol{H}}{\boldsymbol{N}} \times \mathbf{100} \tag{7.1}$$

kjer je *N* število vseh in *H* število pravilno razpoznanih govornih ukazov.

Rezultate ASR smo primerjali z objektivno oceno vrednotenja kakovosti testnih posnetkov. Objektivna ocena kakovosti govornih posnetkov je bila določena z oceno PESQ MOS, kjer *4,5* predstavlja maksimalno in *-0,5* minimalno vrednost. Ocena MOS je bila nato izračunana s translacijo vrednosti definirano v standardu ITU-T P.682.1, da smo dobili vrednosti MOSLQO.

Druga faza simulacije je zajemala tandemski učinek transkodiranja in omrežne degradacije na istem naboru posnetkov govornih ukazov. Da smo definirali model omrežja, smo potrebovali generator napak, ki implementira želene vzorce omrežnega prometa. V ta namen smo omrežne degradacije emulirali z namensko strojno opremo Simena NE100 (slika 7.3).

Slika 7.3: Emulator omrežja Simena NE100.

Testne scenarije smo dobili tako, da smo podatke iz prvega dela simulacije okvarili z izgubo paketov, pri tem pa smo zajeli širok spekter vrednosti izgubljenih paketov (tabela 7.3). Distribucija izgube je bila v vseh scenarijih po neodvisnem Bernoullijevem modelu, kar je dalo povprečen odziv govornega kodeka za dan degradacijski faktor, za katere so pogoji degradacije konstantni.

Tabela 7.3: Količina izgubljenih paketov.

Scenarij	Izgubljeni paketi [%]	Distribucija izgube
1_PL	1	Bernoulli
2_PL	2	Bernoulli
5_PL	5	Bernoulli
10_PL	10	Bernoulli
15_PL	15	Bernoulli
20_PL	20	Bernoulli
25_PL	25	Bernoulli
30_PL	30	Bernoulli
35_PL	35	Bernoulli

Tretja faza simulacije je bila namenjena izdelavi primernega klasifikatorja za določanje vhodne modalnosti storitve IVR. Sestavljena je bila iz *učne* in *testne* faze. V učni fazi sta bila določena dva razreda modalnosti, ki določata govorni, tj. *SPEECH*, in tonski vnos, tj. *DTMF* (tabela 7.4). Iz originalne učne baze je bil po transkodiranju in izgubi paketov naključno izbran nabor množice 15.750 učnih posnetkov. Za klasifikacijo smo uporabili klasifikator z Gaussovimi modeli (GMM), kjer smo kompleksnost modela korakoma povečevali do 1024 funkcij gostote verjetnosti. V test smo kasneje vključili modele s 64 in 1024 funkcijami gostote verjetnosti in opazovali razliko med rezultati za skrajne vrednosti obeh modelov.

Tabela 7.4: Učna faza.

Količina	Vrednost
Posnetkov originalne testne množice	8.491
Učnih posnetkov	15.750
Razredov za klasifikacijo	2 (*SPEECH, DTMF*)

Referenčni posnetki iz baze *SpeechDat(II)* so vsebovali relativno dolgo tišino pred in za govornim signalom. Da bi preučili vpliv tega na klasifikacijo modalnosti smo pripravili tri scenarije, kjer smo tišino pustili v originalni obliki oziroma jo ustrezno skrajšali (tabela 7.5). Pričakovali smo, da to lahko vpliva na lastnosti klasifikatorja, saj ima zvočni signal govora in ne-govora precej drugačno karakteristiko, kar bi posledično pomenilo tudi odstopanje pri uspešnosti delovanja klasifikatorja v testni fazi.

Tabela 7.5: Konfiguracija rezanja posnetkov v učni in testni fazi.

Scenarij	**Rezanje posnetka *pred/po* [ms]**
1	250/250
2	500/500
3	Brez

Za izločanje glasovnih značilk pri učenju klasifikatorja je bil uporabljen algoritem mel kepstralnih koeficientov (MFCC), ki jim je bila dodana še energija. Pri tem sta bila testirana dva primera: prvi s 13 in drugi s 39 dimenzionalnim MFCC vektorjem, kjer smo v drugem primeru dodali še vrednosti prvih in drugih odvodov.

Za predstavljene scenarije in konfiguracije GMM modelov smo izvedli ločene postopke učenja z Baum-Welchevim algoritmom re-estimacije vrednosti parametrov. Pri tem smo za vsako povečanje števila funkcij gostote verjetnosti GMM modela opravili tri iteracije učenja.

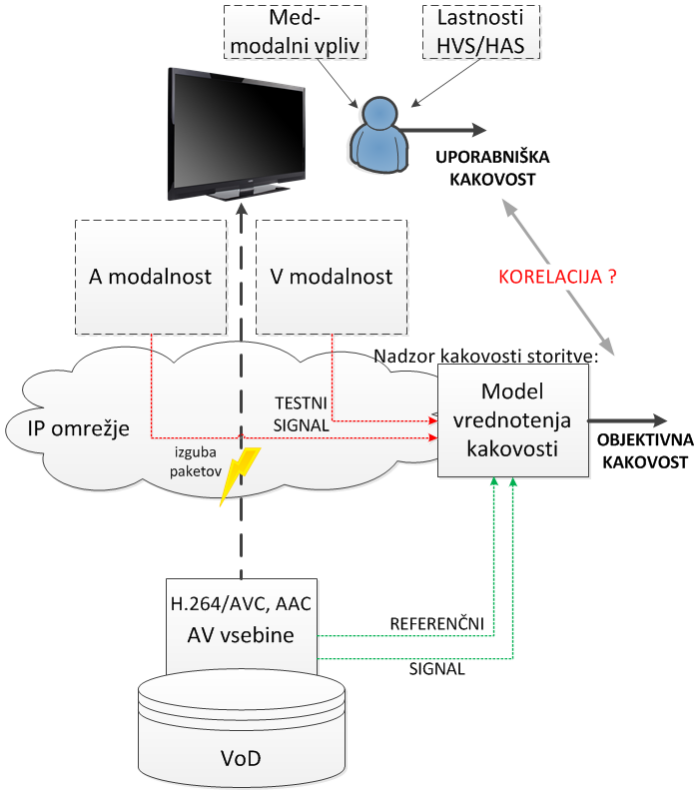
Pri uporabi klasifikatorja prihaja v rezultatih do kratko časovnih prehodov med razredoma zaradi napak klasifikacije. Zato predlagamo uporabo filtra postprocesiranja rezultatov s funkcijo povprečenja na drsečih oknih velikosti 1, 10, 20, 30, 50 ali 100 okvirjev. Velikosti oken smo izbrali tako, da smo lahko primerjali učinek povprečenja na danih govornih signalih. Prag med razredi modalnosti je bil določen s količino zanj detektiranih okvirjev znotraj vsakega okna (algoritem 7.1). Znotraj vsakega okna smo testirali tudi različne pragovne vrednosti (0, 5, 10, 20, 30, 40% velikosti okna) z namenom doseganja statičnosti izbrane modalnosti za vsako okno. Kot rezultat postprocesiranja s predlaganim algoritmom povprečenja smo dobili klasifikacijo celotnega testnega posnetka v eno izmed modalnosti.

Algoritem 7.1: Funkcija povprečenja v filtru postprocesiranja rezultatov.

```
1: procedure AVERAGING(windowSize, threshold)
2:
3:     windowSize ∈ [1, 10, 20, 30, 50, 100], frames
4:     threshold ∈ [0, 5, 10, 20, 30, 40], %windowSize
5:     iFrame = 1
6:     dtmfWindow = 0, speechWindow = 0
7:
8:     for iFrame ≤ ∑ frameInTestFile do
9:         if iFrame mod windowSize == 0 then          ▷ block of windowSize frames
10:            if iSpeech/(iSpeech + iDtmf) ≤ threshold then
11:                dtmfWindow + +
12:            else if iDtmf/(iSpeech + iDtmf) ≤ threshold then
13:                speechWindow + +
14:            end if
15:        end if
16:        iFrame + +
17:    end for
18:
19:    return (speechWindow ≥ dtmfWindow)?"SPEECH" : "DTMF"
20: end procedure
```

7.2. Določanje vpliva degradacij na izhodno modalnost sistema

V tem podpoglavju smo analizirali vpliv omrežne izgube paketov na uporabniško izkušnjo. Pri tem smo obravnavali avditorno, tj. govorno in vizualno, tj. video modalnost kot najpomembnejši modalnosti v sodobnih telekomunikacijskih sistemih. Pri prenosu multimedije (avdio-video) v omrežju prihaja do naključnih degradacij, pri tem pa so, iz stališča uporabnika storitve, še posebej zaznavne izgube IP paketov. Da ohranimo uporabniško izkušnjo, je zato potreben natančen nadzor nad prenosom multimedijskih vsebin. To dosežemo z integracijo napovedovalnega modela vrednotenja kakovosti, ki na podlagi objektivnih meritev odloča o nivojih kakovosti storitve. Pri tem morajo objektivni rezultati kar najbolj relevantno predstavljati povprečen odziv uporabnika, hkrati pa upoštevati omrežne razmere, lastnosti pretočenih vsebin ter biološke karakteristike uporabnika, kot je vpliv medmodalne interakcije in posledice dojemanja kakovosti pri fuziji modalnosti (slika 7.4). Cilj te raziskave je tako predlagati model za objektivno oceno kakovosti večmodalnih storitev.



Slika 7.4: Nadzor kakovosti izhodne modalnosti večmodalne storitve.

Najprej smo izmerili subjektivno kakovost multimedijskih vsebin, ki nam je služila za referenčno točko pri nadaljnji izgradnji modela vrednotenja kakovosti. Referenčna avdio-vizualna (AV) podatkovna baza je nastala na osnovi posnetkov visoke ločljivosti (tabela 7.6). Posnetki so bili za raziskovalne namene pridobljeni iz komercialnega HDTV programa »Slovenija 1« in »Slovenija 2« distributerja RTV SLO.

Tabela 7.6: AV podatkovna baza.

Splošno	
Število različnih scen	4 (film, formula, intervju, nogomet)
Število različnih posnetkov/sceno	4
Dolžina posnetkov	~ 20 sek.
Video kodek	MPEG-4 Part 10/ AVC
Barvni prostor	YUV
Kromatično vzorčenje	4:2:0
Barvna globina	3 × 8 bitov
Št. slik/okvirjev na sekundo	25 fps, konstantno
Način prikaza slik	Progresivni
Ločljivost	1920 × 1080
Razmerje stranic	16:9
Profil kodeka	Main/High
Nivo kodeka	4.2/4.0
Način bitnega toka	VBR
Bitna hitrost podat. toka	~ 14mbit/s
Kodek za zvok	MPEG-4 Part 3/ AAC
Frekvenca vzorčenja	48 kHz
Št. kanalov	2
Profil kodeka	Main
Način bitnega toka	CBR
Bitna hitrost toka	192 kbit/s

Izbrane scene so bile kategorizirane v 4 tipe, razdeljene na podlagi skupnih vizualnih karakteristik (tabela 7.7). Pri izbiri scen smo težili k vključitvi tipičnih lastnosti, ki jih najdemo v medijskem materialu. Pri kategorizaciji smo lastnosti ocenili po stopnji od 1 (najmanj prisotno) do 5 (najbolj prisotno). Različne lastnosti zaradi karakteristik H.264 izražajo drugačne odzive tudi na izgubo paketov, npr. zaradi različne količine sidrnih okvirjev, količine gibalnih vektorjev in podobno. Zato posredno vplivajo na tehnološko komponento vpliva degradacije. Dodatno so različne scene podvržene različnemu subjektivnemu dojemanju, npr. osredotočenost na vizualna polja je drugačna, kar posledično vpliva na zaznavo degradacij v testnem posnetku.

Tabela 7.7: Kategorizacija scenskih tipov.

Lastnost [1-5]	**Scenski tip**			
	film	**formula**	**intervju**	**nogomet**
- gibanje objektov	4	5	1	4
- gibanje kamere	4	5	1	4
- prevlada ene barve	1	2	3 (prevladuje modra)	5 (prevladuje zelena)
- velikost objektov	1-5	3	4	1-2
- govor	1-5	4	5	5
- umetni zvok	1-5	1-2	1	1

Priprava degradiranih posnetkov AV baze je bila emulirana z namensko strojno opremo Simena NE100. Degradirani posnetki so bili razdeljeni v tri skupine:

- **Degradacija avditorne modalnosti (A),**
- **Degradacija vizualne modalnosti (V),**
- **Degradacija avditorno-vizualne modalnosti (AV).**

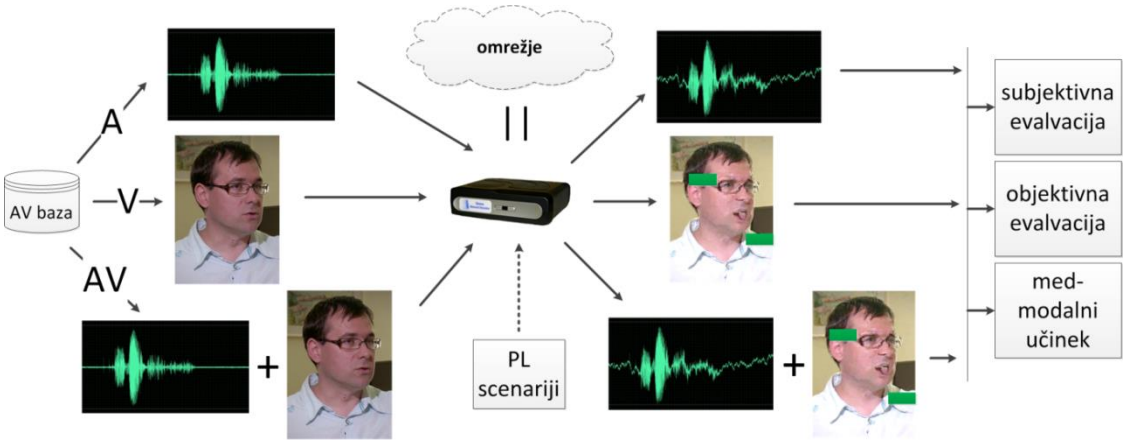
Pri tem je vsaka skupina obravnavala celoten nabor originalnih posnetkov (16), vendar samo za omenjeno modalnost za 5 posameznih scenarijev izgube paketov (tabela 7.8). Pri tem so scenariji A0, V0 in AV0 služili za slepo preverjanje, s katerimi smo vrednotili »verodostojnost« referenčnih ocen testnih oseb. Pri izbiri količine izgube paketov za posamezno modalnost smo izbrali razpon vrednosti, ki sorazmerno ustreza pričakovani subjektivni oceni od zaznavno majhne do zaznavno velike degradacije, ter istočasno ne presega števila robnih vrednosti kot določa standard. A, V in AV imajo skupne 3 scenarije: *PL=0,00%, PL=0,20%* in *PL=0,50%*, kar bo služilo za določanje medmodalnega vpliva.

Uporabljena distribucija izgube paketov in vpliv na H.264 tok podatkov sta predstavljena v prilogi I. V evalvacijo smo tako vključili **240 testnih posnetkov**, ki so bili naključno razporejeni (priloga II). Število testnih posnetkov smo morali ustrezno omejiti, saj bi v nasprotnem primeru subjektivni testi trajali predolgo.

Tabela 7.8: Scenariji degradacije izhodne AV modalnosti.

A		**V**		**AV**	
Scenarij	PL [%]	Scenarij	PL [%]	Scenarij	PL [%]
A0	0,00	**V0**	0,00	**AV0**	0,00
A1	0,20	**V1**	0,01	**AV1**	0,01
A2	0,50	**V2**	0,10	**AV2**	0,10
A3	1,00	**V3**	0,20	**AV3**	0,20
A4	5,00	**V4**	0,50	**AV4**	0,50

Z namenom izgradnje objektivnega modela za vrednotenje kakovosti večmodalnih vsebin smo raziskavo razdelili v tri dele: prvi del je zajemal *ocenjevanje subjektivne kakovosti*, kjer smo ocenili posnetke iz referenčne podatkovne baze. Drugi del je zajemal *ocenjevanje objektivne kakovosti*, kjer smo objektivno ocenili posnetke z objektivnimi metodami ter izbrali najboljšo za posamezno modalnost. Tretji del je obsegal *ocenjevanje medmodalnega vpliva* v katerem smo predlagali model vrednotenja kakovosti zmožen napovedati objektivno večmodalno oceno (slika 7.5).



Slika 7.5: Sistem za evalvacijo vpliva degradacij na izhodno modalnost.

Ocenjevanje subjektivne kakovosti je potekalo v obliki subjektivnih testov, pri čimer so se upoštevala priporočila po standardih ITU-T P.910, ITU-R BT.500 in ITU-R BT.710 (pogoji opazovanja za HDTV, subjektivne metode za HDTV), kot je prikazano v tabeli 7.9.

Tabela 7.9: Subjektivni testi.

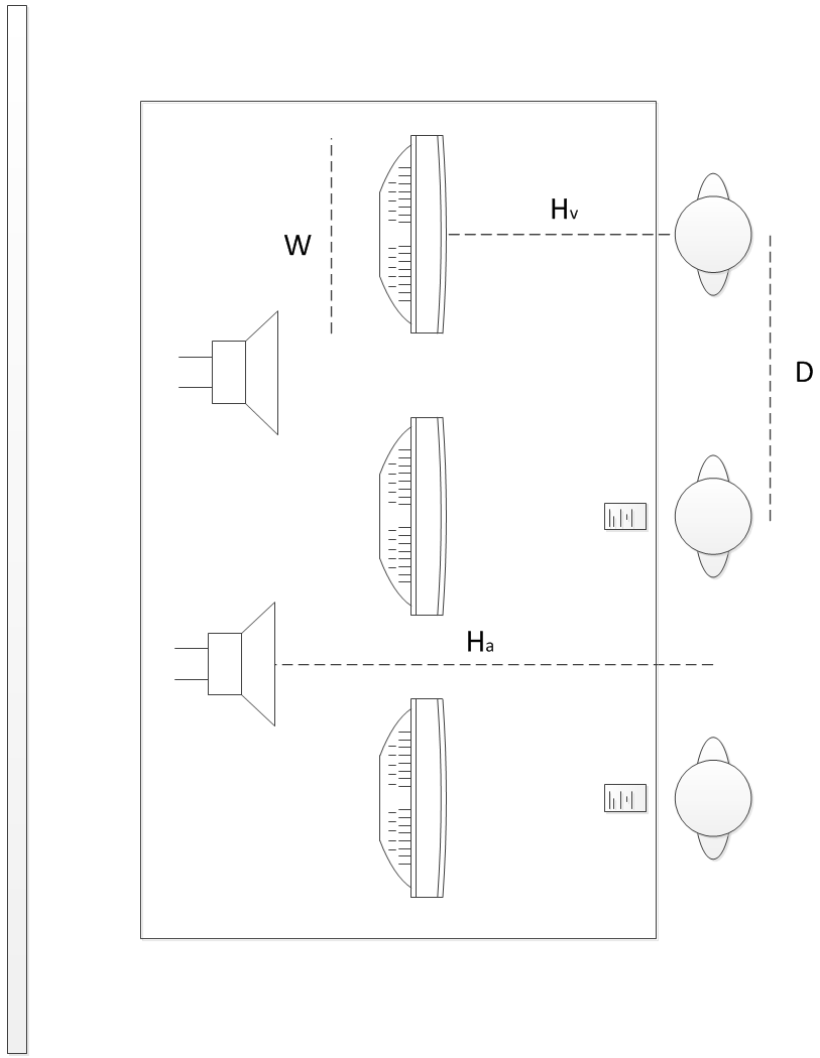
Lastnost	**Vrednost**
Število ocenjevalcev (priloga III)	20
Število testnih posnetkov/ ocenjevalca	240
Razpon vrednosti	5-stopenjski DMOS
Metoda ocenjevanja in lestvica	DCR/ DSIS II (ITU-R BT.500-11)
H_v (oddaljenost od vizualnega izvora)	0,7-1,0 m
H_a (oddaljenost od zvočnega izvora)	1,5 m
D (povprečna oddaljenost med ocenjevalci)	1 m
W (velikost zaslona opazovanja)	24 "
monitor	Philips Brilliance 240BW

V skladu s priporočili smo uporabili 5-stopenjsko lestvico degradacije kakovosti, s katero so ocenjevalci ocenjevali kakovost testnega posnetka v odvisnosti od referenčnega. Ocena posnetkov v anketnih listih je definirala lestvico, ki je vključevala tudi semantične ocene (tabela 7.10):

Tabela 7.10: Subjektivna kategorijska lestvica.

Številska ocena	**Semantična vrednost oz. ocena degradacije**
5	nezaznavno
4	zaznavno, ampak nemoteče
3	rahlo moteče
2	precej moteče
1	zelo moteče

Seje so potekale hkrati za 2 do 3 testne subjekte, ki so bili primerno razporejeni v prostoru (slika 7.6).



Slika 7.6: Postavitev okolja za subjektivno testiranje, pri tem je W – velikost zaslona, D – oddaljenost med opazovalci, H_a – oddaljenost od zvočnega izvora, H_v – oddaljenost od vizualnega izvora.

Ocenjevanje objektivne kakovosti je obsegalo analizo avditorne in vizualne informacije degradiranih posnetkov z objektivnimi metrikami kakovosti (tabela 7.11). Za govorno metriko smo izbrali PESQ, ki je standard za vrednotenje kakovosti govornih signalov. Pri video metrikah je bila naloga težja. Izbrali smo nabor 26 metrik za objektivno vrednotenje kakovosti slik, ki so primerne za zadano nalogo. Ker se uspešnosti posameznih metrik med seboj močno razlikujejo glede na uporabo v specifičnih že obstoječih podatkovnih zbirkah, smo pri analizi izbrali tisto, ki je za naš tip podatkov izkazovala najboljšo korelacijo s subjektivnimi ocenami.

Tabela 7.11: Uporabljene objektivne avdio (A) in video (V) metrike za evalvacijo.

A	**V**	
PESQ	PSNR	VIFP
	MSE	VSNR
	LMSE	WSNR
	NK	scaMeasure
	AD	SSIM
	SC	MSSIM
	MD	NQM
	NAE	IFC
	SFFscore	IWSSIM
	VIF	IWMSE
	ESSIM	IWPSNR
	RFSSIM	CWSSIM
	IQI	GMSD

Za *ocenjevanje medmodalnega vpliva* predlagamo pristop z združitvijo obteženih ocen objektivne enomodalne analize. Pri tem smo izbrali najboljšo slikovno metriko z najvišjo stopnjo korelacije ter PESQ za oceno govorne modalnosti. Večmodalni model vrednotenja kakovosti je zmožen napovedi ocene večmodalne vsebine, pri tem bo dodatno upošteval:

- **tip modalnosti,**
- **tip scene in**
- **vrednost PL.**

Kot izhodišče smo izbrali linearni regresijski model s splošno obliko:

$$\begin{aligned} &\boldsymbol{objMOS} = \boldsymbol{\alpha} * \boldsymbol{f}(\boldsymbol{modalnost}) + \\ &\boldsymbol{\beta} * \boldsymbol{g}(\boldsymbol{scena}) + \\ &\boldsymbol{\gamma} * \boldsymbol{PL} + \\ &\boldsymbol{\delta} * \boldsymbol{objMOS_a} + \\ &\boldsymbol{\varepsilon} * \boldsymbol{objMOS_v} + \\ &\boldsymbol{C} \end{aligned} \tag{7.1}$$

pri tem je *f(modalnost)* funkcija odvisna od tipa modalnosti, *g(scena)* funkcija odvisna od tipa scene, *PL* količina izgube paketov, *objMOSa* objektivna ocena avdia in *objMOSv* objektivna ocena videa. *C* je konstanta, $\alpha, \beta, \gamma, \delta, \varepsilon$ pa utežnostni koeficienti.

Predpostavili smo, da so nelinearne značilnosti HVS/HAS že upoštevane v *objMOSa* in *objMOSv*.

Model smo izgradili na zasnovani podatkovni bazi, ki smo jo razdelili v dva nabora: TRAIN in TEST. Nabor TRAIN je zajemal 75% vseh posnetkov, izbranih po deležu glede na tip scene, modalnosti in PL vrednosti. Ta množica posnetkov je bila učna. Ostali del (25%) je bil namenjen testiranju in performančni analizi modela na modelu neznanih podatkih.

7.3. Določanje vpliva osredotočenosti uporabnika na vizualna polja ROI

Prostorsko pomembnost ROI smo testirali na štirih različnih scenah tipa *intervju* iz AV-podatkovne baze (tabela 7.6). Simulacija je bila razdeljena na *izdelavo programskega detektorja ROI*, *prostorsko porazdeljeno vrednotenje kakovosti slik* in *subjektivno vrednotenje ROI in ne-ROI polj* z namenom ugotovitve pomembnosti pojavitve degradacij v teh območjih.

Izdelava programskega detektorja je bila zasnovana na ogrodju programske opreme iz paketa OpenCV, v katerem smo razvili primerni kaskadni klasifikator Viola-Jones z značilkami Haar za detekcijo strukture obraza iz obstoječe predloge. Izbira značilk je potekala po metodi Adaboost z namenom zmanjšanja nabora možnih značilk. Cilj predlaganega postopka je porazdeljeno vrednotenje kakovosti videa. Uporabili smo dve metriki za vrednotenje kakovosti, ki sta se razlikovali po računski kompleksnosti (tabela 7.12). Za oceno kakovosti znotraj polja ROI smo tako uporabili metodo z boljšo korelacijo s subjektivnimi ocenami, ki pa je računsko kompleksnejša, za oceno kakovosti ozadja pa smo uporabili metodo s slabšo korelacijo, ki pa je računsko preprostejša in zato posledično bistveno hitrejša. Kakovost ozadja smo merili s PSNR, ki enostavno izračuna maksimalno vrednost degradacije oz. razlike med signaloma. Kakovost znotraj polja ROI smo vrednotili z metodo NQM, ki na podlagi Pelijeve kontrastne piramide upošteva tudi različne variacije, s čimer posledično bolje korelira s subjektivno oceno, zaradi česar je primerna za ovrednotenje polj ROI, vendar je računsko bolj zahtevna.

Tabela 7.12: Slikovni metriki kakovosti za porazdeljeno vrednotenje kakovosti videa z detekcijo polja ROI.

Lastnosti	**PSNR**	**NQM**
Računska kompleksnost	Zelo nizka	Zmerna
Povprečna korelacija s *subMOSv*	Srednja	Zelo visoka
Mehanizem delovanja in meritve ovrednotenja kakovosti	Meritev maksimalnega šuma	Variacije kontrastne Senzitivnosti s kontrastno razdaljo, variacije lokalnega povprečja luminance, kontrastna interakcija med prostorskimi frekvencami in maskiranje signala

Predlagani koncept pomembnosti porazdeljenega vrednotenja kakovosti smo primerjali z rezultati subjektivnega vrednotenja, kjer so testne osebe ocenjevale zaznavnost degradacije v in izven polja ROI (tabela 7.13). V ta namen smo zasnovali scenarije z različno količino, dolžino in časovno oz. prostorsko razporeditvijo degradacij (primer scenarijev za en posnetek kaže tabela 7.14). Iz posamičnih posnetkov tipa *intervju* V-modalnosti iz AV-podatkovne baze smo najprej izločili 2 krajša posnetka fiksne dolžine 400 okvirjev. S krajšim časom posnetkov smo zagotovili večjo osredotočenost testnih oseb na testno gradivo. Iz vsakega kratkega posnetka smo nato naredili 6 scenarijev degradacij, pri tem pa je bil *scenarij 0* referenčen, da smo lahko preverili relevantnost vrednotenja testnih oseb. Z izbiro različnih časov degradacije in prostorskočasovno razporeditvijo smo testirali vpliv dolžine in trajanja dražljaja. Scenarije degradacij iz tabele 7.14 smo testirali za pojav degradacij v poljih ROI in izven njih. V scenarijih ROI smo za degradacijo uporabili zamik pravokotnega delca slike velikosti *150 x 50 slikovnih pik*, ki smo ga zamaknili za *30 slikovnih pik* in zamrznili na izbrani prostorski lokaciji za čas degradacije. Pri scenarijih z degradacijo izven polja ROI smo uporabili enak pristop, le da smo prostorsko velikost degradacije spreminjali tako, da je bila ocena PSNR-posnetka z ROI bila skoraj enaka oceni PSNR-posnetka z degradacijo izven polja ROI (absolutna razlika je bila manjša kot 0,1 dB).

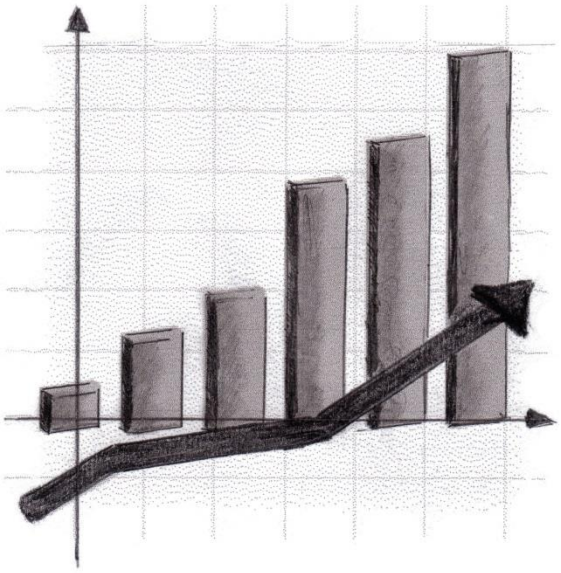
Tabela 7.13: Subjektivni testi porazdeljenega vrednotenja kakovosti.

Lastnost	**Vrednost**
Število ocenjevalcev	10
Število testnih posnetkov/ocenjevalca	88
Čas posnetka	400 okvirjev pri 25 okvirjih/s (8 sekund)
Razpon vrednosti	5-stopenjski DMOS (tabela 7.10)
Metoda ocenjevanja in lestvica	DCR
H_v (oddaljenost od vizualnega izvora)	0,7-1,0 m
H_a (oddaljenost od zvočnega izvora)	1,5 m
D (povprečna oddaljenost med ocenjevalci)	1 m
W (velikost zaslona opazovanja)	24 "
monitor	Philips Brilliance 240BW

Tabela 7.14: Scenariji subjektivnih testov porazdeljenega vrednotenja kakovosti, pridobljeni iz enega posnetka tipa *intervju*.

Št. krajšega posnetka	**Scenarij**	**Število degradacij**	**Čas degradacij [okvirjev]**	**Istoležnost prostorska**	**časovna**
1	0	0			
1	1	1	4		
1	2	1	8		
1	3	2	4/4	Ne	Da
1	4	2	4/4	Ne	Ne
1	5	2	4/8	Ne	Ne
2	0	0			
2	1	1	2		
2	2	1	16		
2	3	2	2/4	Ne	Da
2	4	2	2/4	Ne	Ne
2	5	2	8/4	Ne	Ne

Uporaba takšnega porazdeljenega pristopa vrednotenja kakovosti je lahko posebej učinkovita za vsebine z veliko bitno hitrostjo, kot je na primer format 4K UHD.

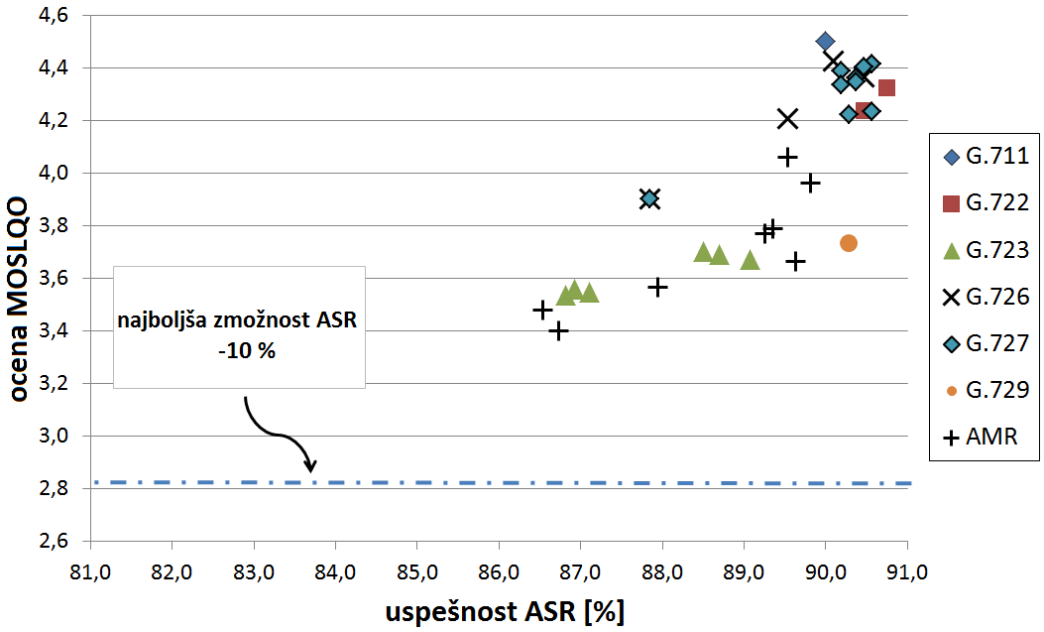


8. Rezultati

V tem poglavju bomo predstavili rezultate eksperimentov, ki smo jih izvedli v okviru doktorske disertacije. Najprej bomo analizirali vpliv degradacije na vhodno modalnost govorno vodene storitve IVR. Na podlagi dobljenih rezultatov bomo predlagali klasifikator izbire vhodne modalnosti, ki odloča med govornim ali DTMF-vnosom. Nato bomo ovrednotili subjektivne rezultate zasnovane večmodalne podatkovne baze. To nam bo služilo za objektivno evalvacijo enomodalnih metrik kakovosti, s pomočjo katere bomo predlagali model vrednotenja kakovosti izhodnih modalnosti večmodalnih storitev. Ta spoznanja nam bodo pomagala pri zadnjem delu naloge, tj. vpeljavi enostavne in kompleksne metrike kakovosti pri porazdeljenem vrednotenju kakovosti izhodne vizualne modalnosti. Tukaj bomo na primeru uspešne detekcije obraza predstavili prednosti uporabljenega predloga.

8.1. Rezultati in analiza vpliva degradacij na vhodno modalnost večmodalnega sistema

V prvi fazi so bile izračunane vrednosti ocen MOSLQO transkodiranih posnetkov (slika 8.1). Pri tem je opaziti statistično signifikantno razliko med rezultati različnih govornih kodekov s stališča percepcije končnega uporabnika, tj. za več kot *1,00* oceno MOSLQO pri primerjavi nekompresiranih podatkov (G.711, najboljša kakovost) s tistimi v npr. AMR-kodiranem bitnem toku (najslabša kakovost). Ta degradacija kakovosti zmanjša uspešnost ASR za 4,8 %, kar predstavlja velik vpliv na delovanje in skupno QoS-storitve, kakršen je govorno voden sistem IVR.



