

# **Teză de doctorat**

**Implementarea unei metode de evaluare  
obiectivă a calității audio percepute,  
cu utilitate în radiodifuziune**

BIBLIOTECA UNIVERSITĂȚII  
Colecția 6th. 703  
369 E

**Conducător științific:**  
**Prof. dr. ing. Corneliu I. Toma**

**Autor:**  
**Ing. Konrad V. Pfaff**

**2004**

# Implementarea unei metode de evaluare obiectivă a calității audio percepute, cu utilitate în radiodifuziune

Teză de doctorat

## Cuprins

<b>1.</b>	<b>Introducere</b>	9
1.1	Aplicații în radiodifuziune ale semnalelor audio digitale	10
<b>1.2</b>	<b>Transmisia, stocarea și emisia semnalelor audio</b>	11
1.2.1	Canale de transmisie	12
1.2.2	Medii de stocare	12
1.2.3	Calitatea serviciului	13
<b>1.3</b>	<b>Internetul ca mediu de transmisie pentru emisiunile radio</b>	14
<b>2.</b>	<b>Psihoacustica</b>	19
<b>2.1</b>	<b>Structura și funcția sistemului auditiv</b>	19
2.1.1	Urechea externă	19
2.1.2	Urechea mijlocie	19
2.1.3	Urechea internă	20
<b>2.2</b>	<b>Psihoacustica</b>	23
2.2.1	Gama auditivă și audiograma	23
2.2.2	Perceperea înălțimii	25
2.2.3	Discriminarea frecvențelor	26
2.2.4	Benzile critice	27
2.2.5	Mascarea auditivă	31
2.2.5.1	Mascare simultană	32
2.2.5.2	Mascare temporală	36
2.2.6	Entropia perceptuală	37
2.2.7	Intensitatea sunetului	39
2.2.8	Audiția binaurală și localizarea sunetului	40
2.2.9	Modelul psihoacustic	41
<b>2.3</b>	<b>Concluzii</b>	44
<b>3.</b>	<b>Compresia semnalelor audio</b>	45
<b>3.1</b>	<b>Compresia semnalelor audio</b>	45
3.1.1	Debite binare	46
3.1.2	Calitatea semnalului	47
3.1.3	Timpu de prelucrare	47
3.1.4	Complexitatea	47

<b>3.2</b>	<b>Tehnici de codare audio</b>	47
3.2.1	Codarea predictivă	48
3.2.2	Codarea în subbenzi	52
3.2.3	Codarea prin transformări	55
3.2.4	Codare sinusoidale	57
<b>3.3</b>	<b>Algoritmi și standarde de codare</b>	61
3.3.1	Standardele de codare MPEG audio	62
3.3.2	Standardul G.722	68
3.3.3	Algoritmul apt-X100	69
3.3.4	Algoritmii Dolby AC-2 și AC-3	70
3.3.5	Tehnica de codare PASC	72
3.3.6	Algoritmul ATRAC	73
3.3.7	Sistemul SDDS	73
3.3.8	Algoritmii Lucent PAC, EPAC și MPAC	73
<b>3.4</b>	<b>Concluzii</b>	75
<b>4.</b>	<b>Metode de evaluare a calității audio</b>	77
<b>4.1</b>	<b>Introducere</b>	77
<b>4.2</b>	<b>Tipuri comune de artefacte</b>	78
4.2.1	Pierderea lățimii de bandă	78
4.2.2	Preecoul	78
4.2.3	Artefacte de alias	80
4.2.4	Artefacte în frecvență și variația lor în timp	81
4.2.5	Artefacte de reverberație vocală	82
4.2.6	Diferența binaurală de nivel de mascare	82
4.2.7	Pierderea imaginii stereofonice	83
4.2.8	Codarea în tandem	84
4.2.9	Audio prin codec vocal	85
<b>4.3</b>	<b>Metode de măsurare a calității audio</b>	85
4.3.1	Metode subiective de evaluare a calității	86
4.3.2	Tehnici simple de măsurare obiectivă	90
4.3.3	Tehnici de măsurare perceptuală	90
	A. Concepte de măsurare perceptuală	90
	B. Metode de măsurare perceptuală	95
<b>4.4</b>	<b>Aplicații posibile ale metodelor obiective de evaluare a calității audio</b>	98
4.4.1	Evaluarea implementărilor	99
4.4.2	Alinierea calității perceptuale	99
4.4.3	Monitorizarea on-line	99
4.4.4	Starea echipamentelor sau conexiunii	99
4.4.5	Identificarea codecurilor	99
4.4.6	Dezvoltarea codecurilor	99
4.4.7	Planificarea rețelelor	100
4.4.8	Ajutor în evaluarea subiectivă	100
<b>4.5</b>	<b>Concluzii</b>	100
<b>5.</b>	<b>Influența codărilor succesive asupra calității</b>	103
<b>5.1</b>	<b>Clasificarea codărilor audio succesive</b>	103
<b>5.2</b>	<b>Practici operaționale</b>	105
<b>5.3</b>	<b>Cauzele deteriorării calității audio</b>	106

5.3.1	Filtre	107
5.3.2	Cadre	107
5.3.3	Factorii care afectează codarea succesivă	107
5.3.4	Alocarea biților	108
5.3.5	Factori de scală	109
<b>5.4</b>	<b>Transcodarea prin modificarea debitului binar</b>	<b>109</b>
<b>6.</b>	<b>Metodă de evaluare obiectivă a calității audio percepute</b>	<b>111</b>
<b>6.1</b>	<b>Descrierea metodei</b>	<b>111</b>
<b>6.2</b>	<b>Modelul urechii periferice</b>	<b>112</b>
6.2.1	Generalități	112
6.2.2	Prelucrarea în domeniul timp	113
6.2.3	Transformarea FFT	114
6.2.4	Urechea externă și mijlocie	115
6.2.5	Gruparea în benzi critice	116
6.2.6	Adăugarea zgomotului intern	120
6.2.7	Împrăștierea	121
6.2.8	Împrăștierea în domeniul timp	122
6.2.9	Pragul de mascare	123
<b>6.3</b>	<b>Preprocesarea mărimilor de excitație</b>	<b>124</b>
6.3.1	Adaptarea nivelului și mărimilor	124
6.3.1.1	Adaptarea nivelului	125
6.3.1.2	Adaptarea modelului	125
6.3.2	Modulația	126
6.3.3	Tăria	127
6.3.4	Calculul semnalului de eroare	128
<b>6.4</b>	<b>Calculul variabilelor de ieșire ale modelului</b>	<b>128</b>
6.4.1	Generalități	128
6.4.2	Diferența de modulație	130
6.4.2.1	WinModDiff1	131
6.4.2.2	AvgModDiff1 și AvgModDiff2	131
6.4.3	Tăria zgomotului	132
6.4.4	Lățimea benzii	134
6.4.5	Raportul semnal-mascare	135
6.4.6	Relative Disturbed Frames	136
6.4.7	Probabilitatea de detecție	136
6.4.7.1	Probabilitatea maximă filtrată de detecție MFPD	138
6.4.7.2	Blocul mediu distorsionat ADB	138
6.4.8	Structura armonică a erorii	139
<b>6.5</b>	<b>Medierea</b>	<b>141</b>
6.5.1	Medierea spectrală	141
6.5.1.1	Media liniară	141
6.5.2	Media temporală	141
6.5.2.1	Media liniară	141
6.5.2.2	Media pătratică	141
6.5.2.3	Media în fereastră	142
6.5.2.4	Alegerea cadrului	142
6.5.2.4.1	Medierea cu întârziere	142
6.5.2.4.2	Pragul de tărie	142

6.5.2.4.3	Pragul energetic	143
6.5.2.4.4	Date limită	143
6.5.3	Medierea pe canale audio	144
<b>6.6</b>	<b>Estimarea calității audio percepute</b>	144
6.6.1	Rețeaua neuronală artificială	145
<b>6.7</b>	<b>Conformitatea implementării</b>	147
<b>6.8</b>	<b>Concluzii</b>	148
<b>7.</b>	<b>Rezultate experimentale</b>	149
<b>7.1</b>	<b>Evaluarea obiectivă a algoritmului ATRAC</b>	149
7.1.1	Scopuri experimentale	149
7.1.2	Metode experimentale	149
7.1.3	Testarea caracteristicilor de bază	149
7.1.3.1	Test 1	150
7.1.3.2	Test 2	153
7.1.3.3	Test 3	154
7.1.4	Evaluarea apariției preecoului	156
7.1.4.1	Test 4	156
7.1.5	Codări ATRAC succesive	157
<b>7.2</b>	<b>Evaluarea obiectivă a efectului frecvenței de eșantionare asupra calității audio percepute</b>	159
<b>7.3</b>	<b>Evaluarea obiectivă a efectului cuantizării asupra calității audio</b>	159
<b>7.4</b>	<b>Evaluarea obiectivă a calității audio pentru sistemele de codare folosite la transmisia prin Internet</b>	160
7.4.1	Obiectiv	160
7.4.2	Semnale de test	160
7.4.3	Rezultatele testelor efectuate	165
<b>7.5</b>	<b>Studiu comparativ asupra caracteristicilor unor codoare .mp3 uzuale</b>	176
7.5.1	Obiectiv	176
7.5.2	Semnale de test	176
7.5.3	Rezultate experimentale	176
<b>7.6</b>	<b>Evaluarea efectului codărilor succesive asupra calității audio</b>	179
7.6.1	Obiectiv	179
7.6.2	Semnale de test	179
7.6.3	Rezultate experimentale	180
<b>8.</b>	<b>Contribuții personale și concluzii</b>	183
<b>ANEXA</b>		187
<b>Bibliografie</b>		215

## Prescurtări utilizate în lucrare

<b>AAC</b>	<b>Advanced Audio Coding</b>
<b>ACR</b>	<b>absolute category rating</b>
<b>ADPCM</b>	<b>Adaptive Differential Pulse Code Modulation</b>
<b>ARCO</b>	<b>Adaptive Resolution Codec</b>
<b>ASAC</b>	<b>Analysis/Synthesis Audio Codec</b>
<b>ASPEC</b>	<b>Adaptive Spectral Entropy Coding of High Quality Music Signals</b>
<b>ATC</b>	<b>Adaptive Transform Coder</b>
<b>ATLANTIC</b>	<b>Advanced Television at Low Bitrates And Networked Transmission over Integrated Communication system</b>
<b>ATRAC</b>	<b>Adaptive TRansform Acoustic Coding</b>
<b>BAQ</b>	<b>Basic Audio Quality</b>
<b>BFU</b>	<b>Block Floating Unit</b>
<b>CB</b>	<b>critical bandwidth</b>
<b>CCR</b>	<b>comparison category rating</b>
<b>CELP</b>	<b>Code Excited Linear Prediction</b>
<b>CS-MSTVQ</b>	<b>constrained storage multistage tree-structural VQ</b>
<b>CS-VQ</b>	<b>constrained storage VQ</b>
<b>DAB</b>	<b>Digital Audio Broadcasting</b>
<b>DAR</b>	<b>Digital Audio Radio</b>
<b>DAT</b>	<b>Digital Audio Tape</b>
<b>DAVIC</b>	<b>Digital Audio-Visual Council</b>
<b>DBS</b>	<b>Direct Broadcast Satellite</b>
<b>DCC</b>	<b>Digital Compact Cassette</b>
<b>DPAC</b>	<b>differential perceptual audio coder</b>
<b>DPCM</b>	<b>Differential Pulse Code Modulation</b>
<b>DWPT</b>	<b>discrete wavelet packet transform</b>
<b>DWT</b>	<b>discrete wavelet transform</b>
<b>EBU</b>	<b>European Broadcasting Union</b>
<b>EPAC</b>	<b>enhanced PAC</b>
<b>ERB</b>	<b>equivalent rectangular bandwidth</b>
<b>FFT</b>	<b>fast Fourier transform</b>
<b>FIR</b>	<b>finite impulse response</b>
<b>HDTV</b>	<b>high definition TV</b>
<b>HILN</b>	<b>Harmonic Individual Lines plus Noise</b>
<b>IIR</b>	<b>infinite impulse response</b>
<b>IRT</b>	<b>Institut für Rundfunktechnik</b>
<b>ISDN</b>	<b>Integrated Services Digital Networks</b>
<b>ISO/IEC</b>	<b>International Standards Organization/International Electrotechnical Commission</b>
<b>JND</b>	<b>just noticeable distortion</b>
<b>JNDF</b>	<b>JND in frequency</b>
<b>KBD</b>	<b>Kaiser-Bessel derived window</b>
<b>LP</b>	<b>linear prediction</b>
<b>LPC</b>	<b>linear predictive coding</b>
<b>MASCAM</b>	<b>Masking Pattern Adapted Subband Coding</b>
<b>MDCT</b>	<b>modified discrete cosine transform</b>
<b>MLT</b>	<b>modulated lapped transform</b>
<b>MOS</b>	<b>mean opinion score</b>