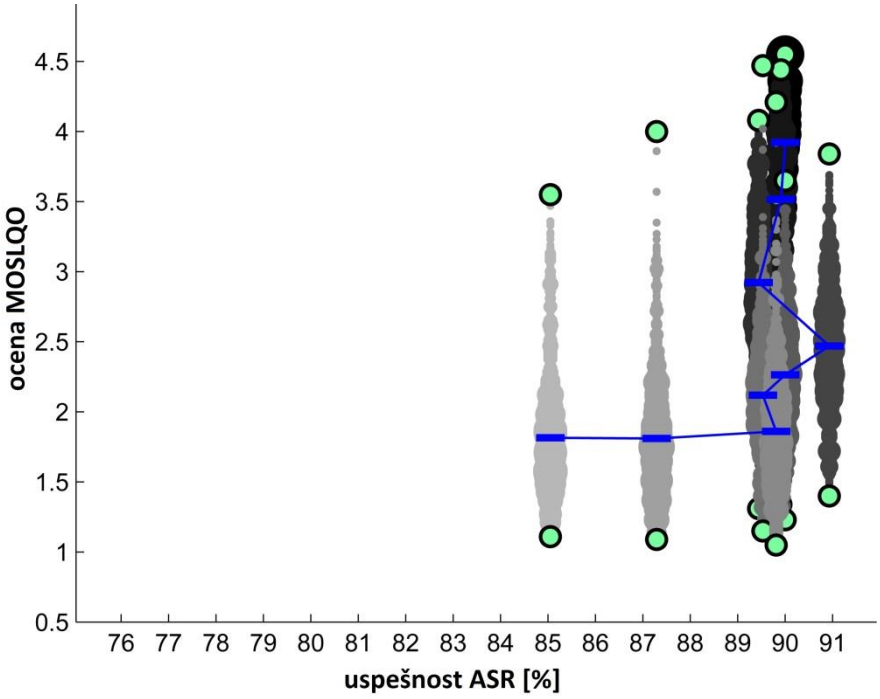
Slika 8.1: Vpliv transkodiranja posnetkov na uspešnost ASR.

V drugi fazi smo naredili primerjavo povprečja MOSLQO-ocene degradiranih posnetkov za nabore različnih konfiguracij kodirnikov in emuliranih PL-scenarijev ter jih primerjali z uspešnostjo ASR na posameznem naboru. Rezultati so združeni glede na *konfiguracijo kodirnika*. Na grafu je vsak nabor posnetkov navpično poravnan k pripadajoči vrednosti uspešnosti ASR-modula za pripadajoč scenarij, pri tem pa:

- debelina navpične črte določa gostoto posnetkov za pripadajoče vertikalno okno velikosti *ΔMOSLQO = 0,01 ocene*,

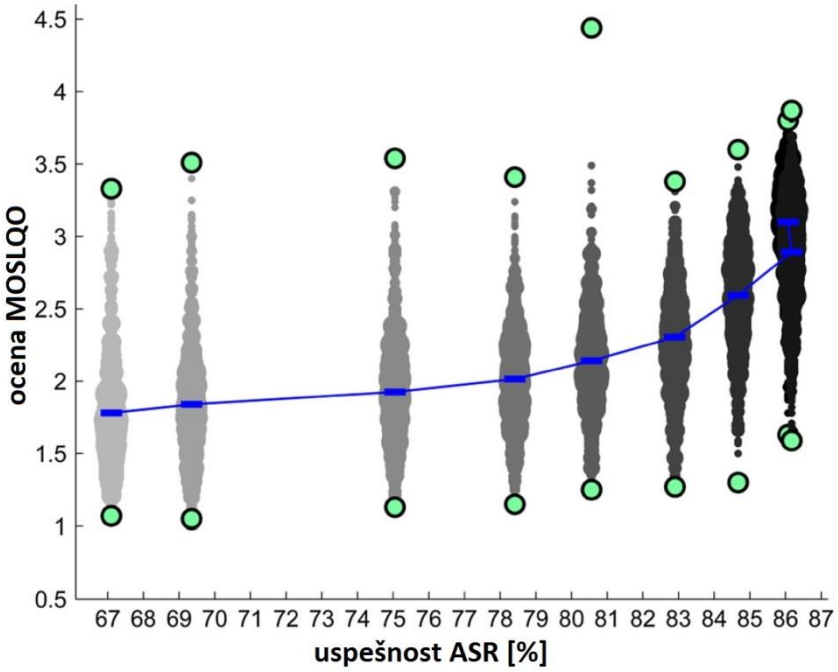
- zeleni piki določata minimalni in maksimalni vrednosti za posamezen PL-scenarij,
- odtenek določa različne PL-scenarije glede na količino izgube paketov, kjer črna predstavlja najmanjšo degradacijo (*1_PL*), najsvetlejša siva pa največjo (*35_PL*).

PL-simulacija G.711 prikazuje dobro uspešnost razpoznavanja besed, kljub temu da se povprečna objektivna ocena približuje *MOSLQO ~ 1,8* (slika 8.2). To označuje, da ima PCM dobro odpornost na naključne izgube paketov tudi za visoko stopnjo povprečne izgube. Pri primerjavi scenarijev *1_PL* in *20_PL* tako opazimo padec uspešnosti ASR za samo *0,5 %*.

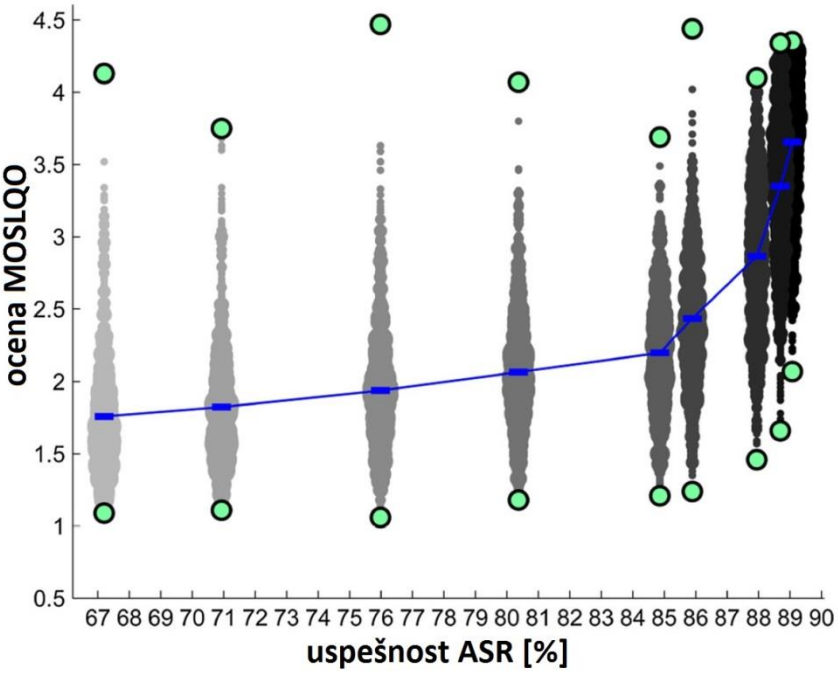


Slika 8.2: PL-scenarij G.711, A-law, 64 kbps.

Vse konfiguracije AMR so pokazale monotoni upad uspešnosti razpoznavanja govora po trendni črti sorazmerno z večanjem izgube paketov. Prišlo je do nekaterih odstopanj, ki pa so bila posledica majhne nominalne hitrosti kodeka (4,75 kbps) v scenarijih z nizkim PL. Razlika med najboljšo in najslabšo konfiguracijo (4,75kbps in 12,2kbps) je bila $\Delta ASR_{AMR12.2,4.75@1\_PL}$ =3,0% in $\Delta ASR_{AMR12.2,4.75@35\_PL}$=0,1% (slika 8.3, slika 8.4).

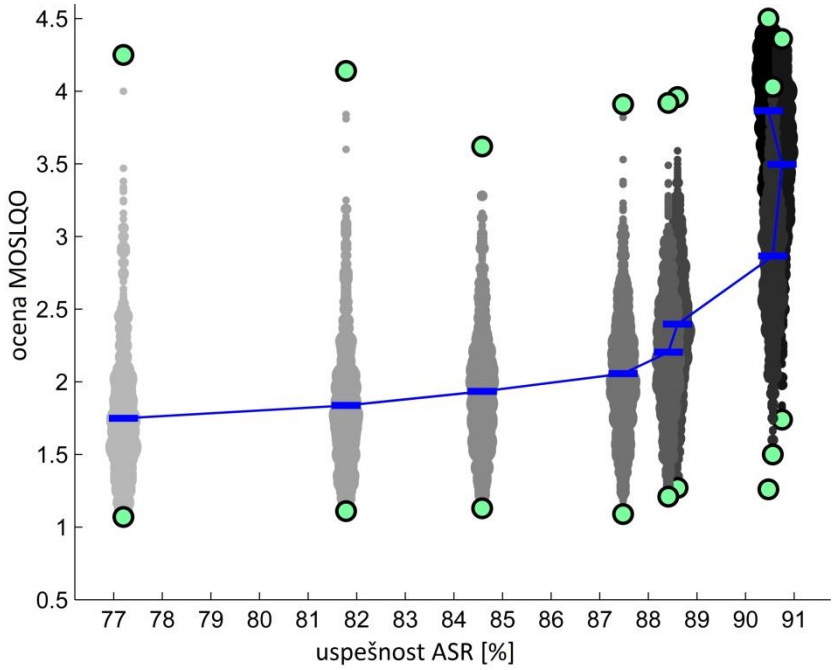


Slika 8.3: PL-scenarij AMR 4,75 kbps.



Slika 8.4: PL-scenarij AMR 12,2 kbps.

Različne konfiguracije kodeka G.722 so prispevale k zelo podobni karakteristiki, neodvisni od pasovne širine kodeka (slika 8.5). Med najboljšo (64 kbps) in najslabšo konfiguracijo (48 kbps) ni opaziti skoraj nobene razlike (tabela 8.1). Podoben odziv je tudi pri hitrosti 56 kbps.

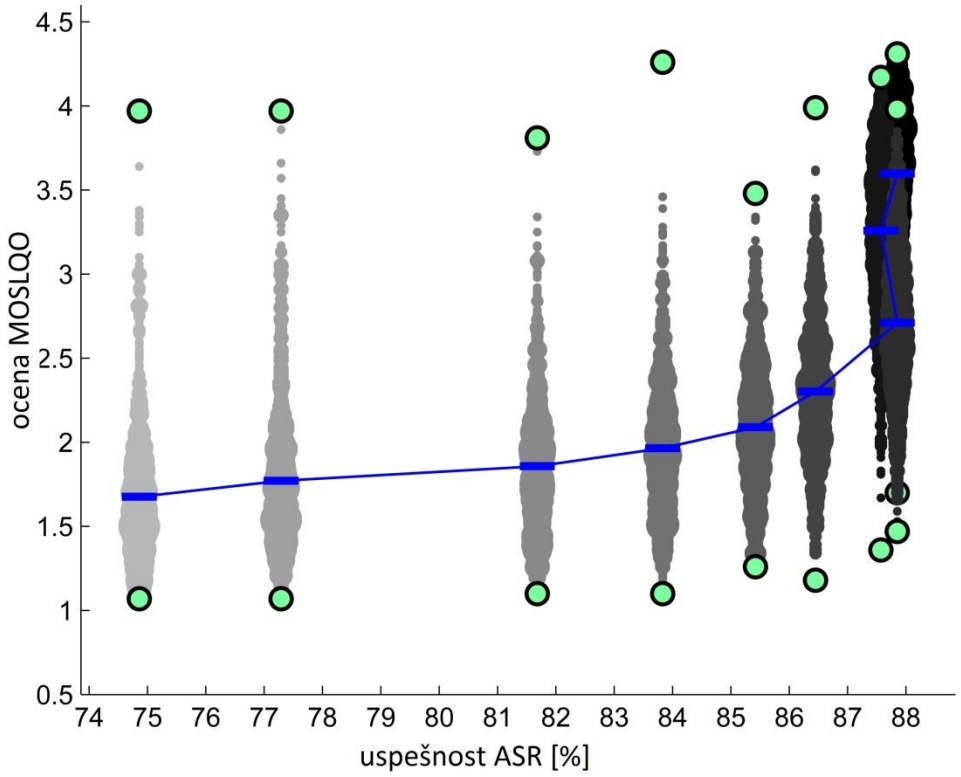


Slika 8.5: PL-scenarij G.722, A-zakon, 64 kbps.

Tabela 8.1: Primerjava uspešnosti ASR za kodek G.722.

PL [%]	**$\Delta ASR_{G.722\ 64,48}$**
1,00	- 0,18
2,00	- 0,18
5,00	+ 1,03
10,00	+ 0,00
15,00	- 0,38
20,00	+ 1,12
25,00	- 0,09
30,00	+ 0,75
35,00	+ 0,37

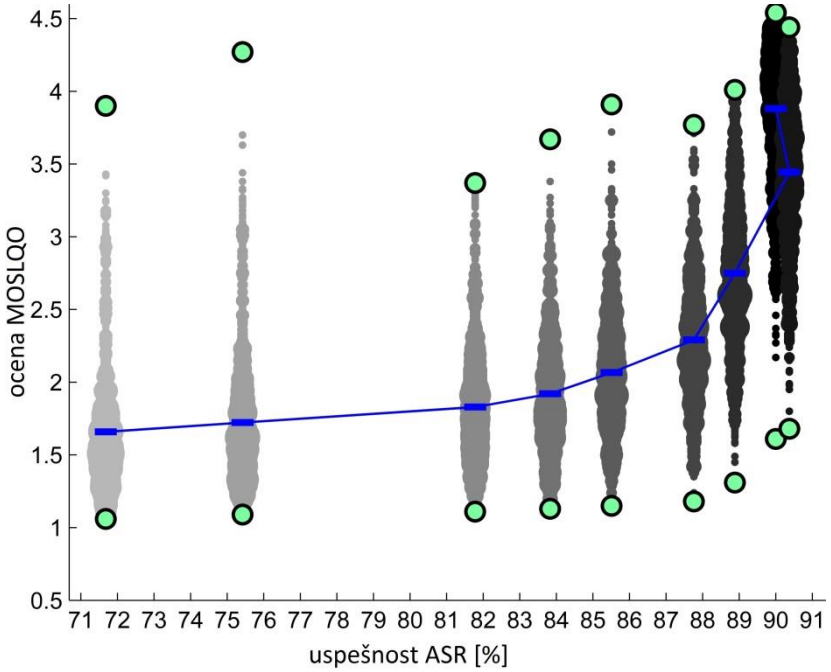
Rezultati nižjih vrednosti PL-kodeka G.726 dajejo skoraj vertikalno trendno črto, kar pomeni, da zmožnosti ASR ostajajo enake (uspešnost ASR ~ 88 %), čeprav povprečna ocena MOSLQO pri tem pada s količino izgubljenih paketov (slika 8.6).



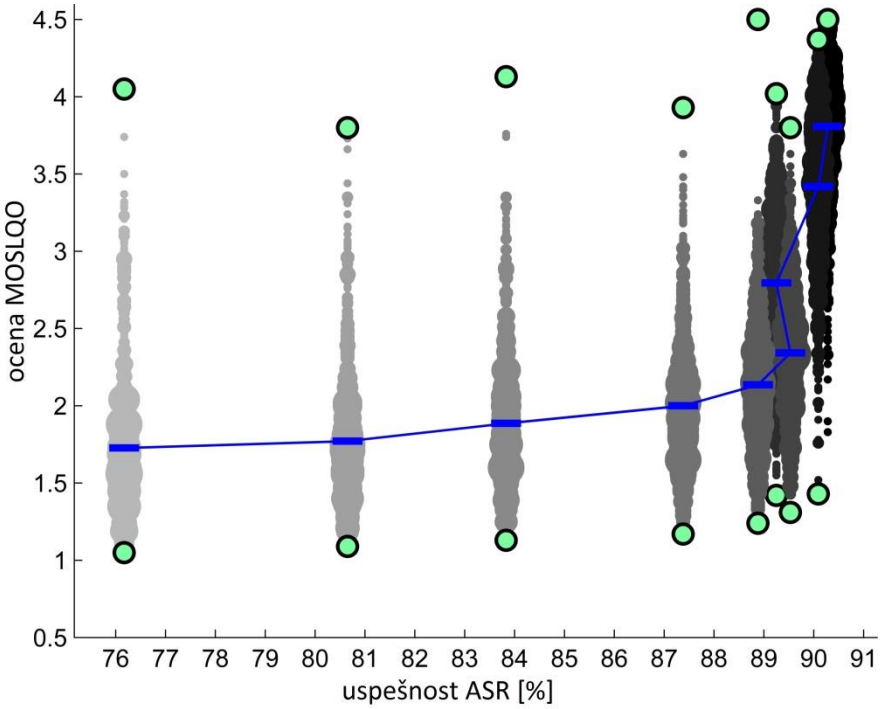
Slika 8.6: PL-scenarij G.726, 16 kbps.

Okrog konfiguracije 10_PL *G.726@16kbps* doseže »zlom« kodeka, kjer se uspešnost ASR precej zmanjša. V primeru scenarijev z nizkim PL je uspešnost ASR boljša za večje vrednosti pasovne širine kodeka (slika 8.7).

Podobne rezultate vidimo pri kodeku G.727 (slika 8.8). Majhna deviacija trendne črte, npr. scenarij pri 1 % PL, ima uspešnost $ASR_{G.723\ 3\_0@1\_PL} = 89{,}72\ \%$, čeprav je pri 2 % PL-uspešnost višja, tj. $ASR_{G.723\ 3\_0@2\_PL} = 90{,}65\ \%$, verjetno zaradi narave statističnega modeliranja znotraj ASR-sistema. Znaten upad uspešnosti ASR je zaznati okrog *10 %-15 % PL* (odvisno od konfiguracije), kjer kodek doživi »zlom«, kar hitro vodi do degradacije kakovosti storitve. Uspešnost ASR se povečuje s številom *bitov/vzorec*, kajti večje število *jedrnih* in *dodatnih bitov* prenaša večjo količino informacije *na vzorec* in izboljša lastnosti algoritma FEC. Pri tem samo *jedrni* biti uporabljajo povratno zanko, kar pomeni, da adaptivni prediktor na sprejemni strani konstantno prilagaja izhodne PCM-kode za naslednjo ADPCM-stopnjo, tj. več kot je uporabljenih jedrnih bitov, manjša je lahko kvantizacijska napaka, kar posledično vodi do boljše uspešnosti ASR. Čeprav se v realnem okolju lahko *dodatni biti* uporabijo v primeru omrežne preobremenjenosti, takšnega načela v naši simulaciji nismo uporabili, saj smo se omejili na paketne, in ne bitne degradacije.



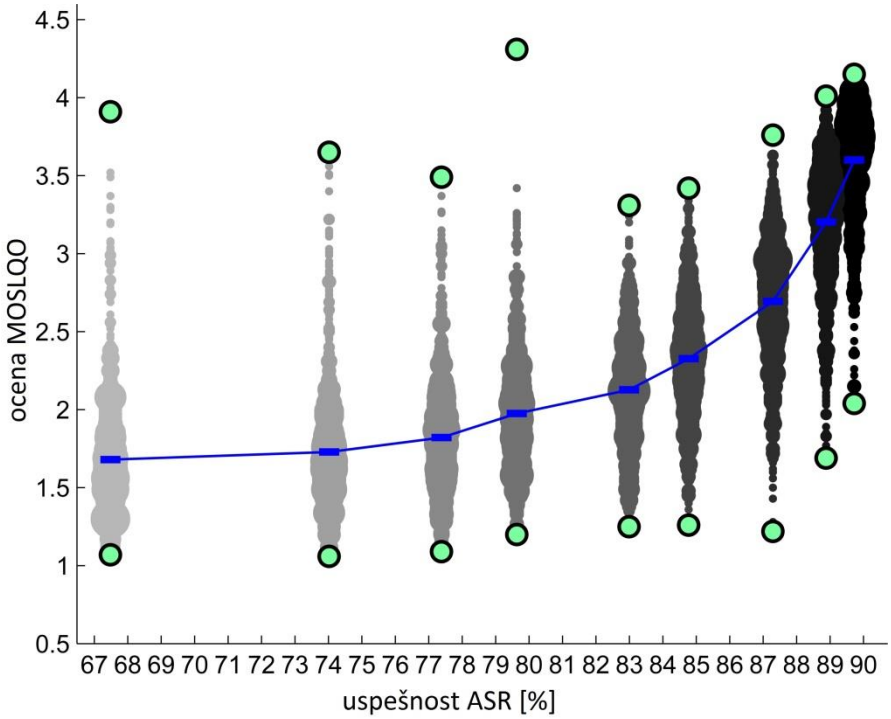
Slika 8.7: PL-scenarij G.726, 40 kbps.



Slika 8.8: PL-scenarij G.727, 4 jedrni biti, 1 dodaten bit.

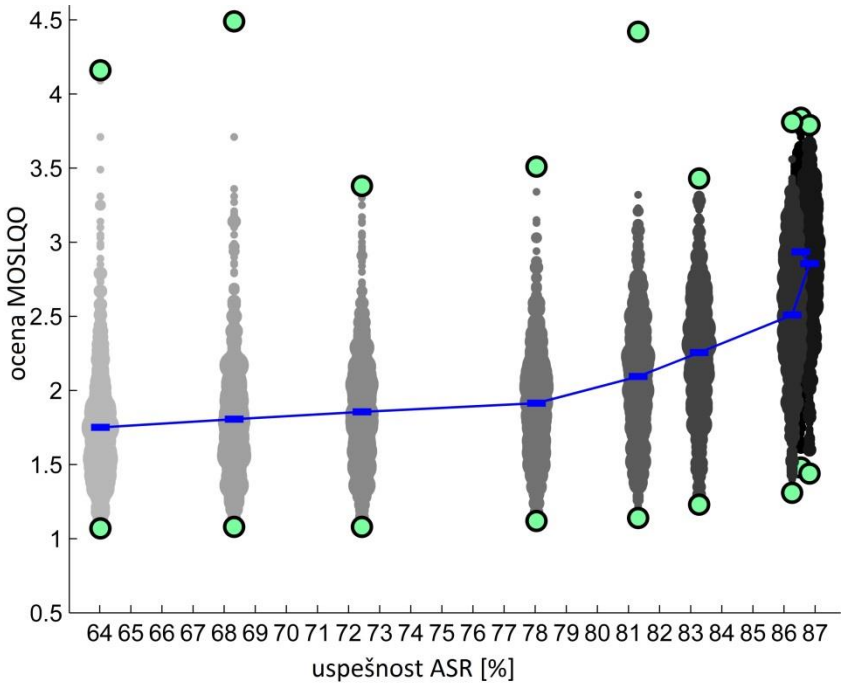
Trendna črta G.729 se dobro prilagaja enakomerno padajoči polinomični funkciji za izbrane nabore PL-scenarijev (slika 8.9). Pri velikem deležu izgube paketov prihaja tudi do velike degradacije kakovosti, kar posledično vodi do degradacije

uspešnosti ASR vse do *67 %*. Verjeten razlog za to je omejitev pasovne širine samega kodeka.

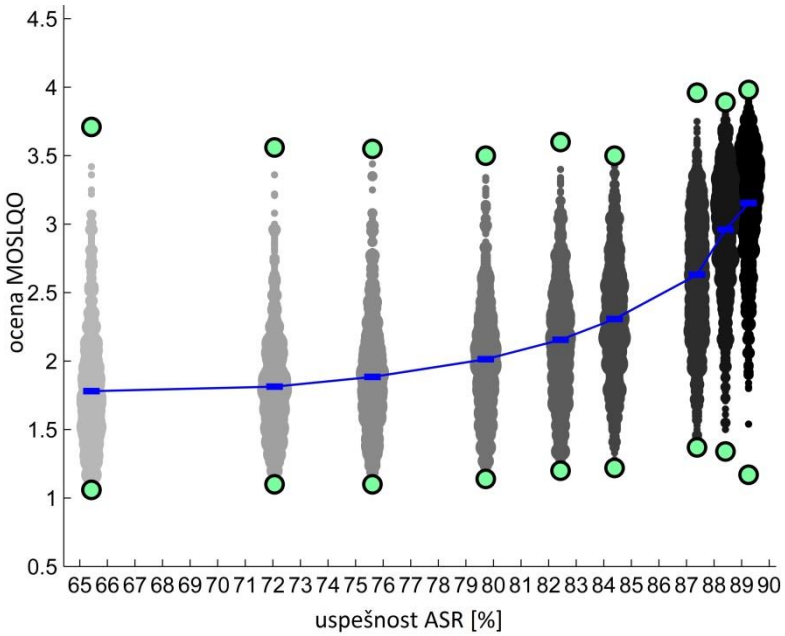


Slika 8.9: PL-scenarij G.729, 8 kbps.

Odziv je podoben tudi pri G.723.1 pri vseh hitrostih, kjer je vpliv izgubljenih (ACELP-) kodiranih paketov najbolj destruktiven (slika 8.10, slika 8.11). Zelo majhno razliko v uspešnosti ASR je opaziti v primeru uporabe ali neuporabe funkcije SID/CNG, postprocesiranja in uporabe visokoprepustnega filtra, saj te funkcije le malo vplivajo na algoritme signalnega procesiranja, ki so del procedure izločanja značilk v sistemu ASR.



Slika 8.10: PL-scenarij G.723.1, 5,3 kbps, HPF, PF, VAD.

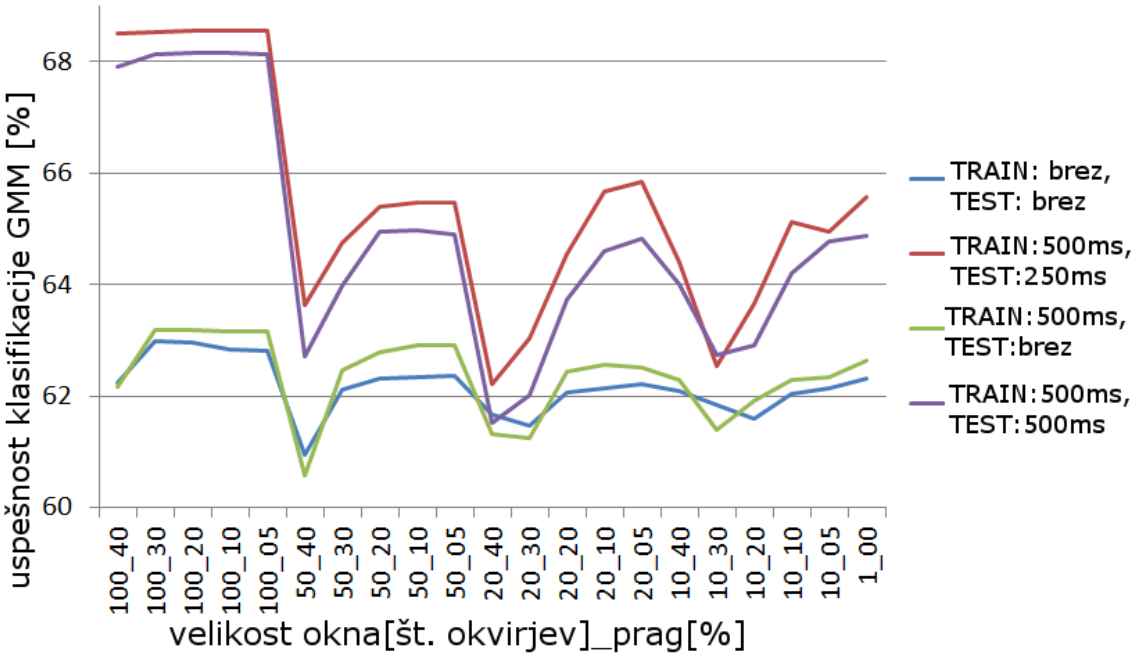


Slika 8.11: PL-scenarij G.723.1, 6,3 kbps, HPF, PF, VAD.

Splošno opažanje in analiza rezultatov dajeta vedenje o različnih odzivih vpliva transkodiranja in omrežnih degradacij na IVR-storitev. Kar se tiče transkodiranja, je vidna znatna razlika med različnimi kodeki, kar se lahko uporabi pri preslikavi ocen kakovosti posameznih kodekov k uspešnosti ASR. Pri predpostavki, da je omejitev *10 % PL*-degradacije na vhodnem avdiu kanalu večmodalnega sistema zadovoljiva za

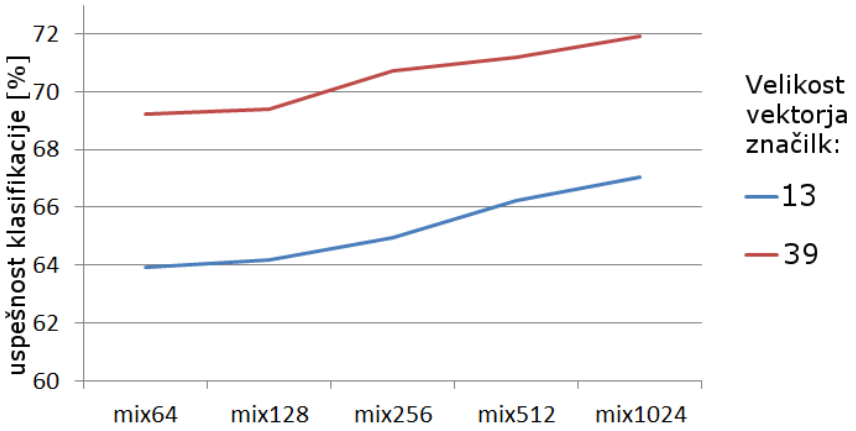
delovanje storitve ASR, kot rezultat dobimo prag kakovosti za uporabo ASR v sistemu IVR, ki znaša **MOSLQO = 2,8**. Ta vrednost predstavlja osnovo za učenje klasifikatorja modalnosti.

V *tretji fazi* določanja vpliva degradacij na vhodno modalnost smo preučili različne parametre učenja modelov GMM na uspešnost *večmodalnega klasifikatorja* vhodne modalnosti. Rezanje posnetkov in odstranitev negovornih signalov pred koristnim signalom in po njem je dalo boljše performance celo pri modelih z nižjo kompleksnostjo. Razlog je v dejstvu, da ne prihaja do možnih nepravilnih ocen značilk v neželenih, degradiranih segmentih avdio signala (slika 8.12).



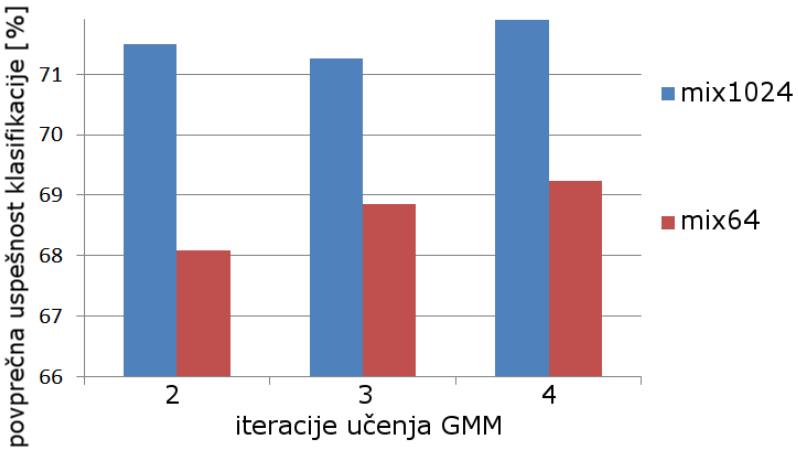
Slika 8.12: Vpliv rezanja posnetkov in odstranitve negovornih signalov na uspešnost klasifikacije GMM (mix64, 2 iteraciji učenja).

Sprememba velikosti vektorja značilk klasifikatorja je imela velik vpliv na čas, potreben za proces učenja: v povprečju je model s 13 značilkami pričakovano potreboval le 25 % časa tistega z 39 značilkami. Čeprav je uspešnost klasifikacije modela z 39 značilkami boljša, je potreben premislek, kateri klasifikator uporabiti, še posebej v primeru omejene procesorske moči in pomembnosti realnočasovnega delovanja (slika 8.13).



Slika 8.13: Vpliv velikosti vektorja značilk na uspešnost klasifikacije.

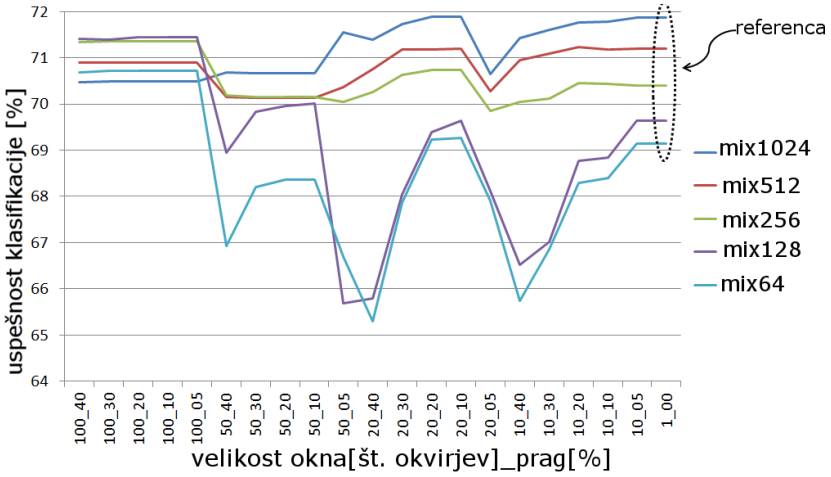
Vsaka naslednja iteracija učenja enako kompleksnih modelov GMM je povzročila dodaten prirastek k povprečni uspešnosti klasifikatorja za manj kompleksne modele GMM, npr. v povprečju 0,86 % izboljšanje za *mix64*. S povečevanjem kompleksnosti se je delež izboljšave zmanjšal, npr. za ~ 0,28 % v primeru *mix1024* (slika 8.14).



Slika 8.14: Vpliv števila iteracij učenja modela GMM.

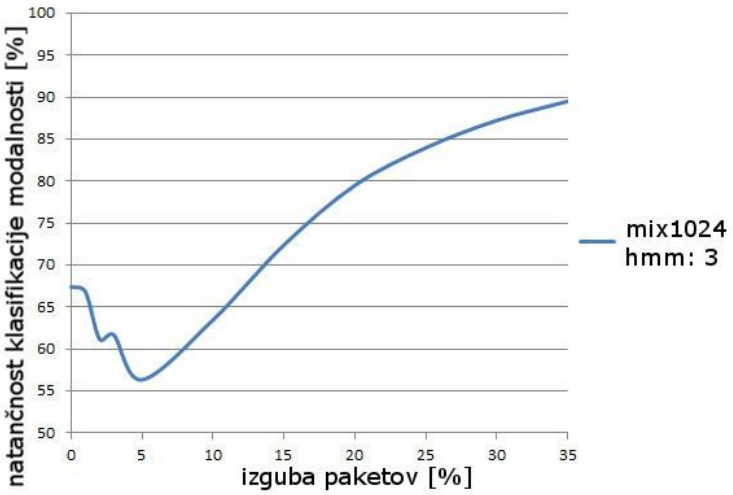
Za manj kompleksne modele GMM je uporaba predlaganega algoritma *povprečenja* bolj učinkovita in takšen pristop je smiselno upoštevati v primeru realnočasovnega delovanja sistema (slika 8.15). Povprečna uspešnost se je v povprečju povečala za 1,7 % za GMM-modele *mix64*. V nasprotju s tem pa je *povprečenje* imelo le majhen prispevek k uspešnosti klasifikatorja pri višjekompleksnih modelih GMM. Kot primer: z oknom velikosti *20 okvirjev* in pragom *10 %* je bilo izboljšanje uspešnosti klasifikacije za

modele *mix1024* manjše od *0,1 %*, gledano v povprečju na vseh PL-scenarijih. Vsi prikazani modeli so bili učeni s 3 HMM-iteracijami.



Slika 8.15: Vpliv algoritma povprečenja.

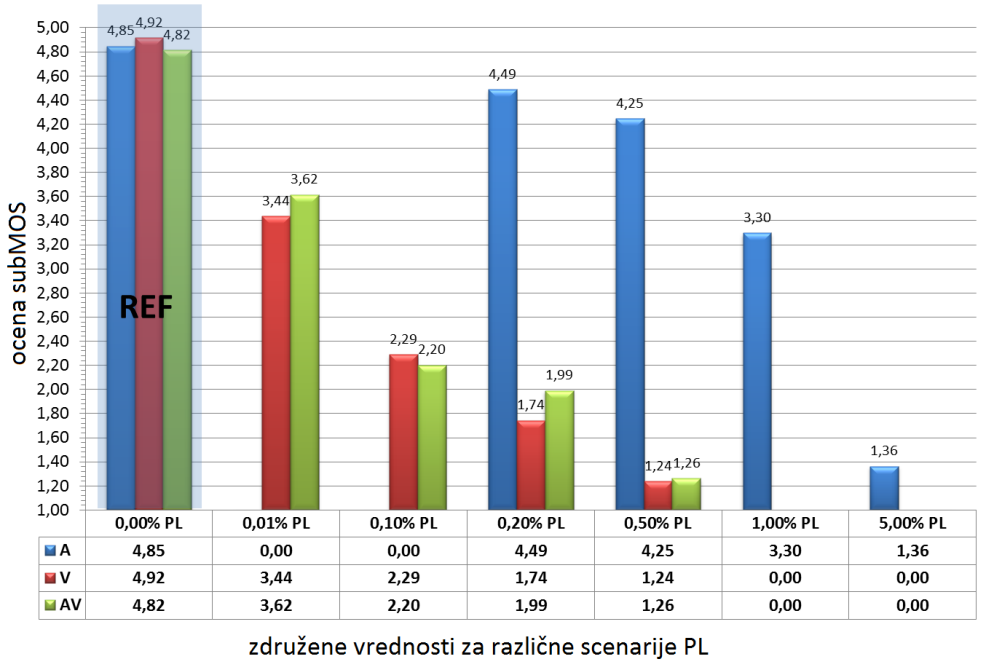
Kot zadnji rezultat klasifikacije vhodne modalnosti glede na kakovost podajamo skupno pravilnost klasifikacije v odvisnosti od izgube paketov, ki kaže celostno sliko predlagane rešitve (slika 8.16). Za visoke PL-vrednosti ima klasifikator dobro natančnost, kar je verjetno posledica izrazitih degradacij v signalu. Na drugi strani pa pri *5_PL* natančnost klasifikacije pade do zgolj 56 %, kar razkriva nekaj težav pri klasifikaciji teh vzorcev. Verjeten vzrok je, da je pri tako majhnem deležu PL-degradacija premalo izrazita, da bi jo bilo mogoče uspešneje klasificirati.



Slika 8.16: Skupna pravilnost klasifikacije modalnosti v odvisnosti od količine PL.

8.2. Rezultati in analiza vpliva degradacij na izhodno modalnost večmodalnega sistema

Rezultati *ocenjevanja subjektivne kakovosti* so pokazali pričakovane rezultate z določenimi odstopanji (slika 8.17). Referenčni scenariji za določanje verodostojnosti opazovalcev (*PL = 0,00 %*) so pokazali nekaj odstopanja pri vseh modalnostih, na podlagi česar lahko sklepamo o pravilni izbiri malo degradiranih PL-vrednosti. Na drugi strani pa zgornji meji za posamezno modalnost, tj. *PL = 0,50 %* za vsebine tipa AV in V ter *PL = 5,00 %* za A, dosegata ustrezno nizko vrednost $subMOS_{[A,V,AV]}$, kar potrjuje vključitev dovolj širokega nabora vrednosti (vse do $subMOS_V = 1{,}24$ pri *PL = 0,50 %*). Vrednosti *0,00* označujejo neveljavno vrednost, npr. $subjMOS_A$ za *A* pri *PL = 0,01 %* in *PL = 0,10 %* ne obstajata zaradi izbire drugega območja izgube paketov ($PL_A \epsilon [0{,}20;\ 0{,}50]$).

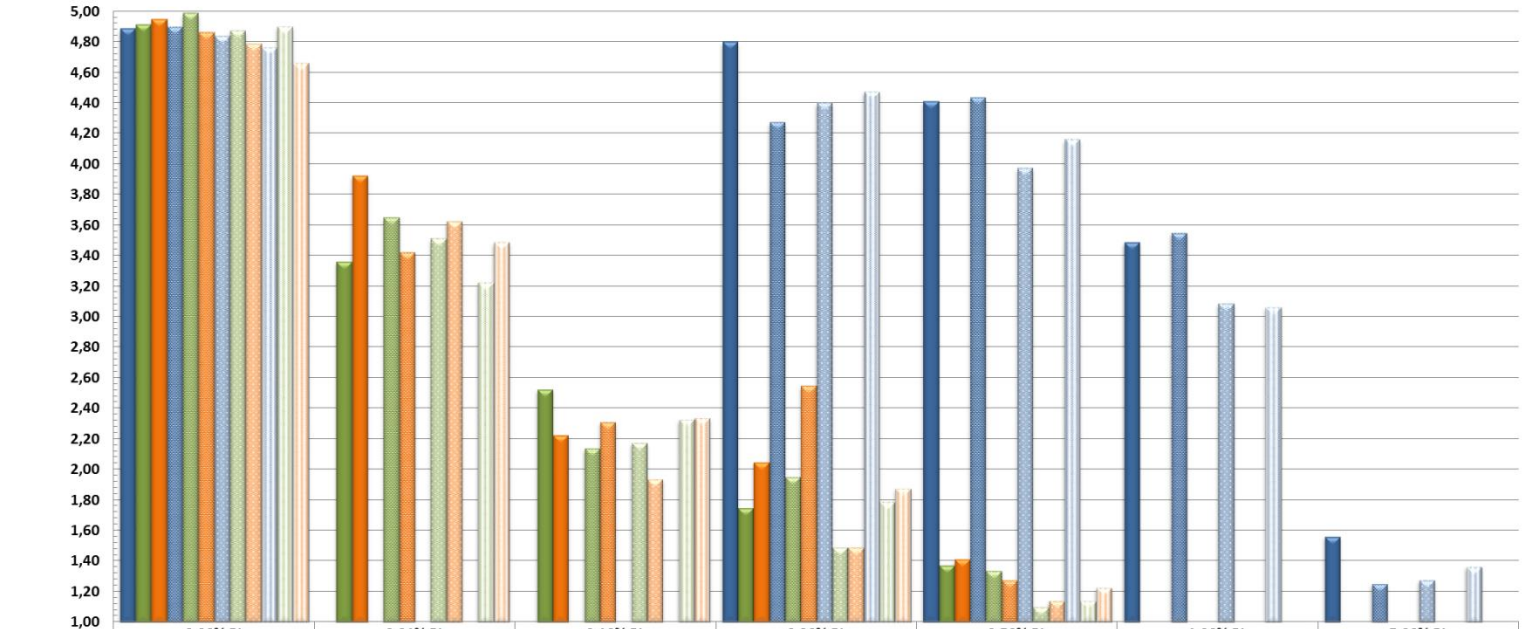


Slika 8.17: Povzetek subjektivnega testiranja in združene vrednosti *subMOS* za posnetke A (modro), V (rdeče) in AV (zeleno).

V nadaljevanju smo pregledali vpliv tipa vsebine na $subMOS_{[A,V,AV]}$ (slika 8.18). Pri modalnosti A je opaziti boljše rezultate za scenska tipa *film* in *intervju*. Razlog se

verjetno nanaša na dejstvo, da je bila v teh primerih degradacija avdio kanala manj slišna, če je do nje prišlo v trenutku negovora, saj je bil v scenah *film* in *intervju* negovorni zvok amplitudno nižji, v scenah *formula in nogomet* pa se je skozi celoten posnetek slišalo ozadje (zvok motorjev formul in zvok navijačev na nogometnem stadionu).

Podobne rezultate smo dobili tudi za modalnosti AV in V. Scenski tip *av_film* ima največje število preklopov scene, pri katerih se popolnoma osveži vsebina (vstavitev okvirja I), zato imajo morebitne degradacije okvirjev v povprečju krajši čas. To privede do navidezno in zaznavno »krajše« nepravilnosti v videu.



	0,00% PL	0,01% PL	0,10% PL	0,20% PL	0,50% PL	1,00% PL	5,00% PL
A_film	4,89	0,00	0,00	4,80	4,41	3,49	1,56
V_film	4,91	3,36	2,53	1,75	1,38	0,00	0,00
AV_film	4,95	3,93	2,23	2,05	1,41	0,00	0,00
A_intervju	4,90	0,00	0,00	4,28	4,44	3,55	1,25
V_intervju	4,99	3,65	2,14	1,95	1,34	0,00	0,00
AV_intervju	4,86	3,43	2,31	2,55	1,28	0,00	0,00
A_nogomet	4,84	0,00	0,00	4,40	3,98	3,09	1,28
V_nogomet	4,88	3,51	2,18	1,49	1,10	0,00	0,00
AV_nogomet	4,79	3,63	1,94	1,49	1,14	0,00	0,00
A_formula	4,76	0,00	0,00	4,48	4,16	3,06	1,36
V_formula	4,90	3,23	2,33	1,79	1,14	0,00	0,00
AV_formula	4,66	3,49	2,34	1,88	1,23	0,00	0,00

združene vrednosti za različne scenarije PL

Slika 8.18: Vpliv tipa vsebine na *subMOS* za 4 scene posnetkov A (modro), V (zeleno) in AV (oranžno).

Najslabše rezultate ima scenski tip *av_nogomet*, kjer se je zaradi karakteristik počasnega premikanja kamere, tj. redkega osveževanja video toka H.264 med okvirji in posledično manjših prostorskih osvežitev makroblokov pogosto razlivala prevladujoča barva (zelena) po celotnem zaslonu (slika 8.19). Za *av_nogomet* so bili za opazovalca ključni ROI nogometni igralci in žoga, ki pa so zaradi tega bili močno degradirani.

Slika 8.19: Razlivanja barvne informacije za sceno *nogomet* pri degradaciji video prenosa (*PL = 0,50 %*).

Slabše rezultati smo dobili tudi pri sceni *av_formula*, kjer pa so prisotni hitri premiki kamere, ki rezultirajo v veliko količino interpoliranih okvirjev (P,B) z velikimi vrednostmi vektorjev premika (slika 8.20). V teh primerih glede na specifikacije H.264 ni bilo možno osveževati referenčnih makroblokov in okvirjev (I), saj je omejitev predstavljala maksimalna prenosna širina multimedijskega vsebnika. V tem primeru se degradacija prenaša na sosednja intra- in inter-področja videa, kar dodatno poslabšuje zaznavno kakovost videa.

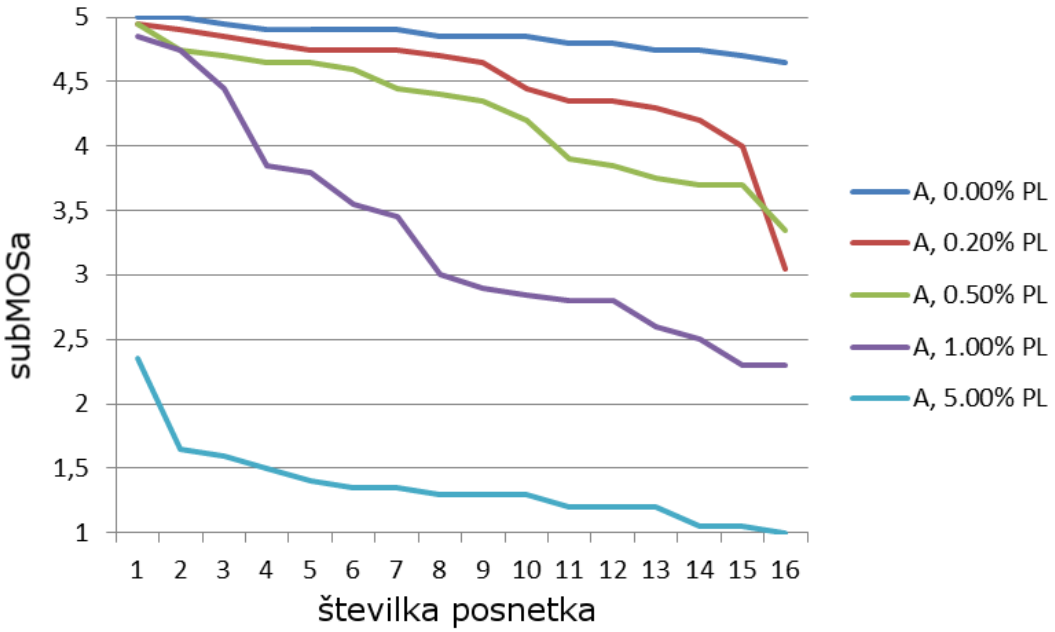
Slika 8.20: Vektorji premika (zgoraj) in vpliv prenašanja degradacije v sosednja področja slike videa (spodaj).

Posnetki videa s scenskim tipom *intervju* so imeli najbolj statično sliko (najmanj vektorjev premika med vsemi scenami) in posledično so bile tudi degradacije »statične«, gledano iz prostorsko-časovne perspektive. Pomanjkanje vektorjev premika in njihova degradacija zato ni močneje vplivala na video (slika 8.21).

Slika 8.21: Izguba paketov z vektorji premika za statično sceno *intervju* ne povzroči večje zaznavne degradacije: originalen posnetek *PL = 0,00 %* (zgoraj), degradiran posnetek *PL = 0,20 %* z označenimi degradacijami (spodaj).

Glede na razporeditev vrednosti različnih posnetkov nas je zanimalo odstopanje *subMOS* različnih modalnosti in scene za scenarije z enakim PL. Za modalnost A opazimo veliko odstopanje na »najslabšem« posnetku s sceno *intervju* za *PL = 0,20 %* (slika 8.22). To je posledica naključne izgube govorne informacije v ključnih trenutkih govora na način, da poslušalec ni uspel dojeti celotnega pomena izgovorjenega. To je

razvidno tudi iz analize okvarjenih besed z dolžino 8 ali več, ki v večjem deležu nosijo ključne informacije v stavku (tabela 8.2).



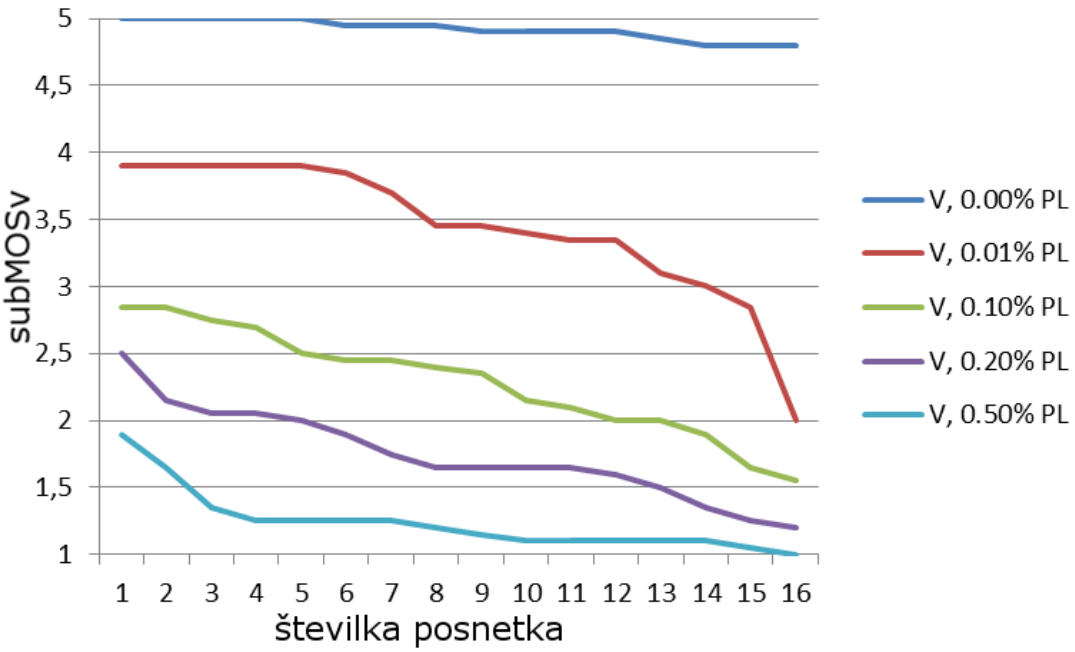
Slika 8.22: Vrednosti *subMOS*$_A$ za posnetke pri istih vrednostih *PL*, posnetki so sortirani padajoče.

Tabela 8.2: Izgubljeni pomeni besed kot posledica izgube paketov v avdio kanalu za sceno *intervju_narator*.

PL [%]	**Delno izgubljene besede dolžine 8 znakov ali več [%]**	**Polno izgubljene besede dolžine 8 znakov ali več [%]**
0,00	0,00	0,00
0,20	0,00	12,50
0,50	6,25	,0,00
1,00	6,25	12,50
5,00	12,50	31,25

Pri modalnosti V so bili razponi PL dobro izbrani, tj. visoke, srednje in nizke vrednosti posameznih PL so precej enakomerno razporejene po celotnem definicijskem območju *subMOS*$_{[V,AV]}$ (slika 8.23). Do odstopanja prihaja le pri *PL = 0,50 %*, kjer se krivulja približa mejni vrednosti *subMOS*$_{[V,AV]}$ *= 1,00*. V tem primeru gre za posnetek *nogomet*, kjer je vzrok naključna degradacija, ki je povzročila močno vizualno napako časovne dolžine ~ 40 % posnetka (slika 8.24). Izstopajoča je tudi relativno visoka vrednost »najboljšega« posnetka najslabšega PL-scenarija, kjer je naključna

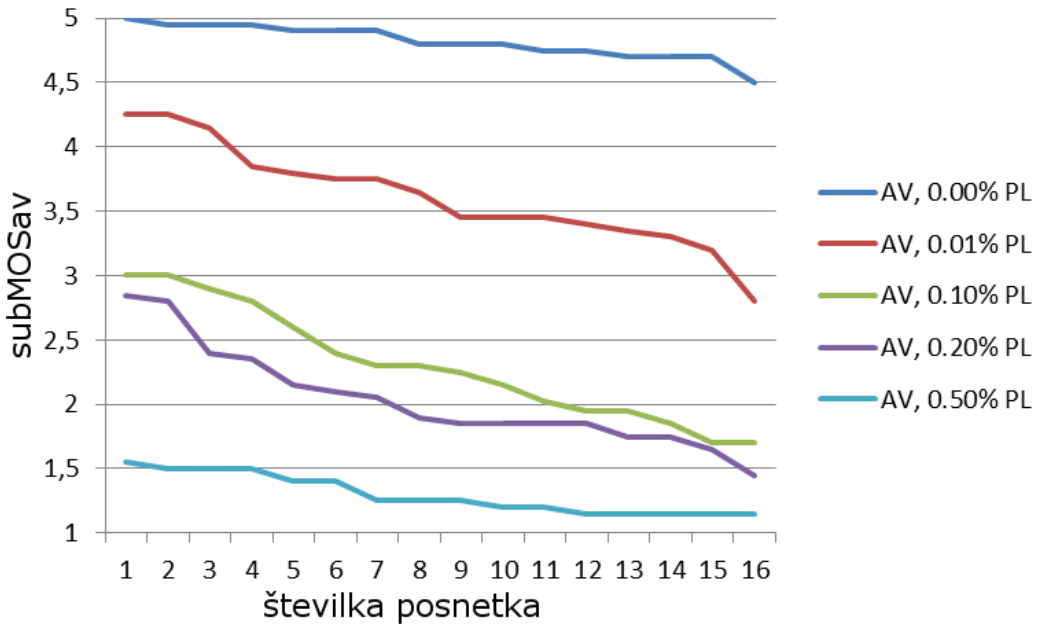
razporeditev PL v povprečju povzročila večjo izgubo paketov, ki so nosili informacijo okvirjev P in B, manjšo pa tistih z vsebnostjo okvirja I.



Slika 8.23: Vrednosti *subMOS*$_V$ za posnetke pri istih vrednostih *PL*, posnetki so sortirani padajoče.

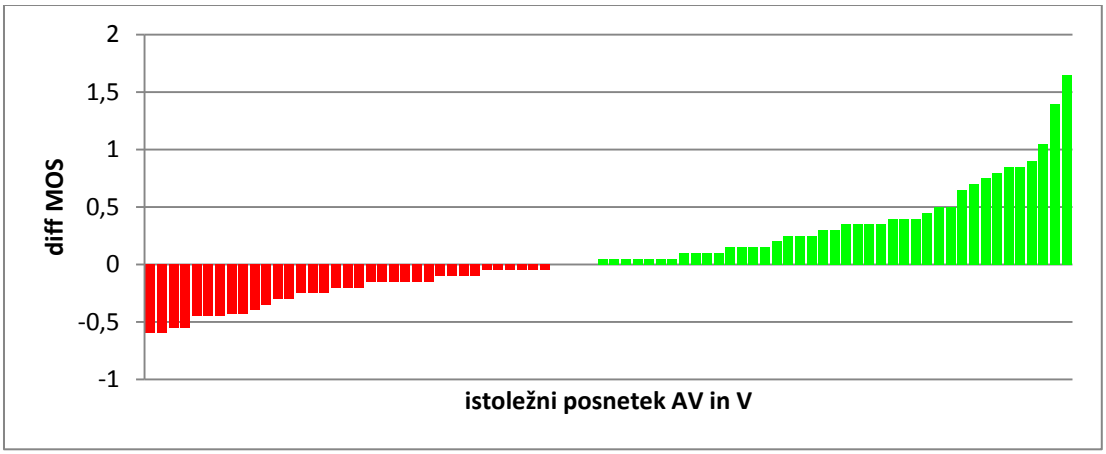
Slika 8.24: Močna degradacija, tj. jerkiness, izguba barvne informacije in izguba strukturne informacije za sceno *v_nogomet* pri *PL = 0,50 %*.

Rezultati modalnosti AV so primerljivi tistim pri modalnosti V (slika 8.25). Pri referenčnih podatkih (*PL = 0,00 %*) sicer graf nakazuje slabše rezultate, vendar je to lahko posledica napačne interpretacije govorne modalnosti testnih oseb v teh scenarijih. Slabše vrednosti izkazujejo posnetki s sceno *av_formula* in delno *av_nogomet*, pri katerih je veliko zvočnega šuma, npr. zvok motorja formule, ki si jih laični ocenjevalec lahko razlaga kot posledico degradacije.



Slika 8.25: Vrednosti $subMOS_{AV}$ za posnetke pri istih vrednostih *PL*, posnetki so sortirani padajoče.

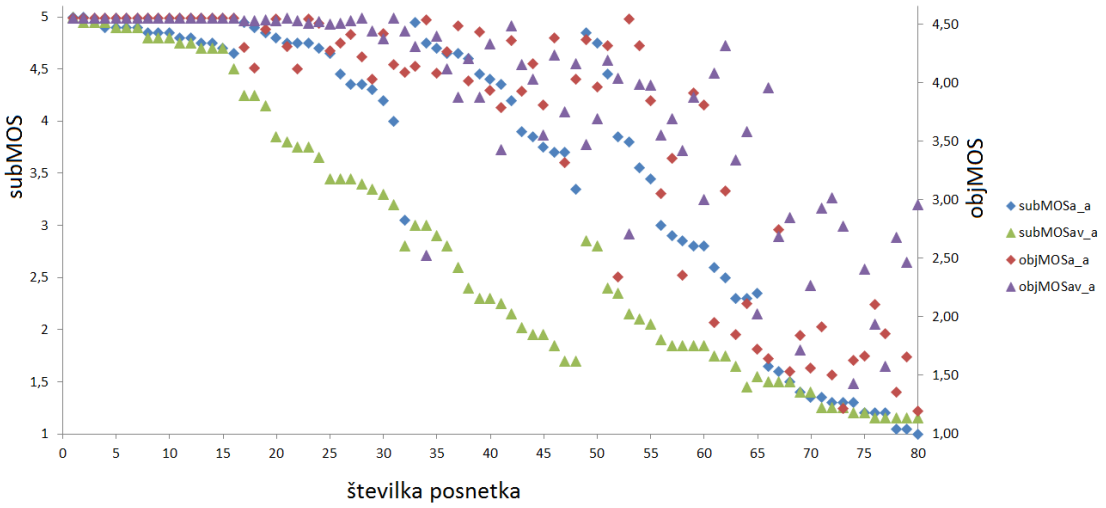
Pri analizi razlike med vrednostmi istoležnih posnetkov AV in V ($diffMOS_{V,AV} = subMOS_{AV} - subMOS_{V}$) opazimo večjo amplitudo in količino pozitivnih vrednosti, kar govori o splošno boljših subjektivnih ocenah modalnosti AV (slika 8.26).



Slika 8.26: Vrednosti $subMOS_{[V,AV]}$ za istoležne posnetke za modalnosti AV in V, sortirane po razliki $(subMOS_{AV} - subMOS_{V})$ naraščajoče.

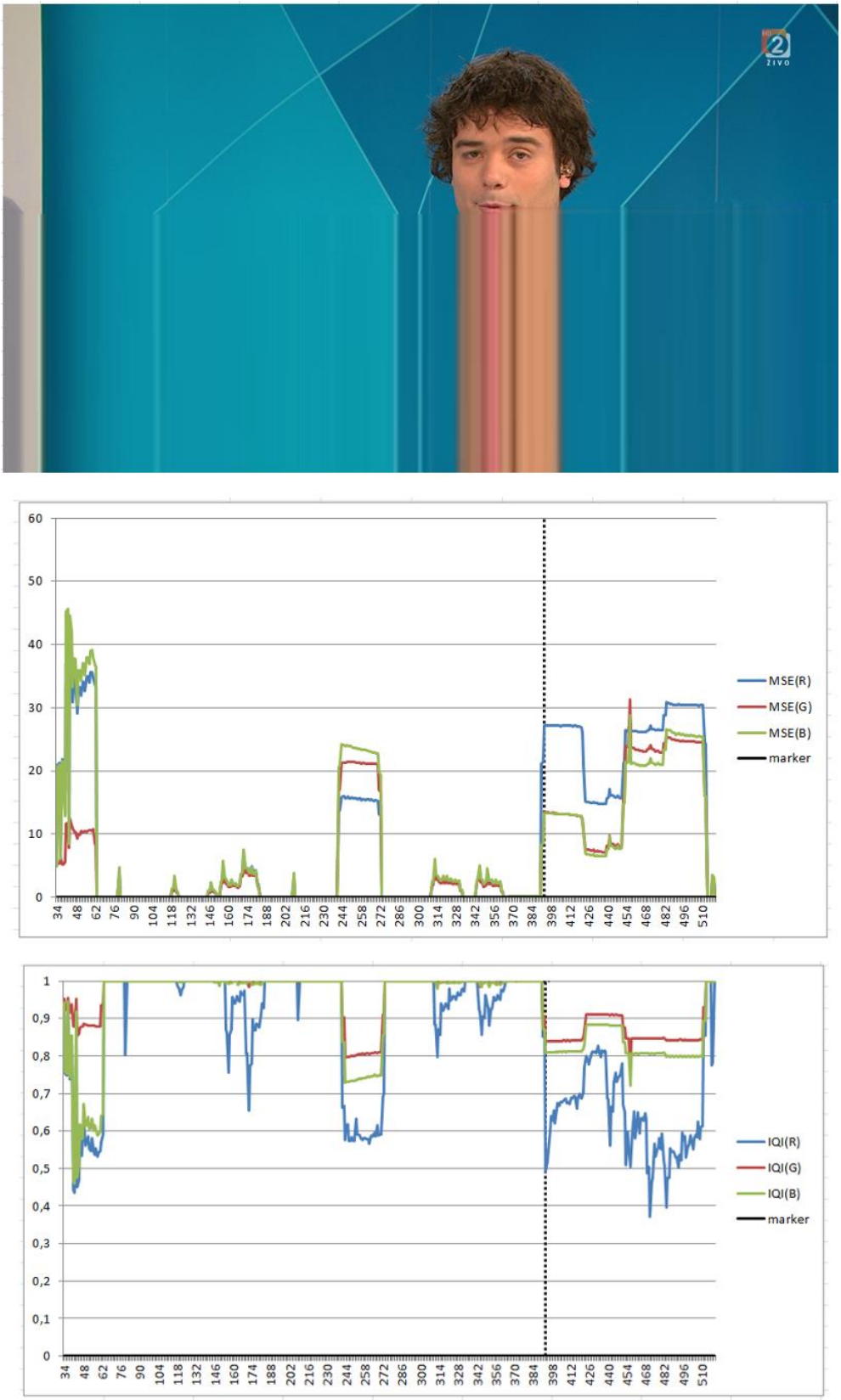
Ocenjevanje objektivne kakovosti je bilo razdeljeno na evalvacijo objektivnih ocen modalnosti posebej za A in posebej za V.

Narejena je bila primerjava subjektivnih in objektivnih vrednosti avdio modalnosti v A- in AV-posnetkih (slika 8.27). Mapirane objektivne vrednosti (MOSLQO, rdeč in vijoličen marker) so izkazovale precejšnje variacije v primerjavi z istoležnimi subjektivnimi vrednostmi (moder in zelen marker). Opazimo povprečno višje vrednosti *objMOSa* pred *subMOSa*, iz česar lahko sklepamo na striktnost subjektivnih ocenjevalcev.

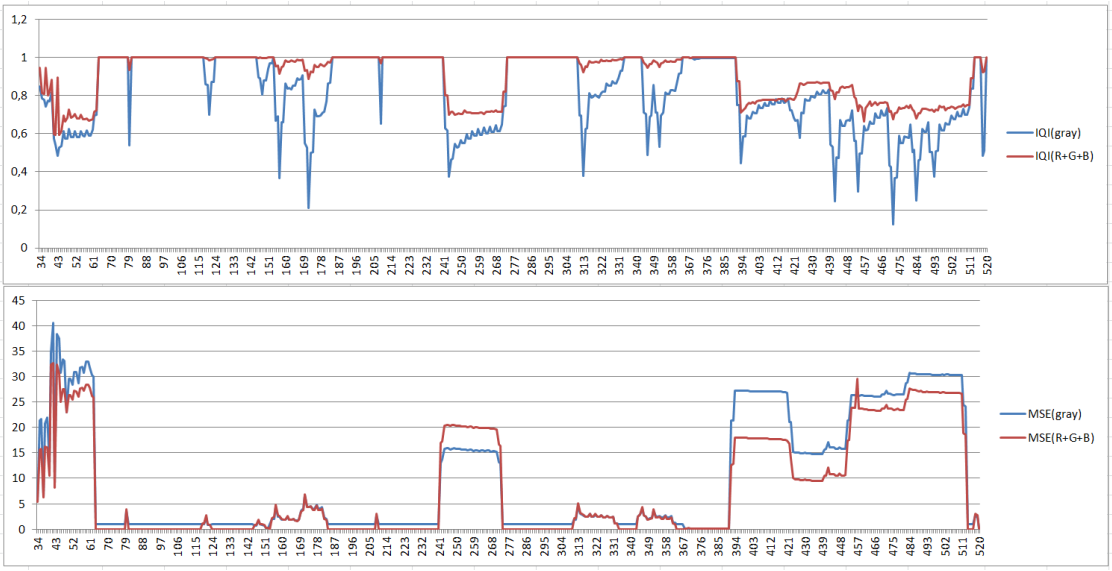


Slika 8.27: Primerjava istoležnih *subjektivnih* in *objektivnih* vrednosti posnetkov za avdio.

Pri evalvaciji kakovosti vizualne modalnosti smo uporabili znane vizualne metrike za objektivno evalvacijo slik. Zaradi uporabe treh kanalov (RGB) smo posamične metrike okvirjev videa združili po načelu matematičnega povprečenja po vseh kanalih. Določeni posnetki z večjo količino PL so izkazovali večja odstopanja rezultatov med RGB-kanali v primeru razlivanja barvne informacije v intraobmočja okvirja (slika 8.28). Uporaba konverzije RGB-modela v GRAY (BT.709) namesto povprečenja vrednosti po posamičnih kanalih je dala slabšo korelacijo s subjektivnimi testi, tj. za slikovne metrike, ki so implementirale to možnost. Primer razlike vrednosti za metriki IQI_{RGB} in IQI_{GRAY} ter MSE_{RGB} in MSE_{GRAY} kaže slika 8.29.



Slika 8.28: Razlika meritev med barvnimi kanali za 394. okvir scene *intervju_narator* pri *PL = 0,10 %* (zgoraj), njegova vrednost MSE_{RGB} (sredina) in IQI_{RGB} (spodaj).



Slika 8.29: Meritev *objTrainMOS - IQI_{RGB}* (zgoraj, rdeče) in *objTrainMOS -IQI_{GRAY}* (zgoraj, modro) ter *objTrainMOS -MSE_{RGB}* (spodaj, rdeče) in *objTrainMOS -MSE_{GRAY}* (spodaj, modro) za sceno *intervju_narator* pri *PL = 0,10 %*.

Kot izhodišče za izdelavo predlaganega večmodalnega modela vrednotenja kakovosti izhodnega sistema smo na naboru TRAIN podatkovne baze izmerili *objTrainMOSa_a*[4], *objTrainMOSa_av*, *objTrainMOSv_v* in *objTrainMOSv_av*. Rezultati so podani v tabeli 8.3 in 8.4. Ker smo za izdelavo modela vrednotenja kakovosti potrebovali le eno A- in eno V-metriko, smo pogledali korelacijo objektivnih in subjektivnih ocen (tabela 8.5). Pri tem smo opazili nekatera odstopanja med posameznimi metrikami in različnimi scenami. Najslabši korelaciji po posameznem scenskem tipu dajeta metriki VIF in SSIM, in sicer obe za sceno *av_formula*. Metriki imata konkavno obliko z maksimalno vrednostjo pri *PL = 0,10 %* (slika 8.30, zgoraj), kar pa ni v korelaciji z monotono padajočo tendenco subjektivnih podatkov (slika 8.30, spodaj). Na drugi strani sta metriki iz nabora metrik, ki računajo »razmerje signal – šum« (*VSNR* in *IWPSNR*), dali najboljši rezultat, in sicer za sceno *av_film*. Nepričakovano je skupna korelacija za *IWPSNR* zelo slaba, kar nakazuje, da je najboljši rezultat v sceni *av_film* lahko le posledica ugodnih parametrov tega videa za mehanizem detekcije metrike IWPSNR.

[4] Parameter kakovosti ocene je sestavljen iz tipa evalvacije (objektivna ali subjektivna), nabora uporabljenega nabora posnetkov (učni ali testni), modalnosti posnetkov v tem naboru (a, v, av) ter meritev enomodalne ocene (a, v). Primer:
objTrainMOSa_av = objektivna ocena kakovosti avdio modalnosti v AV-vsebinah v učnem naboru.

Za izgradnjo modela smo izbrali metriko **NQM**, ki je dala najboljši odziv skozi celoten nabor različnih scenarijev, kot smo ga dobili z meritvijo absolutne vrednosti povprečja vseh korelacij za vse scenske tipe in vrednosti PL, ter najmanjšo deviacijo objektivne ocene na posnetkih TRAIN.

Tabela 8.3: Rezultati slikovnih metrik za posnetke iz nabora TRAIN.