<b>MPAC</b>	<b>multichannel PAC</b>
<b>MPE</b>	<b>mean perceptual error</b>
<b>MPEG</b>	<b>Moving Pictures Experts Group</b>
<b>MSC</b>	<b>multiple adaptive spectral audio coding</b>
<b>MSS</b>	<b>mean subjective score</b>
<b>MSTV</b>	<b>multistage tree-structural VQ</b>
<b>MT</b>	<b>Multi-Transform</b>
<b>MUSICAM</b>	<b>Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing</b>
<b>NADIB</b>	<b>narrow band audio digital broadcasting</b>
<b>NBC/AAC</b>	<b>Non-backward Compatible/AAC</b>
<b>NICAM</b>	<b>near-instantaneously companded audio multiplex</b>
<b>NLIVQ</b>	<b>nonlinear interpolative VQ</b>
<b>NMN</b>	<b>noise-masking-noise</b>
<b>NMR</b>	<b>noise-to-mask ratio</b>
<b>NMT</b>	<b>noise-masking-tone</b>
<b>OCF</b>	<b>Optimum Coding in the Frequency Domain</b>
<b>ODG</b>	<b>Objective Difference Grade</b>
<b>PAC</b>	<b>perceptual audio coding</b>
<b>PASC</b>	<b>Precision Adaptive Sub-band Coding</b>
<b>PCM</b>	<b>pulse code modulation</b>
<b>pdf</b>	<b>probability density function</b>
<b>PE</b>	<b>perceptual entropy</b>
<b>PNS</b>	<b>perceptual noise substitution</b>
<b>PQMF</b>	<b>pseudoquadrature mirror filter</b>
<b>PSD</b>	<b>power spectral density</b>
<b>PXFM</b>	<b>Perceptual Transform Coder</b>
<b>QMF</b>	<b>quadrature mirror filter</b>
<b>RL</b>	<b>run-length</b>
<b>RM</b>	<b>reference model</b>
<b>SACD</b>	<b>Super Audio CD</b>
<b>SAOL</b>	<b>structured audio orchestra language</b>
<b>SDDS</b>	<b>Sony Dynamic Digital Sound</b>
<b>SDG</b>	<b>Subjective Difference Grade</b>
<b>SFM</b>	<b>spectral flatness measure</b>
<b>SFPXFM</b>	<b>Stereophonic PXFM</b>
<b>SL</b>	<b>sensation level</b>
<b>SMR</b>	<b>signal-to-mask ratio</b>
<b>SNR</b>	<b>signal-to-noise ratio</b>
<b>SPIHT</b>	<b>set partitioning in hierarchical trees</b>
<b>SPL</b>	<b>sound pressure level</b>
<b>TDAC</b>	<b>time-domain aliasing cancellation</b>
<b>THD</b>	<b>Total Harmonic Distorsion</b>
<b>TMN</b>	<b>tone-masking-noise</b>
<b>TNS</b>	<b>temporal noise shaping</b>
<b>TTSI</b>	<b>text-to-speech interface</b>
<b>TWIN-VQ</b>	<b>transform-domain weighted interleave Vector Quantization</b>
<b>UEP</b>	<b>unequal error protection</b>
<b>VQ</b>	<b>Vector Quantization</b>
<b>WGN</b>	<b>white Gaussian noise</b>
<b>WLP</b>	<b>warped linear prediction</b>
<b>WP</b>	<b>Wavelet Packet</b>

# CAPITOLUL 1

## INTRODUCERE

Asistăm în ultimii ani la o creștere spectaculoasă a interesului pentru compresia de calitate în domeniul telefoniei și a semnalelor vocale și audio de bandă largă. Aceste clase de semnale diferă nu numai prin lărgimea de bandă și gama dinamică, ci și prin calitatea serviciului ce trebuie oferit utilizatorilor.

Compact discul (CD) este reprezentarea audio standard “de facto”, care stochează semnalele audio cu o bandă de frecvență între 10 – 20.000 Hz, în format PCM, pe 16 biți, având un debit binar de 1.411,2 kbit/s pentru un flux audio stereo. Pentru a transmite sau stoca semnalele audio pe canale de transmisie, respectiv medii de stocare, cu capacitate limitată sunt necesare debite binare mai reduse. Primele propuneri de reducere a debitului de date audio PCM au folosit aceleași principii ca și pentru codarea vocală. Totuși, există mari diferențe între codarea semnalelor audio și codarea semnalelor vocale, deoarece păstrarea calității unui semnal audio implică frecvențe de eșantionare mai mari, o gamă dinamică mai largă (nr. biți/eșantion), mari variații ale densității spectrale de putere, diferențe ale mecanismului de percepție auditivă uman și mai ales pretențiile ridicate ale beneficiarului (ascultătorului) de la capătul lanțului de transmisie/stocare. Codarea semnalelor vocale este eficientă deoarece beneficiază de un model bine pus la punct, ceea ce nu este cazul pentru un semnal audio complex. De asemenea, spre deosebire de semnalele vocale, în cazul semnalelor audio avem de-a face cu reprezentări stereo sau multicanal.

Lucrarea are ca scop implementarea unei metode obiective de evaluare a calității audio percepute, cu grad mare de utilitate în cadrul unei rețele de radiodifuziune.

În acest scop sunt prezentate și analizate metodele moderne de compresie a semnalelor audio, bazate pe principii psihoacustice, utilizate în radiodifuziune, televiziune și multimedia, precum și artefactele de codare specifice și metodele de evaluare a calității audio. Finalitatea acesteia este implementarea unei metode obiective de evaluare a calității audio.

Spre deosebire de domeniul codării vocale, susținut în special de societățile de telefonie mobilă, codarea audio perceptuală, și în special evaluarea calității audio percepute, a fost puțin abordată în țara noastră.

Codarea semnalelor vocale și audio a avut o evoluție spectaculoasă în ultimii ani, ea bazându-se pe înțelegerea aprofundată a mecanismului auditiv uman și pe principii psihoacustice, prezentate în capitolul 2.

Un grad înalt de reducere a datelor se poate obține numai prin folosirea tehnicilor de codare perceptuală, controlate de criterii psihoacustice, prin care sunt transmise numai părțile din semnalul audio percepute de urechea umană. Predicția liniară, codarea în subbenzi, codarea prin transformări, precum și diferite forme de cuantizare vectorială și codarea entropiei sunt tehnici prezentate sintetic în capitolul 3, care au dus la proiectarea unor algoritmi de codare foarte eficienți. O contribuție importantă la codarea informației audio-vizuale o are ISO/MPEG (*International Organization of Standardization/Moving Pictures Expert Group*) prin standardele internaționale din domeniul compresiei de înaltă calitate a semnalelor audio digitale, prezentate de asemenea în capitolul 3.



Pentru a evalua algoritmi de codare audio de înaltă calitate pentru aplicații în radiodifuziune și telecomunicații am avut în vedere mutațiile apărute în domeniul transmisiei, stocării și difuzării semnalelor audio. O atenție deosebită am acordat metodelor de evaluare a calității audio percepute, atât subiective, cât și obiective, prezentate în capitolul 4. Am prezentat într-un subcapitol separat artefactele specifice de codare, a căror descriere să permită unui ascultător expert sau unui tehnician audio să facă o evaluare rapidă a calității audio a semnalului codat.

În drumul său de la sursă și până la emițător, programul audio poate suferi mai multe codări succesive, prelucrări sau transcodări în diferite formate digitale. În capitolul 5 am analizat efectele, dimensiunea și natura distorsiunilor introduse de conversia formatelor digitale, de codările succesive multiple, de operațiile de pre- și post-procesare.

Metoda de evaluare a calității audio percepute, propusă de autor în capitolul 6, are ca punct de plecare versiunea de bază a standardului ITU-R BS.1387-1, aducându-i precizări și interpretări alternative în vederea unei implementări clare și eficiente. Am realizat o implementare experimentală în MATCAD, cu ajutorul căreia am realizat diferite evaluări ale calității audio.

În capitolul 7 sunt prezentate rezultatele experimentale privind evaluarea calității unor algoritmi de codare audio, obținute cu metode obiective:

- evaluarea obiectivă a algoritmului ATRAC;
- evaluarea obiectivă a efectului frecvenței de eșantionare asupra calității audio percepute;
- evaluarea obiectivă a efectului cuantizării asupra calității audio percepute;
- evaluarea obiectivă a calității audio pentru sistemele de codare folosite la transmisia prin Internet;
- studiu comparativ asupra caracteristicilor unor codoare .mp3 uzuale;
- evaluarea efectului codărilor succesive asupra calității audio.

În afară de evaluarea algoritmului ATRAC, toate celelalte rezultate experimentale au fost obținute cu ajutorul metodei de evaluare a calității audio percepute propusă de autor.

Problemele care apar în evaluarea calității audio percepute pentru diferite combinații posibile de scheme de codare și debite binare au fost identificate de autor în urma contactelor directe cu mai multe studiouri de radiodifuziune locale, regionale și naționale din țară și străinătate (Danemarca, Germania, Italia, Ungaria) și în urma participării la diferite conferințe, sesiuni de comunicări științifice (unde am prezentat și rezultatele cercetărilor științifice elaborate de mine) și expoziții de specialitate (*Tonmeistertagung* – Germania, Karlsruhe-1998, Hanovra-2002 și cea de-a 106 *Conferință Internațională a AES – Audio Engineers Society*, München 1999).

Capitolul 8 prezintă o sinteză a tezei, a contribuțiilor personale și a rezultatelor obținute. Experiența acumulată de-a lungul anilor, prin utilizarea unor metode diferite de codare audio, la diferite debite binare și niveluri de calitate, mi-a permis să realizez o comparație a performanțelor acestora, iar concluziile rezultate m-au determinat să propun un mod de organizare a circulației și prelucrării fișierelor audio, care să mențină la un nivel acceptabil calitatea audio a unui material audio pe tot parcursul său, de la contribuție și până la emisie. Am prezentat posibilitățile ulterioare de utilizare și dezvoltare a metodei obiective de evaluare a calității audio în cadrul posturilor de radio și a altor instituții specializate în transportul semnalelor audio.

## 1.1 Aplicații în radiodifuziune ale semnalelor audio digitale

Aplicațiile tipice pentru semnalele audio digitale sunt: producția audio, distribuția și schimbul de programe, emisia digitală DAB (*Digital Audio Broadcasting*), stocarea digitală (arhive, studiouri sau amatori) și chiar comunicațiile interpersonale, precum video-conferința,

aplicațiile multimedia sau televiziunea de înaltă definiție. Codecurile audio cu debit binar redus sunt larg utilizate pentru transmisii în timp real sau transferul materialelor audio de la un punct la altul prin rețele de telecomunicații terestre, prin satelit sau rețele digitale interne.

Uniunea Internațională de Telecomunicații - secțiunea de radiocomunicații ITU-R (*International Telecommunications Union Radiocommunication Sector*) a definit următoarele categorii de aplicații pentru codecurile cu debit binar redus (dbr) utilizate în domeniul radiodifuziunii [BEA94]:

**Studiou** - înregistrări proprii, în studiou, editări și arhivarea sunetului. În acest domeniu codarea cu dbr nu este prea larg acceptată, în ciuda creșterii raportului randament/calitate față de reprezentarea liniară PCM. Această reținere este explicabilă datorită fenomenelor ireversibile care se produc la înregistrare. Metodele de compresie a datelor, matematic fără pierderi, sunt acceptate de inginerii de sunet, dar nu oferă același grad de reducere a datelor ca metodele de reducere a datelor bazate pe principii psihoacustice.

**Contribuție** - transmiterea materialului sonor de la studiou la studiou sau de la sursă la studiou păstrând o margine de codare suplimentară pentru codări succesive sau prelucrarea (post-procesarea) semnalului audio. Codecurile de contribuție sunt necesare pentru a menține calitatea de studiou a semnalului la un debit de 180 kbit/s/canal. Mai precis, semnalul audio nu trebuie să sufere nici o modificare subiectivă de calitate la trecerea prin cinci codecuri succesive cu debitul de 180 kbit/s/canal.

**Distribuție** - liniile de transmisie a semnalului de la studiou la emițător sau distribuția programelor la sau de la studiouri regionale sau locale. Rolul codecurilor de distribuție este de a permite transmisii punct-la-punct cu margini de codare reduse pentru prelucrare sau codări succesive. Calitatea unui semnal audio trecut prin trei codecuri de distribuție succesive la un debit de 120 kbit/s/canal nu trebuie să scadă cu mai mult de 0,5 puncte față de semnalul original pe scara subiectivă ITU-R de apreciere a calității audio, cu 5 trepte.

**Emisie** - relația emițător - consumator în cadrul radiodifuziunii digitale (DSB - *Digital Signal Broadcasting*). Codecurile de emisie trebuie să permită transmiterea economică într-un singur sens a unui semnal audio de calitate la un debit de 128 kbit/s/canal, fără să includă vreo margine de codare pentru postprocesare sau codări succesive. Calitatea subiectivă a semnalului audio trecut printr-un codec de emisie nu trebuie să fie sub 4 (perceptibil, dar nederanjant) pe scara subiectivă ITU-R de apreciere a calității audio, cu 5 trepte.

**Comentariu** - știri sau comentarii asupra conținutului unui program. Codecul de comentariu trebuie să permită transmisii vocale sau muzicale de calitate la debite binare reduse specifice canalelor ISDN (60 kbit/s/canal). Se cere o lărgime de bandă de 15 kHz și sunt permise deteriorări perceptibile ale semnalelor muzicale.