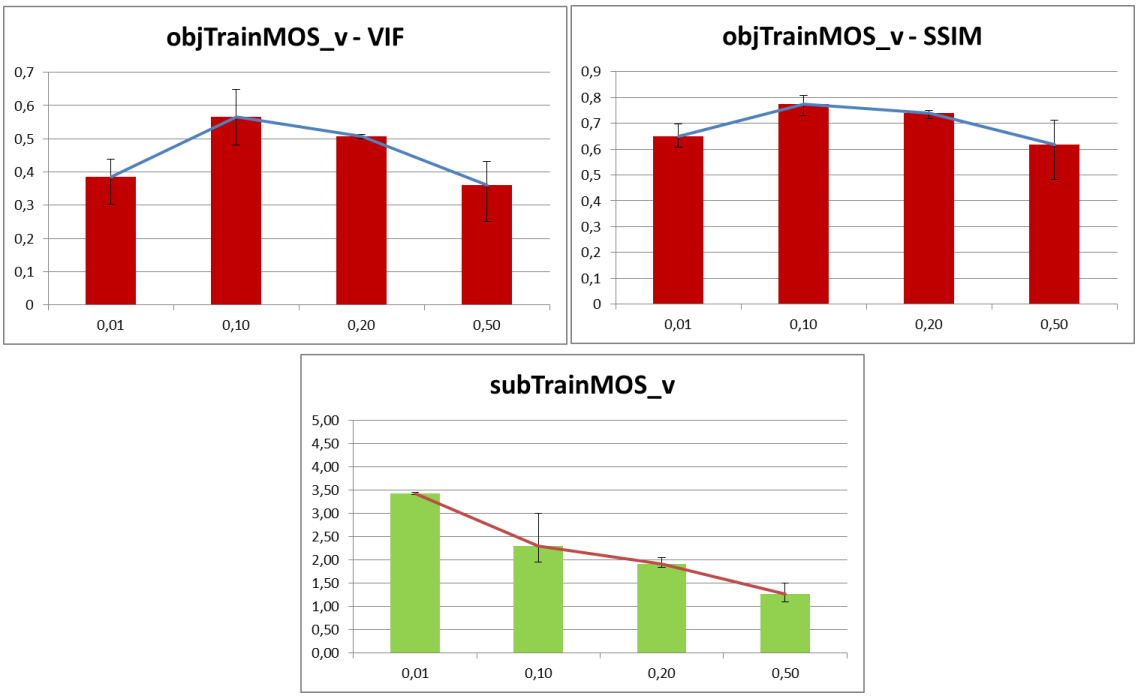
TIP SCENE	% PL	IQI	MSE	LMSE	PSNR	NK	AD	SC	MD	NAE	SFF	scaM	VIF	VIFP	VSNR	WSNR	UQI	SSIM	MSSIM	NQM	IFC	IWMSE	IWPSNR	IWSSIM	CWSSIM	ESSIM	RFSIM	GMSD
av_film	0,01	0,94	10,8	0,62	40,9	0,9840	4,3	1,009	151	0,062	0,95	36,8	0,743	0,753	18,9	18,8	0,865	0,93	0,936	16,04	31,55	978,3	19,9	0,87	0,97	0,99951	0,537	0,099
av_film	0,10	0,81	16,3	1,41	32,4	0,9350	14,7	1,088	224	0,227	0,64	108,1	0,475	0,475	11,4	9,9	0,665	0,78	0,774	6,23	17,31	2653,8	14,5	0,66	0,78	0,99853	0,143	0,169
av_film	0,20	0,84	15,2	1,89	33,2	0,9785	11,2	0,985	234	0,171	0,61	90,8	0,485	0,483	10,9	10,3	0,668	0,75	0,748	6,87	20,46	3192,3	14,3	0,65	0,77	0,99839	0,189	0,173
av_film	0,50	0,71	22,0	2,67	31,0	0,9011	18,3	1,438	240	0,277	0,39	115,4	0,296	0,292	8,3	7,0	0,475	0,60	0,617	3,62	9,36	4531,6	12,2	0,48	0,66	0,99746	0,078	0,216
v_film	0,01	0,95	5,1	1,04	38,4	0,9988	2,2	0,967	168	0,032	0,90	53,7	0,717	0,723	13,7	15,7	0,865	0,91	0,910	12,42	31,22	1285,6	18,7	0,85	0,94	0,99928	0,471	0,111
v_film	0,10	0,90	7,9	1,21	35,4	0,9958	4,4	0,976	209	0,063	0,78	72,1	0,535	0,538	11,8	12,6	0,745	0,83	0,809	9,42	21,46	3069,4	15,2	0,71	0,85	0,99869	0,264	0,153
v_film	0,20	0,80	12,7	2,00	33,3	0,9624	8,8	0,989	228	0,141	0,57	104,4	0,434	0,439	8,4	8,9	0,620	0,73	0,731	5,49	15,95	3494,0	13,6	0,62	0,75	0,99819	0,150	0,186
v_film	0,50	0,69	23,4	2,36	30,9	0,9036	20,5	1,482	239	0,325	0,34	129,4	0,286	0,283	8,0	6,1	0,467	0,62	0,614	2,40	8,04	4135,2	12,5	0,48	0,63	0,99756	0,046	0,210
av_formula	0,01	0,93	18,6	1,07	31,6	0,9781	13,7	1,030	243	0,124	0,58	89,5	0,385	0,374	11,8	14,9	0,465	0,65	0,659	7,15	22,11	2799,8	15,0	0,55	0,70	0,99789	0,135	0,199
av_formula	0,10	0,93	17,2	0,61	33,3	0,9982	11,1	1,001	238	0,090	0,62	73,1	0,566	0,533	11,8	15,8	0,658	0,77	0,736	6,93	33,61	2281,8	18,7	0,65	0,71	0,99832	0,198	0,174
av_formula	0,20	0,93	14,6	0,70	33,6	0,9930	9,8	1,016	224	0,081	0,60	73,7	0,507	0,476	12,9	15,8	0,602	0,74	0,702	6,59	29,37	2752,4	32,7	0,61	0,60	0,99796	0,148	0,185
av_formula	0,50	0,87	25,4	0,95	30,2	0,9723	20,1	1,042	243	0,154	0,34	108,8	0,360	0,324	8,2	11,8	0,446	0,62	0,555	2,73	18,17	4096,4	15,8	0,44	0,56	0,99686	0,043	0,230
v_formula	0,01	0,98	8,7	0,51	35,3	0,9998	4,4	0,998	235	0,037	0,77	37,5	0,665	0,613	13,0	19,1	0,760	0,85	0,841	10,63	41,56	1684,8	17,1	0,76	0,86	0,99899	0,324	0,143
v_formula	0,10	0,94	12,1	0,54	34,2	1,0002	7,8	0,998	220	0,064	0,65	70,3	0,623	0,583	12,2	16,6	0,689	0,80	0,759	6,87	39,21	2178,7	18,3	0,68	0,77	0,99837	0,211	0,163
v_formula	0,20	0,92	17,0	0,73	32,7	0,9979	11,5	1,000	235	0,094	0,51	75,2	0,504	0,463	10,6	14,7	0,577	0,72	0,667	5,43	30,25	3109,0	18,6	0,58	0,68	0,99775	0,152	0,188
v_formula	0,50	0,87	24,7	1,06	30,5	0,9910	17,9	1,016	242	0,147	0,24	106,1	0,305	0,262	7,1	10,6	0,386	0,59	0,517	1,96	14,85	4074,6	14,0	0,40	0,52	0,99658	0,031	0,236
av_intervju	0,01	0,95	5,0	0,71	37,1	0,9991	3,7	1,005	230	0,045	0,88	48,7	0,668	0,552	16,0	20,3	0,722	0,90	0,916	14,19	42,67	871,6	20,3	0,82	0,87	0,99942	0,401	0,108
av_intervju	0,10	0,88	11,8	1,00	35,1	0,9867	8,4	1,033	230	0,093	0,67	49,9	0,638	0,426	15,4	18,6	0,583	0,84	0,831	12,49	38,81	1472,5	48,3	0,72	0,84	0,99897	0,319	0,134
av_intervju	0,20	0,95	6,7	0,77	37,6	0,9963	4,8	1,010	228	0,052	0,83	32,5	0,704	0,577	17,6	21,2	0,737	0,90	0,888	15,18	45,28	1210,8	52,1	0,80	0,91	0,99928	0,428	0,116
av_intervju	0,50	0,86	17,6	1,12	32,4	0,9929	13,4	1,031	237	0,146	0,51	69,5	0,514	0,333	11,6	12,6	0,515	0,80	0,750	5,74	29,48	2179,6	16,4	0,59	0,74	0,99839	0,124	0,171
v_intervju	0,01	0,97	2,2	0,51	43,1	1,0001	1,0	0,999	175	0,011	0,96	25,0	0,723	0,692	23,6	29,0	0,826	0,94	0,956	23,80	38,16	326,1	43,2	0,91	0,98	0,99974	0,576	0,066
v_intervju	0,10	0,91	10,0	0,73	36,0	0,9952	7,6	1,029	228	0,091	0,76	55,2	0,566	0,528	15,9	18,5	0,704	0,88	0,861	12,79	28,93	1445,4	28,8	0,75	0,81	0,99908	0,354	0,123
v_intervju	0,20	0,91	10,6	0,81	36,2	0,9866	8,5	1,034	227	0,099	0,73	56,2	0,572	0,540	16,2	17,9	0,701	0,88	0,861	11,46	31,27	1420,7	68,5	0,75	0,83	0,99909	0,340	0,126
v_intervju	0,50	0,82	22,7	1,39	31,0	0,9904	17,2	1,051	241	0,186	0,41	96,1	0,301	0,253	10,6	11,0	0,398	0,76	0,703	3,83	10,87	2468,4	15,8	0,53	0,60	0,99814	0,089	0,184
av_nogomet	0,01	0,98	5,7	0,87	40,8	0,9946	3,2	1,009	171	0,039	0,82	37,4	0,564	0,555	37,9	25,7	0,602	0,88	0,885	14,44	25,84	760,8	107,0	0,80	0,97	0,99938	0,381	0,118
av_nogomet	0,10	0,95	6,3	0,39	37,8	0,9971	3,1	1,004	220	0,038	0,74	35,8	0,548	0,548	13,3	20,1	0,583	0,90	0,854	9,33	19,68	1400,1	20,1	0,74	0,78	0,99916	0,291	0,125
av_nogomet	0,20	0,94	9,3	0,47	35,9	0,9939	4,6	1,009	231	0,055	0,62	46,3	0,495	0,492	9,5	17,2	0,509	0,86	0,789	5,62	14,81	1862,3	17,7	0,67	0,72	0,99887	0,185	0,155
av_nogomet	0,50	0,79	24,4	0,85	31,7	0,8825	16,7	1,354	247	0,201	0,41	69,7	0,273	0,263	6,7	12,8	0,291	0,73	0,638	1,73	5,70	3061,7	14,4	0,47	0,57	0,99801	0,068	0,209
v_nogomet	0,01	0,98	3,7	0,10	40,4	1,0006	1,8	0,998	177	0,023	0,81	17,9	0,802	0,783	16,9	24,0	0,741	0,95	0,908	11,83	38,69	774,2	24,4	0,84	0,89	0,99943	0,451	0,098
v_nogomet	0,10	0,93	9,3	0,27	39,1	0,9593	5,9	1,112	193	0,070	0,78	34,1	0,651	0,640	32,7	23,3	0,655	0,92	0,886	11,98	22,98	1029,4	24,4	0,80	0,81	0,99934	0,381	0,108
v_nogomet	0,20	0,93	10,6	0,50	35,6	0,9954	5,8	1,007	223	0,067	0,59	54,7	0,441	0,422	9,1	16,2	0,484	0,84	0,769	5,28	14,00	1851,6	17,6	0,64	0,69	0,99871	0,160	0,161
v_nogomet	0,50	0,84	19,2	0,91	32,0	0,9740	12,2	1,057	248	0,139	0,34	77,0	0,240	0,223	5,1	11,9	0,262	0,70	0,590	0,11	4,84	3404,4	13,8	0,43	0,46	0,99762	0,041	0,220

Tabela 8.4: Rezultati govorne metrike za posnetke iz nabora TRAIN.

SCENA	% PL	MOSLQO
av_film	0,01	4,53
av_film	0,10	4,28
av_film	0,20	3,86
av_film	0,50	2,87
a_film	0,20	4,30
a_film	0,50	3,99
a_film	1,00	3,13
a_film	5,00	1,56
av_formula	0,01	4,53
av_formula	0,10	4,16
av_formula	0,20	3,76
av_formula	0,50	2,68
a_formula	0,20	4,367
a_formula	0,50	4,16
a_formula	1,00	3,86
a_formula	5,00	2,14
av_intervju	0,01	4,47
av_intervju	0,10	3,62
av_intervju	0,20	3,79
av_intervju	0,50	2,21
a_intervju	0,20	4,29
a_intervju	0,50	4,21
a_intervju	1,00	2,71
a_intervju	5,00	1,54
av_nogomet	0,01	4,55
av_nogomet	0,10	4,01
av_nogomet	0,20	3,63
av_nogomet	0,50	2,56
a_nogomet	0,20	4,33
a_nogomet	0,50	3,90
a_nogomet	1,00	3,256
a_nogomet	5,00	1,71

Tabela 8.5: Korelacija objektivnih slikovnih ocen nabora TRAIN s subjektivnimi ocenami, združeno po scenskem tipu in skupna korelacija ter standardna deviacija za posamezno slikovno metriko.

TIP SCENE	IQI	MSE	LMSE	PSNR	NK	AD	SC	MD	NAE	SFF	scaM	VIF	VIFP	VSNR	WSNR	SSIM	MSSIM	NQM	IFC	IWMSE	IWPSNR	IWSSIM	CWSSIM	ESSIM	RFSIM	GMSD
av_film	0,951	-0,926	-0,951	0,980	0,749	-0,951	-0,652	-0,977	-0,949	0,989	-0,959	0,986	0,987	0,999	0,994	0,957	0,979	0,991	0,965	-0,975	0,999	0,982	0,994	0,967	0,979	-0,994
v_film	0,960	-0,905	-0,963	0,990	0,874	-0,879	-0,700	-0,983	-0,882	0,963	-0,982	0,985	0,983	0,989	0,994	0,977	0,978	0,990	0,986	-0,964	0,984	0,983	0,980	0,985	0,992	-0,999
av_formula	0,683	-0,428	0,335	0,216	0,071	-0,409	-0,205	0,213	-0,202	0,628	-0,322	-0,004	0,107	0,564	0,540	0,055	0,427	0,774	0,123	-0,598	-0,269	0,385	0,825	0,586	0,505	-0,418
v_formula	0,957	-0,935	-0,867	0,942	0,774	-0,947	-0,731	-0,344	-0,946	0,935	-0,975	0,894	0,878	0,888	0,954	0,927	0,950	0,986	0,879	-0,946	0,426	0,935	0,948	0,947	0,981	-0,931
av_intervju	0,915	-0,957	-0,949	0,873	0,657	-0,950	-0,884	-0,712	-0,939	0,958	-0,629	0,794	0,878	0,748	0,835	0,940	0,966	0,828	0,826	-0,972	0,064	0,946	0,789	0,947	0,857	-0,937
v_intervju	0,911	-0,926	-0,866	0,989	0,795	-0,946	-0,998	-0,985	-0,965	0,921	-0,944	0,907	0,904	0,989	0,995	0,888	0,914	0,993	0,858	-0,974	0,318	0,938	0,925	0,932	0,961	-0,969
av_nogomet	0,675	-0,636	0,383	0,889	0,534	-0,584	-0,534	-0,996	-0,581	0,812	-0,636	0,667	0,645	0,997	0,944	0,579	0,756	0,937	0,879	-0,850	0,980	0,779	0,954	0,769	0,884	-0,733
v_nogomet	0,861	-0,898	-0,890	0,893	0,412	-0,874	-0,315	-0,926	-0,871	0,834	-0,938	0,940	0,935	0,440	0,863	0,854	0,824	0,823	0,989	-0,825	0,839	0,846	0,876	0,805	0,914	-0,852
ZDRUŽENE OCENE KORELACIJ ZA POSAMEZEN TIP SCENE:																										
ABS(AVERAGE)	0,864	0,826	0,596	0,847	0,608	0,818	0,627	0,714	0,792	0,880	0,798	0,771	0,790	0,827	0,890	0,772	0,849	0,915	0,813	0,888	0,543	0,849	0,911	0,867	0,884	0,854
ABS(MEDIAN)	0,913	0,916	0,879	0,918	0,703	0,913	0,676	0,951	0,910	0,928	0,941	0,901	0,891	0,938	0,949	0,907	0,932	0,962	0,879	0,955	0,633	0,937	0,937	0,939	0,937	0,934
STDEV	0,119	0,191	0,591	0,259	0,264	0,206	0,268	0,437	0,269	0,119	0,243	0,331	0,296	0,221	0,154	0,316	0,189	0,092	0,286	0,131	0,483	0,200	0,074	0,138	0,161	0,197



Slika 8.30: Korelacija metrik *VIF* in *SSIM* za sceno *av_formula*: povprečne vrednosti in odstopanja vrednosti od povprečja objMOS za posamezen scenarij TRAIN za metriko kakovosti VIF (levo zgoraj) in SSIM (desno zgoraj) ter subMOSv za iste scenarije.

Povprečni odziv *objTrainMOSa_a*, *objTrainMOSv_v*, *objTrainMOSa_av* in *objTrainMOSv_av* smo aproksimirali po sledečih enačbah:

$$objTrainMOS_{a\_a}(PL) = \begin{cases} -0{,}432 * \ln(PL) + 2{,}994, & ko\ je\ 0{,}001 \leq PL \leq 100 \\ 5, & ko\ je\ PL < 0{,}001 \end{cases} \quad (8.1)$$

$$objTrainMOS_{v\_v}(PL) = \begin{cases} -2{,}626 * \ln(PL) + 2{,}409, & ko\ je\ 0{,}373 \leq PL \leq 1{,}710 \\ 5, & ko\ je\ PL < 0{,}373 \\ 1, & ko\ je\ PL > 1{,}710 \end{cases} \quad (8.2)$$

$$objTrainMOS_{a\_av}(PL) = \begin{cases} -1{,}008 * \ln(PL) + 3{,}087, & ko\ je\ 0{,}150 \leq PL \leq 7{,}930 \\ 5, & ko\ je\ PL < 0{,}150 \\ 1, & ko\ je\ PL > 7{,}930 \end{cases} \quad (8.3)$$

$$objTrainMOS_{v\_av}(PL) = \begin{cases} -0{,}617 * \ln(PL) + 0{,}768, & ko\ je\ 0{,}001 \leq PL \leq 0{,}687 \\ 5, & ko\ je\ PL < 0{,}001 \\ 1, & ko\ je\ PL > 0{,}687 \end{cases} \quad (8.4)$$

Determinacijski koeficient omenjenih logaritemskih aproksimacij prikazuje tabela 8.6. Kljub temu da je ujemanje za $objTrainMOS_{a\_a}$ in $objTrainMOS_{v\_v}$ nizko, ima tak model boljšo učinkovitost za predikcijo PL-vrednosti izven obstoječih podatkov, npr. *PL > 5,00 %* in *PL < 0,01 %*, od ostalih aproksimacij, tj. eksponentne, linearne ali polinomične, z *n* členi.

Tabela 8.6: Determinacijski koeficienti slikovnih in govornih metrik na naboru TRAIN.

Model	**R^2**
$objTrainMOS_{a\_a}(PL)$	0,634
$objTrainMOS_{v\_v}(PL)$	0,558
$objTrainMOS_{a\_av}(PL)$	0,904
$objTrainMOS_{v\_av}(PL)$	0,970

S statističnim modeliranjem smo nato izdelali linearni regresijski model na podlagi podatkov TRAIN za objektivno vrednotenje kakovosti v primeru AV-modalnosti:

$$\begin{aligned} objMOS_{[A,V,AV]} = \quad & -1{,}8904 * V \\ & + 0{,}2282 * A \\ & - 0{,}1027 * S \\ & - 0{,}1253 * PL \\ & + 0{,}7088 * objMOS_a \\ & + 0{,}0856 * objMOS_v \\ & + 0{,}7839 \end{aligned} \quad (8.5)$$

pri tem je *V* prisotnost videa ($V \epsilon [0, 1]$), *A* prisotnost avdia ($A \epsilon [0, 1]$), *S* tip scene (tabela 8.7), *PL* izguba podatkovnih paketov (v odstotkih), $objMOS_a$ ocena avditorne modalnosti izražena kot ocena MOSLQO, in $objMOS_v$ ocena video modalnosti izražena kot ocena NQM, samo za degradirane posnetke.

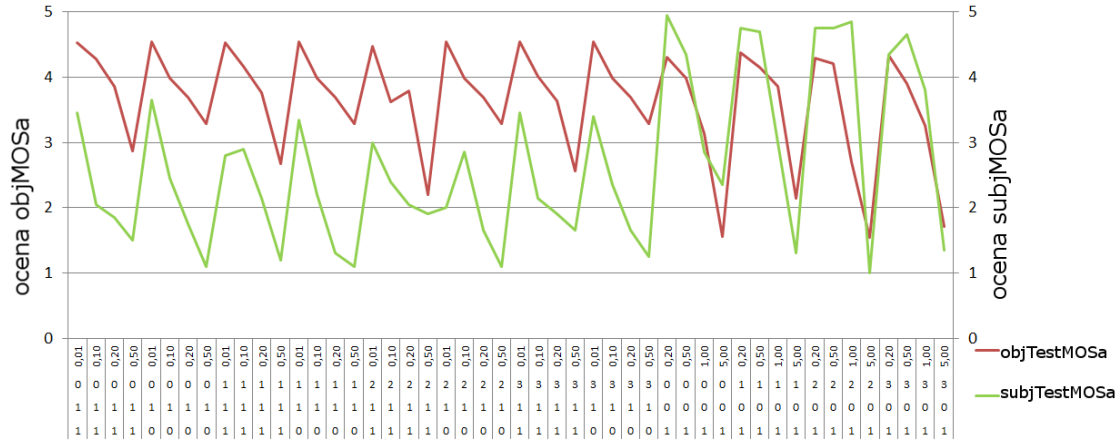
Tabela 8.7: Tip scene večmodalnega modela kakovosti.

S	**Tip scene**
0	Film
1	Formula
2	Intervju
3	Nogomet

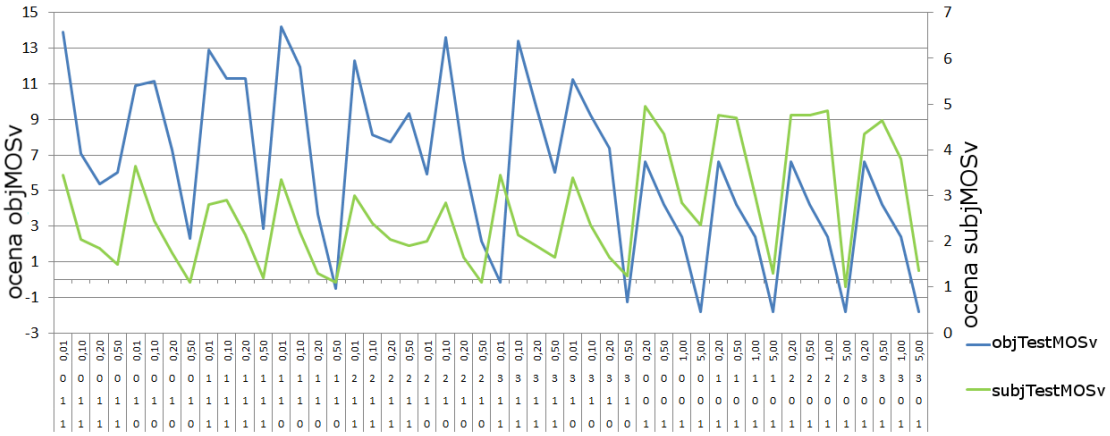
Ker je model mogoče uporabljati tudi samo za 1 modalnost, je potrebno v enačbo 8.5 implementirati preddefinirane vrednosti (konstante) za primere, ko objektivna ocena modalnosti ni na voljo. To so vrednosti ustreznih logaritemskih funkcij, ki definirajo *povprečen odziv enomodalne metrike* v danih pogojih, definirane s funkcijami 8.1, 8.2, 8.3 in 8.4.

Testiranje na naboru TEST je pokazalo dobro korelacijo za modalnost A in dobro korelacijo za V (slika 8.31, slika 8.32). Paersonov korelacijski koeficient znaša:

$$\boldsymbol{R(subTestMOS, objTestMOS) = 0,892} \qquad (8.6)$$



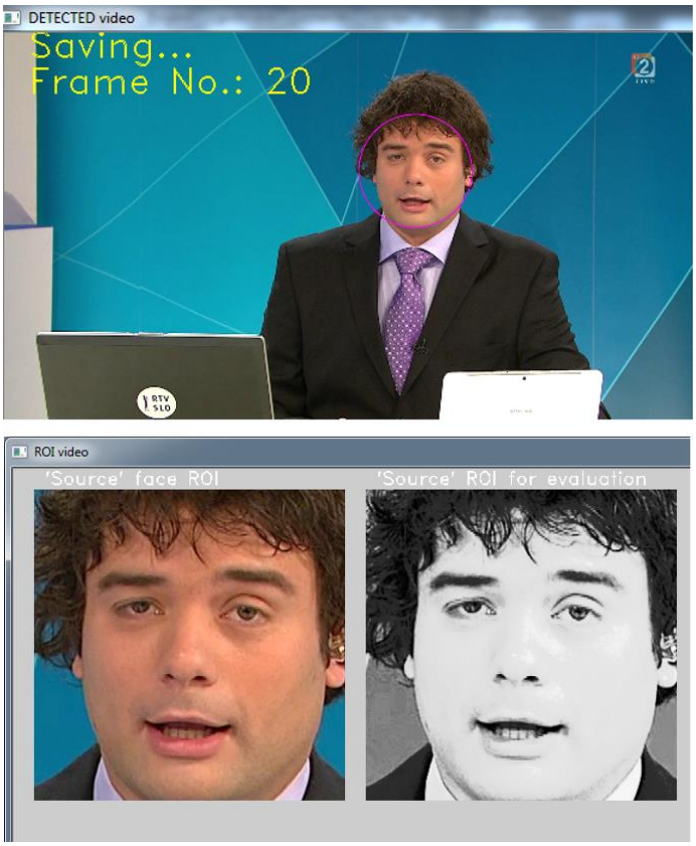
Slika 8.31: *ObjTestMOSa* in *subTestMOSa* vrednosti na naboru TEST: na abscisi so (od zgoraj navzdol): količina PL (v odstotkih), tip scene, prisotnost vizualne modalnosti in prisotnost zvočne modalnosti.



Slika 8.32: *ObjTestMOSv* in *subTestMOSv* vrednosti na naboru TEST, na abscisi so (od zgoraj navzdol): količina PL (v odstotkih), tip scene, prisotnost vizualne modalnosti in prisotnost zvočne modalnosti.

8.3. Rezultati določanja vpliva osredotočenosti uporabnika na vizualna polja ROI

Pri detekciji polj ROI smo najprej preverili uspešnost detekcije obraza za dani detektor, ki je bil razvit na sivinski podatkovni bazi (slika 8.33).



Slika 8.33: Primer detekcije obraza v posnetku *Intervju_narator* (zgoraj), ujemanje strukture na posnetku (spodaj, levo) in obdelava DSP strukture obraza (histogramska poravnava), primerna za primerjavo s podatkovno bazo (spodaj, desno).

Privzete vrednosti so dale nezadovoljive rezultate, saj detektor ni bil optimiziran za HD-vsebine in je zaznaval le zelo drobne strukture. Konfiguracijo detektorja smo zato priredili za velikosti obraza, večje od *150 x 150 slikovnih pik* (privzeto min. okno kaskadnega klasifikatorja Haar je sicer *20 x 20 slikovnih pik*). S tem smo še vedno bili znotraj mej povprečne velikosti obraza in smo tudi dodatno izločili morebitne (prostorsko manjše) *obraze* v ozadju, ki niso bili fokus posnetka na scenah tipa *intervju*.

Hkrati s tem smo zmanjšali časovno kompleksnost detekcijskega algoritma zaradi manjšega števila potrebnih obhodov slike. Rezultati so bili zadovoljivi (tabela 8.8).

Tabela 8.8: Detekcija polj ROI z oknom obraza *150 x 150 slikovnih pik*, rezultati so normirani na posamezen posnetek.

Okvir: / **Posnetek**	**brez detekcije [%]**	**s pravilno detekcijo [%]**	**z večkratno detekcijo [%]**
intervju_narator	1,05	95,83	3,12
intervju_naratorka	1,12	96,85	2,03
intervju_robert	1,30	85,98	12,72
intervju_sara	0,81	79,09	20,10

Opazili smo težave zaradi večkratne detekcije obrazu podobnih struktur ter posamičnih okvirjev brez detekcije, ki so verjetno posledica napake pri statističnem modeliranju samega algoritma in vrednosti blizu mejnih nivojev detekcije (slika 8.34).

Slika 8.34: Primer napačne večkratne detekcije strukture *obraz*.

Težave smo reševali z uporabo dveh mehanizmov:

- povečanjem parametra **minNeighbors** iz programskega paketa OpenCV in
- predpostavko, da mora vsaka struktura *obraz* imeti tudi strukturo *usta*.

Povečanje parametra *minNeighbors* pomeni združevanje detektiranih struktur, ki so zelo blizu. S tem smo tvegali nastanek okvirjev *brez detekcije*, na drugi strani pa smo v primeru pravilne detekcije zmanjšali število multipliciranih detekcij iste strukture.

Na podlagi meritev oblike povprečnega obraza smo definirali idealni center pozicije ust ter ga priključili k vektorjem Haar za strukturo *usta*, kar je služilo kot referenčna točka za detekcijo obraza:

$$\boldsymbol{idealniCenterUst}(\boldsymbol{x},\boldsymbol{y}) = \left[\left(\frac{\boldsymbol{sirinaObrazaROI}}{\mathbf{2}}\right),\left(\frac{\boldsymbol{visinaObrazaROI}}{\mathbf{1,304}}\right)\right] \quad (8.7)$$

Kadar so bile napake prisotne zaradi posamičnih *okvirjev brez detekcije* smo predpostavili, da so v videu razmere med zaporednima okvirjema statične in se isti obraz v naslednjem okvirju nahaja na približno istem mestu kot v prejšnjem. S kratkim časovnim oknom velikosti 5 okvirjev (200 ms) smo tako omejili pojav ne-zaznanih in delno tudi dislociranih struktur. S temi predpostavkami smo dosegli 100% detekcijo strukture obraza na danem naboru testnih podatkov.

Naslednji korak je bila performančna analiza diferenciranega objektivnega vrednotenja kakovosti. Najprej smo izmerili hitrosti izbranih kakovostnih metrik. Nato smo za vse posnetke izračunali povprečno prostorsko velikost polja ROI. Ta je znašala **380 x 380 slikovnih pik**. Predpostavili smo okroglo strukturo *obraza*, kar je predstavljalo ~ 5,5 % celotne površine video okvirja (slika 8.35).

Slika 8.35: Razmerje strukture *obraz* in ozadja.

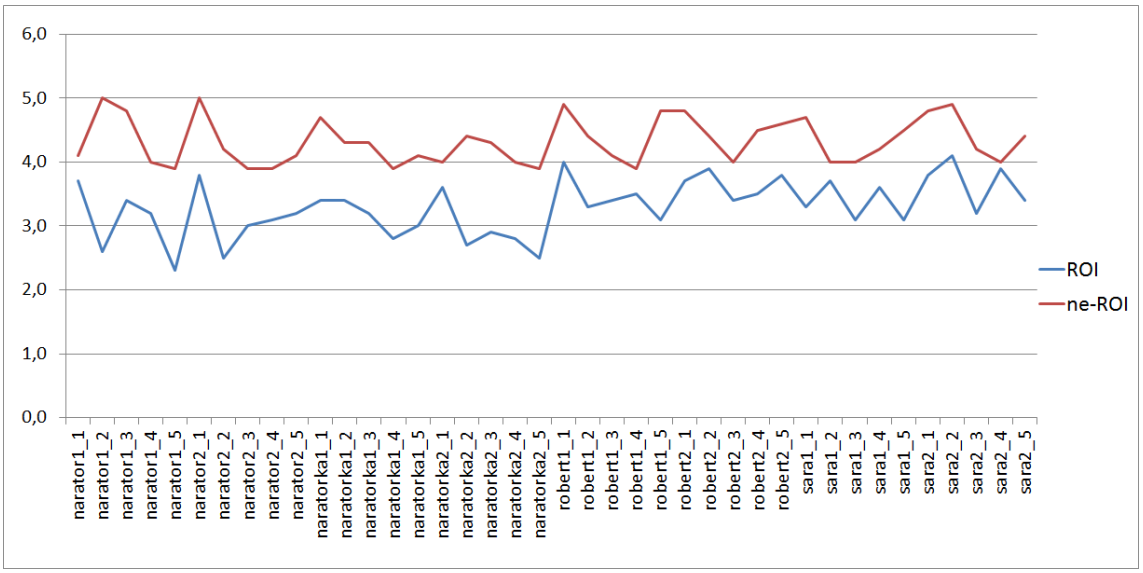
Pri tej predpostavki smo izračunali povprečno časovno kompleksnost za dana algoritma evalvacije kakovosti, v primeru, da se za bolj pomemben predel slike (polje

ROI) uporabi natančnejša (in kompleksnejša) kakovostna metrika, za ozadje pa enostavnejša (tabela 8.9).

Tabela 8.9: Časovna kompleksnost vrednotenja kakovosti video okvirja pri porazdeljenem vrednotenju.

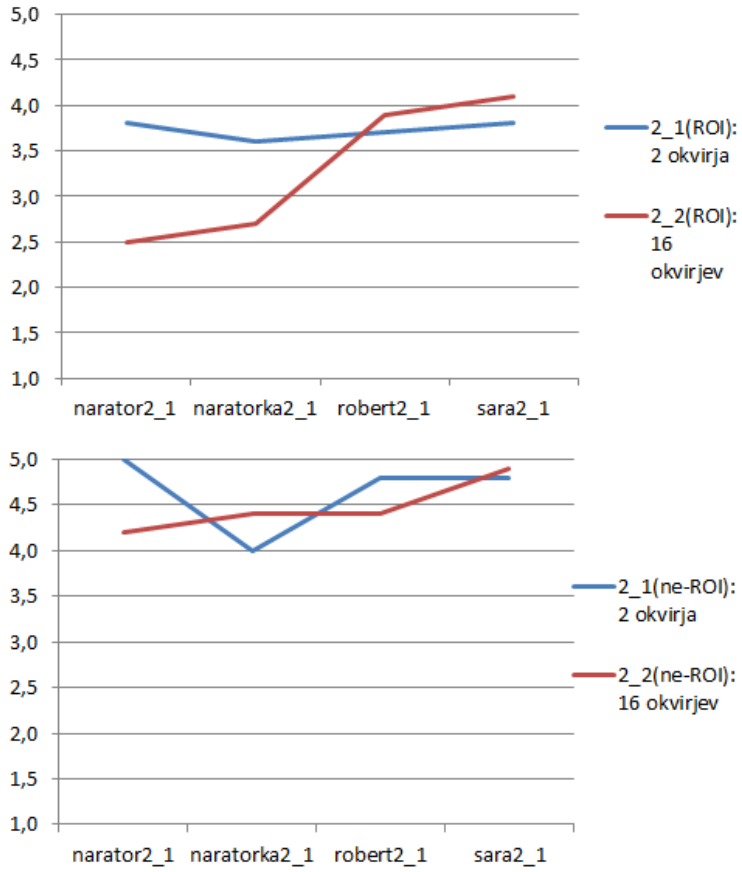
Časovna kompleksnost enega video okvirja	**PSNR čas [sek.]**	**NQM čas [sek.]**
Brez porazdelitve	0,019	1,625
S porazdelitvijo(5,5% ROI)	**0,107**	

Pri analizi *subjektivnega vrednotenja ROI in ne-ROI polj* smo analizirali pomembnost polj ROI na podlagi rezultatov subjektivnih testov (priloga V). Kljub skoraj enaki objektivni oceni PSNR smo opazili razliko v rezultatih povprečja subjektivnih ocen med posnetki z degradacijo v polju ROI in izven polja ROI v enakih scenarijih (slika 8.36). Skupna povprečna razlika za vse posnetke je znašala **1,02 ocene MOS**. To dokazuje precej večjo osredotočenost uporabnikov na polja ROI, tj. strukturo *obraza,* saj je bila zaznavnost degradacij v ne-ROI območjih precej manjša, kljub večji povprečni prostorski velikosti degradacij. Ta je znašala 228,75 % povprečne prostorske velikosti degradacije v scenarijih s polji ROI, kar bi bilo seveda pogojeno z večjo izgubo omrežnih paketov. Razlog potrebne povečave polja degradacije v ne-ROI je posledica manjših kontrastnih razlik na teh območjih, kar posledično vodi v manjšo razliko v oceni PSNR za isto površino vizualne napake.



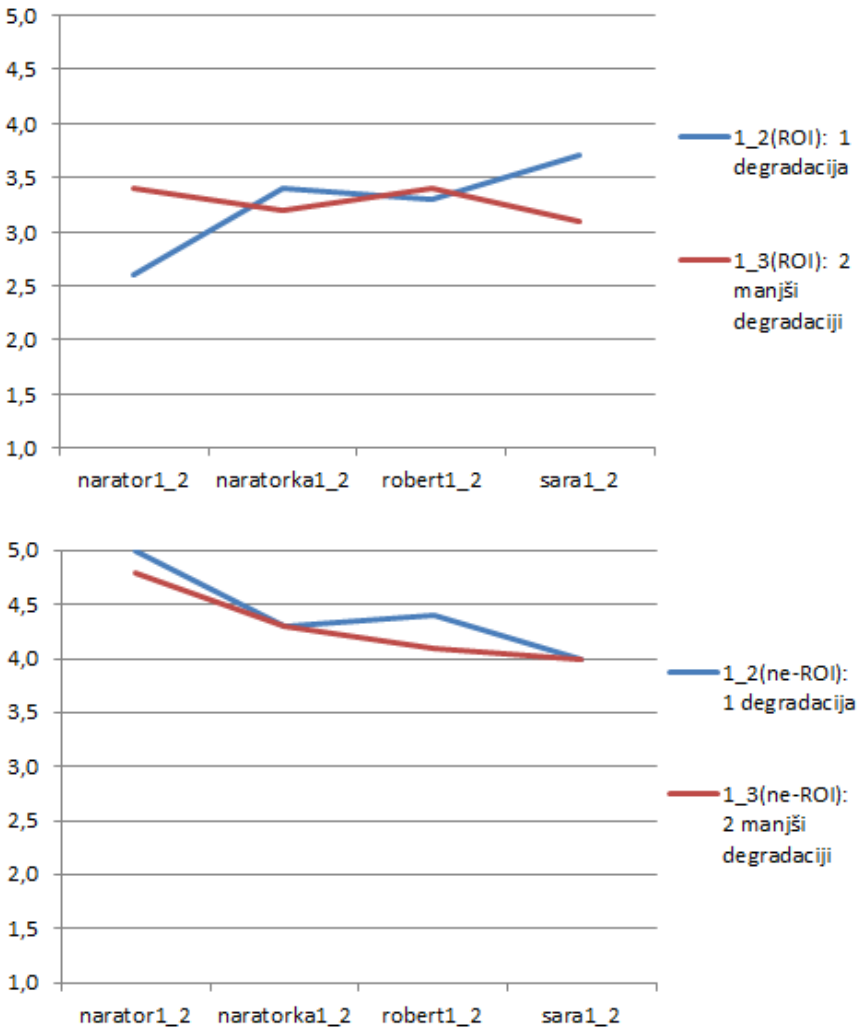
Slika 8.36: Razlika subjektivne ocene za pojav degradacije v polju ROI in ne-ROI za enake scenarije degradacij.

V nadaljevanju smo preverili vpliv dolžine ene degradacije na povprečno subjektivno oceno (slika 8.37). Ugotovili smo, da v testiranih pogojih dolžina ni imela večjega vpliva, ampak je bila pomembna zgolj prisotnost degradacije. To je verjetno posledica relativno kratkega časovnega okna degradacije (2/25 oz. 16/25 sekunde) ter majhne razlike med vrednostmi dolžine okna za ta scenarija.

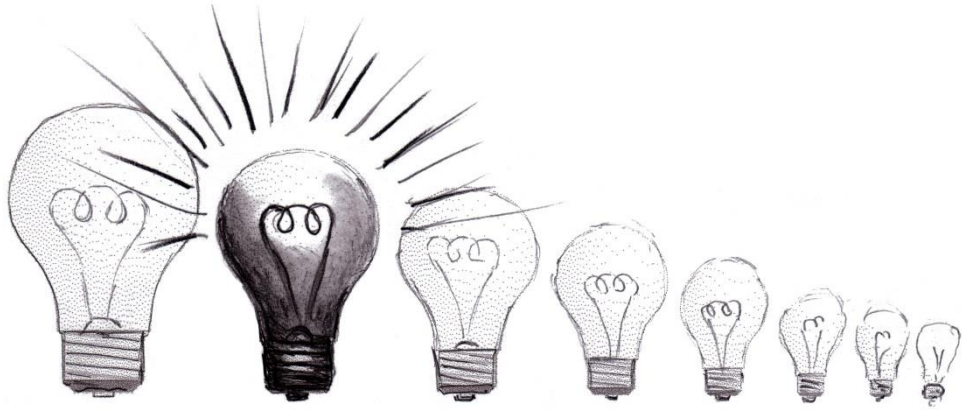


Slika 8.37: Vpliv dolžine degradacije na subMOS za območja ROI (zgoraj) in ne-ROI (spodaj).

Preverili smo tudi vpliv števila degradacij pri enaki količini okvarjenih okvirjev. Primerjali smo scenarij z eno degradacijo dolžine 8 okvirjev in scenarij z dvema degradacijama dolžine 2 × 4 okvirje (slika 8.38). V scenarijih degradacij v polju ROI tudi v tem primeru ni bilo opaziti konstantnega trenda vrednosti. Nasprotno pa je v scenarijih z degradacijami izven polj ROI v vseh scenah *intervju* prihajalo do slabših subjektivnih ocen v primeru scenarija z dvema krajšima degradacijama.



Slika 8.38: Vpliv števila degradacij na subjektivno oceno.



9. Zaključek

V doktorski disertaciji smo se posvetili vrednotenju kakovosti sodobnih večmodalnih storitev. Skladno z vidiki obravnave smo raziskavo razdelili na tri dele. V prvem delu smo obravnavali vpliv degradacij na vhodno modalnost večmodalne storitve IVR, kjer smo izpostavili učinke omrežnih degradacij na uspešnost razpoznavanja govornih ukazov v modulu ASR. Analiza je pokazala precejšnja odstopanja med uporabljenimi govornimi kodeki in njihovimi konfiguracijami. Na naboru transkodiranih posnetkov, degradiranih z izgubo paketov med 1 % in 35 %, smo analizirali uspešnost razpoznavanja govora v storitvi IVR. Pričakovano je bila uspešnost razpoznavanja degradiranega govora za različne kodeke v veliki meri odvisna od nominalne hitrosti kodeka in uporabljenega algoritma. To je v skladu s teorijo informacij, saj stisnjeni podatki nosijo več informacij na enoto in je zato tudi izguba bolj destruktivna. Z analizo smo določili mejno vrednost objektivne ocene kakovosti govora, ki še zadostuje za uporabo razpoznavalnika govora in je predstavljala osnovo za učenje klasifikatorja vhodne modalnosti ASR/DTMF. Za predlagani koncept klasifikatorja smo uporabili Gaussove modele. S konfiguracijo različnih parametrov učnega nabora smo izdelali

nabor modelov. Rezultati kažejo precej nelinearno pravilnost klasifikacije v odvisnosti od količine izgubljenih paketov. Zgolj 56% natančnost klasifikacije vhodne modalnosti pri vrednosti izgube paketov ~5 % kaže na pomanjkljivosti uporabe modela pri teh vrednostih izgube paketov. Slabše rezultate klasifikacije vhodne modalnosti v tem območju izgube paketov omili dejstvo, da je uspešnost razpoznavanja govora pri takšnih degradacijah še vedno visoka. Po drugi strani pa se uspešnost pravilno klasificirane vhodne modalnosti povečuje z večanjem vrednosti izgube paketov, kjer doseže do 90% pravilnost klasifikacije modalnosti. Predpostavljamo, da je slabše delovanje klasifikatorja na določenih območjih lahko posledica izbire pragovne meje, saj v tem področju prihaja do preklopov pri izbiri vhodne modalnosti sistema. Menimo, da bi bilo možno izboljšati rezultat klasifikacije na območju majhne izgube paketov z uporabo dodatne baze učnih posnetkov, ki bi vsebovala večji delež nedegradiranega govora. Predpostavka je podana na opažanju, da smo najboljše rezultate dobili z odstranitvijo negovornega signala v posnetkih. Prihodnje delo na področju klasifikacije vhodne modalnosti v odvisnosti od kakovosti bi se lahko osredotočilo na podatkovno voden način adaptivne izbire pragovne meje ter na uporabo drugih načinov klasifikacije.

Drugi del disertacije se je nanašal na vrednotenje večmodalnih vsebin. Izdelava lastne referenčne baze je pripomogla k pridobitvi dragocenih podatkov subjektivnega testiranja, kjer smo uporabili izključno izvorno gradivo visoke ločljivosti. To pripomore k obravnavi sodobnih vsebin, k evalvaciji objektivnih enomodalnih metrik kakovosti za avdio in video ter k združevanju enomodalnih metrik v večmodalno metodo vrednotenja. V tem delu predlagani linearni model za objektivno vrednotenje kakovosti večmodalnih vsebin je pokazal dobro korelacijo s testnimi posnetki na širokem območju vrednosti izgube paketov, tudi na področju mejnih vrednosti, kjer je prihajalo do izrazitih degradacij. Model je zgrajen na osnovi rezultatov enomodalnega pristopa ter dodatno upošteva širok nabor vplivnih parametrov. Eden izmed pomembnejših je medmodalni učinek, ki ga enomodalne metode vrednotenja kakovosti ne morejo zaznati. Prednost modela je njegova široka uporabnost: zaradi utežnostnih koeficientov je uporaba možna ne samo v večmodalnih storitvah, temveč je model zmožen tudi vrednotenja storitev, kjer je prisotna samo avdio ali samo video modalnost. Paersonov korelacijski koeficient primerjave s subjektivnimi ocenami znaša 0,892. Prednost modela je tudi v njegovi matematični učinkovitosti, saj je nelinearna kompleksnost HAS/HVS implicitno modelirana v vključenih enomodalnih algoritmih. Predlagana metoda ima zaradi tega lahko določene prednosti v aplikativni uporabi. Zaradi uporabe

sodobnih kodekov (H.264/AVC, AAC) predpostavljamo, da so spoznanja aktualna tudi za prihodnje raziskovalno delo in s tem odpirajo možnosti za nadaljnje izboljšave modela vrednotenja kakovosti.

V tretjem delu disertacije smo obravnavali porazdeljeno vrednotenje kakovosti z detekcijo vizualnih polj ROI. Predpostavili smo, da je osredotočenost opazovalca večja na polje ROI, zato je tam smiselna uporaba računsko kompleksnejše metrike vrednotenja kakovosti, ki daje boljšo subjektivno korelacijo, na ostalih vizualnih območjih pa se uporabi računsko manj zahtevna metrika z meritvijo maksimalne vrednosti napake. Težave z nepravilno detekcijo polja ROI smo reducirali z modifikacijo detektorja. Potrebno povečanje drsečega okna je bila logična posledica zaradi obdelave vsebin visoke ločljivosti. Zaradi lastnosti šibkih klasifikatorjev Haar so bile iste strukture na sosednjih območjih večkrat detektirane kar smo preprečili z omejitvijo izbire. Vpliv dislociranih struktur smo omejili z uporabo časovnega okna petih okvirjev, kjer smo predpostavili skoraj statične lastnosti. To lahko sicer predstavlja omejitev v primeru hitrih prehodov scene. Na testnih posnetkih smo z omenjenimi postopki dosegli zelo dobro detekcijo strukture obraza, kar predstavlja dobro izhodišče za porazdeljeno vrednotenje kakovosti. V ta namen smo nato izvedli performančno analizo in pokazali časovno kompleksnost vrednotenja kakovosti za primer polja ROI velikosti 5,5 % celotne površine okvirja, kar je izhajalo iz povprečne velikosti strukture obraza. V prihodnje bi bilo smiselno detektor polja ROI preizkusiti še na posnetkih v sorodnih bazah videa. Sledila je subjektivna evalvacija, s katero smo potrdili pomembnost polj ROI in posledično tudi smiselnost porazdeljenega vrednotenja kakovosti storitve. Skupna povprečna razlika med posnetki z degradacijo v ne-ROI območjih in tistimi z degradacijo v polju ROI je znašala več kot 1 oceno MOS v prid scenarijem ne-ROI. Ker smo normirali vrednosti PSNR za enake tipe scenarijev degradacij, je bilo v scenarijih ne-ROI potrebno povečati velikost vizualne degradacije. Ta je v povprečju znašala 228,75 % velikosti napake v scenarijih ROI, kar kaže na to, da kljub večji degradaciji, gledano iz stališča omrežja, zaradi bioloških lastnosti HVS, tj. manjše sposobnosti zaznave s perifernim vidom, pogojuje boljšo oceno MOS. Prihodnje delo na tem področju bi se lahko osredotočilo na podrobnejšo analizo vizualne pozornosti, npr. z raziskavo sledenja pogledu oči. Smiselno bi bilo tudi preveriti neposredno relacijo ocene kakovosti ROI-območij v odvisnosti od omrežnih parametrov, tj. izgube omrežnih paketov.

V nadaljevanju podajamo analizo v uvodu doktorske disertacije zastavljenih hipotez.

Hipoteza 1:

Definiramo lahko klasifikator vhodne modalnosti, ki na podlagi objektivne ocene kakovosti uporabniškega vnosa določa tip vhodne modalnosti.

Glede na rezultate analize lahko sklepamo o skupni uspešnosti klasifikacije vhodnega govornega signala (slika 8.16). Ker na uporabniško izkušnjo najpomembneje vpliva pravilnost vnosa, je nujen preklop na robustnejšo modalnost v primeru, da se stopnja degradacij poveča do te mere, da storitev uporabniku ne zagotavlja ustreznega nivoja kakovosti. V scenarijih, kjer degradacija presega nivo zadovoljivega razpoznavanja govora in bi lahko odločilno vplivala na uporabniško izkušnjo, predlagan klasifikator modalnosti izkazuje dobre lastnosti. Zaradi tega lahko trdimo, da smo potrdili hipotezo 1.

Hipoteza 2:

Objektivno oceno kakovosti večmodalnih vsebin lahko napovemo s primerno združitvijo enomodalnih ocen kakovosti ter z upoštevanjem medmodalnega učinka. Takšen model kakovosti bo izkazoval dobro korelacijo za različne tipe posnetkov in degradacij. Pri tem lahko definiramo dodatne funkcije, ki zmanjšajo časovno kompleksnost evalvacije z upoštevanjem vpliva polja ROI.

Na podlagi rezultatov govorne metrike in izbire uspešne metrike kakovosti slik smo predlagali model vrednotenja izhodne modalnosti večmodalnih storitev. Ta združuje dve enomodalni metriki kakovosti, pri tem pa je ena namenjena evalvaciji avdio modalnosti, druga pa evalvaciji vizualne modalnosti. S parametrizacijo koeficientov, ki določajo medmodalni vpliv in tip scene posnetka ter upoštevajo delovanje v širokem območju vrednosti izgube paketov, smo dobili skupno korelacijo (za avdio, video in avdio-video scenarije) s subjektivno oceno, ki znaša 0,892. Glede na število vplivnih parametrov predstavlja to dober rezultat. S predlogom vpeljave porazdeljenega vrednotenja kakovosti vizualne modalnosti smo zasnovali detektor obraza. Dobri rezultati uspešnosti detekcije so pokazali smiselnost uporabe takšnega

pristopa, s performančno analizo pa smo potrdili optimizacijo kompleksnosti. Na podlagi predstavljenih dejstev lahko povzamemo, da smo potrdili tudi hipotezo 2.

10. Literatura

[1] W. C. Hardy, *QoS Measurement and Evaluation of Telecommunications Quality of Service*: Wiley, 2001.

[2] K. M. Lee in J. Lai, "Speech versus touch: A comparative study of the use of speech and DTMF keypad for navigation" v *International Journal of Human-Computer Interaction,* št. 19, str. 343-360, 2005.

[3] C. Delogu, A. Di Carlo, P. Rotundi in D. Sartori, "A comparison between DTMF and ASR IVR services through objective and subjective evaluation" na *IEEE 4th Workshop Interactive Voice Technology for Telecommunications Applications (IVTTA '98)*, str. 145 - 150, 1998.

[4] S. Furui, "Generalization problem in ASR acoustic model training and adaptation" na *IEEE Workshop Automatic Speech Recognition & Understanding (ASRU 2009)*, str. 1-10, 2009.

[5] J. P. Haton, "Automatic speech recognition: A review" v *Enterprise Information Systems V,* str. 6-11, 2004.

[6] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi in C. Wellekens, "Automatic speech recognition and speech variability: A review" v *Speech Communication,* št. 49, str. 763-786, okt. - nov. 2007.

[7] C. H. Lee, F. K. Soong, K. K. Paliwal, *Automatic Speech and Speaker Recognition: Advanced Topics*: Springer, 1996.

[8] D. O'Shaughnessy, *Speech Communications: human and Machine*: Universities Press (India) Pvt. Limited.

[9] K. Takagi, S. Miyaji, S. Sakazawa in Y. Takishima, "Conversion of MP3 to AAC in the Compressed Domain" na *IEEE 8th Workshop Multimedia Signal Processing*, str. 132-135, 2006.

[10] B. Duysburgh, S. Vanhastel, B. De Vreese, C. Petrisor in P. Demeester, "On the influence of best-effort network conditions on the perceived speech quality of VoIP connections" na *Tenth International Conference of Computer Communications and Networks*, str. 334-339, 2001.

[11] Wenyu Jiang H. Schulzrinne, "Modeling of Packet Loss and Delay and their Effect on Real-Time Multimedia Service Quality", v *PROCEEDINGS OF NOSSDAV,* 2000.

[12] D. Pratsolis, N. Tsourakis in V. Digalakis, "Degradation of Speech Recognition Performance over Lossy Data Networks" v *Proceedings of the Third Acm Workshop on Wireless Multimedia Networking and Performance Modeling,* str. 88 - 91, 2007.

[13] J. Kenny. *IVR: Speech Recognition vs. DTMF*. Dostopno na: http://ivr.tmcnet.com/topics/ivr/articles/168556-ivr-speech-recognition-vs-dtmf.htm, obiskano: nov. 2014.

[14] S. Hanwu, L. Shue in C. Jianfeng, "Investigations into the relationship between measurable speech quality and speech recognition rate for telephony speech", v *Proceedings of Acoustics, Speech, and Signal Processing (ICASSP '04)*, str. 865 - 868, zbr. 1.

[15] S. Mohamed, F. Cervantes-Perez in H. Afifi, "Audio quality assessment in packet networks: An "Inter-Subjective" Neural Network model" na *15th International Conference on Information Networking*, str. 579 - 586, 2001.

[16] A. E. Mahdi, "Voice Quality Measurement v Modern Telecommunication Networks", na *14th International Workshop on Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, str. 25 - 32, 2007.

[17] A. E. Conway, "A passive method for monitoring voice-over-IP call quality with ITU-T objective speech quality measurement methods", na *IEEE International Conference on Communications (ICC 2002)*, str. 2583 - 2586, zbr. 4, 2002.

[18] A. E. Conway, "Output-based method of applying PESQ to measure the perceptual quality of framed speech signals", na *Wireless Communications and Networking Conference (WCNC 2004)*, str. 2521 - 2526, št. 4, 2004.

[19] D. Campbell, E. Jones in M. Glavin, "Audio quality assessment techniques-A review, and recent developments" v *Signal Processing*, št. 89, str. 1489 - 1500, avg. 2009.

[20] A. W. Rix, J. G. Beerends, K. Doh-Suk, P. Kroon in O. Ghitza, "Objective Assessment of Speech and Audio Quality - Technology and Applications" v *IEEE Transactions on Audio, Speech, and Language Processing*, št. 14, str. 1890 - 1901, 2006.

[21] A. Raake, J. Gustafsson, S. Argyropoulos, M. Garcia, D. Lindegren, G. Heikkila, M. Pettersson, P. List in B. Feiten, "IP-Based Mobile and Fixed Network Audiovisual Media Services" v *IEEE Signal Processing Magazine*, št. 28, str. 68 - 79, 2011.

[22] A. Watson in M. A. Sasse, "Multimedia Conferencing via Multicast: Determining the Quality of Service Required by the End User", str. 189 - 194, 1997.

[23] M. Carnec, P. Le Callet in D. Barba, "Full reference and reduced reference metrics for image quality assessment" na *Seventh International Symposium on Signal Processing and Its Applications*, št. 1, str. 477 - 480, 2003.

[24] W. Zhou in A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures" v *IEEE Signal Processing Magazine*, št. 26, str. 98 - 117, 2009.

[25] E. P. Simoncelli, W. T. Freeman, E. H. Adelson in D. J. Heeger, "Shiftable Multiscale Transforms" v *Ieee Transactions on Information Theory*, št. 38, str. 587 - 607, mar. 1992.

[26] W. Zhou, A. C. Bovik, H. R. Sheikh in E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity" v *IEEE Transactions on Image Processing*, št. 13, str. 600 - 612, 2004.
[27] E. P. Simoncelli in E. H. Adelson, "Subband Transforms", MIT, 1990.
[28] G. Bernd, H. Frank in H. Uwe, "Subband Image Coding" v *Design and Applications of Subbands and Wavelets*, ed Norwell, MA: Kluwer Academic Publishers, 1995, str. 213 - 250.
[29] S. Winkler, "A perceptual distortion metric for digital color video" v *Human Vision and Electronic Imaging Iv*, št. 3644, str. 175 - 184, 1999.
[30] A. Ninassi, O. Le Meur, P. Le Callet in D. Barba, "Which Semi-Local Visual Masking Model for Wavelet Based Image Quality Metric?" na *15th IEEE International Conference on Image Processing*, str. 1180 - 1183, 2008.
[31] S. Rimac-Drlje, D. Žagar in G. Martinović, "Spatial Masking and Perceived Video Quality in Multimedia Applications" na *16th International Conference on Systems, Signals and Image Processing (IWSSIP 2009)*, str. 1 - 4, 2009.
[32] A. Eden, "Studies of the Relevance of Spatial and Temporal Masking Effects in Video Quality Evaluation", na *13th IEEE International Symposium on Consumer Electronics*, str. 469 - 473, 2009.
[33] M. Slanina, T. Kratochvil, L. Polak in V. Ricny, "Analysis of Temporal Effects in Quality Assessment of High Definition Video" v *Radioengineering*, št. 21, str. 63 - 69, apr. 2012.
[34] Z. Wang in A. C. Bovik, "Contrast Sensitivity Functions" v *Modern Image Quality Assessment*, urednik: Morgan & Claypool publishers, str. 23, 2006.
[35] Z. Wang, A. C. Bovik, "Light Adaptation" v *Modern Image Quality Assessment*, urednik: Morgan & Claypool publishers, 2006, str. 23-25.
[36] Z. Wang, A. C. Bovik in L. G. Lu, "Why is image quality assessment so difficult?" na IEEE *International Conference on Acoustics, Speech, and Signal Processing*, str. 3313 - 3316, 2002.
[37] M. Usher in E. Niebur, "Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention" v *Journal of Cognitive Neuroscience*, št. 8, str. 311 - 327, jul. 1996.
[38] P. Le Callet, O. Le Meur, D. Barba in D. Thoreau, "Bottom-up visual attention modeling: Quantitative comparison of predicted salience maps with observers eye-tracking data" v *Perception*, št. 33, str. 120 - 121, 2004.
[39] J. B. Hopfinger, M. H. Buonocore in G. R. Mangun, "The neural mechanisms of top-down attentional control" v *Nature Neuroscience*, št. 3, str. 284 - 291, mar. 2000.
[40] S. Kastner in L. G. Ungerleider, "Mechanisms of visual attention in the human cortex" v *Journal of Cognitive Neuroscience*, str. 315–341, 2000.
[41] Z. K. Lu, W. S. Lin, X. K. Yang, E. P. Ong in S. S. Yao, "Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation" v *IEEE Transactions on Image Processing*, št. 14, str. 1928 - 1942, nov. 2005.
[42] ITU-T, "Objective quality measurement of telephorie-band (300-3400 Hz) speech codecs" v *ITU-T Recommendation P.861*, Geneva, Švica: ITU-T, 1998.

[43] S. Voran, "Estimation of perceived speech quality using measuring normalizing blocks" na *IEEE Workshop on Speech Coding for Telecommunications,* str. 83 - 84, 1997.

[44] S. H. Wang, A. Sekey in A. Gersho, "An Objective-Measure for Predicting Subjective Quality of Speech Coders" v *IEEE Journal on Selected Areas in Communications,* št. 10, str. 819 - 829, jun. 1992.

[45] A. W. Rix in M. P. Hollier, "The perceptual analysis measurement system for robust end-to-end speech quality assessment" na *IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings,* str. 1515 - 1518, 2000.

[46] A. Rix, R. Reynolds in M. Hollier, "Perceptual Measurement of End-to-End Speech Quality Over Audio and Packet-Based Networks" na *Audio Engineering Society Convention,* št. 106, 1999.

[47] A. W. Rix, J. G. Beerends, M. P. Hollier in A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs" na *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01),* str. 749 - 752, št. 2, 2001.

[48] ITU-T, "P.862.2 : Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs", urednik: ITU-T, 2007.

[49] P. Pocta in J. Holub, "Predicting the Quality of Synthesized and Natural Speech Impaired by Packet Loss and Coding Using PESQ and P.563 Models" v *Acta Acustica United with Acustica,* št. 97, str. 852 - 868, sept. - okt. 2011.

[50] P. Pocta in T. Terpak, "Packet Loss and Coding Impact on Quality of Synthesized Speech Predicted by PESQ and P.563 Models" v *Measurement of Speech, Audio and Video Quality in Networks,* str. 26 - 36, 2010.

[51] F. Mousavipour in M. J. Khosravipour, "VoIP Quality Enhancement with Wideband Extension Method in Broadband Networks" v *IEEE Latin America Transactions,* št. 10, str. 1190 - 1194, jan. 2012.

[52] T. H. Falk in W. Y. Chan, "Performance Study of Objective Speech Quality Measurement for Modern Wireless-VoIP Communications" v *Eurasip Journal on Audio Speech and Music Processing,* 2009.

[53] H. Q. Zhang, J. Y. Zhao v O. Yang, "Adaptive rate control for VoIP in wireless ad hoc networks" na *IEEE International Conference on Communications,* str. 3166 - 3170, 2008.

[54] H. G. Zhang, M. Boutabia, H. Nguyen in L. N. Xia, "Field Performance Evaluation of Voip in 4g Trials" na *IEEE International Conference on Multimedia and Expo (ICME),* 2011.

[55] S. Singh, H. P. Singh in J. Singh, "Spectral Analysis of Speech Quality in VoIP for G.729A and AMR-WB Speech Coders" na *Second International Conference on Computational Intelligence, Communication Systems and Networks (Cicsyn),* str. 182 - 187, 2010.

[56] A. W. Rix, J. G. Beerends, M. P. Hollier in A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs" na *IEEE International Conference on Acoustics, Speech, and Signal Processing*, str. 749 - 752, 2001.

[57] M. Karjalainen, "A new auditory model for the evaluation of sound quality of audio systems" na *EEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '85)*, str. 608 - 611, 1985.

[58] N. Avadhanam in V. R. Algazi, "Evaluation of a human vision system based image fidelity metric for image compression" v *Applications of Digital Image Processing Xxii*, št. 3808, str. 569 - 579, 1999.

[59] Z. Wang in A. C. Bovik, "A universal image quality index" v *IEEE Signal Processing Letters*, št. 9, str. 81 - 84, mar. 2002.

[60] D. M. Chandler in S. S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images" v *IEEE Transactions on Image Processing*, št. 16, str. 2284 - 2298, 2007.

[61] S. Winkler in P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics" v *IEEE Transactions on Broadcasting*, št. 54, str. 660 - 668, 2008.

[62] Y. Zhenghua, W. Hong Ren, S. Winkler in C. Tao, "Vision-model-based impairment metric to evaluate blocking artifacts in digital video" v *Proceedings of the IEEE*, št. 90, str. 154 - 169, 2002.

[63] J. G. Beerends, A. R. Hekstra, A. W. Rix in M. P. Hollier, "Perceptual evaluation of speech quality (PESQ) - The new ITU standard for end-to-end speech quality assessment - Part II - Psychoacoust model" v *Journal of the Audio Engineering Society*, št. 50, str. 765 - 778, okt. 2002.

[64] H. Yi in P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement" v *IEEE Transactions on Audio, Speech, and Language Processing*, št. 16, str. 229 - 238, 2008.

[65] K. Kondo, "Estimation of speech intelligibility using objective measures" v *Applied Acoustics*, št. 74, str. 63 - 70, jan. 2013.

[66] L. Malfait, J. Berger in M. Kastner, "P.563 - The ITU-T standard for single-ended speech quality assessment" v *IEEE Transactions on Audio Speech and Language Processing*, št. 14, str. 1924 - 1934, nov. 2006.

[67] D. S. Hands, "A basic multimedia quality model" v *IEEE Transactions on Multimedia*, št. 6, str. 806 - 816, 2004.

[68] J. Gustafsson, G. Heikkila, M. Pettersson in IEEE, "MEASURING MULTIMEDIA QUALITY IN MOBILE NETWORKS WITH AN OBJECTIVE PARAMETRIC MODEL" na *15th IEEE International Conference on Image Processing*, str. 405 - 408, 2008.

[69] K. Okarma, "Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment" na *Artificial Intelligence and Soft Computing*, Springer Berlin Heidelberg, str. 539 - 546, 2010.

[70] C. D. Creusere, K. D. Kallakuri in R. Vanam, "An Objective Metric of Human Subjective Audio Quality Optimized for a Wide Range of Audio Fidelities" v

IEEE Transactions on Audio, Speech, and Language Processing, št. 16, str. 129 - 136, 2008.

[71] L. Songnan, Z. Fan, M. Lin in N. King Ngi, "Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments" v *IEEE Transactions on Multimedia*, št. 13, str. 935 - 949, 2011.

[72] D. John in A. Boucouvalas, "The effect of the Visual-Auditory dimension of cognitive style" na *24th International Conference on Information Technology Interfaces*, str. 179 - 184, 2002.

[73] R. Song, X. H. Jiang in H. Shi, "Objective Quality Assessment of HD Video Based on Compressed Domain" na *International Conference on Graphic and Image Processing (ICGIP 2011)*, str. 141 - 144, apr. 2011.

[74] B. Belmudez, S. Moeller, B. Lewcio, A. Raake in A. Mehmood, "Audio and Video Channel Impact on Perceived Audio-visual Quality in Different Interactive Contexts" na *IEEE International Workshop on Multimedia Signal Processing (MMSP 2009)*, str. 256 - 260, 2009.

[75] A. Borowiak, U. Reiter in U. P. Svensson, "Quality evaluation of long duration audiovisual content" na *IEEE Consumer Communications and Networking Conference (CCNC)*, str. 337 - 341, 2012.

[76] U. Reiter, "Overall perceived audiovisual quality - What people pay attention to" na *IEEE 15th International Symposium on Consumer Electronics (ISCE)*, str. 513 - 517, 2011.

[77] Y. Junyong, J. Korhonen in U. Reiter, "Audiovisual quality fusion based on relative multimodal complexity" na *18th IEEE International Conference on Image Processing (ICIP)*, str. 3337 - 3340, 2011.

[78] M. Garcia, R. Schleicher in A. Raake, "Impairment-Factor-Based Audiovisual Quality Model for IPTV: Influence of Video Resolution, Degradation Type, and Content Type" v *Eurasip Journal on Image and Video Processing*, št. 2011, str. 629284, 2011.

[79] J. Vroomen, "Causal inference in audiovisual speech Comment on "Crossmodal influences on visual perception" by L. Shams" v *Physics of Life Reviews*, št. 7, str. 289 - 290, sep. 2010.

[80] S. Jumisko-Pyykko, J. Hakkinen in G. Nyman, "Experienced quality factors - Qualitative evaluation approach to audiovisual quality" v *Multimedia on Mobile Devices 2007*, št. 6507, 2007.

[81] J. G. Beerends and F. E. de Caluwe, "The influence of video quality on perceived audio quality and vice versa" v *Journal of the Audio Engineering Society*, št. 47, str. 355 - 362, maj 1999.

[82] U. Reiter and Y. Junyong, "Estimating perceived audiovisual and multimedia quality; a survey" na *14th International IEEE Symposium on Consumer Electronics (ISCE 2010)*, str. 1 - 6, 2010.

[83] J. X. Maier, M. Di Luca in U. Noppeney, "Audiovisual Asynchrony Detection in Human Speech" v *Journal of Experimental Psychology-Human Perception and Performance*, št. 37, str. 245 - 256, feb. 2011.

[84] C. Griwodz, "Perceived audiovisual synchrony in distorted speech", *predavanja*, okt. 2012.

[85] L. Lijie in F. Guoliang, "A new JPEG2000 region-of-interest image coding method: partial significant bitplanes shift" v *IEEE Signal Processing Letters*, št. 10, str. 35 - 38, 2003.

[86] F. Boulos, C. Wei, B. Parrein in P. Le Callet, "Region-of-Interest intra prediction for H.264/AVC error resilience" na *16th IEEE International Conference on Image Processing (ICIP 2009)*, str. 3109 - 3112, 2009.

[87] W. Osberger in A. J. Maeder, "Automatic identification of perceptually important regions in an image" na *Fourteenth International Conference on Pattern Recognition, št. 1,* str. 701 - 704, avg. 1998.

[88] L. Itti, C. Koch in E. Niebur, "A model of saliency-based visual attention for rapid scene analysis" v *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 20, str. 1254 - 1259, nov. 1998.

[89] C. Wen-Huang, C. Wei-Ta, K. Jin-Hau in W. Ja-Ling, "Automatic video region-of-interest determination based on user attention model" na *IEEE International Symposium on Circuits and Systems (ISCAS 2005)*, št. 4, str. 3219 - 3222, 2005.

[90] K. Wonjun in K. Changick, "Automatic region of interest determination in music videos" na *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers (ACSSC 2007)*, str. 485 - 489, 2007.

[91] M. C. Chi, C. H. Yeh in M. J. Chen, "Robust Region-of-Interest Determination Based on User Attention Model Through Visual Rhythm Analysis" v *IEEE Transactions on Circuits and Systems for Video Technology*, št. 19, str. 1025 - 1038, jul. 2009.

[92] M. Clauss, P. Bayerl in H. Neumann, "A statistical measure for evaluating regions-of-interest based attention algorithms" v *Pattern Recognition*, št. 3175, str. 383 - 390, 2004.

[93] U. Engelke, A. Maeder in H. J. Zepernick, "Visual Attention Modelling for Subjective Image Quality Databases" na *IEEE International Workshop on Multimedia Signal Processing (MMSP 2009)*, str. 25 - 30, 2009.