## 1.2 Transmisia, stocarea și emisia semnalelor audio

Deși canalele și rețelele de mare viteză sunt din ce în ce mai răspândite și accesibile, codarea cu debit binar redus (dbr) are o deosebită importanță pentru a minimiza costurile de transmisie sau stocare și pentru a permite transmisia prin canale de capacitate limitată. Mai mult, s-ar putea ca un canal de o anumită capacitate să fie împărțit pentru diferite servicii (voce, audio, date, grafică, imagini) într-o rețea de servicii integrate. Condițiile principale de care trebuie ținut cont în proiectarea codoarelor cu dbr sunt:

- menținerea calității semnalului audio reconstruit la variații ale componentei spectrale și ale nivelului;
- rezistența la erori aleatoare de canal și pierderi de pachete;
- complexitate redusă și consum mic de energie;
- întârzieri mici în procesul de codare/decodare pentru aplicații în timp real (transmisii directe);

- degradarea cât mai redusă a calității audio în cazul codărilor multiple (tandem, transcodări).

Toate aceste condiții, dintre care unele antagoniste, trebuie luate în considerare la alegerea algoritmului de codare potrivit pentru o anumită aplicație.

### 1.2.1 Canale de transmisie

Transmisia semnalelor vocale și audio este posibilă prin diferite sisteme digitale terestre sau prin satelit. Un exemplu îl reprezintă sistemul ISDN (*Integrated Services Digital Networks*) prin care utilizatorii au acces, cu costuri relativ reduse, la canale de telecomunicații digitale comutate. Tabelul 1.1 prezintă câteva canale puse la dispoziție de ISDN.

**Tabelul 1.1** Canale ISDN care pot fi utilizate în codarea audio

Canal B	64 kbit/s	B
Canal D	16 kbit/s	D (date)
Canal H0	384 kbit/s	6 x B
Canal H11	1.536 kbit/s	(=4 H0) SUA
Canal H12	1.920 kbit/s	(=5 H0) Europa

Interfața ISDN de bază constă din două canale B și un canal D, care permite distribuția unor semnale audio stereofonice sau multicanal împreună cu alte informații utile.

O mare parte din materialele audio care până acum erau transmise folosind linii dedicate, analogice sau digitale, de bandă largă, se pot transmite acum pe linii digitale comutate, cu preț scăzut. Acest lucru se datorează tehnologiilor de codare audio cu dbr, transmisiei pe linii ISDN sau prin Internet. Un exemplu concret îl poate constitui transmiterea spoturilor publicitare de la un studio la altul sau de la realizator la beneficiar, codate MPEG Nivel-3 și transmise sub formă de e-mail prin Internet.

### 1.2.2 Medii de stocare

Compact discul (CD) și banda audio digitală (DAT) permit stocarea unor semnale audio cu frecvențe între 10 și 20.000 Hz, cu o rezoluție de 16 biți/eșantion. Pentru reprezentarea unui eșantion audio de 16 biți, CD-ul folosește o reprezentare pe 49 de biți, rezultând un debit binar total de  $1,41 \times 49/16 = 4,32$  Mbit/s. În tabelul 1.2 sunt prezentate debitele binare pentru CD, DAT și alte medii.

**Tabel 1.2** Debitele binare pentru diferite medii de stocare (semnale stereofonice eșantionate la 44,1 kHz; de remarcat că sistemul DAT suportă și ratele de eșantionare de 32 și 48 kHz)

Dispozitiv de stocare	Format	Debit audio		Debit total
Compact disc (CD)	PCM	1,41 MBit/s	2,91 Mbit/s	4,32 Mbit/s
Banda audio digitală (DAT)	PCM	1,41 Mbit/s	1,67 Mbit/s	3,08 Mbit/s
Caseta compactă digitală (DCC)	PASC	384 kbit/s	384 kbit/s	768 kbit/s
MiniDisc (MD)	ATRAC	292 kbit/s	718 kbit/s	1,01 Mbit/s

Alte medii digitale pentru stocarea informației audiovizuale sunt discurile magneto-optice (MOD), CD-ROM-ul, CD-ul interactiv sau caseta digitală compactă (DCC) produsă de Philips (vezi §3.3.5), care utilizează algoritmul PASC (*Precision Adaptive Sub-band Coding*). Un alt exemplu de mediu digital pentru stocarea comprimată a informației audio este minidiscul (MD). Este un suport optic sau magneto-optic, creat de firma SONY, cu diametrul de 64mm, care folosește algoritmul ATRAC (*Adaptive TRansform Acoustic Coding*) pentru obținerea unui debit

de 292 kbit/s (vezi §3.3.6). În sfârșit, există diferite dispozitive de înregistrare și redare a sunetului (în special pentru semnale vocale) cu cartele de memorie EPROM.

### 1.2.3 Calitatea serviciului

Calitatea serviciului în sistemele de transmisie digitale este, în general, mult mai constantă decât în cazul sistemelor analogice cu modulație de amplitudine sau modulație de frecvență. Într-un sistem de transmisie digitală se folosesc două niveluri de codare, a canalului și a sursei. Codarea canalului permite un canal de date foarte robust și transparent. Măsurile de corecție și de protecție împotriva erorilor sunt adaptate condițiilor specifice de transmisie, precum receptoare mobile, în mișcare și prezența unor ecouri puternice. Codarea sursei este adaptată naturii datelor, adică semnalul audio și are ca scop reducerea capacității necesare a canalului fără o degradare majoră a calității audio. Protecția și corecția erorilor nu au un rol important la proiectarea metodelor de codare a sursei.

Calitatea audio rezultată depinde numai de codarea audio, deoarece nu există o degradare datorată erorilor. De aceea, calitatea audio într-un sistem DAB este determinată de algoritmul de codare a sursei, implementarea și parametrii operaționali: debit binar, frecvența de eșantionare și modul de codare.

În cazul transmisiei analogice, orice modificare a canalului de transmisie de la parametrii optimi influențează direct calitatea audio. Efectele tipice sunt interferențele dintre stații, fading-ul datorat reflecțiilor în clădiri și zgomotele datorate perturbațiilor industriale (motoare, sisteme de aprindere, sisteme electronice de comandă etc.). Aceste perturbații au devenit atât de comune încât câteodată nici nu le mai remarcăm. În general, nu există soluții împotriva acestor perturbații.

Într-un sistem de transmisie digital, separarea clară dintre codarea canalului și codarea sursei permite reducerea la minim sau chiar eliminarea acestor perturbații. Protecția la erori a codării canalului este atât de puternică încât furnizează un canal transparent pentru datele sursei în aproape orice condiții. Modificările în timp ale căii de transmisie nu afectează datele sursei, decât în cazul întreruperii codării canalului, ceea ce se întâmplă destul de rar.

În concluzie, calitatea audio în cazul unui sistem de transmisie digitală este determinată de parametrii schemei de codare a sursei și, în general, independentă de caracteristicile căii de transmisie. De aici rezultă că, și în cazul unei căi de transmisie perfecte, calitatea audio poate fi limitată dacă parametrii operaționali nu sunt aleși corespunzător [HOE03].

În epoca transmisiilor digitale, calitatea audio nu mai este definită prin măsurători obiective (raport semnal-zgomot, liniaritate, distorsiuni și bandă de frecvențe). În cazul codării audio perceptuale, calitatea audio se definește prin diferențele percepute între un semnal de referință și semnalul codat transmis ascultătorilor. Metodele tradiționale de măsurare a calității furnizează informații limitate asupra calității audio reale și pot conduce la concluzii eronate. Din această cauză au fost dezvoltate noi metode, subiective și obiective, de evaluare a calității audio (cap. 4).

Și calitatea referinței este relativă. În ultimii 10 ani calitatea înregistrărilor audio de pe CD este considerată drept referință, în timp ce cu 20 de ani în urmă discurile de vinil se bucurau de această calitate. Odată cu introducerea noilor formate de calitate superioară, Super Audio CD (SACD) și DVD-Audio, probabil că referința se va schimba din nou. În același timp, mulți ascultători consideră perfect acceptabilă calitatea audio obținută cu diverse metode de compresie a datelor audio, ca de exemplu MPEG-2 nivel-3, cunoscută sub numele de MP3, deși calitatea audio percepută este sub a standardului CD. De aceea, este evident că orice măsurare a calității audio trebuie să se bazeze pe diferența dintre semnalul original necompresat și versiunea codată.

Degradările introduse de un sistem de codare, numite și artefacte de codare, includ distorsiuni liniare, zgomot de cuantizare, pre-ecou, limitarea benzii de frecvențe, componente de alias, modificarea imaginii stereofonice și a timbrului. Ele vor fi prezentate mai detaliat în §4.2.

### 1.3 Internetul ca mediu de transmisie pentru emisiunile radio

În ultimii 10 ani, tehnologiile de codare audio au făcut progrese enorme, fiind folosite cu succes în transmisii radio, medii de stocare (CD, MiniDisc, CD-ROM și DVD) și în Internet. S-au făcut progrese semnificative pentru reducerea debitului binar și îmbunătățirea constantă a calității sunetului (vocal sau muzical) transmis. Cele mai mari progrese în codarea audio cu debit binar redus (dbr) au avut loc însă în ultimii 10 ani, datorită dezvoltării rapide a Internetului, unde sunt necesare debite binare extrem de scăzute și menținerea calității subiective a semnalului original. La rândul său, dezvoltarea Internetului a avut loc datorită creșterii vitezei de acces, în special pentru utilizatorii casnici. La ora actuală se poate considera că nivelul minim de acces este printr-un modem de 28,8 kbit/s, având ca medie 33 – 56 kbit/s pentru liniile analogice comutate și 64 kbit/s pentru liniile ISDN. Sunt deja accesibile și alte forme de legături de bandă largă: cablu, satelit sau telefonie mobilă (GPRS, CDMA).

Răspândirea utilizării Internetului ca mediu de difuzare a Radioului a dus la apariția unor noi termeni, ca de exemplu “Webcasting”, care înseamnă, într-o traducere adaptată, difuzare prin Internet (*World Wide Web*) și are două componente: producția și distribuția informației digitale multimedia către unul sau mai mulți receptori, către o anumită audiență sau comunitate. Conținutul poate fi de diferite tipuri: știri, informații economice sau divertisment, iar formatul poate fi text, grafică, audio sau video, la care se pot adăuga diverse facilități interactive. Există trei tipuri de tehnologii pentru Webcasting:

- *Streaming Technology* – este procesul de transmitere on-line a semnalelor audio/video, care se bazează pe fluxuri (*stream*) continue de informație. Această tehnologie permite redarea aproape instantanee a informației audio/video.

- *Downloading Technology* – este procesul prin care informația este descărcată în întregime și stocată înainte de a putea fi redată. Necesită un spațiu mare de stocare.

- *Push/Pull Technology* – este procesul prin care informația audio, video sau text sau orice combinație între cele trei ajunge prin Internet la receptor. Informația a fost cerută de receptor sau trimisă de altcineva care a considerat că receptorul trebuie să o aibe.

Multe stații de radio se află într-o perioadă de tranziție spre era digitală, care va adăuga o latură multimedia componentei de distribuție a radioului. Având în vedere că peste 5000 de stații de radio din lumea întreagă sunt conectate într-un fel sau altul la World Wide Web, este necesară o analiză cât mai profundă a modului de interacțiune dintre radio și Internet. Transportul fișierelor audio prin Internet a devenit o modalitate uzuală de contribuție la producția de programe. Materialele audio nu mai sunt recepționate doar prin canalele de comunicații terestre sau prin satelit, ci sunt descărcate sau recepționate în timp real prin Internet.

Poziția Internetului față de posturile de radio s-a schimbat, de la funcțiunea inițială de cutie poștală pentru informația transmisă, până la un nou mediu de transmisie, care poate atrage noi clienți prin noi concepte interactive. Tehnologiile de transmisie audio și video prin Internet reprezintă o nouă cale pentru posturile de radio de a-și mări oferta față de public. Deocamdată, majoritatea posturilor de radio oferă servicii bazate pe programele existente.

Motivele pentru care un post de radio trebuie să-și îndrepte atenția către Internet sunt:

- distribuție nelimitată din punct de vedere geografic;
- știrile și informațiile pot fi distribuite instantaneu sau la cerere;
- chiar și posturile mici de radio, cu resurse financiare limitate pot fi recepționate în toată lumea;
- se pot crea și distribui mai eficient emisiuni destinate unui grup țintă de audiență;
- tehnologia push/pull permite personalizarea informației în funcție de sfera de interes.

O categorie importantă de ascultători ai radioului prin Internet o reprezintă funcționarii, care, în timpul programului de lucru la birou, dispun de conexiuni bune la Internet, iar recepția programelor radiodifuzate devine din ce în ce mai dificilă datorită ecranării clădirilor și interferențelor produse de tehnica de calcul și iluminatul fluorescent. Expatriații reprezintă o altă categorie de ascultători de radio prin Internet. Ei pot asculta posturi de radio din țările lor, pentru știri sau divertisment în limba maternă. În mod tradițional acesta era rolul programelor pentru străinătate emise pe unde scurte [PFA03].