[94] L. Hantao in I. Heynderickx, "Studying the added value of visual attention in objective image quality metrics based on eye movement data" na *16th IEEE International Conference on Image Processing (ICIP 2009)*, str. 3097 - 3100, 2009.

[95] H. T. Liu in I. Heynderickx, "Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data" v *IEEE Transactions on Circuits and Systems for Video Technology*, št. 21, str. 971 - 982, jul. 2011.

[96] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli in T. S. Chua, "An Eye Fixation Database for Saliency Detection in Images" v *Computer Vision-Eccv 2010*, št. 6314, str. 30 - 43, 2010.

[97] K. Fliegel, "Eyetracking Based Approach to Objective Image Quality Assessment" na *42nd Annual 2008 IEEE International Carnahan Conference on Security Technology*, str. 371 - 376, 2008.

[98] E. Kidron, Y. Y. Schechner in M. Elad, "Cross-modal localization via sparsity" v *IEEE Transactions on Signal Processing*, št. 55, str. 1390 - 1404, Apr 2007.
[99] U. Engelke in H.-J. Zepernick, "Optimal region-of-interest based visual quality assessment" iz *Electronic Research Archive of Blekinge Institute of Technology*, dosegljivo na: http://www.bth.se/fou/, obiskano: nov. 2014.
[100] M. Vranješ, S. Rimac-Drlje in O. Nemčić, "Influence of foveated vision on video quality perception" na *International Symposium ELMAR (ELMAR '09)*, str. 29 - 32, 2009.
[101] W. Osberger, N. Bergmann in A. Maeder, "An automatic image quality assessment technique incorporating higher level perceptual factors" na *International Conference on Image Processing*, št. 3, str. 414 - 418, 1998.
[102] I. Himawan, S. Wei in D. Tjondronegoro, "Impact of Region-of-Interest Video Coding on Perceived Quality in Mobile Video" na *IEEE International Conference on Multimedia and Expo (ICME 2012)*, str. 79 - 84, 2012.
[103] B. Ciubotaru, G. M. Muntean in G. Ghinea, "Objective Assessment of Region of Interest-Aware Adaptive Multimedia Streaming Quality" v *IEEE Transactions on Broadcasting*, št. 55, str. 202 - 212, jun. 2009.
[104] U. Engelke in H. J. Zepernick, "Framework for optimal region of interest-based quality assessment in wireless imaging" v *Journal of Electronic Imaging*, št. 19, jan. - mar. 2010.
[105] U. Engelke in H. J. Zepernick, "Psychophysical assessment of perceived interest in natural images: The ROI-D database" na *IEEE Visual Communications and Image Processing (VCIP 2011)*, str. 1 - 4, 2011.
[106] ISO, "ISO 8402:1986: Quality - Vocabulary", 1986.
[107] A. Parasuraman, V. A. Zeithaml in L. L. Berry, "Servqual - a Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality" v *Journal of Retailing*, št. 64, str. 12 - 40, sept. 1988.
[108] W. Boulding, A. Kalra, R. Staelin in V. A. Zeithaml, "A Dynamic Process Model of Service Quality - from Expectations to Behavioral Intentions" v *Journal of Marketing Research*, št. 30, str. 7 - 27, feb. 1993.
[109] ITU-T, "E.800 : Definitions of terms related to quality of service", 2008.
[110] M. Volk, J. Sterle, U. Sedlar in A. Kos, "An approach to modeling and control of QoE in next generation networks [Next Generation Telco IT Architectures]" v *IEEE Communications Magazine*, št. 48, str. 126 - 135, 2010.
[111] R. T. Rust in A. J. Zahorik, "Customer Satisfaction, Customer Retention, and Market Share" v *Journal of Retailing*, št. 69, str. 193 - 215, 1993.
[112] T. M. Fabricio Carvalho de Gouveia, *Telecommunication Systems and Technologies*, št. 2, 2009.
[113] M. Molinari, "On the Monitoring and Measuring of Quality of Experience (QoE) in IP Networks" objavljeno na Facoltà di Ingegneria, Univerza v Neaplu, Italija, 2012.
[114] A. van Moorsel, "Metrics for the Internet Age: Quality of Experience and Quality of Business" na *Fifth Performability Workshop*, Erlangen, Nemčija, 2001.

[115] M. Siller in J. C. Woods, "QoS arbitration for improving the QoE in multimedia transmission" na *International Conference on Visual Information Engineering (VIE 2003)*, str. 238 - 241, 2003.
[116] A. S. Patrick in B. Bauer, "A Human Factors Extension to the Seven-Layer OSI Reference Model", dostopno na: http://www.andrewpatrick.ca/OSI/10layer.html, obiskano: nov. 2014
[117] ITU-T, "P.10 : New Appendix I - Definition of Quality of Experience (QoE)", 2008.
[118] S. Moller, l. A. Perkis in P. Le Callet, "Qualinet White Paper on Definitions of Quality of Experience Output version of the Dagstuhl seminar 12181", Dagstuhl, jun. 2012.
[119] U. Jekosch, *Voice and Speech Quality Perception*: Springer, 2005.
[120] ETSI, "ETSI TR 102 643: Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services", Francija, 2009.
[121] S. Baraković, J. Baraković in H. Bajrić, "QoE Dimensions and QoE Measurement of NGN Services" na *18th Telecommunications forum, TELFOR 2010*, Beograd, Srbija, 2010.
[122] P. Barthelmess in S. Oviatt, "Multimodal Interfaces: Combining Interfaces to Accomplish a Single Task" v *Hci Beyond the Gui: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*, str. 391 - 444, 2008.
[123] S. Oviatt, "Breaking the robustness barrier: Recent progress on the design of robust multimodal systems" v *Advances in Computers*, št. 56, str. 305 - 341, 2002.
[124] V. Radu-Daniel, "Interfaces That Should Feel Right: Natural Interaction with Multimedia Information" v *Recent Advances in Multimedia Signal Processing and Communications*, št. 231, str. 145 - 170, 2009.
[125] L. Nigay in J. Coutaz, "A Design Space for Multimodal Systems - Concurrent Processing and Data Fusion" v *Human Factors in Computing Systems*, str. 172 - 178, 1993.
[126] T. Halliday, *The senses and communication*: Springer, in association with the Open University, 1998.
[127] R. Wasinger, *Multimodal Interaction with Mobile Devices: Fusing a Broad Spectrum of Modality Combinations*: IOS Press, 2006.
[128] W3C, *EMMA: Extensible MultiModal Annotation markup language*, dostopno na: http://www.w3.org/TR/emma/, obiskano: nov. 2014.
[129] P. Grifoni, *Multimodal Human Computer Interaction and Pervasive Services*, IGI Global, ZDA, maj 2009.
[130] J. N. L. Schomaker, A. Camurri, F. Lavagetto, P. Morasso, C. Benoît, T. Guiard marigny, B. Le Goff, J. Robert ribes, A. Adjoudani, I. Defée, S. Münch, K. Hartung in J. Blauert, "A Taxonomy of Multimodal Interaction in the Human Information Processing System", report, feb. 1995.
[131] S. Anastopoulou, C. Baber in M. Sharples, "Multimedia and multimodal systems: commonalities and differences" na *5th Human Centred Technology Postgraduate Workshop*, University of Sussex, Anglija, 2001.

[132] W3C (2003). *W3C Multimodal Interaction Framework*, dostopno na: http://www.w3.org/TR/mmi-framework/, obiskano: nov. 2014.
[133] M. T. Maybury in W. Wahlster, *Readings in Intelligent User Interfaces*: Morgan Kaufmann Publishers, 1998.
[134] R. Raisamo, *Multimodal Human-computer Interaction: A Constructive and Empirical Study*: University of Tampere, 1999.
[135] S. L. Oviatt, "Multimodal interactive maps: Designing for human performance" v *Human-Computer Interaction*. št. 12, str. 93 - 129, 1997.
[136] S. Oviatt, R. Coulston in R. Lunsford, "When do we interact multimodally?: cognitive load and multimodal communication patterns" na *6th international conference on Multimodal interfaces*, State College, Pensilvanija, ZDA, 2004.
[137] S. Oviatt, A. DeAngeli in K. Kuhn, "Integration and synchronization of input modes during multimodal human-computer interaction" na *ACM SIGCHI Conference on Human factors in computing systems*, Atlanta, Georgia, ZDA, 1997.
[138] M. Perakakis in A. Potamianos, "A study in efficiency and modality usage in multimodal form filling systems" v *IEEE Transactions on Audio Speech and Language Processing*, št. 16, str. 1194 - 1206, avg. 2008.
[139] B. Dumas, D. Lalanne in S. Oviatt, "Multimodal Interfaces: A Survey of Principles, Models and Frameworks" v *Human Machine Interaction: Research Results of the MMI Program*., št. 5440, Springer, str. 3 - 26, 2009.
[140] S. Mandal, B. Das in P. Mitra, "Shruti-II: A vernacular speech recognition system in Bengali and an application for visually impaired community" na *IEEE Students' Technology Symposium (TechSym) 2010*, str. 229 - 233, 2010.
[141] E. Verdurand, G. Coppin, F. Poirier in O. Grisvard, "Modeling multimodal interaction for performance evaluation" na *13th International conference on Human-Computer Interaction*, str. 103 - 112, 2009.
[142] S. Oviatt, "Ten myths of multimodal interaction" v *Communications of the ACM*, št. 42, str. 74 - 81, nov. 1999.
[143] K. Jokinen in A. Raike, "Multimodality - technology, visions and demands for the future" na *1st Nordic Symposium on Multimodal Interfaces*, Copenhagen, Danska, 2003.
[144] G. Stollnberger, A. Weiss in M. Tscheligi, "Input Modality and Task Complexity: Do they Relate?" v *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (Hri 2013)*, str. 233 - 234, 2013.
[145] K. Barton. (2013, sept. 2014). *STATE OF THE IVR: INDUSTRY EXPERTS WEIGH IN, insights and best practices for getting the most out of your IVR interactions*, dostopno na: http://www.strategiccontact.com/articles/State-Of-IVR-Industry-Nov2013.pdf, obiskano: nov. 2014.
[146] M. Matassoni, M. Omologo in C. Zieger, "Experiments of in-car audio compensation for hands-free speech recognition" na *IEEE Workshop on Automatic Speech Recognition and Understanding (Asru '03)*, str. 369 - 374, 2003.

[147] A. V. Anand, P. S. Devi, J. Stephen in V. K. Bhadran, "Malayalam Speech Recognition System and Its Application for visually impaired people" na *Annual IEEE India Conference (Indicon 2012)*, str. 619 - 624, 2012.
[148] Y. F. Gong, "Speech Recognition in Noisy Environments - a Survey" v *Speech Communication*, št. 16, str. 261 - 291, apr. 1995.
[149] P. Mermelstein, "Distance measures for speech recognition, psychological and instrumental" v *Pattern recognition and artificial intelligence*, urednik: C. H. Chen, Elsevier Science, 2013.
[150] J. P. Olive, "Mixed Spectral Representation - Formants and Linear Predictive Coding (Lpc)" v *Journal of the Acoustical Society of America*, št. 92, str. 1837 - 1840, okt. 1992.
[151] H. Hermansky, "Perceptual Linear Predictive (Plp) Analysis of Speech" v *Journal of the Acoustical Society of America*, št. 87, str. 1738 - 1752, apr. 1990.
[152] H. Hermansky in N. Morgan, "RASTA Processing of Speech" v *IEEE Transactions on Speech and Audio Processing*, št. 2, str. 578 - 589, okt. 1994.
[153] S. Young, "A review of large-vocabulary continuous-speech recognition" v *IEEE Signal Processing Magazine*, št. 13, str. 45 - 57, sept. 1996.
[154] T. H. Falk, Q. F. Xu in W. Y. Chan, "Non-intrusive GMM-based speech quality measurement" na *IEEE International Conference on Acoustics, Speech, and Signal Processing*, str. 125 - 128, 2005.
[155] M. Kos, M. Grašič in Z. Kačič, "Online Speech/Music Segmentation Based on the Variance Mean of Filter Bank Energy" v *Eurasip Journal on Advances in Signal Processing*, 2009.
[156] H. Jiang, S. Y. Chen, Y. Yang, Z. Z. Jie, H. R. Leung, J. Xu in L. Wang, "Estimation of Packet Loss Rate at Wireless Link of Vanet-Rple" na *6th International Conference on Wireless Communications Networking and Mobile Computing (Wicom)*, 2010.
[157] M. Štrucl, *Fiziologija živčevja*: Medicinski razgledi, 1999.
[158] W. M. Osberger, A. J. Maeder in N. W. Bergmann, "A Technique For Image Quality Assessment Based On A Human Visual System Model" na *Signal processing -European conference*, 1998.
[159] R. L. Gregory, "Organization in Vision - Essays on Gestalt Perception - Kanizsa,G" v *Perception*, št. 13, str. 226 - 226, 1984.
[160] R. Jackendoff, "Consciousness and the Computational Mind" v *Science*, št. 239, str. 1546 - 1547, mar. 1988.
[161] S. Treue, "Visual attention: the where, what, how and why of saliency" v *Current Opinion in Neurobiology*, št. 13, str. 428 - 432, avg. 2003.
[162] I. Biederman, "Recognition-by-Components - a Theory of Human Image Understanding" v *Psychological Review*, št. 94, str. 115 - 147, apr. 1987.
[163] M. J. Tarr, "Rotating Objects to Recognize Them - a Case-Study on the Role of Viewpoint Dependency in the Recognition of 3-Dimensional Objects" v *Psychonomic Bulletin & Review*, št. 2, str. 55 - 82, mar. 1995.
[164] M. J. Farah, K. D. Wilson, M. Drain in J. N. Tanaka, "What is "special" about face perception?" v *Psychological Review*, št. 105, str. 482 - 498, jul. 1998.

[165] M. Moscovitch, G. Winocur in M. Behrmann, "What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition" v *Journal of Cognitive Neuroscience,* št. 9, str. 555 - 604, sept. 1997.
[166] D. Purves, *Principles of Cognitive Neuroscience*: Sinauer Associates, 2013.
[167] M. D. Fairchild, "Human Color Vision" v *Color Appearance Models, 3rd Edition,* M. D. Fairchild, John Wiley & Sons, 2013.
[168] L. T. Maloney, "Human perception of objects: Early visual processing of spatial form defined by luminance, color, texture, motion, and binocular disparity" v *Perception,* št. 31, str. 1031 - 1032, 2002.
[169] J. L. Mannos and D. J. Sakrison, "Effects of a Visual Fidelity Criterion on Encoding of Images" v *IEEE Transactions on Information Theory,* št. 20, str. 525 - 536, 1974.
[170] A. B. Watson, G. Y. Yang, J. A. Solomon in J. Villasenor, "Visibility of wavelet quantization noise" v *IEEE Transactions on Image Processing,* št. 6, str. 1164 - 1175, avg. 1997.
[171] M. W. Cannon, "Evoked-Potential Contrast Sensitivity as a Function of Luminance" v *Journal of the Optical Society of America a-Optics Image Science and Vision,* št. 1, str. 1260 - 1260, 1984.
[172] S. Westland, H. Owens, V. Cheung in I. Paterson-Stephens, "Model of luminance contrast-sensitivity function for application to image assessment" v *Color Research and Application,* št. 31, str. 315 - 319, avg. 2006.
[173] P. G. J. Barten, "Spatiotemporal Model for the Contrast Sensitivity of the Human Eye and Its Temporal Aspects" v *Human Vision, Visual Processing, and Digital Display,* št. 1913, str. 2 - 14, 1993.
[174] B. R. Wooten, L. M. Renzi, R. Moore in B. R. Hammond, "A practical method of measuring the human temporal contrast sensitivity function" v *Biomedical Optics Express,* št. 1, str. 47 - 58, avg. 2010.
[175] R. Li, J. J. Ge in Z. Q. Wang, "The study on neural contrast sensitivity function at temporal frequencies" v *Optik,* št. 123, str. 343 - 347, 2012.
[176] Z. Wang. (2001). *Human Visual Foveation Model*. Dostopno: https://ece.uwaterloo.ca/~z70wang/research/fovea/fovea_model.html, obiskano: nov. 2014.
[177] S. Daly, "Application of a Noise-Adaptive Contrast Sensitivity Function to Image Data-Compression" v *Optical Engineering,* št. 29, str. 977 - 987, avg. 1990.
[178] C. S. Furmanski in S. A. Engel, "An oblique effect in human primary visual cortex" v *Nature Neuroscience,* št. 3, str. 535 - 536, jun. 2000.
[179] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong in S. S. Yao, "Just noticeable distortion model and its applications in video coding" v *Signal Processing-Image Communication,* št. 20, str. 662 - 680, avg. 2005.
[180] E. S. Ferry, O. W. Silvey, G. W. Sherman in D. C. Duncan, *A Handbook of Physics Measurements*: John Wiley & Sons, 1918.

[181] J. D. Conner, "The Temporal Properties of Rod Vision" v *Journal of Physiology-London*, št. 332, str. 139 - 155, 1982.
[182] S. Rimac-Drlje, D. Žagar, Martinović, x in G., "Spatial Masking and Perceived Video Quality in Multimedia Applications" na *16th International Conference on Systems, Signals and Image Processing (IWSSIP 2009)*, str. 1 - 4, 2009.
[183] A. Valberg, *Light Vision Color*: Wiley, 2005.
[184] T. Noll, "Tone Apperception, Relativity and Weber-Fechner's Law" na *2nd International Conference of Understanding and Creating Music*, Napoli, Italija, 2002.
[185] E. H. Weber, D. J. Ross, H. E. Ross, "E. H. Weber on the tactile senses" v *Perception*, št. 26, str. 120 - 122, 1997.
[186] G. E. Legge in J. M. Foley, "Contrast Masking in Human-Vision" v *Journal of the Optical Society of America*, št. 70, str. 1458 - 1471, 1980.
[187] G. H. Chen, C. L. Yang, L. M. Po in S. L. Xie, "Edge-based structural similarity for image quality assessment" na *IEEE International Conference on Acoustics, Speech and Signal Processing*, str. 2181 - 2184, 2006.
[188] G. W. Larson, H. Rushmeier in C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes" v *IEEE Transactions on Visualization and Computer Graphics*, št. 3, str. 291 - 306, okt. - dec. 1997.
[189] Y. C. Liu in J. P. Allebach, "A Computational Texture Masking Model for Natural Images Based on Adjacent Visual Channel Inhibition" na *Image Quality and System Performance Xi*, št. 9016, 2014.
[190] M. D. Gaubatz, D. M. Chandler in S. S. Hemami, "Spatial quantization via local texture masking" na *Human Vision and Electronic Imaging X*, št. 5666, str. 95 - 106, 2005.
[191] A. H. Munsell, *A Color Notation*: G. H. Ellis Company, 1905.
[192] Adobe Systems (2000). *Color Models*, dostopno na: http://dba.med.sc.edu/price/irf/Adobe_tg/models/cieluv.html, obiskano: nov. 2014.
[193] R. W. G. Hunt, "The Physiological-Basis of a Model of Color-Vision and Its Applications in Color Reproduction" v *Journal of Photographic Science*, št. 38, str. 105 - 108, 1990.
[194] Sakurambo (2014). *Rec. 709*, dostopno na: http://commons.wikimedia.org/wiki/File:CIExy1931_Rec_709.svg, obiskano: nov. 2014.
[195] Sakurambo (2014). *Rec. 2020*, dostopno na: http://commons.wikimedia.org/wiki/File:CIExy1931_Rec_2020.svg, obiskano: nov. 2014.
[196] M. Hassan and C. Bhagvati, "Color Image Quantization Quality Assessment" na *Wireless Networks and Computational Intelligence (ICIP 2012)*, št. 292, str. 139 - 148, 2012.
[197] X. Q. Zhang, "A Novel Quality Metric for Image Fusion Based on Color and Structural Similarity" v *Proceedings of the 2009 International Conference on Signal Processing Systems*, str. 258 - 262, 2009.

[198] S. Kaya, T. Bennett, M. Milanova, J. Talburt, B. Tsou, M. Altynova in H. Y. Xu, "Perception-Based Image/Video Quality Metric using CIELAB color space" v *Sensors, and Command, Control, Communications, and Intelligence (C3i) Technologies for Homeland Security and Homeland Defense X*, št. 8019, 2011.

[199] U. Rajashekar, Z. Wang in E. P. Simoncelli, "Perceptual quality assessment of color images using adaptive signal representation" v *Human Vision and Electronic Imaging Xv*, št. 7527, 2010.

[200] H. Fletcher, "Auditory Patterns" v *Reviews of Modern Physics*, št. 12, str. 47 - 65, 1940.

[201] B. C. J. Moore and B. R. Glasberg, "A model of loudness perception applied to cochlear hearing loss" v *Auditory Neuroscience*, št. 3, str. 289 - 311, 1997.

[202] M. S. D. Ze-Nian Li, *Fundamentals of Multimedia*: Prentice-Hall, 2004.

[203] C. C. Wier, W. Jesteadt in D. M. Green, "Frequency Discrimination as a Function of Frequency and Sensation Level" v *Journal of the Acoustical Society of America*, št. 61, str. 178 - 184, 1977.

[204] C. E. Jack and W. R. Thurlow, "Effects of Degree of Visual Association and Angle of Displacement on Ventriloquism Effect" v *Perceptual and Motor Skills*, št. 37, str. 967 - 979, 1973.

[205] H. Mcgurk and J. Macdonald, "Hearing Lips and Seeing Voices" v *Nature*, št. 264, str. 746 - 748, 1976.

[206] L. Shams, Y. Kamitani in S. Shimojo, "Visual illusion induced by sound" v *Cognitive Brain Research*, št. 14, str. 147 - 152, jun. 2002.

[207] J. S. Lee, F. De Simone in T. Ebrahimi, "Influence of Audio-Visual Attention on Perceived Quality of Standard Definition Multimedia Content" na *Qomex: International Workshop on Quality of Multimedia Experience*, str. 13 - 18, 2009.

[208] B. Belmudez in S. Moller, "Audiovisual quality integration for interactive communications", *Eurasip Journal on Audio Speech and Music Processing*, nov. 2013.

[209] S. Tasaka in H. Yoshimi, "Enhancement of QoE in Audio-Video IP Transmission by Utilizing Tradeoff between Spatial and Temporal Quality for Video Packet Loss" na *IEEE Global Telecommunications Conference (Globecom 2008)*, 2008.

[210] A. Vetro, C. Christopoulos in T. Ebrahimi, "Universal multimedia access" v *IEEE Signal Processing Magazine*, št. 20, str. 16, mar. 2003.

[211] I. Ahmad, X. Wei, Y. Sun in Y. Q. Zhang, "Video transcoding: An overview of various techniques and research issues" v *IEEE Transactions on Multimedia*, št. 7, str. 793 - 804, okt. 2005.

[212] M. G. Sandro Moiron, Pedro Assuncao in Sergio Faria, "Video Transcoding Techniques" v *Recent Advances in Multimedia Signal Processing and Communications*, št. 231, str. 245 - 270, 2009.

[213] H. F. Sun, W. Kwok in J. W. Zdepski, "Architectures for MPEG compressed bitstream scaling" v *IEEE Transactions on Circuits and Systems for Video Technology*, št. 6, str. 191 - 199, apr. 1996.

[214] X. Jun, S. Ming-Ting in K. KouSou, "Bit allocation for joint transcoding of multiple MPEG coded video streams" na *IEEE International Conference on Multimedia and Expo (ICME 2001)*, str. 8 - 11, 2001.
[215] IHS. (2013). *UHD TV Panel Market Forecast - Units*, dostopno na: http://www.ihs.com, obiskano: nov. 2014.
[216] P. F. Correia, V. M. Silva in P. A. Assuncao, "A method for improving the quality of mobile video under hard transcoding conditions" na *IEEE International Conference on Communications*, str. 928 - 932, 2003.
[217] C. W. Tang, "Spatiotemporal visual considerations for video coding" v *IEEE Transactions on Multimedia*, št. 9, str. 231 - 238, feb. 2007.
[218] T. Sikora, "The MPEG-4 video standard verification model" v *IEEE Transactions on Circuits and Systems for Video Technology*, št. 7, str. 19 - 31, feb. 1997.
[219] T. Sikora in L. Chiariglione, "MPEG-4 video and its potential for future multimedia services" na *IEEE International Symposium on Circuits and Systems (Iscas '97)*, str. 1468 - 1471, 1997.
[220] Y. Q. Shi in H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*: Taylor & Francis, 1999.
[221] B. Waggoner, *Compression for Great Video and Audio: Master Tips and Common Sense*: Taylor & Francis, 2013.
[222] S. Wichman, "A comparison of speech coding algorithms ADPCM vs CELP" na The University of Texas, Dallas, ZDA, 1999.
[223] C. Mantel, P. Ladret in T. Kunlin, "Measurement of compression-induced temporal artifacts in subjective and objective video quality assessment" v *Human Vision and Electronic Imaging Xvi*, št. 7865, 2011.
[224] A. K. Moorthy in A. C. Bovik, "H.264 visually lossless compressibility index: Psychophysics and algorithm design" na *10th IEEE Workshop on Image, Video, and Multidimensional Signal Processing, (IVMSP 2011)*, str. 111 - 116, 2011.
[225] Y. K. Thomas Wedi, "Subjective quality evaluation of H.264/AVC FRExt for HD movie content", University of Washington, ZDA, 2004.
[226] Y. Luo in R. K. Ward, "Removing the blocking artifacts of block-based DCT compressed images" v *IEEE Transactions on Image Processing*, št. 12, str. 838 - 842, jul. 2003.
[227] H. W. Paik in A. Khubchandani, "Quantization scheme for JPEG baseline sequential encoding of still images" na *35th Midwest Symposium on Circuits and Systems*, št. 2, str. 976 - 979, 1992.
[228] M. Antonini, M. Barlaud, P. Mathieu in I. Daubechies, "Image coding using vector quantization in the wavelet transform domain" na *International Conference on Acoustics, Speech, and Signal Processing (ICASSP '90)*, št. 4, str. 2297 - 2300, 1990.
[229] C. Christopoulos, A. Skodras in T. Ebrahimi, "The JPEG2000 still image coding system: An overview" v *IEEE Transactions on Consumer Electronics*, št. 46, str. 1103 - 1127, nov. 2000.

[230] A. Katharotiya, S. Patel in M. Goyani, "Comparative Analysis between DCT & DWT Techniques of Image Compression" v *Journal of Information Engineering and Applications*, št. 1, 2011.

[231] S. A. Pradeep in R. Manavalan, "Image Compression Using Radon Transform With DCT : Performance Analysis" v *International Journal of Scientific Engineering and Technology*, št. 2, str. 759 - 765, 2013.

[232] H. Olkkonen, P. Pesola in J. T. Olkkonen, "Computation of Hilbert Transform via Discrete Cosine Transform" v *Journal of Signal and Information Processing*, št. 1, str. 18 - 23, 2010.

[233] S. P. Maity in S. Maity, "Wavelet based Hilbert transform with digital design and application to QCM-SS watermarking" v *Radioengineering*, št. 17, str. 64 - 72, apr. 2008.

[234] R. K. Rashmi Agarwal, M. S. Santhanam, K. Srinivas in K. Venugopalan. (2010). *Digital watermarking : An approach based on Hilbert transform*, dostopno na: http://arxiv.org/pdf/1012.2965v1.pdf, obiskano: nov. 2014.

[235] N. Chen, J. Xiuhua, W. Caihong in S. Jia, "Study on relationship between network video packet loss and video quality" na *4th International Congress on Image and Signal Processing (CISP 2011)*, str. 282 - 286, 2011.

[236] J. M. Mwela in O. E. Adebomi, "Impact of packet loss on the quality of video stream transmission", magisterska naloga, Blekinge Institute of Technology, School of Computing, Ronneby, Švedska, 2010.

[237] T. Wallingford, *Switching to VoIP*: O'Reilly Media, 2005.

[238] R. V. Babu in A. Perkis, "An HVS-based no-reference perceptual quality assessment of JPEG coded images using neural networks" na *International Conference on Image Processing (ICIP 2005)*, str. 881 - 884, 2005.

[239] P. Marziliano, F. Dufaux, S. Winkler in T. Ebrahimi, "A no-reference perceptual blur metric" na *International Conference on Image Processing*, št. 3, str. 57 - 60, 2002.

[240] P. Marziliano, F. Dufaux, S. Winkler in T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000" v *Signal Processing-Image Communication*, št. 19, str. 163-172, feb. 2004.

[241] A. Punchihewa, J. Armstrong, S. Hangai in T. Hamamoto, "Objective Evaluation of Components of Colour Distortions due to Image Compression" v *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, št. E92a, str. 3307 - 3312, dec. 2009.

[242] S. Li, O. C. Au, L. Sun, W. Dai in R. Zou, "Color Bleeding Reduction in Image and Video Compression" na *International Conference on Computer Science and Network Technology (ICCSNT 2011)*, str. 665 - 669, 2012.

[243] T. Meier, K. N. Ngan in G. Crebbin, "Reduction of blocking artifacts in image and video coding" v *IEEE Transactions on Circuits and Systems for Video Technology*, št. 9, str. 490 - 500, apr. 1999.

[244] M. Erne, "Perceptual Audio Coders 'What to listen for'" na *Audio Engineering Society Convention 111*, 2001.

[245] C. M. Liu, H. W. Hsu in W. C. Lee, "Compression artifacts in perceptual audio coding" v *IEEE Transactions on Audio Speech and Language Processing*, št. 16, str. 681 - 695, maj 2008.
[246] C.-M. Liu. *Audio Artifacts in Perceptual Audio Coding*, dostopno na: http://people.cs.nctu.edu.tw/~cmliu/Courses/Compression/Artifacts.pdf, obiskano: nov. 2014.
[247] ITU-T, "P.910 : Subjective video quality assessment methods for multimedia applications" v *ITU-T: P Series*, 2008.
[248] ITU-T, "P.911 : Subjective audiovisual quality assessment methods for multimedia applications" v *ITU-T: P Series*, 1998.
[249] ITU-T, "P.913 : Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment" v *ITU-T: P Series*, 2014.
[250] ITU-R, "Methodology for the subjective assessment of the quality of television pictures" v *ITU-R Recommendations*, 2012.
[251] ITU-R, "BT.710 : Subjective assessment methods for image quality in high-definition television" v *ITU-R Recommendations*, 1998.
[252] ITU-T, "P.800 : Methods for subjective determination of transmission quality" v *ITU-T: P Series*, 1996.
[253] ITU-T, "P.830 : Subjective performance assessment of telephone-band and wideband digital codecs" v *ITU-T: P Series*, 1996.
[254] ITU-R, "Methodology for the subjective assessment of video quality in multimedia applications " v *ITU-R Recommendations*, 2007.
[255] R. Lukac, *Perceptual Digital Imaging: Methods and Applications*: Taylor & Francis, 2012.
[256] O. Nemethova, M. Ries, A. Dantcheva, S. Fikar in M. Rupp, "Test equipment of time-variant subjective perceptual video quality in mobile terminals" v *Proceedings of the IASTED International Conference on Human-Computer Interaction*, str. 72 - 76, 2005.
[257] S. Buchinger, W. Robitza, M. Nezveda, M. C. Sack, P. Hummelbrunner in H. Hlavacs, "Slider or glove? Proposing an alternative quality rating methodology" na *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2010)*, Scottsdale, Arizona, ZDA, 2010.
[258] T. Liu, G. Cash, N. Narvekar, in J. Bloom, "Continuous mobile video subjective quality assessment using gaming steering wheel" na *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2012)*, Scottsdale, Arizona, ZDA, 2012.
[259] S. Voran in A. Catellier, "Gradient Ascent Paired-Comparison Subjective Quality Testing" na *Qomex: International Workshop on Quality of Multimedia Experience*, str. 133 - 138, 2009.
[260] Z. Miličević in Z. Bojković, "Subjective video quality assessment in H.264/AVC video coding standard" na *19th Telecommunications Forum (TELFOR 2011)*, str. 1183 - 1186, 2011.

[261] H. R. Wu in K. R. Rao, *Digital Video Image Quality and Perceptual Coding*: Taylor & Francis, 2005.
[262] ITU-T, "P.800.2: Mean opinion score interpretation and reporting" v *ITU-T: P Series*, 2013.
[263] L. S. Eisenberg, D. D. Dirks in J. A. Gornbein, "Subjective judgements of speech clarify measured by paired comparisons and category rating" v arhivu UCLA School of Medicine, Division of Head and Neck Surgery, ZDA, 1997.
[264] A. M. Eskicioglu in P. S. Fisher, "Image quality measures and their performance" v *IEEE Transactions on Communications*, št. 43, str. 2959 - 2965, 1995.
[265] S. R. Nirmala, S. Dandapat in P. K. Bora, "Image quality assessment in retinal image compression systems" na *IET-UK International Conference on Information and Communication Technology in Electrical Sciences (ICTES 2007)*, str. 737-742, 2007.
[266] J. Du, Y. Yu in S. Xie, "A new image quality assessment based on HVS" v *Journal of Electronics (China)*, št. 22, str. 315 - 320, 2005.
[267] Y. J. Wang, J. H. Li, Y. Lu, Y. Fu in Q. Z. Jiang, "Image quality evaluation based on image weighted separating block peak signal to noise ratio" na *International Conference on Neural Networks & Signal Processing*, str. 994 - 997, 2003.
[268] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti in M. Carli, "A new full-reference quality metrics based on HVS" na *Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, ZDA, 2006.
[269] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola in V. Lukin, "On between-coefficient contrast masking of DCT basis functions" na *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-07)*, Scottsdale, Arizona, ZDA, 2007.
[270] Center for Neutral Science, Univerza v New Yorku (2001). *"Lena" Image with Different Types of Distortions*, dostopno na: http://www.cns.nyu.edu/~zwang/files/research/quality_index/demo_lena.html, obiskano nov. 2014.
[271] A. C. Brooks, X. N. Zhao in T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions" v *IEEE Transactions on Image Processing*, št. 17, str. 1261 - 1273, avg. 2008.
[272] Z. Wang, E. P. Simoncelli in A. C. Bovik, "Multi-scale structural similarity for image quality assessment" v *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, str. 1398 - 1402, 2003.
[273] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik in M. K. Markey, "Complex Wavelet Structural Similarity: A New Image Similarity Index" v *IEEE Transactions on Image Processing*, št. 18, str. 2385 - 2401, nov. 2009.
[274] D. Cochran, "Phase and Magnitude in Normalized Images" v *IEEE Transactions on Image Processing*, št. 3, str. 858 - 862, nov. 1994.

[275] S. G. Mallat, "Multifrequency Channel Decompositions of Images and Wavelet Models" v *IEEE Transactions on Acoustics Speech and Signal Processing*, št. 37, str. 2091 - 2110, dec. 1989.

[276] A. C. Brooks in T. N. Pappas, "Structural similarity quality metrics in a coding context: exploring the space of realistic distortions" v *Human Vision and Electronic Imaging*, št. 6057, 2006.

[277] L. Zhang, L. Zhang in X. Q. Mou, "Rfsim: A Feature Based Image Quality Assessment Metric Using Riesz Transforms" na *IEEE International Conference on Image Processing*, str. 321 - 324, 2010.

[278] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans in A. C. Bovik, "Image quality assessment based on a degradation model" v *IEEE Transactions on Image Processing*, št. 9, str. 636 - 650, apr. 2000.

[279] E. Peli, "Contrast in Complex Images" v *Journal of the Optical Society of America a-Optics Image Science and Vision*, št. 7, str. 2032 - 2040, okt. 1990.

[280] H.-J. Z. Ulrich Engelke, "Optimal Region-of-interest based visual quality assessment" na *Human vision and electronic imaging XIV*, San Jose, 2009.

[281] C. M. Privitera in L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations" v *IEEE Transactions on Pattern Analysis and Machine Intelligence*, št. 22, str. 970 - 982, sept. 2000.

[282] A. Ninassi, O. Le Meur, P. Le Callet in D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric" na *IEEE International Conference on Image Processing*, str. 733 - 736, 2007.

[283] M. Castrillon-Santana, O. Deniz-Suarez, L. Anton-Canalis in J. Lorenzo-Navarro, "Face and facial feature detection evaluation - Performance evaluation of public domain Haar detectors for face and facial feature detection" na *Third International Conference on Computer Vision Theory and Applications (VISAPP 2008)*,št. 2, str. 167 - 172, 2008.

[284] D. N. Chandrappa, G. Akshay in M. Ravishankar, "Face Detection Using a Boosted Cascade of Features Using OpenCV" na *Wireless Networks and Computational Intelligence (ICIP 2012)*, št. 292, str. 399 - 404, 2012.

[285] M. Ratsch, S. Romdhani in T. Vetter, "Efficient face detection by a cascaded support vector machine using Haar-like features" v *Pattern Recognition*, št. 3175, str. 62 - 70, 2004.

[286] Z. Wang in A. C. Bovik, "Mean Squared Error: Love It or Leave It? A new look at signal fidelity measures" v *IEEE Signal Processing Magazine*, št. 26, str. 98 - 117, jan. 2009.

[287] M. H. Pinson in S. Wolf, "New standardized method for objectively measuring video quality" v *IEEE Transactions on Broadcasting*, št. 50, str. 312 - 322, sept. 2004.

[288] Z. Wang in E. P. Simoncelli, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics" v *Human Vision and Electronic Imaging*, št. 5292, str. 99 - 108, 2004.

[289] Q. Ma in L. M. Zhang, "Image Quality Assessment with Visual Attention" na *19th International Conference on Pattern Recognition*, str. 2783 - 2786, 2008.
[290] ANSI, "ANSI T1.801.03: Digital Transport of One-Way Video Signals - Parameters for Objective Performance Assessment", 2003.
[291] ITU-T, "J.144 : Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference ", 2004.
[292] D. Hasler in S. Susstrunk, "Measuring colourfulness in natural images" v *Human Vision and Electronic Imaging Viii*, št. 5007, str. 87 - 95, 2003.
[293] S. Winkler, "Analysis of Public Image and Video Databases for Quality Assessment" v *IEEE Journal of Selected Topics in Signal Processing*, št. 6, str. 616 - 625, okt. 2012.
[294] K. Fliegel, "QUALINET Multimedia Databases v4.0", Češka tehniška univerza v Pragi, Češka, 2013.
[295] H. R. Sheikh, M. F. Sabir in A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms" v *IEEE Transactions on Image Processing*, št. 15, str. 3440 - 3451, nov. 2006.
[296] Z. W. H.R. Sheikh, L. Cormack in A.C. Bovik, *LIVE Image Quality Assessment Database Release 2*, dostopno na: http://live.ece.utexas.edu/research/quality, obiskano: nov. 2014.
[297] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro in T. Ebrahimi, "Subjective Assessment of H.264/Avc Video Sequences Transmitted over a Noisy Channel" na *Qomex: International Workshop on Quality of Multimedia Experience*, str. 204 - 209, 2009.
[298] F. De Simone, M. Tagliasacchi, M. Naccari, S. Tubaro in T. Ebrahimi, "A H.264/Avc Video Database for the Evaluation of Quality Metrics" na *IEEE International Conference on Acoustics, Speech, and Signal Processing*, str. 2430 - 2433, 2010.
[299] H.-J. Z. U. Engelke, M. Kusuma, *Wireless Imaging Quality Database [Online]*, dostopno na: http://www.bth.se/tek/rcg.nsf/pages/wiq-db, obiskano: nov. 2014.
[300] H.-J. Z. U. Engelke in T. M. Kusuma, "Subjective Quality Assessment for Wireless Image Communication: The Wireless Imaging Quality Database" na *Video Processing and Quality Metrics (VPQM 2010)*, 2010.
[301] O. I. N. Ponomarenko, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti in C.-C. Jay Kuo, "Color Image Database TID2013: Peculiarities and Preliminary Results" na *4th Europian Workshop on Visual Information Processing (EUVIP 2013)*, Pariz, Francija, 2013.
[302] T. Liu, Y. Wang, J. M. Boyce, H. Yang in Z. Y. Wu, "A Novel Video Quality Metric for Low Bit-Rate Video Considering Both Coding and Packet-Loss Artifacts" v *IEEE Journal of Selected Topics in Signal Processing*, št. 3, str. 280 - 293, apr. 2009.
[303] S. R. Quackenbush, T. P. Barnwell in M. A. Clements, *Objective measures of speech quality*: Prentice Hall PTR, 1988.

[304] M. Varela, I. Marsh in B. Gronvall, "A systematic study of PESQ's behavior (from a networking perspective)" v *Measurement of Speech and Audio Quality in Networks*, str. 45 - 55, 2006.

[305] J. G. Beerends, C. Schmidmer, J. Berger, M. Obermann, R. Ullmann, J. Pomy in M. Keyhl, "Perceptual Objective Listening Quality Assessment (POLQA), The Third Generation ITU-T Standard for End-to-End Speech Quality Measurement Part II-Perceptual Model" v *Journal of the Audio Engineering Society*, št. 61, str. 385 - 402, jun. 2013.

[306] VQEG, "Objective perceptual assessment of video quality: Full reference television", *ITU*, 2004.

[307] e. a. Van den Heuvel, "Annotation in the SpeechDat Projects" v *International Journal of Speech Technology*, št. 4, str. 127 - 143, 2001.

[308] A. Žgank, Z. Kačič, F. Diehl, K. Vicsi, G. Szaszak, J. Juhar in S. Lihan, "The COST 278 MASPER initiative - crosslingual speech recognition with large telephone database" na *International Conference on Language Resources and Evaluation (LREC 2004)*, Lizbona, Portugalska, 2004.

PRILOGA I. Distribucija PL in vpliv na dekodirnik

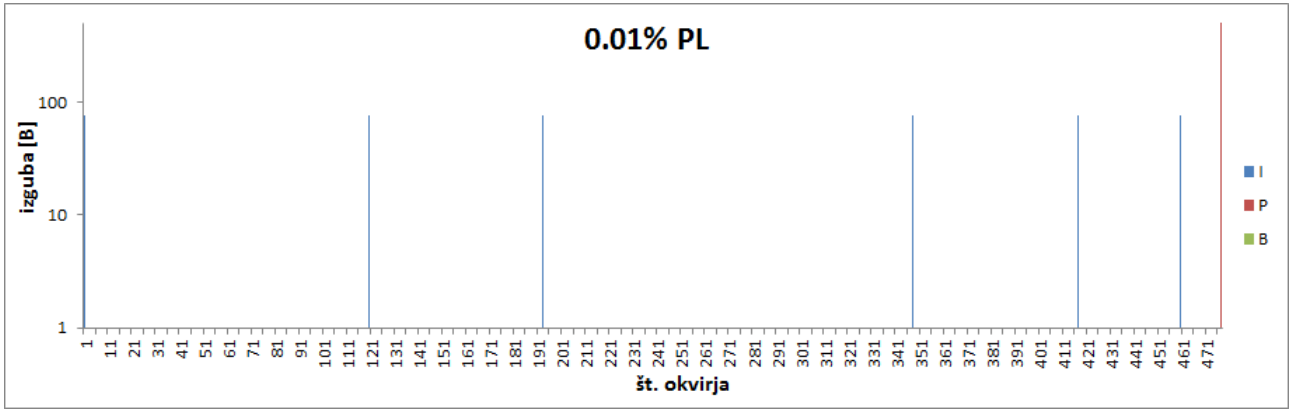
Naključna izguba IP-paketov daje široko porazdeljen vzorec, ki dekodirniku multimedijskega vsebnika onemogoča uspešno dekodiranje tako posamične slike (prostorska degradacija) kot tudi časovnih komponent. Na nivoju dekodirnika prihaja do:

- Nepravilnega ujemanja PES-velikosti paketov (A in V),
- Nepravilnosti na vhodnem pomnilniku (A),
- Nepravilno zaznanih konfiguracij avdio kanala (A),
- Napak pri dekodiranju makro-blokov.

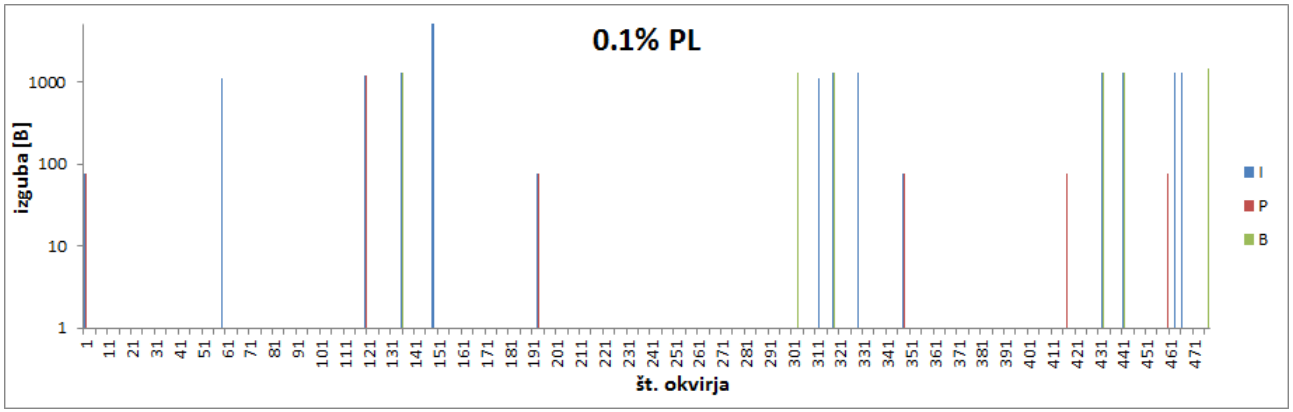
Pri modalnosti V je nastanek degradacije na kakovost časovno pogojena zaradi VBR narave multimedijskega vsebnika in prihaja do izgube različnih tipov okvirjev v GOP strukturi (I, P, B). Primer količine neuspešno dekodiranih okvirjev in njihovo časovno pomnilniško pozicijo za različne stopnje PL prikazujejo slike: slika I.1 za AV vsebino , slika I.2, slika I.3 in slika I.4. Slike prikazujejo v katerem tipu okvirja je bila izgubljena informacija ter njena amplituda. V scenarijih z nizkim PL je okvara po večini v okvirjih tipa I, saj je zaradi večje velikosti teh tudi večja verjetno, da pride do izgube paketov iz tega tipa okvirja. V scenarijih z višjim PL se odziv razporedi tudi med ostala tipa (P, B). Pri modalnosti A in AV je avditorna modalnost vsebovala 20 sekund zaključenega govora (kratek filmski odsek, komentar nogometne tekme, komentatorski odstavek ali komentar formule) s pomensko zaključeno vsebino.

Izbira vrednosti PL za posamezne tipe modalnosti je bila izbrana na podlagi pričakovane zaznavne kakovosti tako, da razpon PL vrednosti ustreza polnemu razponu MOS vrednosti. S to predpostavko so bili izbrani razponi vrednosti izgube paketov:

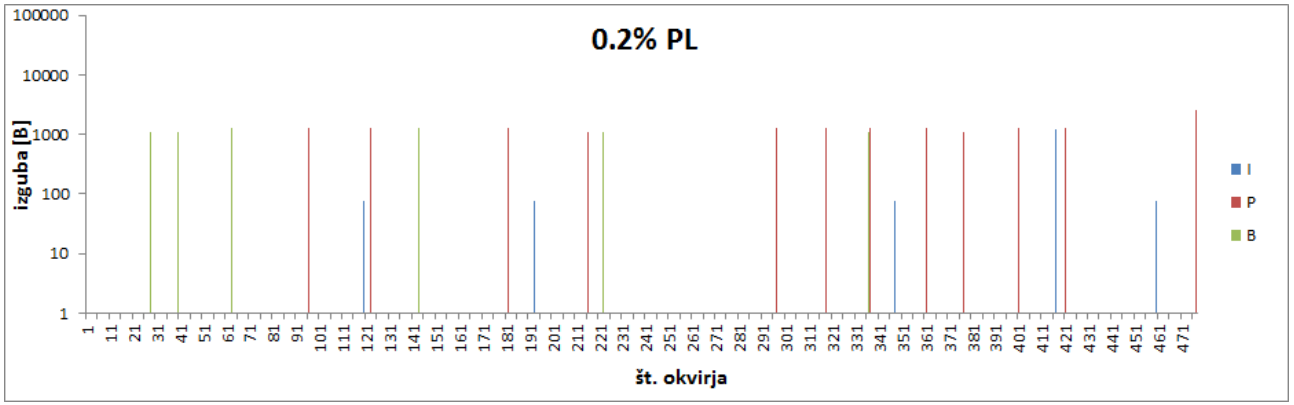
- za A vsebine: 0.00 – 5.00 % PL,
- za V vsebine: 0.00 – 0.50% PL in
- za AV vsebine: 0.00 – 0.50%.



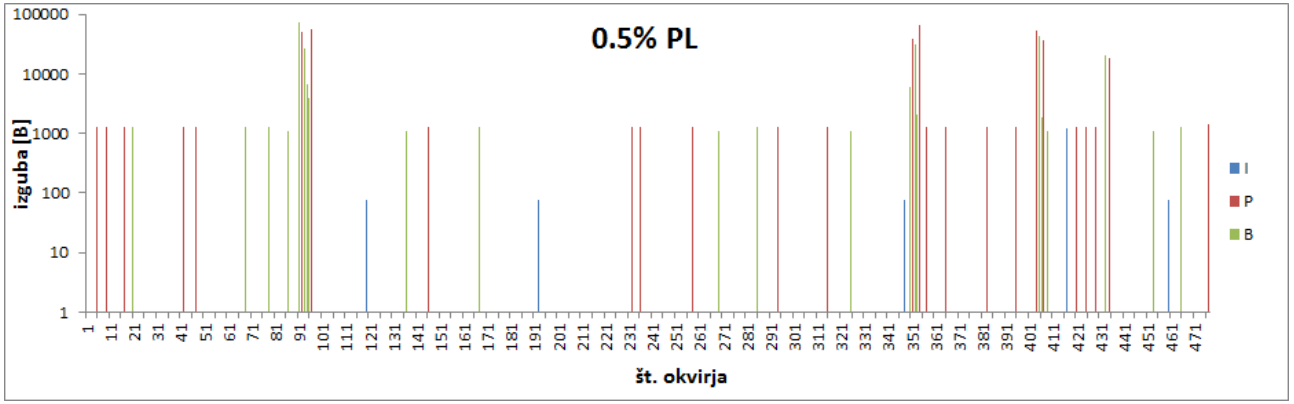
Slika I.1: Surova izguba V podatkov v medijskem vsebniku, *v_film_1* (0.01 % PL).



Slika I.2: Surova izguba V podatkov v medijskem vsebniku, *v_film_1* (0.1 % PL).

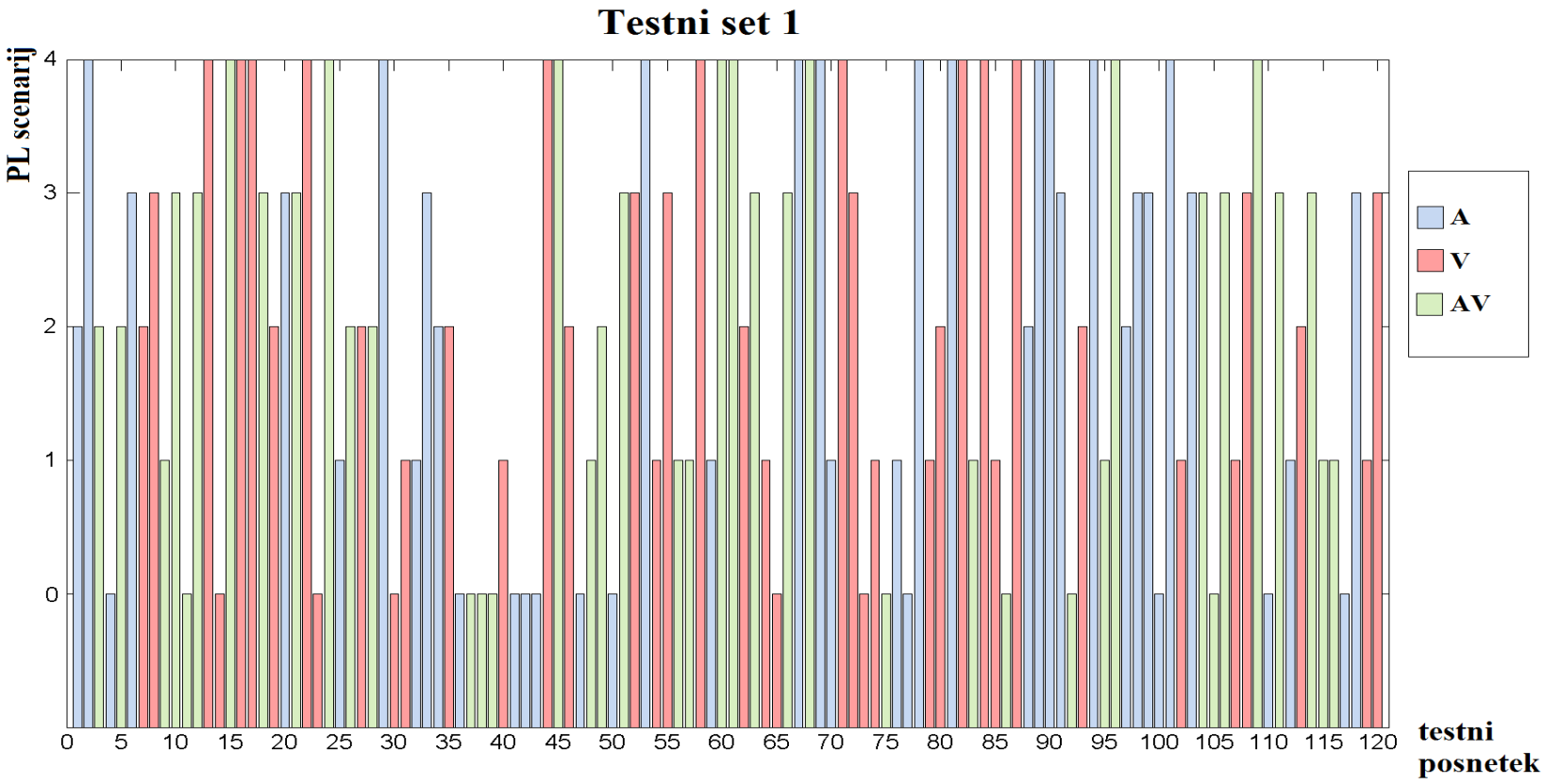


Slika I.3: Surova izguba V podatkov v medijskem vsebniku, *v_film_1* (0.2 % PL).

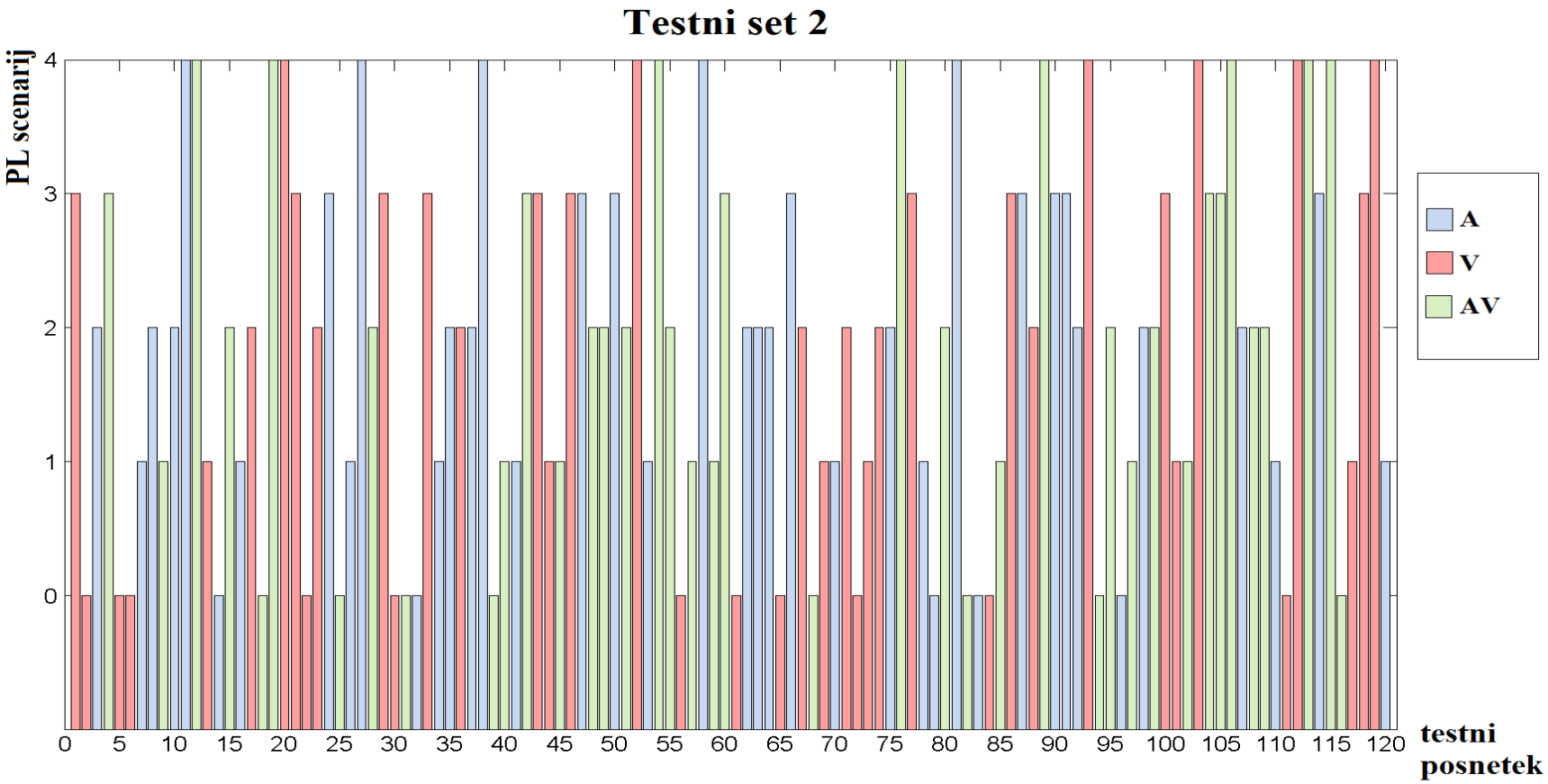


Slika I.4: Surova izguba V podatkov v medijskem vsebniku, *v_film_1* (0.5 % PL).

PRILOGA II. Razporeditev testnih vrednosti pri subjektivnih testih



Slika II.1: Testni nabor 1.



Slika II.2: Testni nabor 2.

PRILOGA III. Testni ocenjevalci

Tabela III.1: Lastnosti ocenjevalcev AV-podatkovne baze.

št. ocenjevalca	spol	starost	korekcija vida	daltonizem
1	M	21	ne	ne
2	Ž	27	ne	ne
3	M	26	ne	ne
4	M	38	ne	ne
5	M	28	ne	ne
6	M	28	ne	ne
7	Ž	26	očala	ne
8	M	28	ne	ne
9	M	31	očala	ne
10	Ž	21	očala	ne
11	Ž	39	očala	ne
12	M	24	ne	ne
13	M	23	ne	ne
14	M	25	ne	ne
15	Ž	40	očala	ne
16	M	27	očala	ne
17	M	23	ne	ne
18	M	22	ne	ne
19	Ž	27	ne	ne
20	M	27	ne	ne

PRILOGA IV. Rezultati subjektivnih testov

Tabela IV.1: Subjektivne ocene MOS-ocenjevalcev AV-podatkovne baze.

#	posnetek (modal_scena_št.)	PL [%]	modal.	scena	ocenjevalec																			MOS skupaj
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	a_film_4	0,50	A	film	3	3	5	5	4	5	5	3	5	5	4	5	5	5	5	4	4	4	4	4	**4,35**
2	a_formula_4	5,00	A	formula	1	1	1	2	1	2	1	1	2	1	1	1	1	1	2	2	2	1	1	1	**1,30**
3	av_nogomet_2	0,10	AV	nogomet	2	1	1	1	2	2	1	1	2	1	1	3	1	1	1	1	1	1	1	1	**1,30**
4	a_intervju_robert	0,00	A	intervju	5	5	4	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	**4,90**
5	av_nogomet_4	0,10	AV	nogomet	3	1	2	2	2	3	2	3	2	1	3	3	2	2	2	3	2	2	3	1	**2,20**
6	a_intervju_sara	1,00	A	intervju	4	5	5	5	5	4	5	5	5	5	5	5	5	4	5	5	5	5	5	5	**4,85**
7	v_film_2	0,10	V	film	3	4	2	2	2	3	2	2	2	2	3	4	2	2	3	3	1	2	4	1	**2,45**
8	v_formula_3	0,20	V	formula	2	1	1	1	1	2	1	2	1	1	1	2	1	2	1	1	3	1	1	1	**1,35**
9	av_film_4	0,01	AV	film	4	4	3	4	3	4	4	4	3	1	3	4	3	4	4	3	4	4	4	2	**3,45**
10	av_film_4	0,20	AV	film	3	2	2	2	1	3	1	2	2	2	2	2	2	2	1	2	2	2	1	1	**1,85**
11	av_formula_3	0,00	AV	formula	5	5	5	5	5	5	5	3	3	5	3	5	5	3	4	5	5	4	5	5	**4,50**
12	av_film_1	0,20	AV	film	2	4	2	2	2	3	2	2	2	2	2	3	3	2	2	2	3	3	3	1	**2,35**
13	v_intervju_narator	0,50	V	intervju	2	1	2	1	1	2	1	1	2	1	2	1	2	2	1	1	1	1	1	1	**1,35**
14	v_film_3	0,00	V	film	5	5	5	5	5	5	5	4	5	5	5	5	5	4	4	5	5	4	5	5	**4,80**
15	av_formula_4	0,50	AV	formula	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	**1,10**
16	v_intervju_sara	0,50	V	intervju	3	1	2	1	1	2	1	1	2	1	2	2	2	2	2	1	2	2	2	1	**1,65**
17	v_nogomet_4	0,50	V	nogomet	2	2	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	2	1	1	**1,25**
18	av_intervju_robert	0,20	AV	intervju	4	3	3	3	2	3	3	2	3	3	3	3	3	2	3	3	3	3	3	2	**2,85**
19	v_film_1	0,10	V	film	4	3	3	3	2	2	2	3	3	2	3	2	3	1	3	2	3	2	3	1	**2,50**
20	a_film_4	1,00	A	film	4	3	2	4	2	3	2	3	3	3	3	3	3	3	3	3	3	3	3	1	**2,85**
21	av_film_2	0,20	AV	film	3	2	3	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	1	1	**1,90**

22	v_formula_1	0,50	V	formula	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	3	1	1	1	**1,15**
23	v_film_2	0,00	V	film	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
24	av_intervju_narator	0,50	AV	intervju	2	1	2	2	1	2	1	1	2	1	1	2	2	2	2	2	1	2	1	1	**1,55**
25	a_film_2	0,20	A	film	5	4	5	5	5	4	5	5	4	5	5	3	5	5	5	5	4	5	5	5	**4,70**
26	av_nogomet_3	0,10	AV	nogomet	4	3	3	3	3	2	2	2	3	2	3	3	2	2	3	2	3	3	3	1	**2,60**
27	v_intervju_naratorka	0,10	V	intervju	2	3	3	2	2	2	2	2	2	1	2	3	3	1	2	2	2	2	1	1	**2,00**
28	av_formula_3	0,10	AV	formula	4	4	3	4	3	3	2	2	3	3	3	4	3	2	3	3	3	3	3	2	**3,00**
29	a_film_4	5,00	A	film	2	3	2	3	1	3	3	1	3	2	3	2	2	2	3	3	4	1	1	3	**2,35**
30	v_formula_2	0,00	V	formula	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	**4,95**
31	v_nogomet_3	0,01	V	nogomet	4	4	4	4	3	3	3	3	3	1	3	4	3	4	4	3	4	4	4	2	**3,35**
32	a_nogomet_3	0,20	A	nogomet	4	2	4	4	5	4	5	4	5	4	5	5	5	3	5	4	5	4	5	4	**4,30**
33	a_intervju_naratorka	1,00	A	intervju	4	2	2	3	2	3	3	1	3	2	3	2	4	3	3	3	3	3	2	1	**2,60**
34	a_formula_2	0,50	A	formula	3	3	3	4	4	4	4	2	4	3	4	4	5	4	4	4	4	4	4	3	**3,70**
35	v_nogomet_4	0,10	V	nogomet	3	3	2	3	2	2	1	2	3	3	3	3	3	3	3	2	1	2	2	1	**2,35**
36	a_film_2	0,00	A	film	5	5	5	5	5	4	5	5	5	5	4	5	5	5	5	5	4	5	5	5	**4,85**
37	av_intervju_naratorka	0,00	AV	intervju	5	5	5	5	5	5	5	5	4	4	5	4	5	5	5	5	5	5	4	5	**4,80**
38	av_nogomet_1	0,00	AV	nogomet	5	5	5	5	5	5	4	5	4	4	5	5	4	5	5	5	5	5	4	5	**4,75**
39	av_intervju_robert	0,00	AV	intervju	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	**4,95**
40	v_intervju_sara	0,01	V	intervju	3	5	4	4	3	3	4	3	3	1	3	5	3	4	4	3	4	4	3	3	**3,45**
41	a_film_1	0,00	A	film	5	4	5	5	5	5	5	5	4	4	5	5	5	5	5	5	4	5	4	5	5	**4,75**
42	a_nogomet_3	0,00	A	nogomet	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	4	4	5	5	**4,85**
43	a_intervju_naratorka	0,00	A	intervju	5	5	5	5	5	5	5	5	5	4	5	5	4	5	5	4	5	5	4	5	**4,80**
44	v_film_2	0,50	V	film	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	**1,10**
45	av_film_2	0,50	AV	film	1	1	1	2	1	2	1	1	2	1	1	1	2	1	2	1	1	1	1	1	**1,25**
46	v_intervju_robert	0,10	V	intervju	2	4	3	3	3	3	3	2	3	3	3	3	3	2	3	3	3	3	3	2	**2,85**
47	a_nogomet_4	0,00	A	nogomet	5	5	5	5	5	5	5	5	4	5	5	5	5	4	5	5	5	4	5	5	**4,85**
48	av_formula_3	0,01	AV	formula	4	5	4	4	4	3	3	3	3	1	2	4	4	3	4	4	4	4	3	3	**3,45**
49	av_intervju_naratorka	0,10	AV	intervju	3	1	3	3	2	2	2	1	2	3	2	3	2	2	3	2	3	2	3	1	**2,25**
50	a_intervju_sara	0,00	A	intervju	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
51	av_formula_2	0,20	AV	formula	2	2	2	3	2	2	1	1	2	2	2	2	2	1	2	2	2	2	2	1	**1,85**
52	v_nogomet_4	0,20	V	nogomet	3	2	2	2	1	2	1	1	2	1	2	2	2	1	2	2	2	1	1	1	**1,65**

53	a_intervju_robert	5,00	A	intervju	1	2	1	2	1	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	**1,20**
54	v_intervju_robert	0,01	V	intervju	4	4	4	4	4	4	4	3	4	1	3	4	4	5	4	4	4	4	5	5	**3,90**
55	v_intervju_sara	0,20	V	intervju	2	2	2	2	2	2	1	1	2	2	3	2	2	2	2	2	2	2	2	1	**1,90**
56	av_formula_4	0,01	AV	formula	4	4	4	4	4	4	4	4	4	1	3	4	4	5	4	3	3	3	3	4	**3,65**
57	av_intervju_robert	0,01	AV	intervju	4	4	4	4	3	4	4	3	3	1	3	4	4	5	4	4	4	4	5	4	**3,75**
58	v_nogomet_3	0,50	V	nogomet	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**1,00**
59	a_intervju_sara	0,20	A	intervju	4	5	5	5	5	5	5	4	4	4	4	5	5	5	5	5	5	5	5	5	**4,75**
60	av_nogomet_3	0,50	AV	nogomet	1	1	1	2	2	1	1	1	2	2	1	1	2	1	1	2	2	1	2	1	**1,40**
61	av_formula_1	0,50	AV	formula	2	1	1	1	2	2	1	1	2	1	1	2	1	2	2	2	2	2	1	1	**1,50**
62	v_nogomet_2	0,10	V	nogomet	2	1	1	2	2	2	1	1	2	1	2	1	3	1	2	2	2	3	1	1	**1,65**
63	av_intervju_narator	0,20	AV	intervju	3	3	3	3	2	3	2	2	3	3	2	3	3	3	3	3	3	3	4	2	**2,80**
64	v_film_2	0,01	V	film	3	3	3	3	3	3	2	3	3	1	2	4	3	5	3	3	3	3	3	1	**2,85**
65	v_intervju_naratorka	0,00	V	intervju	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
66	av_nogomet_3	0,20	AV	nogomet	1	3	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	2	1	**1,85**
67	a_film_3	5,00	A	film	1	2	1	3	1	1	1	1	3	1	1	1	3	1	3	2	3	2	1	1	**1,65**
68	av_film_4	0,50	AV	film	1	1	1	2	1	1	1	1	2	1	2	2	3	2	2	2	1	2	1	1	**1,50**
69	a_film_2	5,00	A	film	1	1	1	2	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1	**1,20**
70	a_film_1	0,20	A	film	2	5	5	4	5	5	5	4	4	5	5	5	5	5	5	4	5	5	5	5	**4,65**
71	v_film_1	0,50	V	film	2	2	1	1	1	1	1	1	1	1	1	2	1	2	1	1	2	1	1	1	**1,25**
72	v_nogomet_2	0,20	V	nogomet	1	2	2	2	1	1	1	1	2	2	2	1	1	1	1	2	3	3	2	1	**1,60**
73	v_nogomet_2	0,00	V	nogomet	4	5	5	4	5	5	5	5	4	5	4	5	5	5	5	5	5	5	5	5	**4,80**
74	v_nogomet_1	0,01	V	nogomet	5	4	4	4	4	4	4	4	4	1	3	4	3	4	4	4	5	4	4	4	**3,85**
75	av_nogomet_3	0,00	AV	nogomet	5	5	5	5	5	5	5	3	5	5	5	5	5	3	5	5	5	5	3	5	**4,70**
76	a_film_4	0,20	A	film	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	**4,95**
77	a_nogomet_2	0,00	A	nogomet	3	5	5	5	5	4	5	5	5	4	5	5	5	5	5	4	5	5	5	5	**4,75**
78	a_nogomet_1	5,00	A	nogomet	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	**1,05**
79	v_formula_2	0,01	V	formula	3	4	3	4	3	3	3	3	3	1	2	4	3	4	4	3	3	3	3	3	**3,10**
80	v_formula_4	0,10	V	formula	3	3	3	3	3	2	2	3	3	3	2	4	2	3	3	3	4	3	3	2	**2,85**
81	a_formula_1	5,00	A	formula	1	2	1	2	2	1	1	1	2	1	1	1	3	2	2	2	3	2	1	1	**1,60**
82	v_intervju_naratorka	0,50	V	intervju	1	1	1	1	1	1	1	3	2	1	1	1	1	1	1	1	3	1	1	1	**1,25**