Majoritatea diferențelor dintre radiodifuziunea tradițională și cea prin Internet sunt de ordin tehnologic. Pentru ca transmisia prin Internet să poată deservi mai mulți ascultători sunt necesare calculatoare mai mari. Aceasta se poate compara cu un emițător mai mare în cazul radiodifuziunii tradiționale, care ar acoperi o arie geografică mai mare. Cu cât lățimea benzii semnalului este mai mare sau cu creșterea numărului de ascultători, cu atât trebuie să crească și puterea de calcul a serverului pentru a transmite mai multe fluxuri de semnal în același timp.

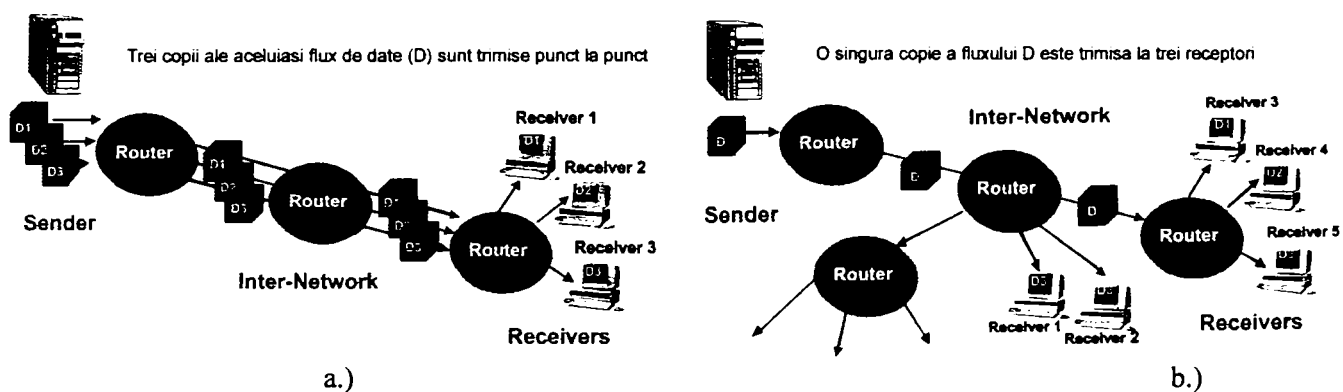


Figura 1.1 Principiul de transmitere Unicast (a) și Multicast (b)

Transmiterea informației audio în timp real dintr-un punct în mai multe direcții necesită o anumită lățime de banda continuă a rețelei. Aceasta este una din problemele majore în multe părți ale lumii pentru Webcasting. În prezent, majoritatea transferurilor în Internet se fac prin conexiuni punct-la-punct, procedeu numit “unicasting” (fig.1.1.a). Aceasta înseamnă că dacă 200 de oameni doresc să asculte același program simultan, serverul postului trebuie să furnizeze același flux de 200 de ori. Evident că aceste fluxuri încarcă întreaga rețea de la emițător la receptor și nu este un mod eficient de a transmite un mesaj prin Internet. De aceea s-a dezvoltat un protocol numit “multicasting” care are posibilitatea de a lua un singur semnal și de a-l transmite la mai mulți receptori prin rețea. Semnalul își caută o cale prin rețea până la un router sau un splitter care îl transmite mai departe către mai multe terminale, ca în fig.1.1.b. Aceasta înseamnă că fiecare ascultător trimite o solicitare de recepție a programului, apoi clientul și serverul negociază pentru a găsi calea cea mai ușoară (de obicei calea cea mai eficientă din punct de vedere al capacității) pentru a trimite fluxul. Este ca și în cazul radiodifuziunii tradiționale când un singur semnal poate fi recepționat de mai multe persoane, cu condiția să se afle în raza de acoperire a emițătorului [WPG99].

Din păcate Internetul este încă foarte îngust, necesitând rate de compresie foarte mari, de exemplu, în cazul unui modem de 28,8 kbit/s, rata de transfer a unui program audio este de 50 de ori mai mică decât în cazul CD-ului. Aceasta înseamnă că, înainte de a fi transmis prin Internet, semnalul audio trebuie supus unui procedeu sofisticat de compresie. Aceste tehnici de compresie audio care permit un raport de compresie mare păstrând totuși o calitate audio acceptabilă, se dezvoltă continuu, găsindu-și aplicarea în Internet.

Radioul prin Internet necesită o tehnologie de transmitere a fluxului de informație audio în timp real sau la cerere. Mai mulți ascultători (până la ordinul miilor) pot asculta același program

în același timp. Pentru aceasta sunt necesare programe speciale de transmisie, inclusiv protocoale de transfer adecvate, precum și un program de redare la locul de recepție. În comparație cu descărcarea informației audio, transmisia audio este mai critică, datorită emisie și recepției în timp real a informației audio.

Tabelul 1.3 prezintă aprecierea subiectivă a calității sunetului pentru diferite fluxuri și caracteristicile lor.

Tabel 1.3 Aprecierea subiectivă a calității audio în funcție de debitul binar

Calitatea sunetului	Banda de frecvențe kHz	Mod	Debit kbit/s	Raport de compresie
sunet telefonic	2,5	mono	8	96:1
unde scurte	4,5	mono	16	48:1
radio MA	7,5	mono	32	24:1
radio MF	11	stereo	56-64	26-24:1
aproape CD	15	stereo	96	16:1
CD	>15	stereo	128	14-12:1

Operațiunea standard de transfer a unui fișier audio (de ex. în format “.wav”) este mai puțin pretențioasă, este nevoie numai de un codor audio potrivit care să reducă informația astfel încât costurile de transfer să fie rezonabile. Informația audio trebuie comprimată și codată astfel încât să poată fi utilizată de decodor. Se alege un protocol de transmisie cu corecția erorilor, deoarece fișierul audio trebuie transferat complet înainte de redare. Astfel, transferul informației audio nu depinde în mod deosebit de un anumit debit binar pentru materialul audio codat și beneficiază de corecția erorilor din cadrul protocolului de transport ales, încât, chiar și cu un modem de 28,8 kbit/s se poate obține un fișier audio de o calitate suficientă pentru a fi introdus într-un program de radio. Dezavantajul acestui procedeu este durata ridicată de transfer a materialului și dacă în timpul transferului materialului apare vreun incident, operația trebuie reluată de la capăt.

Aplicațiile tipice de radio în Internet care permit transmiterea unor evenimente în timp real (“live”) necesită o tehnologie de transport a fluxurilor audio (*streaming audio*). Dacă lățimea benzii rețelei nu este foarte limitată datorită traficului și dacă ascultătorul este dispus să accepte o calitate a sunetului sub cea a CD-ului, se poate spune că programele de transmitere a sunetului prin Internet funcționează destul de bine. Totuși, stabilitatea unei transmisii audio live prin Internet este destul de fragilă datorită variațiilor mari de trafic. Majoritatea programelor de streaming-audio sunt funcționale începând cu o viteză de cel puțin 28,8 kbit/s.

Lărgimea benzii unui sistem de streaming-audio reprezintă lărgimea de bandă maximă care se poate obține în cazul unei conexiuni în rețea, de exemplu cu ajutorul unui modem de 28,8 kbit/s. Debitul binar utilizabil pentru o transmisie audio va fi mai mic decât debitul maxim al conexiunii. De exemplu, în cazul unei conexiuni prin intermediul unui modem de 28,8 kbit/s, pentru o transmisie audio putem conta pe o lărgime de bandă corespunzătoare unui debit de aproximativ 75% din cel nominal, adică, în acest caz putem dispune de aprox. 20 kbit/s. Debitul binar total cuprinde două părți:

- debitul binar maxim utilizat pentru transmisia audio;
- debitul necesar pentru prevenirea problemelor de transmisie, pierderi de pachete etc., care este de aprox. 25% din debitul total; el depinde de tipul conexiunii și de încărcarea rețelei.

Debitele binare maxime disponibile pentru o transmisie audio sunt prezentate în tabelul 1.4.

Toate programele de streaming-audio folosesc codoare pentru compresia semnalului audio pentru a putea transporta mari cantități de date și a crea fișiere audio compacte. Debitul binar în cazul unui CD audio este de 1,5 Mbit/s, iar în cazul transmisiei printr-un modem de 28,8 kbit/s debitul disponibil este de aprox. 50 ori mai mic, fiind necesară o compresie corespunzătoare a datelor audio. Mai mult, modemurile pentru linii analogice nu ating niciodată parametrii maximi,

datorită zgomotului și altor distorsiuni. De aceea, în cazul transmisiei printr-un modem de 28,8 kbit/s, codarea audio trebuie să se facă cu un raport de compresie mai mare de 50:1. În cazul unei transmisii pe linii ISDN printr-un canal B raportul de compresie este 24:1, iar în cazul utilizării ambelor canale este de 12:1.

Tabelul 1.4 Debitul binare disponibile pentru transmiterea sunetului prin Internet

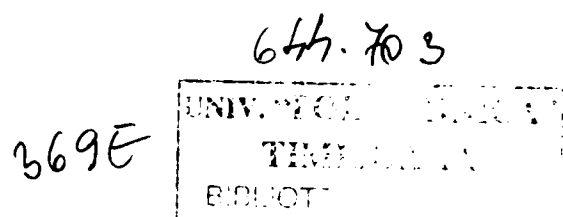
Viteza maximă de conectare	Debitul maxim recomandat pt. streaming audio
14,4 kbit/s Modem	10 kbit/s
28,8 kbit/s Modem	20 kbit/s
56,0 kbit/s Modem	32 kbit/s
56 kbit/s ISDN	45 kbit/s
64,0 kbit/s ISDN	56 kbit/s
112 kbit/s ISDN	80 kbit/s
128 kbit/s ISDN	96 kbit/s

Majoritatea codoarelor audio permit utilizatorului să dimensioneze fișierele codate pentru un anumit tip de audiență. Există optimizări specifice pentru modemi analogice, linii ISDN sau modemi de cablu de mare viteză, precum și în funcție de conținutul materialului audio.

În capitolul 7 am efectuat o evaluare obiectivă a calității audio pentru sistemele de codare folosite la transmisia prin Internet, în funcție de debitele binare disponibile.

Deoarece încărcarea rețelei variază în funcție de numărul utilizatorilor din rețea și debitul binar disponibil pentru o transmisie audio variază semnificativ. Pentru o recepție optimă a materialului audio, serverul trebuie să se adapteze în mod dinamic pentru a trimite în orice moment fluxul maxim posibil. Această adaptare dinamică este posibilă folosind rapoartele generate de programul de redare (*media player*) în cadrul protocolului de control în timp real RTCP (*Real-Time Control Protocol*), prin care se măsoară încărcarea rețelei și se transmite varianta optimă corespunzătoare fluxului disponibil. Există două tehnologii care permit codarea combinată cu diferite debite binare într-un singur fișier media: prima se numește “*Sure Stream*” de la Real Networks, iar a doua este “*Intelligent Streaming*” de la Windows Media. La începutul redării, serverul și playerul negociază pentru a alege debitul cel mai potrivit. În timpul redării, playerul și serverul comunică în mod repetat pentru a comuta, dacă este nevoie, între variantele corespunzătoare diferitelor debite binare [KOZ02].

În final, trebuie ținut cont de faptul că Internetul nu a fost conceput de creatorii săi ca un mediu de transmisie audio în timp real, ci ca un mediu de transport al fișierelor și al poștei electronice. Cu tehnologia actuală, nu putem avea garanția unor transmisii audio neîntrerupte, problemă pe care toți ascultătorii radioului prin Internet au întâmpinat-o. De cele mai multe ori recepția unui program radio prin Internet este fragmentată și nu putem ști exact cine este de vină: serverul de pe care este transmis programul, vreunul din numeroasele noduri pe care semnalul trebuie să le traverseze, furnizorul local de Internet, rețeaua interioară sau chiar propriul calculator, care este receptorul final. La un anumit moment dat, cauza întreruperilor de recepție poate oricare sau orice combinație dintre cauzele enumerate mai sus [THT98].







# CAPITOLUL 2

## PSIHOACUSTICA

### 2.1 Structura și funcția sistemului auditiv

Aparatul auditiv uman captează undele sonore, le analizează, le sortează, le clasifică după frecvență (senzația de înălțime), după intensitate (senzația de tărie), după compoziția spectrală (senzația de timbru) și transmite informația auditivă către sistemul nervos central.

Urechea, compusă din trei părți - *urechea externă* (organ colector), *urechea mijlocie* (organ transmițător), *urechea internă* (organ senzorial) - transformă undele sonore în vibrații mecanice și fluide, le analizează compoziția spectrală și le transformă în semnale electrice și neuronale pentru a le transmite creierului. Semnalele neuronale suferă alte prelucrări în nucleul cohlear și nivelurile superioare ale creierului, permițându-ne să recunoaștem și să utilizăm informația complexă purtată de semnalul sonor original.