83	av_film_2	0,01	AV	film	4	5	4	4	4	4	4	5	4	5	3	5	5	5	4	3	4	5	4	4	**4,25**
84	v_intervju_robert	0,50	V	intervju	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	**1,10**
85	v_film_3	0,01	V	film	3	4	4	4	4	4	4	3	4	1	3	4	4	4	4	4	4	4	4	4	**3,70**
86	av_film_2	0,00	AV	film	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	**4,95**
87	v_formula_2	0,50	V	formula	1	2	1	2	1	1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	**1,20**
88	a_formula_1	0,50	A	formula	4	4	5	5	5	4	5	4	4	4	5	4	5	5	5	4	4	4	4	4	**4,40**
89	a_nogomet_2	5,00	A	nogomet	1	1	2	2	1	1	1	1	2	1	1	1	1	2	2	1	2	1	1	1	**1,30**
90	a_nogomet_4	5,00	A	nogomet	1	1	1	2	2	2	1	1	2	1	1	1	2	1	2	2	1	1	1	1	**1,35**
91	a_nogomet_2	1,00	A	nogomet	2	3	2	4	3	3	4	2	3	3	3	2	3	2	3	3	3	3	3	2	**2,80**
92	av_intervju_narator	0,00	AV	intervju	5	5	5	5	5	5	5	3	5	4	5	5	4	5	5	5	5	5	5	5	**4,80**
93	v_formula_3	0,10	V	formula	1	2	2	3	2	2	2	1	3	3	1	3	2	3	2	2	2	2	3	1	**2,10**
94	a_intervju_narator	5,00	A	intervju	1	2	1	2	1	1	1	1	2	2	1	2	2	2	2	2	2	1	1	1	**1,50**
95	av_nogomet_2	0,01	AV	nogomet	5	4	4	4	5	3	3	3	3	4	2	4	4	5	3	3	4	5	4	4	**3,80**
96	av_film_3	0,50	AV	film	1	1	1	1	1	1	1	1	2	1	1	2	1	1	5	2	1	1	2	1	**1,40**
97	a_intervju_robert	0,50	A	intervju	5	5	5	5	5	5	5	5	4	4	4	5	5	4	1	5	4	4	5	4	**4,45**
98	a_formula_3	1,00	A	formula	3	3	4	4	4	4	3	2	4	4	3	4	3	4	3	4	4	4	4	3	**3,55**
99	a_formula_2	1,00	A	formula	2	3	1	3	3	4	3	1	3	3	3	3	3	3	4	4	3	3	3	1	**2,80**
100	a_nogomet_1	0,00	A	nogomet	5	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**4,90**
101	a_formula_2	5,00	A	formula	1	2	1	1	1	1	1	1	2	1	1	1	1	2	2	2	2	2	1	1	**1,35**
102	v_formula_1	0,01	V	formula	4	4	4	4	4	5	4	3	4	1	4	4	4	4	4	4	4	5	5	3	4	**3,90**
103	a_nogomet_4	1,00	A	nogomet	3	3	2	3	5	5	5	2	4	3	4	5	3	3	5	4	5	3	4	5	**3,80**
104	av_intervju_naratorka	0,20	AV	intervju	3	3	2	2	2	3	2	2	3	2	2	3	2	3	3	2	3	3	2	1	**2,40**
105	av_formula_1	0,00	AV	formula	5	5	4	5	5	5	5	5	4	5	5	4	5	5	5	5	4	5	4	5	**4,75**
106	av_formula_4	0,20	AV	formula	2	1	2	3	1	2	1	2	2	1	1	2	1	2	2	2	3	2	2	1	**1,75**
107	v_formula_3	0,01	V	formula	4	4	4	4	5	4	4	4	4	1	3	4	4	4	4	5	4	4	4	4	**3,90**
108	v_film_3	0,20	V	film	1	1	2	2	1	1	1	1	2	1	1	1	2	1	2	1	1	1	1	1	**1,25**
109	av_nogomet_2	0,50	AV	nogomet	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**1,05**
110	a_formula_4	0,00	A	formula	4	4	5	5	5	5	5	2	5	4	5	5	5	5	5	5	5	5	5	5	**4,70**
111	av_nogomet_1	0,20	AV	nogomet	2	2	1	2	2	3	2	1	2	2	1	2	1	2	2	2	1	2	2	1	**1,75**
112	a_intervju_narator	0,20	A	intervju	4	4	2	4	3	4	3	2	4	2	3	4	4	3	3	3	2	3	3	1	**3,05**
113	v_formula_2	0,10	V	formula	1	2	2	2	2	2	1	2	2	2	1	3	2	2	2	1	3	3	2	1	**1,90**

114	av_formula_3	0,20	AV	formula	2	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	1	**1,85**
115	av_film_3	0,01	AV	film	5	4	4	4	4	3	4	3	3	1	4	4	5	4	4	4	4	4	4	3	**3,75**
116	av_intervju_naratorka	0,01	AV	intervju	5	5	4	3	4	3	4	2	3	5	2	4	5	5	4	4	3	5	4	3	**3,85**
117	a_formula_3	0,00	A	formula	5	4	5	5	5	5	5	3	4	4	5	4	5	5	5	5	5	5	4	5	**4,65**
118	a_formula_4	1,00	A	formula	2	2	2	4	2	3	3	3	3	3	3	3	4	4	4	3	4	3	3	2	**3,00**
119	v_nogomet_4	0,01	V	nogomet	3	4	3	4	4	3	4	3	4	2	4	4	3	4	4	3	3	3	3	3	**3,40**
120	v_film_1	0,20	V	film	2	1	2	3	1	1	1	1	2	1	2	2	1	2	2	2	2	2	2	1	**1,65**
121	v_formula_4	0,20	V	formula	2	2	2	2	1	1	1	1	2	2	1	2	2	2	2	2	2	1	2	1	**1,65**
122	v_nogomet_1	0,00	V	nogomet	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5	5	5	**4,90**
123	a_film_2	0,50	A	film	4	4	3	4	3	4	4	3	4	4	4	3	3	4	4	4	4	4	4	4	**3,75**
124	av_intervju_sara	0,20	AV	intervju	3	2	2	2	2	3	2	2	2	2	2	3	2	2	2	2	3	2	2	1	**2,15**
125	v_formula_4	0,00	V	formula	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
126	v_formula_3	0,00	V	formula	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	4	5	5	5	**4,85**
127	a_nogomet_1	0,20	A	nogomet	4	5	4	4	4	4	5	4	3	5	4	4	5	5	4	4	4	4	4	4	**4,20**
128	a_film_1	0,50	A	film	5	5	5	5	5	5	5	5	4	4	4	4	4	4	5	3	5	5	5	5	**4,60**
129	av_formula_1	0,01	AV	formula	4	5	4	4	4	4	4	4	0	0	0	4	4	4	4	3	5	4	4	4	**3,45**
130	a_nogomet_2	0,50	A	nogomet	3	3	3	4	3	4	4	3	4	4	3	3	4	4	3	3	3	3	3	3	**3,35**
131	a_intervju_sara	5,00	A	intervju	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**1,00**
132	av_intervju_robert	0,50	AV	intervju	1	2	1	1	1	1	1	1	1	1	1	2	2	1	1	2	1	2	1	1	**1,25**
133	v_nogomet_2	0,01	V	nogomet	4	4	4	4	4	4	4	4	0	0	0	4	4	4	4	4	5	4	4	4	**3,45**
134	a_film_4	0,00	A	film	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	**4,95**
135	av_film_3	0,10	AV	film	3	2	2	3	1	2	1	1	2	1	1	3	2	2	2	2	1	2	2	2	**1,85**
136	a_formula_2	0,20	A	formula	4	4	4	4	3	4	4	3	4	4	4	5	5	5	4	4	4	4	4	3	**4,00**
137	v_nogomet_3	0,10	V	nogomet	2	2	3	3	3	3	2	3	3	3	2	3	3	4	2	2	3	3	3	2	**2,70**
138	av_formula_4	0,00	AV	formula	5	5	5	5	5	5	5	5	4	5	5	5	2	3	5	5	5	5	5	5	**4,70**
139	av_film_1	0,50	AV	film	2	2	1	2	1	1	1	3	2	1	1	2	1	1	2	1	1	2	2	1	**1,50**
140	v_formula_4	0,50	V	formula	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**1,10**
141	v_intervju_robert	0,20	V	intervju	2	3	3	3	3	2	2	2	3	3	3	3	2	3	2	2	2	2	3	2	**2,50**
142	v_intervju_sara	0,00	V	intervju	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
143	v_film_4	0,10	V	film	3	3	2	3	2	2	2	2	3	3	2	3	2	2	2	2	3	3	2	2	**2,40**

144	a_film_1	1,00	A	film	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	4	5	5	4	4	**4,75**
145	av_film_1	0,00	AV	film	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
146	a_nogomet_4	0,20	A	nogomet	4	4	4	5	4	4	5	3	4	4	4	5	5	5	5	5	5	4	4	5	4	**4,35**
147	a_formula_3	5,00	A	formula	1	1	1	2	1	2	1	1	1	1	1	1	1	1	3	1	1	1	1	1	**1,20**
148	av_formula_2	0,10	AV	formula	2	3	2	3	2	2	1	2	2	2	1	2	2	2	2	2	2	2	2	1	**1,95**
149	v_film_2	0,20	V	film	1	3	1	3	3	2	1	2	3	3	2	2	2	2	2	2	2	2	2	1	**2,05**
150	v_film_1	0,00	V	film	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**4,95**
151	av_nogomet_4	0,00	AV	nogomet	5	5	5	5	5	5	5	4	5	5	5	5	5	4	5	5	5	5	5	5	**4,90**
152	a_formula_2	0,00	A	formula	5	4	5	5	5	5	5	5	4	4	4	5	5	5	5	5	5	5	5	5	**4,80**
153	v_intervju_naratorka	0,20	V	intervju	2	3	1	3	1	1	1	1	2	2	1	2	2	1	2	2	1	2	2	1	**1,65**
154	a_intervju_robert	0,20	A	intervju	4	4	4	5	4	5	5	4	5	5	5	4	4	5	5	4	5	4	4	4	**4,45**
155	a_nogomet_1	0,50	A	nogomet	4	3	4	5	4	5	4	4	4	4	4	4	4	5	4	4	4	4	5	5	**4,20**
156	v_film_3	0,10	V	film	3	3	3	3	3	2	2	2	3	3	3	3	2	2	3	3	4	3	3	2	**2,75**
157	a_nogomet_4	0,50	A	nogomet	5	4	5	5	5	5	5	5	3	3	3	5	5	5	5	5	5	5	5	5	**4,65**
158	a_intervju_naratorka	5,00	A	intervju	1	3	1	1	1	1	1	1	2	2	1	1	1	2	1	2	1	1	1	1	**1,30**
159	av_nogomet_2	0,00	AV	nogomet	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	4	5	5	4	5	**4,80**
160	av_nogomet_3	0,01	AV	nogomet	5	4	5	5	5	5	5	5	0	0	0	5	5	5	5	4	5	5	5	5	**4,15**
161	a_intervju_naratorka	0,20	A	intervju	5	5	4	5	5	5	5	5	5	4	5	5	5	5	5	5	4	5	5	5	**4,85**
162	av_formula_1	0,20	AV	formula	2	2	2	2	3	2	1	2	2	3	1	3	2	2	2	2	3	2	2	1	**2,05**
163	v_formula_2	0,20	V	formula	1	3	2	3	2	2	1	1	2	2	4	2	2	2	1	3	3	2	1	1	**2,00**
164	v_intervju_narator	0,01	V	intervju	5	5	4	5	5	5	4	4	0	0	0	4	5	5	4	5	4	4	5	5	**3,90**
165	av_film_1	0,01	AV	film	5	5	5	5	5	5	5	5	0	0	0	5	5	5	5	5	5	5	5	5	**4,25**
166	v_film_4	0,20	V	film	3	3	2	3	2	2	1	2	2	2	2	3	2	1	2	2	2	2	2	1	**2,05**
167	a_nogomet_1	1,00	A	nogomet	4	3	3	4	4	4	4	3	3	3	4	5	4	2	4	4	1	3	3	4	**3,45**
168	av_intervju_sara	0,10	AV	intervju	3	3	3	3	3	3	3	2	3	3	2	3	2	4	3	3	4	3	4	1	**2,90**
169	av_film_2	0,10	AV	film	3	3	2	2	2	2	2	2	3	3	1	3	2	2	2	2	3	2	3	2	**2,30**
170	a_intervju_robert	1,00	A	intervju	5	4	4	5	5	5	5	5	4	4	3	5	5	4	5	4	4	4	5	4	**4,45**
171	av_formula_4	0,10	AV	formula	3	3	4	3	2	3	2	2	2	3	2	3	2	1	2	2	2	2	2	4	**2,45**
172	v_film_4	0,50	V	film	2	2	2	2	2	2	2	2	1	2	1	3	2	1	2	2	1	2	3	2	**1,90**
173	a_formula_1	0,20	A	formula	5	4	5	5	5	5	5	5	4	5	5	5	5	5	5	4	4	5	5	5	**4,80**
174	av_nogomet_1	0,50	AV	nogomet	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**1,00**

175	av_nogomet_1	0,10	AV	nogomet	2	1	2	2	2	1	1	1	2	2	1	2	1	2	2	2	2	2	2	1	**1,65**
176	v_nogomet_4	0,00	V	nogomet	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	4	**4,90**
177	av_nogomet_1	0,01	AV	nogomet	4	4	4	4	4	4	4	1	0	0	0	5	3	5	4	4	5	4	4	1	**3,20**
178	a_nogomet_3	5,00	A	nogomet	1	1	1	1	1	1	1	3	2	1	2	1	2	2	2	1	1	1	1	2	**1,40**
179	av_formula_2	0,01	AV	formula	3	3	4	3	4	3	3	3	5	4	5	4	3	3	3	3	4	3	3	2	**3,40**
180	av_film_3	0,20	AV	film	2	3	2	2	3	2	1	1	2	3	1	3	2	2	2	3	2	2	2	2	**2,10**
181	v_intervju_robert	0,00	V	intervju	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
182	a_intervju_narator	0,50	A	intervju	4	5	4	5	5	5	5	4	5	4	5	5	5	5	4	5	4	4	5	5	**4,65**
183	a_formula_3	0,50	A	formula	4	4	4	4	3	4	4	3	3	4	4	4	4	5	4	4	3	4	4	4	**3,85**
184	a_formula_4	0,50	A	formula	5	3	5	5	5	5	5	3	4	5	5	5	5	5	5	4	5	5	5	5	**4,70**
185	v_nogomet_3	0,00	V	nogomet	5	4	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	**4,90**
186	a_nogomet_3	1,00	A	nogomet	3	2	2	3	2	3	2	1	2	3	2	3	3	2	3	3	1	3	2	1	**2,30**
187	v_intervju_sara	0,10	V	intervju	2	2	3	2	2	2	2	2	2	3	1	3	2	1	2	2	1	2	3	4	**2,15**
188	av_film_4	0,00	AV	film	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	4	5	5	5	5	**4,90**
189	v_film_1	0,01	V	film	5	5	4	4	5	5	4	4	0	0	0	5	5	5	5	4	4	4	5	5	**3,90**
190	a_formula_3	0,20	A	formula	4	4	5	4	5	5	4	5	4	4	5	5	3	5	5	4	4	4	4	4	**4,35**
191	v_formula_1	0,10	V	formula	3	2	3	3	2	3	2	2	3	3	1	3	2	2	3	2	2	2	3	3	**2,45**
192	v_film_4	0,00	V	film	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	**4,90**
193	v_formula_4	0,01	V	formula	3	3	2	3	3	3	2	1	0	0	0	3	3	0	3	2	2	3	2	2	**2,00**
194	v_nogomet_1	0,10	V	nogomet	2	2	2	2	2	2	1	1	2	3	2	3	2	3	2	2	1	2	3	1	**2,00**
195	a_intervju_naratorka	0,50	A	intervju	3	4	4	4	4	4	4	3	4	4	4	4	4	5	4	4	3	4	4	4	**3,90**
196	av_intervju_naratorka	0,50	AV	intervju	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,10**
197	v_nogomet_1	0,20	V	nogomet	1	3	1	2	2	1	1	1	2	3	1	2	1	1	1	2	1	1	2	1	**1,50**
198	a_film_3	0,20	A	film	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	4	5	5	5	**4,90**
199	a_film_3	0,00	A	film	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**5,00**
200	av_formula_1	0,10	AV	formula	2	3	2	2	2	2	1	1	2	3	1	3	1	2	2	2	3	2	2	1	**1,95**
201	a_film_1	5,00	A	film	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	**1,05**
202	av_formula_2	0,00	AV	formula	4	5	5	4	5	5	5	3	5	5	5	5	5	3	5	5	5	5	5	5	**4,70**
203	a_formula_1	0,00	A	formula	5	5	5	5	5	5	5	5	4	5	4	5	5	5	5	5	5	5	5	5	**4,90**
204	v_intervju_narator	0,00	V	intervju	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	**4,95**

205	av_intervju_sara	0,01	AV	intervju	3	1	4	4	3	3	4	2	0	0	0	4	3	4	3	3	4	3	4	4	**2,80**
206	v_nogomet_3	0,20	V	nogomet	1	1	1	1	2	1	1	1	2	1	1	2	1	1	1	1	1	2	1	1	**1,20**
207	a_intervju_narator	1,00	A	intervju	2	2	2	4	2	2	2	1	2	3	2	3	2	3	3	2	2	1	3	3	**2,30**
208	v_intervju_narator	0,10	V	intervju	3	1	2	2	1	1	1	1	2	2	1	2	2	2	1	2	1	1	2	1	**1,55**
209	av_formula_2	0,50	AV	formula	1	1	2	2	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,20**
210	a_formula_1	1,00	A	formula	3	4	3	3	3	3	3	3	3	3	2	3	3	2	3	3	3	3	3	2	**2,90**
211	a_film_3	1,00	A	film	2	3	2	3	1	3	3	2	3	3	2	3	1	2	3	3	2	2	5	2	**2,50**
212	a_film_3	0,50	A	film	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	**4,95**
213	v_nogomet_2	0,50	V	nogomet	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,10**
214	av_film_3	0,00	AV	film	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	**4,95**
215	av_film_4	0,10	AV	film	3	2	3	2	2	2	2	2	0	0	0	3	3	2	2	2	2	3	3	3	**2,05**
216	a_intervju_narator	0,00	A	intervju	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5	5	5	**4,90**
217	av_intervju_narator	0,01	AV	intervju	4	3	4	4	4	3	4	5	0	0	0	4	5	5	4	3	4	3	4	3	**3,30**
218	a_nogomet_3	0,50	A	nogomet	3	4	4	4	3	4	4	3	4	4	3	4	4	4	4	4	3	4	4	3	**3,70**
219	av_intervju_narator	0,10	AV	intervju	3	2	2	3	2	2	2	1	2	2	1	2	1	1	2	2	2	2	3	2	**1,95**
220	v_intervju_narator	0,20	V	intervju	2	1	2	2	2	2	2	2	2	2	1	2	2	1	2	2	1	2	2	1	**1,75**
221	v_film_4	0,01	V	film	5	3	4	4	3	3	3	3	0	0	0	4	5	5	3	3	3	3	3	3	**3,00**
222	av_nogomet_4	0,01	AV	nogomet	4	4	4	4	4	4	4	3	0	0	0	4	4	4	4	4	4	4	4	4	**3,35**
223	v_formula_3	0,50	V	formula	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,10**
224	av_nogomet_4	0,20	AV	nogomet	1	2	1	2	1	1	1	1	2	1	1	2	1	1	1	1	1	2	2	1	**1,30**
225	av_nogomet_2	0,20	AV	nogomet	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,05**
226	av_nogomet_4	0,50	AV	nogomet	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	**1,10**
227	a_intervju_sara	0,50	A	intervju	5	5	5	5	5	5	5	4	5	5	5	5	5	5	2	5	5	4	5	5	5	**4,75**
228	av_film_1	0,10	AV	film	3	2	2	3	2	2	2	2	3	3	4	3	3	3	3	2	3	3	3	4	2	**2,70**
229	av_intervju_robert	0,10	AV	intervju	3	3	3	2	2	2	2	1	2	2	1	3	2	2	3	2	2	2	2	2	**2,15**
230	a_nogomet_2	0,20	A	nogomet	4	4	5	5	5	5	5	4	5	4	5	5	5	5	5	4	5	5	5	5	**4,75**
231	v_formula_1	0,00	V	formula	5	4	5	5	5	5	5	5	4	5	4	5	5	5	5	4	5	5	5	5	**4,80**
232	v_nogomet_1	0,50	V	nogomet	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	**1,05**
233	av_intervju_sara	0,50	AV	intervju	2	2	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,20**
234	a_film_2	1,00	A	film	4	3	4	4	4	4	4	3	4	4	4	4	4	4	4	4	3	4	4	4	**3,85**
235	av_formula_3	0,50	AV	formula	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	**1,10**

236	av_intervju_sara	0,00	AV	intervju	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	4	5	5	5	5	**4,90**
237	v_intervju_naratorka	0,01	V	intervju	5	3	4	4	3	4	4	3	0	0	0	5	4	4	4	4	4	4	4	4	**3,35**
238	v_formula_1	0,20	V	formula	2	2	3	2	2	1	2	2	2	3	2	2	2	2	2	2	2	2	3	3	**2,15**
239	v_film_3	0,50	V	film	2	1	1	1	1	1	1	1	1	3	1	2	1	1	1	1	1	1	2	1	**1,25**
240	a_formula_4	0,20	A	formula	5	3	4	5	5	5	5	4	5	5	5	5	5	5	5	5	4	5	5	5	**4,75**

PRILOGA V. Rezultati subjektivnih testov porazdeljenega vrednotenja

Tabela V.1: Subjektivne ocene MOS ocenjevalcev pri porazdeljenem vrednotenju kakovosti.

Scena posnetka	Tip degradacije v posnetku	Posnetek	Scenarij degradacije	Zap. št. subj. testa	Številka ocenjevalca 1	2	3	4	5	6	7	8	9	10	Povprečna subMOS	Standardna deviacija
narator	ROI	1	0	**31**	4	5	5	5	5	5	5	5	5	5	4,9	0,316
narator	ROI	1	1	**83**	5	3	4	4	3	4	4	4	3	3	3,7	0,675
narator	ROI	1	2	**65**	3	3	3	1	2	3	2	4	2	3	2,6	0,843
narator	ROI	1	3	**59**	3	3	4	2	3	5	4	4	3	3	3,4	0,843
narator	ROI	1	4	**21**	4	3	3	2	3	4	3	4	3	3	3,2	0,632
narator	ROI	1	5	**78**	2	3	2	1	1	3	2	4	2	3	2,3	0,949
narator	ROI	2	0	**73**	5	5	5	5	5	5	5	5	5	5	5,0	0,000
narator	ROI	2	1	**63**	4	4	4	3	3	5	4	4	3	4	3,8	0,632
narator	ROI	2	2	**64**	3	2	2	1	2	4	3	4	2	2	2,5	0,972
narator	ROI	2	3	**22**	4	3	3	1	2	4	3	4	3	3	3,0	0,943
narator	ROI	2	4	**40**	4	3	4	2	2	4	3	3	3	3	3,1	0,738
narator	ROI	2	5	**84**	4	3	3	2	2	4	4	4	3	3	3,2	0,789
naratorka	ROI	1	0	**11**	5	5	5	5	5	4	5	5	5	5	4,9	0,316
naratorka	ROI	1	1	**62**	4	3	3	3	3	4	4	4	3	3	3,4	0,516
naratorka	ROI	1	2	**32**	4	3	3	2	3	5	4	4	3	3	3,4	0,843
naratorka	ROI	1	3	**34**	3	3	3	4	2	4	3	4	3	3	3,2	0,632
naratorka	ROI	1	4	**19**	4	3	2	2	2	3	2	4	3	3	2,8	0,789
naratorka	ROI	1	5	**9**	4	3	2	2	3	3	3	4	3	3	3,0	0,667
naratorka	ROI	2	0	**42**	5	5	5	5	5	5	5	5	4	5	4,9	0,316
naratorka	ROI	2	1	**41**	4	3	4	4	3	4	4	4	3	3	3,6	0,516
naratorka	ROI	2	2	**43**	3	3	2	1	2	4	3	4	2	3	2,7	0,949
naratorka	ROI	2	3	**20**	4	3	3	2	2	3	2	4	3	3	2,9	0,738
naratorka	ROI	2	4	**50**	3	3	3	2	2	3	3	3	3	3	2,8	0,422

naratorka	ROI	2	5	**79**	2	3	2	1	2	3	2	4	3	3	2,5	0,850
robert	ROI	1	0	**30**	5	5	5	5	5	5	5	5	5	5	5,0	0,000
robert	ROI	1	1	**86**	4	5	4	4	3	4	3	4	5	4	4,0	0,667
robert	ROI	1	2	**68**	4	3	3	3	2	4	4	4	3	3	3,3	0,675
robert	ROI	1	3	**33**	5	3	4	2	2	4	4	4	3	3	3,4	0,966
robert	ROI	1	4	**29**	4	3	3	3	3	5	4	4	3	3	3,5	0,707
robert	ROI	1	5	**80**	3	3	4	2	3	3	3	4	3	3	3,1	0,568
robert	ROI	2	0	**52**	5	5	5	5	5	5	5	5	5	5	5,0	0,000
robert	ROI	2	1	**72**	4	4	4	4	3	4	3	4	4	3	3,7	0,483
robert	ROI	2	2	**17**	4	4	4	4	3	4	4	4	4	4	3,9	0,316
robert	ROI	2	3	**10**	5	3	3	2	3	4	4	4	3	3	3,4	0,843
robert	ROI	2	4	**77**	4	4	4	3	3	3	3	4	3	4	3,5	0,527
robert	ROI	2	5	**88**	4	3	4	3	3	5	5	5	3	3	3,8	0,919
sara	ROI	1	0	**2**	5	5	5	5	5	4	4	4	5	5	4,7	0,483
sara	ROI	1	1	**55**	3	3	4	2	3	4	4	4	3	3	3,3	0,675
sara	ROI	1	2	**67**	4	4	3	3	3	4	4	4	5	3	3,7	0,675
sara	ROI	1	3	**8**	3	3	3	3	3	3	3	4	3	3	3,1	0,316
sara	ROI	1	4	**28**	4	3	4	4	3	4	4	4	3	3	3,6	0,516
sara	ROI	1	5	**66**	3	3	3	3	3	3	3	4	3	3	3,1	0,316
sara	ROI	2	0	**76**	5	5	5	5	5	5	4	4	5	5	4,8	0,422
sara	ROI	2	1	**46**	3	4	4	4	3	4	4	4	4	4	3,8	0,422
sara	ROI	2	2	**24**	5	4	4	4	4	4	4	4	4	4	4,1	0,316
sara	ROI	2	3	**26**	4	3	3	3	3	4	3	4	2	3	3,2	0,632
sara	ROI	2	4	**36**	4	4	4	3	4	4	4	4	4	4	3,9	0,316
sara	ROI	2	5	**25**	4	3	3	3	3	5	3	4	3	3	3,4	0,699
narator	ne-ROI	1	1	**51**	4	4	4	4	4	5	4	4	4	4	4,1	0,316
narator	ne-ROI	1	2	**71**	5	5	5	5	5	5	5	5	5	5	5,0	0,000
narator	ne-ROI	1	3	**4**	5	5	5	4	4	5	5	5	5	5	4,8	0,422
narator	ne-ROI	1	4	**39**	4	5	4	3	3	4	4	4	5	4	4,0	0,667
narator	ne-ROI	1	5	**3**	4	4	4	4	4	4	4	4	4	3	3,9	0,316

narator	ne-ROI	2	1	**5**	5	5	5	5	5	5	5	5	5	5	5,0	0,000
narator	ne-ROI	2	2	**60**	4	4	4	4	4	5	5	4	4	4	4,2	0,422
narator	ne-ROI	2	3	**12**	5	4	4	3	4	4	3	4	4	4	3,9	0,568
narator	ne-ROI	2	4	**15**	4	4	4	4	4	4	4	4	3	4	3,9	0,316
narator	ne-ROI	2	5	**69**	5	4	4	4	4	4	4	4	4	4	4,1	0,316
naratorka	ne-ROI	1	1	**35**	5	5	5	5	5	5	4	4	4	5	4,7	0,483
naratorka	ne-ROI	1	2	**37**	5	4	5	4	4	5	4	4	4	4	4,3	0,483
naratorka	ne-ROI	1	3	**23**	5	5	4	4	4	4	4	4	5	4	4,3	0,483
naratorka	ne-ROI	1	4	**7**	4	4	4	4	4	4	4	4	4	3	3,9	0,316
naratorka	ne-ROI	1	5	**38**	5	4	4	4	4	4	4	4	4	4	4,1	0,316
naratorka	ne-ROI	2	1	**61**	4	4	4	4	4	4	4	4	4	4	4,0	0,000
naratorka	ne-ROI	2	2	**49**	5	4	5	5	5	4	4	4	4	4	4,4	0,516
naratorka	ne-ROI	2	3	**54**	4	4	5	4	5	5	4	4	4	4	4,3	0,483
naratorka	ne-ROI	2	4	**53**	4	4	4	4	4	4	4	4	4	4	4,0	0,000
naratorka	ne-ROI	2	5	**14**	5	4	4	4	4	4	3	4	3	4	3,9	0,568
robert	ne-ROI	1	1	**27**	4	5	5	5	5	5	5	5	5	5	4,9	0,316
robert	ne-ROI	1	2	**18**	5	4	4	4	4	5	5	5	4	4	4,4	0,516
robert	ne-ROI	1	3	**44**	4	4	5	4	4	4	4	4	4	4	4,1	0,316
robert	ne-ROI	1	4	**87**	4	4	4	4	4	4	3	4	4	4	3,9	0,316
robert	ne-ROI	1	5	**1**	5	5	5	5	5	5	4	4	5	5	4,8	0,422
robert	ne-ROI	2	1	**81**	5	5	5	5	5	5	4	4	5	5	4,8	0,422
robert	ne-ROI	2	2	**16**	5	4	4	5	5	5	4	4	4	4	4,4	0,516
robert	ne-ROI	2	3	**58**	4	4	4	4	4	4	4	4	4	4	4,0	0,000
robert	ne-ROI	2	4	**82**	5	5	4	5	4	5	4	4	4	5	4,5	0,527
robert	ne-ROI	2	5	**75**	5	5	5	5	5	4	3	4	5	5	4,6	0,699
sara	ne-ROI	1	1	**56**	5	5	5	5	5	4	4	4	5	5	4,7	0,483
sara	ne-ROI	1	2	**6**	4	4	5	5	5	3	3	3	4	4	4,0	0,816
sara	ne-ROI	1	3	**13**	4	4	4	4	4	4	4	4	4	4	4,0	0,000
sara	ne-ROI	1	4	**70**	5	4	5	4	4	4	4	4	4	4	4,2	0,422
sara	ne-ROI	1	5	**57**	5	5	5	5	4	4	4	4	5	4	4,5	0,527
sara	ne-ROI	2	1	**85**	5	5	5	5	5	5	4	4	5	5	4,8	0,422

sara	ne-ROI	2	2	**47**	5	5	5	5	5	5	5	5	4	5	4,9	0,316
sara	ne-ROI	2	3	**45**	4	5	4	4	4	4	4	4	4	5	4,2	0,422
sara	ne-ROI	2	4	**74**	5	4	4	4	4	4	3	4	4	4	4,0	0,471
sara	ne-ROI	2	5	**48**	5	5	4	4	4	5	4	4	5	4	4,4	0,516

ŽIVLJENJEPIS

Ime in priimek: Tomaž Lovrenčič

Rojen:	13. 7. 1984	v Mariboru
Šolanje:	1991 – 1999	Osnovna šola Draga Kobala v Mariboru
	1999 – 2003	II. gimnazija v Mariboru
	2003 – 2009	dodiplomski študij na Fakulteti za elektrotehniko, računalništvo in informatiko, smer Telekomunikacije
	2009	diplomiral na Fakulteti za elektrotehniko, računalništvo in informatiko
	2009 – 2014	podiplomski študij na Fakulteti za elektrotehniko, računalništvo in informatiko v Mariboru, smer Elektrotehnika
Zaposlitev:	2010 – 2011	mladi raziskovalec iz gospodarstva pri podjetju Mobitel d.d.
	2011 – 2014	mladi raziskovalec iz gospodarstva pri podjetju Telekom Slovenije d.d.

Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko
Smetanova ulica 17
2000 Maribor, Slovenija

FERI

IZJAVA DOKTORSKEGA KANDIDATA

Podpisani Tomaž Lovrenčič,

vpisna številka E9500381.

izjavljam,

da je doktorska disertacija z naslovom
Vrednotenje kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih

- rezultat lastnega raziskovalnega dela,
- da predložena disertacija v celoti ali v delih ni bila predložena za pridobitev kakršnekoli izobrazbe po študijskem programu druge fakultete ali univerze,
- da so rezultati korektno navedeni in
- da nisem kršil-a avtorskih pravic in intelektualne lastnine drugih.

Podpis doktorskega kandidata:

Lovrenčič

Obrazec RŠZ

Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko
Smetanova ulica 17
2000 Maribor, Slovenija

FERI

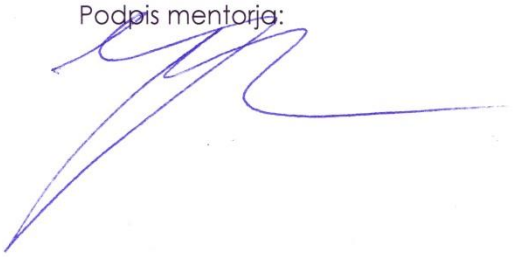
IZJAVA KANDIDATOVEGA MENTORJA O USTREZNOSTI DOKTORSKE DISERTACIJE

Podpisani-a izr. prof. dr. Andrej Žgank, mentor doktorskemu kandidatu,
izjavljam, da je doktorska disertacija z naslovom
Vrednotenje kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih,
ki jo je izdelal doktorski kandidat Tomaž Lovrenčič,
v skladu z odobreno temo, Pravilnikom o pripravi in zagovoru doktorske disertacije ter mojimi navodili in predstavlja izviren prispevek k razvoju znanstvene discipline.

Datum in kraj:

11. 11. 2014 v Mariboru

Podpis mentorja:



Obrazec RŠZ

Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko
Smetanova ulica 17
2000 Maribor, Slovenija

FERI

IZJAVA O OBJAVI ELEKTRONSKE VERZIJE DOKTORSKE DISERTACIJE IN OSEBNIH PODATKOV, VEZANIH NA ZAKLJUČEK ŠTUDIJA

Ime in priimek doktoranda: Tomaž Lovrenčič
Vpisna številka: E9500381
Študijski program: Elektrotehnika
Naslov doktorskega dela:
Vrednotenje kakovosti večmodalnih storitev v sodobnih telekomunikacijskih sistemih

Mentor: izr. prof. dr. Žgank Andrej
Somentor:

Podpisani soglašam z objavo doktorske disertacije v Digitalni knjižnici Univerze v Mariboru.

Tiskana verzija doktorske disertacije je istovetna elektronski verziji, ki sem jo oddal-a v Digitalno knjižnico Univerze v Mariboru.

Podpisani hkrati izjavljam, da dovoljujem objavo osebnih podatkov, vezanih na zaključek študija (ime, priimek, leto in kraj rojstva, datum diplomiranja, naslov diplomskega dela) na spletnih straneh in v publikacijah Univerze v Mariboru.

Datum in kraj: 11. 11. 2014 v Mariboru

Podpis doktoranda: Lovrenčič

Obrazec RŠZ