#### 2.1.1 Urechea externă

**Urechea externă** este segmentul aparatului auditiv care culege și concentrază undele sonore din mediul înconjurător spre sistemul timpano-osicular și este formată din pavilionul urechii și conductul auditiv extern. Pavilionul urechii, împreună cu cavitatea sa internă, *conca*, și conductul auditiv extern, acționează parțial ca un concentrator și parțial ca un tub rezonator deschis cu frecvența de rezonanță în jurul a 3 – 4 kHz (lungimea obișnuită a conductului este de aprox. 2,25 cm, ceea ce reprezintă  $\lambda/4$ , adică lungimea de undă  $\lambda = 9$  cm, corespunzătoare unei frecvențe de 3.800 Hz). Forma asimetrică complexă a pavilionului (*anfractuozitățile complexe pavilionare*) produce reflexia sau difracția vibrațiilor sonore venite din diferite direcții, concentrarea și dirijarea lor spre timpan, dar numai pentru frecvențe înalte, peste 4 – 6 kHz, când lungimea de undă a sunetului devine comparabilă cu dimensiunile pavilionului (340 m/s : 6.000 Hz = 57 mm). De aceea pavilioanele au un rol important în localizarea sursei sonore, adică în aprecierea direcției de incidență a sunetului pe orizontală (azimut) și pe verticală (elevație).

#### 2.1.2 Urechea mijlocie

Timpanul desparte urechea externă de **urechea mijlocie** și are o funcție auditivă și una de protecție. Timpanul are forma unei elipse cu o suprafață de aproximativ 55mm<sup>2</sup> și este o membrană foarte subțire, formată din trei straturi. Stratul mijlociu cuprinde două grupuri de fibre, unul radial și altul circular, asemănător cu pânza de paianjen. Prin această structură, timpanul este o membrană tipic aperiodică, adică poate vibra cu orice frecvență. Timpanul vibrează sub efectul stimulului sonor, prin presiune, amplitudinea vibrației lui mecanice în zona percepției optime (aprox. 2.000 Hz) fiind de 0,0045 microni [URM82]. Amplitudinea vibrațiilor timpanului variază cu frecvența, membrana timpanică având o frecvență proprie de rezonanță între 1.000 – 1.500 Hz, iar o undă sonoră sinusoidală nu este reprodusă fidel. Vibrațiile timpanului față de poziția de echilibru nu sunt identice, datorită *asimetriei* (prin forma sa conică), *fixării oscioarelor* și *tensiunii* determinate de mușchii urechii medii. Astfel, amplitudinea vibrațiilor diferă în raport cu diferitele sectoare timpanice: partea posterioară este mai sensibilă decât cea anterioară, amplitudinea cea mai mare fiind la distanță egală între ombilic și marginea postero-superioară a inelului timpanal, iar cea mai mică la nivelul mânerului ciocanului. Această dissimetrie timpanală explică producerea de tonuri adiționale, cu frecvențe

diferite față de sunetul inițial, numite armonici aurale. Ele sunt pur subiective, nu pot fi analizate cu metode obiective și variază cu frecvența și intensitatea sunetului inițial [GĂR80].

Presiunea sonoră se exercită numai pe partea exterioară a membranei timpanice, ca în cazul unui microfon prin presiune. Numai în cazuri patologice, perforații timpanice, unda sonoră se repartizează și pe fața internă, opunându-se presiunii normale de pe fața externă și explicând astfel deficitul auditiv.

În continuare vibrațiile ajung în cavitatea timpanică cu un volum de aprox.  $1\text{cm}^3$  ( $3 \times 15 \times 15\text{mm}$ ), în care se găsesc 3 oscioare: ciocanul, nicovala și scărița. Ciocanul se sprijină pe timpan în timp ce talpa scăriței se sprijină pe fereastra ovală, realizând un transformator cu un raport de aprox 30:1 între timpan și fereastra ovală.

Cavitatea care formează urechea mijlocie se află în legătură cu mediul exterior prin intermediul cavității bucale și al unui canal numit trompa lui Eustache. Datorită aerului care pătrunde prin acest canal se stabilește un echilibru între presiunea aerului de ambele părți ale timpanului.

### 2.1.3 Urechea internă

Procesul de transmitere a undelor sonore se continuă și în **urechea internă**, de la fereastra ovală până la membrana bazilară și organul lui Corti, prin lichidele labirintice, care trebuie considerate ca făcând parte din aparatul de transmisie. Urechea internă reprezintă partea esențială a organului auditiv și echilibrului, constituind porțiunea terminală a aparatului conductor complex al energiei sonore și debutul aparatului de recepție. La acest nivel undele sonore sosite la nivelul ferestrei ovale sub formă de energie mecanică sunt transformate în impulsuri nervoase și mesaje auditive, care sunt transmise pe căile cohleo-vestibulare spre sistemul nervos central.

Urechea internă are trei părți: vestibulul, canalele semicirculare și melcul (fig.2.1). Canalele semicirculare nu au rol în mecanismul audierii, funcția lor fiind aceea de a crea senzația de echilibru. Melcul (cochleea) are cea mai importantă contribuție la formarea senzației auditive și este format dintr-un tub principal, lung de aproximativ 30 mm, răsucit în spirală de 2,75 ori în jurul unui stâlp osos central.

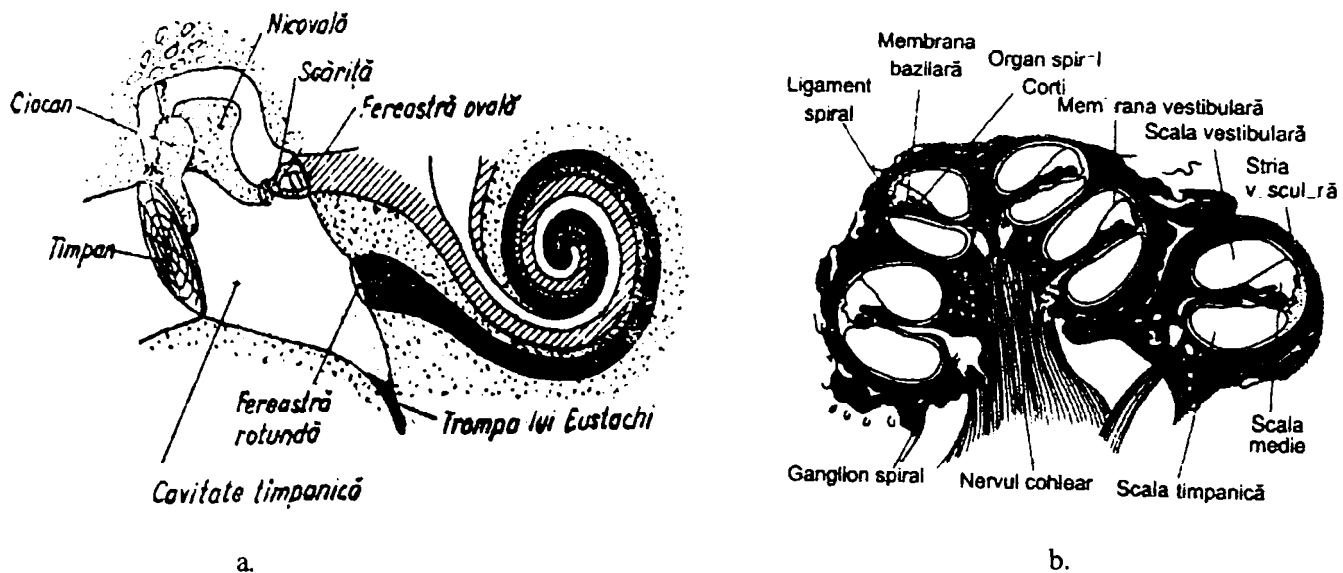


Figura 2.1 Urechea internă: a. Membrana timpanică, sistemul osicular al urechii medii și urechea internă [URM82] b. Melcul (cochleea) - secțiune transversală [GUY96]

Cochleea este un sistem de tuburi spiralate, compus din trei tuburi diferite, spiralate și așezate unul lângă altul: *rampa vestibulară*, *rampa medie* și *rampa timpanică*. Rampa vestibulară și rampa medie sunt separate una de cealaltă prin *membrana Reissner* (denumită și *membrana vestibulară*) așa cum se vede în fig.2.2; rampa timpanică este separată de rampa

medie prin *membrana basilară*. Pe suprafața membranei basilară se găsește organul lui Corti, care conține mai multe celule senzitive electromecanice, numite celule cu cili. Acestea sunt organe terminale receptoare care generează impulsuri nervoase ca răspuns la vibrațiile sonore.

O secțiune prin melc (fig. 2.2) prezintă divizarea lui în cele trei rampe (canale, scale). Acestea sunt umplute cu lichide de compoziție specială, care facilitează transmiterea vibrațiilor.

Desfășurarea în plan a melcului (cohlee despiralată - fig.2.3) permite să ne dăm seama mai bine de modul cum se transmit vibrațiile de la baza scăriței la fereastra ovală și de aici la rampa vestibulară, canalul cohlear și rampa timpanică până la fereastra rotundă. Mai întâi se observă că din figură lipsește membrana Reissner. Ea este atât de subțire și ușor deplasată de vibrațiile sonore încât nu obstrucționează de loc trecerea acestora din rampa vestibulară către rampa medie. De aceea, neafectând conducerea sunetelor, se consideră că rampa vestibulară și cea medie formează o singură cameră. Importanța membranei Reissner rezultă din faptul că ea menține lichidul special din rampa medie, necesar celulelor ciliate care au funcția de recepție a sunetelor.

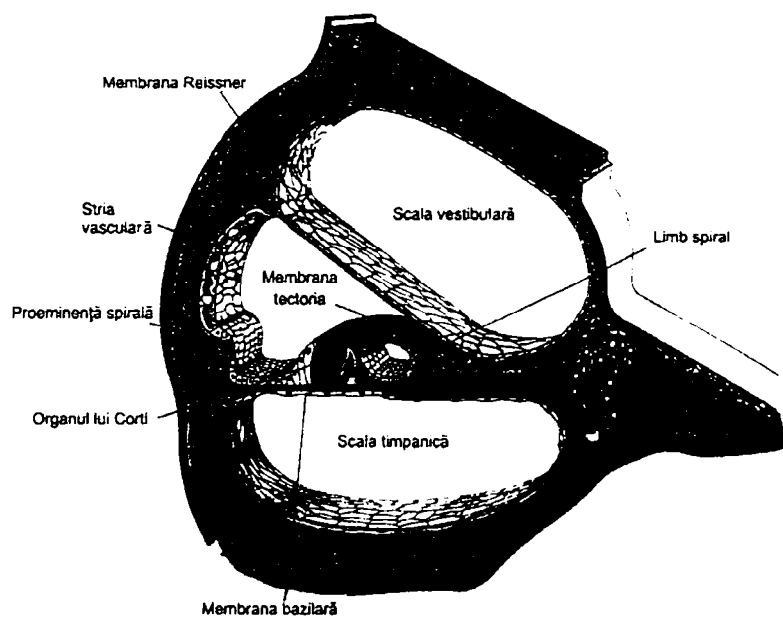


Figura 2.2 Secțiune prin una din spiralele cohleei [GUY96]

Atașarea suprafeței articulare a scăriței de fereastra ovală permite vibrațiilor sonore să pătrundă în rampa vestibulară. Fața articulară a scăriței acoperă fereastra ovală și este legată de de marginea ei printr-un ligament inelar relativ lax care permite tălpii scăriței să se miște în afară și înăuntru odată cu vibrațiile sonore. Deplasarea spre interior produce mișcarea lichidului înspre rampele vestibulară și medie, iar deplasarea înafară determină revenirea lichidului la poziția inițială.

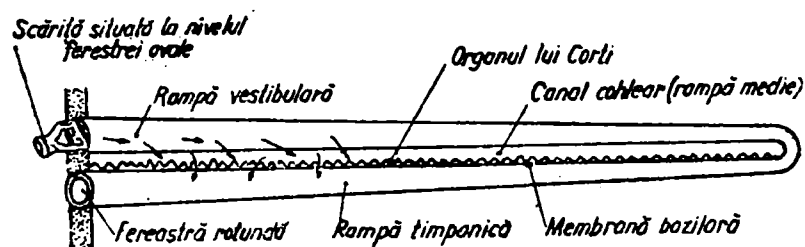


Figura 2.3 Melcul - desfășurare în plan [URM82]

În rampa medie (canalul cohlear) există *membrana bazilară*, o membrană fibroasă care desparte canalul cohlear de rampa timpanică. Ea conține 20.000 - 30.000 de fibre bazilare, numite și coarde auditive. Aceste fibre sunt tari, elastice, cu o structură asemănătoare cu a unui pai, cu baza fixată pe axul osos al cohleei (modiol) și celălalt capăt liber. Lungimea fibrelor bazilare crește progresiv, de la 12 ori, de la aproximativ 0,04 mm la baza cohleei, în apropierea ferestrei rotunde și a celei ovale, până la 0,5 mm în vârful cohleei. Fibrele scurte, din apropierea ferestrei ovale, vibrează la frecvențe înalte, în timp ce fibrele lungi din vârful cohleei vibrează la frecvențe joase.

Astfel rezonanța cu frecvențele înalte a membranei bazilare se produce în apropierea bazei, acolo unde undele sonore pătrund, prin fereastra ovală, în cohlee; rezonanța cu frecvențele joase apare în apropierea vârfului cohleei, datorită în principal diferenței de rigiditate a fibrelor, dar și a creșterii încărcării membranei bazilare cu cantități suplimentare de lichid care trebuie să vibreze odată cu membrana din regiunea apexului.

Pe membrana bazilară, lungă de 30 mm și înfășurată în spirala melcului, se află organul spiral al lui Corti, care cuprinde celulele cu peri (cili), elementele senzoriale auditive esențiale. Organul lui Corti este organul receptor care generează impulsuri nervoase ca răspuns la vibrația membranei bazilare. Se observă că organul lui Corti se întinde pe suprafața fibrelor bazilare și a membranei bazilare. Receptorii senzoriali existenți în organul lui Corti sunt de două tipuri de *celule ciliate*: *celule ciliate interne*, pe un singur rând (aprox. 3.500), cu un diametru de 12  $\mu\text{m}$  și *celule ciliate externe*, pe 3 sau 4 rânduri (aproximativ 15.000) mai lungi și cu un diametru de aproximativ 8  $\mu\text{m}$ . La polul bazal și pe laturile celulelor ciliate există sinapse cu o rețea de nervi cohleari, mai precis cu terminațiile lor. Acestea merg la *ganglionul spiral al lui Corti* care se găsește în modiol (axul osos al cohleei). La rândul lui, ganglionul spiral trimite axoni în nervul cohlear și apoi în sistemul nervos central, la nivelul părții superioare a bulbului.

Din cele de mai sus rezultă că undele sonore cu o frecvență joasă determină activarea membranei bazilare de lângă apexul cohlear, sunetele cu frecvență înaltă activează membrana bazilară de la baza cohleei, iar sunetele cu frecvență intermediară activează porțiunea membranei bazilare situată între cele două extreme. Mai mult, pe calea nervoasă acustică, de la cohlee la cortexul cerebral există permanent o anumită organizare spațială a fibrelor nervoase. Înregistrarea semnalelor din tracturile auditive din trunchiul cerebral și din câmpurile receptoare auditive din cortexul cerebral arată că neuroni specifici sunt activați de către frecvențe sonore specifice. Astfel metoda principală utilizată de către sistemul nervos pentru detectarea diferitelor frecvențe sonore este să determine locul unde membrana bazilară este cel mai mult stimulată. Acesta se numește *principiul localizării* pentru determinarea înălțimii sunetului (fig. 2.4).

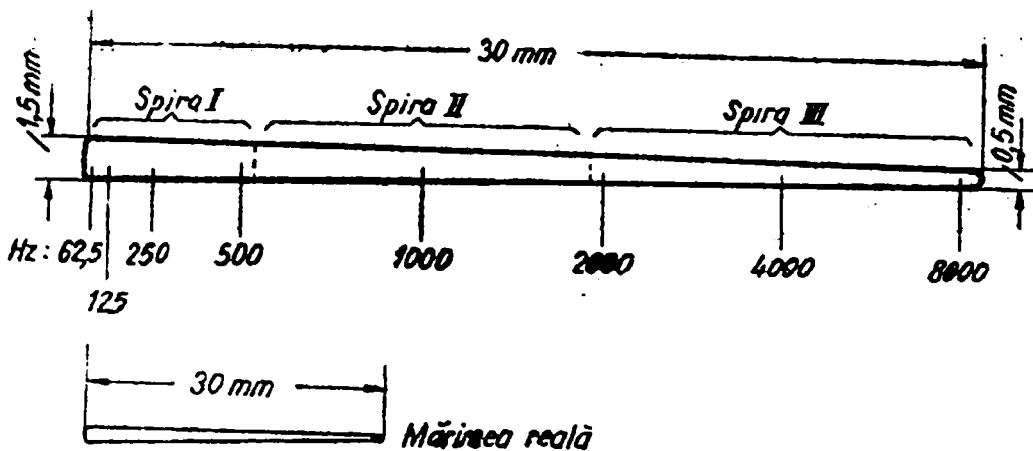


Figura 2.4 Localizarea înălțimii sunetelor pe fibrele membranei bazilare (desfășurată în plan) [URM82]

## 2.2 Psihoacustica

Psihoacustica se referă la relația dintre stimulul fizic și senzația subiectivă pe care o provoacă. Psihofizica clasică tratează observatorul ca un detector de nul, care sesizează când tăria unui semnal este detectabilă, în mărime absolută sau relativ la un fundal (prag) de mascare. Psihofizica senzorială acordă valori numerice întregi sau fracționare amplitudinii senzației.

Prin termenul de *prag* se înțelege în psihologie nivelul sau mărimea stimulului la care apare sau dispare un fenomen psihic sau devine posibilă o discriminare. În esteziologie se disting două categorii de praguri: *absolute* și *relative (diferențiale)*. Condiția de bază la determinarea oricărui prag în raport cu un parametru (frecvență, intensitate, durată etc.) este aceea de a menține constantă valoarea altor parametri care-l pot influența pe cel ales.

Pragul diferențial este exprimat prin cantitatea cu care trebuie să varieze (în plus sau în minus) un stimul pentru a determina o senzație *distinctă* de cea inițială.

Sistemul auditiv efectuează o analiză spectrografică a stimulilor auditivi. Membrana basilară poate fi considerată ca un grup de filtre cu ieșirile ordonate tonotopic, astfel încât se realizează o transformare frecvență-loc.

Sistemul auditiv efectuează, de asemenea, și o analiză oscilografică temporală a setului de semnale neuronale produse în canalul cohlear ca răspuns la stimulul auditiv. Acest proces este important pentru frecvențele sub 500 Hz și contribuie la discriminarea frecvențelor până la aproximativ 1,5 kHz.

Codoarele audio perceptuale realizează compresia datelor exploataând faptul că într-un semnal există informație nerelevantă din punct de vedere acustic, care nu poate fi detectată nici de cel mai sensibil și antrenat ascultător. Informația nerelevantă este identificată în timpul analizei semnalului prin introducerea în codor a mai multor principii psihoacustice: pragul de audibilitate, analiza frecvenței în benzi critice, mascarea simultană, împrăștierea mascării de-a lungul membranei basilare și mascarea temporală. Combinarea acestor noțiuni de psihoacustică cu proprietățile de bază ale cuantizării semnalului a condus la teoria entropiei perceptuale, o estimare cantitativă a limitelor fundamentale ale compresiei transparente a semnalelor audio.

Intensitatea stimulilor acustici este cuantificată printr-o mărime denumită nivel de presiune sonoră SPL (*sound pressure level*). SPL măsoară nivelul presiunii sonore în dB relativ la un nivel de referință.

$$L_{SPL} = 20 \log_{10}(p/p_0) \text{ dB} \quad (2.1)$$

unde  $L_{SPL}$  este nivelul presiunii sonore a unui sunet,  $p$  este presiunea sonoră a sunetului ( $\text{N/m}^2$ ) și  $p_0 = 2 \times 10^{-5} \text{ N/m}^2$  este nivelul de referință standard. Nivelul de referință al SPL este calibrat astfel încât pragul absolut de audibilitate, dependent de frecvență, să fie în vecinătatea lui 0 dB SPL (nivel 0 acustic). Pe de altă parte, un stimul cu nivelul de 140 dB SPL se află la pragul durerii [PAI00].

### 2.2.1 Gama auditivă și audiograma

Gama auditivă este mărginită la limita inferioară de pragul de audibilitate sau audiograma, care este de aproximativ 0 dB SPL în domeniul 1 – 4 kHz și crește spre 70 dB la 20 Hz. Frecvența inferioară limită se consideră 20 Hz, deoarece sub această frecvență se sesizează numai o pulsație fizică fără o scădere a înălțimii. Limita superioară de frecvență depinde foarte mult de vârstă, fiind de aproximativ 20 kHz la 15 ani și scăzând în medie cu 1 Hz/zi, ajungând ca la 70 de ani să atingă o valoare tipică de 2 – 4 kHz [MTS94]. În general, se consideră că un ascultător otologic normal (o persoană între 18 și 25 de ani care are auzul normal și nu a suferit afecțiuni ale organului auditiv) poate sesiza sunete cuprinse în gama 20 Hz – 20 kHz, având maximul de sensibilitate între 2 – 4 kHz. De asemenea, se poate remarca că vocea umană

normală este cuprinsă între 500 Hz și 2 kHz, domeniu în care frecvențele joase sunt date de vocale și bași, iar frecvențele înalte de consoane.

Prin experimente asupra unor grupuri diferite de ascultători, s-a constatat că intensitatea sunetului perceput la un nivel de tărie constant variază cu frecvența (și este diferit în general de la o persoană la alta). Trasând pentru diferite frecvențe curba intensității minime percepute s-a obținut curba pragului de audibilitate (0 foni). Refăcând aceeași curbă pentru alte niveluri de tărie s-au obținut celelalte curbe din fig. 2.5 [BER71]. Curbele de niveluri egale de tărie, numite și curbele lui Fletcher-Munson au fost realizate în 1933 și reactualizate în 1956 de Robinson și Dadson. Curba superioară se numește pragul durerii și este limitată de durerea provocată, depășirea ei putând deteriora auzul. La frecvența de 1.000 Hz, pragul audibilității corespunde unei intensități sonore de  $10^{-12} \text{ W/m}^2$  sau presiunii de  $2 \times 10^{-5} \text{ N/m}^2$ . În cazul sunetelor inferioare frecvenței de 1.000 Hz, precum și al celor ce depășesc 3.000 Hz (zona de sensibilitate maximă a urechii), pentru a obține o senzație auditivă de același nivel de tărie, intensitatea sonoră trebuie să crească.

Pragul absolut de audibilitate se definește ca fiind energia necesară unui ton pur pentru a putea fi detectat de un ascultător într-un mediu lipsit de alte zgomote. El se măsoară în dB și poate fi aproximat printr-o funcție neliniară [PAI00]:

$$T_q(f) = 3,64(f/1000)^{-0,8} - 6,5e^{-0,6(f/1000-3,3)^2} + 10^{-3}(f/1000)^4 \text{ dB} \quad (2.2)$$

și este reprezentativă pentru un ascultător tânăr cu auz sănătos.

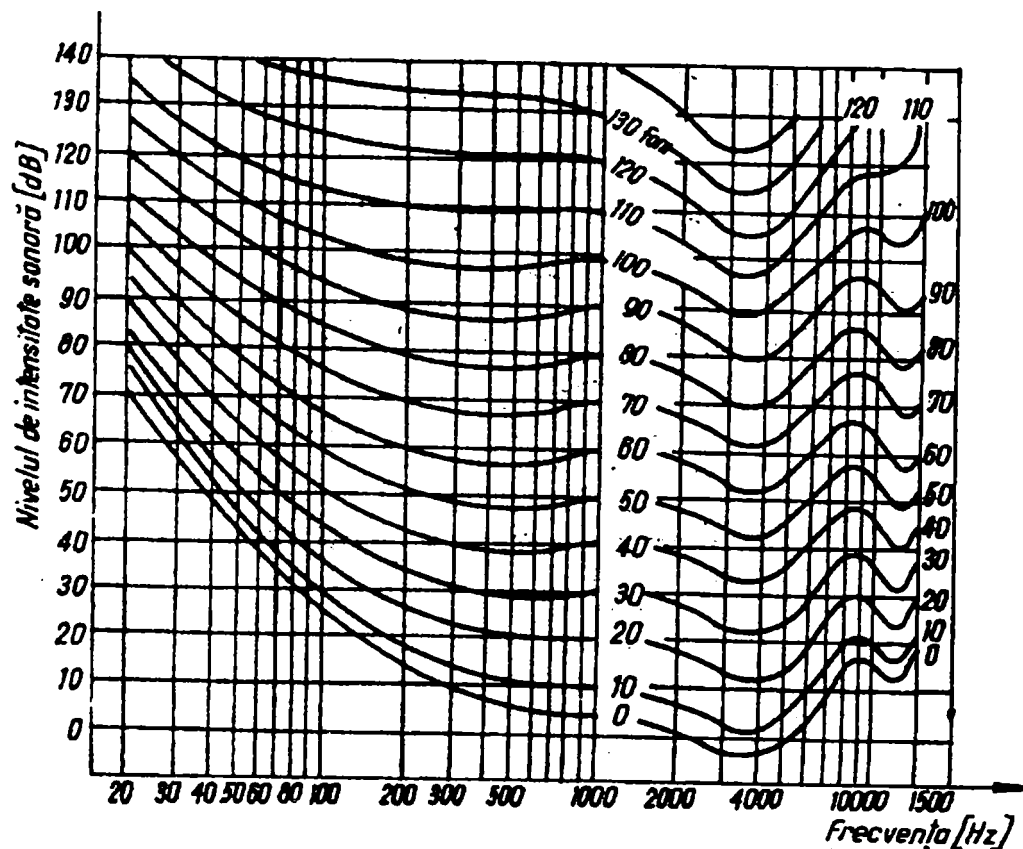


Figura 2.5 Curbele lui Fletcher-Munson (curbele de niveluri egale de tărie) [BER71]

Urechea este sensibilă pentru întreaga zonă cuprinsă între pragul de audibilitate și pragul de durere, care se numește suprafața de audibilitate. Din această suprafață se folosește practic o zonă mai redusă pentru muzică și una și mai redusă pentru voce, așa după cum rezultă din fig. 2.6. Din examinarea acestei diagrame rezultă că domeniul intensității audibile are o întindere extrem de mare, raportul dintre pragul de durere și cel de audibilitate fiind de aprox.  $10^{12}$  (între  $1 \text{ W/m}^2$  și  $10^{-12} \text{ W/m}^2$ ).

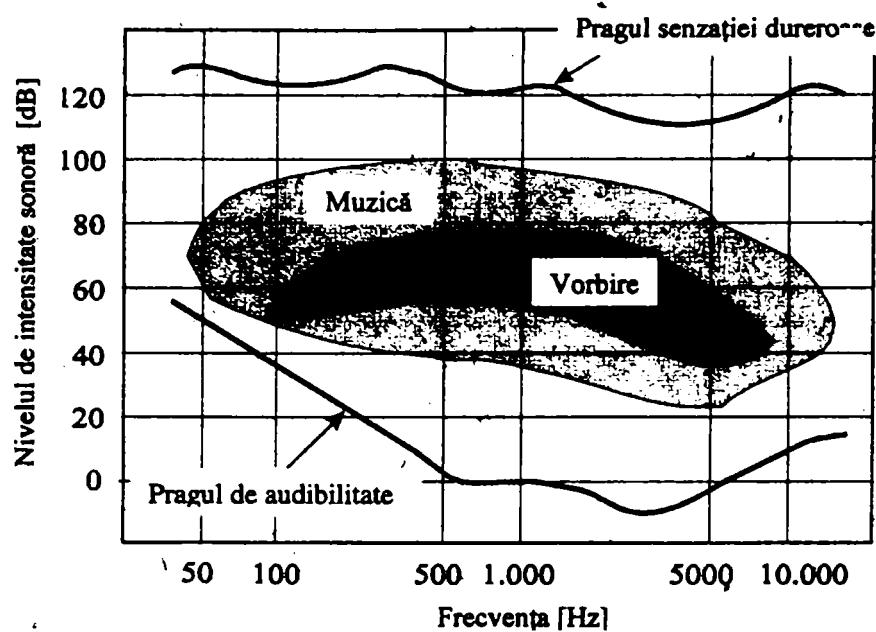


Figura 2.6 Suprafața de audibilitate

### 2.2.2 Perceperea înălțimii

Pentru a înțelege percepția înălțimii trebuie avut în vedere faptul că în natură sunetele pure sunt aproape inexistente și deci trebuie ținut cont de structura armonică a sunetelor complexe. Senzația de înălțime a sunetului are două forme: înălțimea fizică și înălțimea muzicală.

*Înălțimea fizică* este strict legată de frecvență și de locul din membrana bazilară unde are loc stimularea. Se poate considera astfel, că înălțimea unui sunet complex este egală cu frecvența unui sunet pur (judecat subiectiv ca având aceeași înălțime cu sunetul complex), dar pentru modelele percepției auditive sunt necesare alte scări. Chiar și în cazul unui sunet sinusoidal înălțimea percepută nu depinde numai de frecvență ci, într-o oarecare măsură, și de intensitatea sunetului.

Mărimea fiziologică care definește înălțimea unui sunet se numește *mel*, a fost stabilită experimental și are ca reper înălțimea de 1.000 de meli corespunzătoare unui sunet cu frecvența de 1.000 Hz și un nivel de presiune de 60 dB. Între 1.000 Hz și 5.000 Hz legătura dintre frecvență și mel este liniară, dar pentru frecvențe joase are o pronunțată neliniaritate.

Relațiile empirice între frecvență ( $f$ ) și mel ( $m$ ) sunt:

$$m = 1.127 \ln(1 + f/700) \quad (2.3)$$

$$f = (\exp(m/1127) - 1) \quad (2.4)$$

*Înălțimea muzicală* sau senzația subiectivă de înălțime se poate defini ca fiind însușirea senzației auditive prin care sunetele pot fi ordonate după o scară de la sunete grave la sunete acute. Această ordonare se face crescând frecvența sunetelor într-un anumit raport, cel mai utilizat în muzică fiind 2/1, care caracterizează intervalul numit *octavă*. Parcurgând șirul frecvențelor din octavă în octavă, începând de la 2 Hz se obține dispunerea:

Frecvența	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	16348
Octava	1	2	3	4	5	6	7	8	9	10	11	12	13	



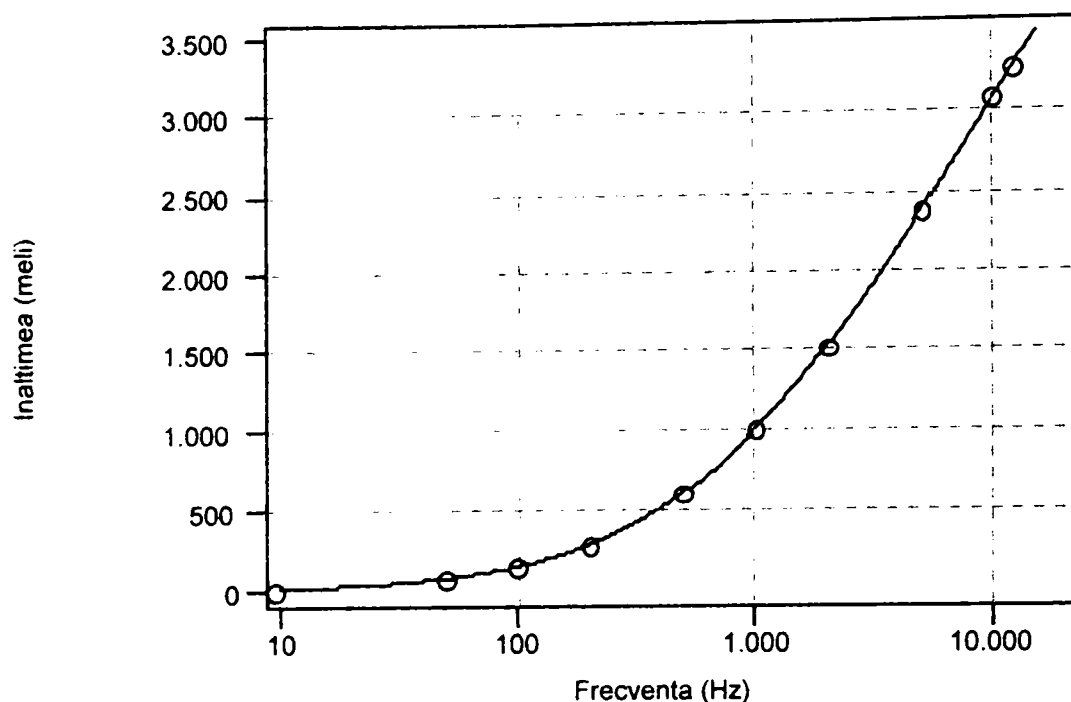


Figura 2.7 Legătura între frecvență și mel

Din aceste 13 octave sunt audibile numai aproximativ 10 (între 16 Hz și 16 kHz) [BER71].

În ceea ce privește senzația de înălțime a sunetului, pragul diferențial este exprimat în numărul de hertzi cu care trebuie să crească sau să scadă frecvența unui sunet pentru a putea afirma că noul sunet este mai sus sau mai jos decât cel inițial.

### 2.2.3 Discriminarea frecvențelor

Sensibilitatea diferențială a frecvențelor este definită prin variația de frecvență relativă minimă,  $\Delta f/f$ , necesară pentru ca două sunete imediat consecutive (ca frecvență) să nu pară a avea aceeași înălțime. Modificarea minimă în frecvență care poate fi detectată pentru stimuli succesivi se numește *diferență de frecvență limită (dl)* sau prag diferențial de percepție a înălțimii. Deși depinde de mulți factori (frecvență, intensitate, durată, interval între stimuli), această limită este de aprox. 1Hz la frecvențe sub 1 – 2 kHz, la niveluri mai mari de 20 dB peste prag și durată peste 0,25 s. La frecvențe peste 2 kHz, *dl* crește aproape proporțional cu frecvența.

În urma unor experimente s-a ajuns la concluzia că în domeniul 500 – 3.000 Hz, pragul diferențial de percepție a înălțimii este aproape constant și are valoarea medie de 3‰. Aceasta înseamnă că două sunete de tărie egală sunt percepute ca având înălțime *diferită* numai dacă frecvențele lor diferă cu cel puțin  $\pm 3\%$ . Considerând ca exemplu sunetul de 1.000 Hz (ceva mai înalt de decât  $si_6 = 987,8$  Hz), cele mai apropiate sunete de înălțime apreciată ca diferită trebuie să aibă frecvența de 1003Hz sau 997 Hz. Pentru comparație, cele mai apropiate note muzicale de  $si_6$  sunt (la pian)  $si_6$  bemol = 932,3 Hz și  $do_7 = 1.046,5$  Hz. Rezultă că intervalul determinat de pragul diferențial de 3‰ în domeniul 500 – 3.000 Hz este cu mult mai mic decât intervalul de semiton temperat, urechea putând distinge în acest interval un număr de 20 de sunete [URM82].

În afara acestui domeniu, pragul diferențial crește destul de repede, ceea ce înseamnă că sensibilitatea diferențială pentru înălțimi a organului auditiv scade, la frecvențe joase ajungând până la 12‰, iar la frecvențe înalte până la 7 - 8‰. Variația pragului diferențial al înălțimii sunetului în raport cu frecvența este reprezentată în fig. 2.8.

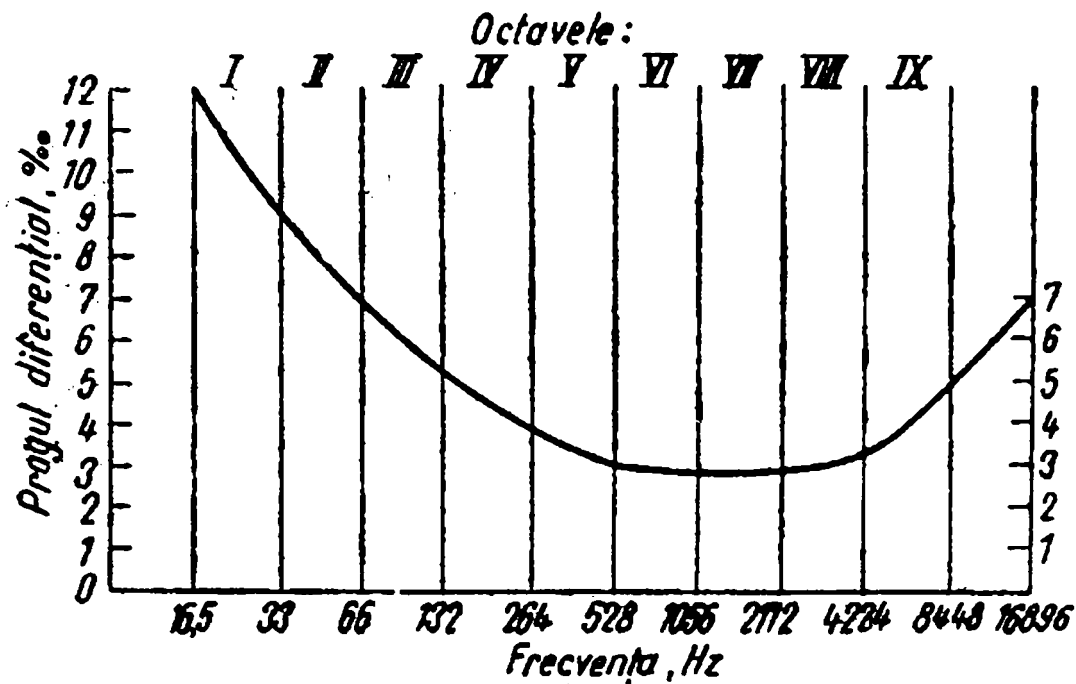


Figura 2.8 Variația pragului diferențial al senzației de înălțime a sunetelor în raport cu frecvențele din câmpul audibil (forma curbei poate diferi sensibil de la persoană la persoană)

De asemenea, pragul diferențial de frecvențe depinde de intensitatea sonoră și de durata stimulului, el crescând cu scăderea intensității și duratei. Pragul diferențial (relativ) și cel absolut depind de caracteristicile individuale ale analizorului auditiv și, într-o foarte mare măsură, de antrenament.

#### 2.2.4 Benzile critice

Prelucrarea semnalelor audio, în special în vederea reducerii debitului de date, a condus în ultimul timp la utilizarea tehnicilor în domeniul frecvență, bazate pe diferite măsurări spectrale psihoacustice.

Percepția auditivă se bazează pe analiza benzilor critice de frecvență în urechea internă unde, de-a lungul membranei basilare are loc o transformare frecvență-loc. Această transformare funcționează în felul următor: o undă sonoră, generată de un stimul acustic, parcurge tractul auditiv până la fereastra ovală, de unde sunt induse unde călătoare de-a lungul membranei basilare. Undele călătoare generează vârfuri de răspuns dependente de frecvență în anumite poziții ale membranei. Rezultă că receptorii neuronali sunt acordați pe diferite frecvențe în funcție de poziție. Pentru stimuli sinusoidali, unda călătoare se propagă de la fereastra ovală până când se apropie de o regiune cu o frecvență de rezonanță apropiată de frecvența stimulului. Viteza undei scade, amplitudinea crește până la atingerea unui vârf și apoi scade rapid. Locul în care apare vârful se numește “loc caracteristic”, iar frecvența care excită cel mai bine un anumit loc se numește “frecvență caracteristică” [GRE90]. Figura 2.9 prezintă un exemplu de transformare frecvență-loc pentru un stimul tritonal.

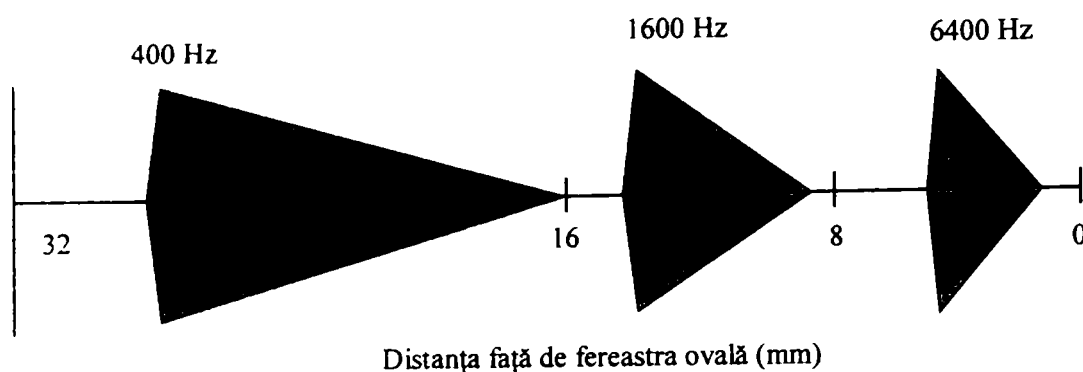


Figura 2.9 Transformarea frecvență-loc de-a lungul membranei basilare

#### A. Scara Bark

Spectrul energetic nu se reprezintă pe o scară liniară de frecvență, ci prin benzi de frecvențe cu diferite lățimi, denumite **benzi critice**, reprezentate pe așa numita **scară Bark** (după Barkhausen). Ea este o scară empirică, bazată pe numeroase experimente psihoacustice, definită astfel încât benzile critice ale auzului uman au lățimea de un Bark. Reprezentarea energiei spectrale (în dB) pe scara Bark este în strânsă corespondență cu informația spectrală prelucrată de ureche. Scala Bark se întinde de la 1 la 24 Bark, corespunzător primelor 24 de benzi critice ale auzului uman. Multe efecte psihoacustice pot fi puse în evidență în cadrul benzilor critice. În tabelul 2.1 sunt prezentate valorile standard rotunjite (descrise de Zwicker) ale scării Bark cu frecvențele limită ale benzilor critice, frecvența centrală și lățimea benzii. Frecvențele centrale și lățimile benzilor reprezintă eșantioane ale răspunsului în frecvență a urechii la un semnal sinusoidal sau un zgomot de bandă îngustă.

Tabelul 2.1 Valorile standard rotunjite ale scalei Bark

Bark	Frecvența	Lățimea benzii (Hz)	Frecvența centrală (Hz)
0	20	80	50
1	100	100	150
2	200	100	250
3	300	100	350
4	400	110	450
5	510	120	570
6	630	140	700
7	770	150	840
8	920	160	1.000
9	1.080	190	1.170
10	1.270	210	1.370
11	1.480	240	1.600
12	1.720	280	1.850
13	2.000	320	2.150
14	2.320	380	2.500
15	2.700	450	2.900
16	3.150	550	3.400
17	3.700	700	4.000
18	4.400	900	4.800
19	5.300	1.100	5.800
20	6.400	1.300	7.000
21	7.700	1.800	8.500
22	9.500	2.500	10.500
23	12.000	3.500	13.500
24	15.500		

De remarcat că, deoarece scara Bark este definită până la 15,5 kHz, cea mai mare rată de eșantionare pentru care scara Bark este definită fără extrapolare, conform criteriului lui Nyquist, este 31 kHz. Cea de-a 25-a bandă Bark se poate extinde peste 19 kHz (suma dintre limita celei de-a 24-a benzi plus lățimea celei de-a 23-a benzi), permițând frecvențe de eșantionare de 40kHz. Limitele benzilor critice se pot extrapola, adăugând valorile de 20.500 și 27.000, astfel încât se pot defini rate de eșantionare până la 54 kHz. Deoarece limita superioară a auzului uman nu depășește 20 kHz, se utilizează rate de eșantionare de până la 48 kHz.

Valorile standard rotunjite ale scării Bark se bazează pe observații empirice, dar în aplicații este necesară utilizarea unor relații de conversie. Există șapte relații care aproximează curba Bark/Hertz:

$$\text{Zwicker \& Terhardt (1980)} \quad B = 13 \tan^{-1}(0,76f/1000) + 3,5 \tan^{-1}(f/7500)^2 \quad (2.5)$$

$$B = 8,7 + 14,2 \log_{10}(f/1000) \quad (2.6)$$

$$\text{Terhardt (1979)} \quad B = 13,3 \tan^{-1}(0,75f/1000) \quad (2.7)$$

$$B = 12,82 \tan^{-1}(0,78f/1000) + 0,17(f/1000)^{1,4} \quad (2.8)$$

$$\text{Wang, Sekey \& Gersho (1992)} \quad B = 6 \sinh^{-1}(f/600) \quad (2.9)$$

$$\text{Schroeder (1977)} \quad B = 7 \sinh^{-1}(f/650) \quad (2.10)$$

$$\text{Traunmüller (1990)} \quad B = 26,81 / (1 + (1960/f)) - 0,53 \quad (2.11)$$

unde  $B$  este în bark,  $f$  în Hertzi și unghiurile în radiani.

În tabelul 2.2 sunt prezentate valorile obținute cu fiecare formulă pentru fiecare punct standard al curbei:

Tabelul 2.2

Bark (rotunjit)	Frecvența (Hz)	Z & Te (1980)		Te (1979)		W, Se & G (1992)	Sc (1977)	Tr (1990)
		(2.3)	(2.4)	(2.5)	(2.6)	(2.7)	(2.8)	(2.9)
0	20	*0	-	*0	*0	*0	*0	
1	100	0,987	-5,500	*0,996	1,005	0,995	1,073	0,771
2	200	1,963	-1,225	1,980	*2,002	1,965	2,121	1,952
3	300	2,920	1,275	2,943	*2,978	2,887	3,126	3,029
4	400	3,847	3,049	3,876	3,924	3,751	4,074	*4,014
5	510	4,823	4,548	4,859	4,920	4,627	5,044	*5,006
6	630	5,830	5,851	5,871	5,944	5,498	6,016	*5,991
7	770	6,920	7,088	6,965	7,052	6,409	7,043	*7,032
8	920	*7,985	8,186	8,033	8,131	7,279	8,028	8,034
9	1080	9,006	9,175	9,055	9,164	8,103	8,968	*8,995
10	1270	10,079	10,174	10,122	10,246	8,968	9,960	*10,011
11	1480	11,106	11,118	11,139	11,280	9,809	10,927	*11,005
12	1720	12,109	12,045	12,121	12,289	10,653	11,901	*12,001
13	2000	13,093	12,975	13,071	13,278	11,513	12,897	*13,010
14	2320	14,028	13,890	13,954	14,218	12,371	13,892	*14,003
15	2700	14,943	14,825	14,791	15,137	13,256	14,920	*15,004
16	3150	15,830	15,776	15,566	16,030	14,162	15,973	*15,997
17	3700	16,714	16,768	16,292	16,923	15,113	17,079	*16,996
18	4400	17,628	17,837	16,978	17,856	16,141	18,277	*18,018
19	5300	18,582	*18,985	17,614	18,851	17,249	19,568	19,042
20	6400	19,531	20,148	18,160	19,889	18,375	20,880	*19,994
21	7700	20,452	21,288	18,611	*20,984	19,480	22,168	20,840
22	9500	21,482	22,584	19,037	22,392	20,738	23,635	*21,695
23	12000	*22,587	24,024	19,420	24,285	22,137	25,267	22,516
24	15500	*23,711	25,603	19,750	26,967	23,671	27,056	23,270

Valorile din tabel au fost înscrise cu următoarele aproximații:

- caractere îngroșate =  $\pm 0,1$  bark
- caractere italice =  $\pm 0,2$  bark
- \* = cea mai bună aproximație

Rezultatele din acest tabel arată că relația lui Traunmüller (2.11) are o precizie de  $\pm 0,05$  bark în intervalul 200 Hz la 6,7 kHz și este cea mai potrivită pentru aplicații de analiză vocală. Vocea este rareori digitalizată la o rată mai mare de 16 kHz. Aceasta permite o analiză până la 8 kHz, domeniu în care relația lui Traunmüller dă cele mai bune aproximații, deși este cea mai simplă.

Relația inversă (Traunmüller) este:

$$f = 1960 / (26,81 / (B + 0,53) - 1) \quad (2.12)$$

Lățimea benzii critice (în Hz) este:

$$B_C = 52548 / (B^2 - 52,56B + 690,39) \quad (2.13)$$

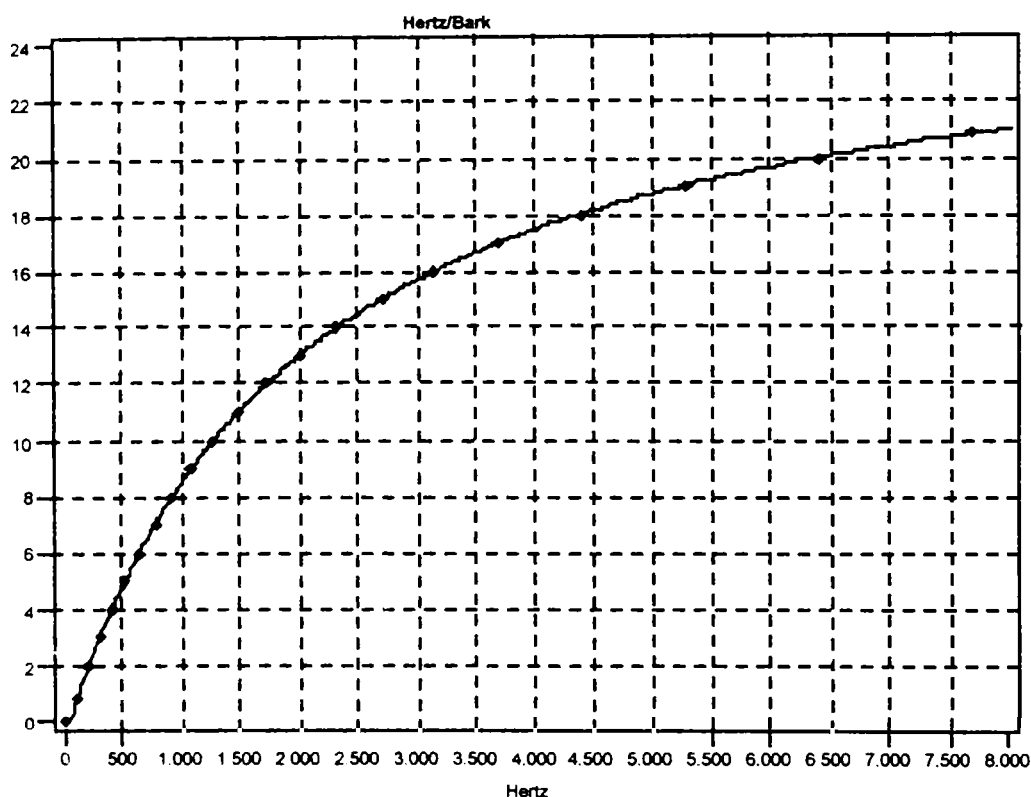


Figura 2.9 Scara Bark în funcție de frecvență, conform aproximației lui Traunmüller

### B. Scara ERB

Moore și Glasberg au revizuit modelul tăriei lui Zwicker pentru a explica mai bine cum se modifică curbele de egală tărie în funcție de intensitate, de ce tăria rămâne constantă în timp ce lățimea de bandă a unui sunet cu intensitate fixă crește până la lățimea benzii critice și tăria sunetelor mascate parțial. Principala modificare este înlocuirea scării Bark cu scara ERB (*equivalent rectangular bandwidth*). Lățimea benzii dreptunghiulare echivalente a filtrului auditiv este foarte apropiată de lățimea benzilor critice, dar a fost măsurată cu metoda “*notched-noise*” și nu prin experimentele clasice de mascare. Drept urmare determinarea scării ERB nu este afectată de fenomenul de bătăi sau de produsele de intermodulație dintre mascator și mascat. Această scală are un comportament mai uniform decât scala Bark, lățimea benzilor fiind calculată cu relația:

$$\text{ERB}(f) = 0,108f + 24,7 \quad (2.14)$$

unde  $f$  este frecvența centrală în Hz, de obicei în domeniul 100 Hz – 10 kHz. Lățimea de bandă echivalentă ERB este în general mai îngustă decât lățimea benzilor critice, fiind de aproximativ 11% din frecvența centrală, pentru frecvențe înalte, și coborând până la 25 Hz pentru frecvențe joase. Pe de altă parte, lățimea benzilor critice este de aproximativ 20% din frecvența centrală, mergând până la 100 Hz sub 500 Hz. În figura 2.10 sunt prezentate comparativ 4 curbe:

- lățimile benzilor critice;
- lățimile benzilor echivalente;
- lățimile benzilor critice determinate cu regula clasică (100 Hz pentru frecvențe până la 500 Hz și 20% din frecvența centrală peste 500 Hz);
- formula empirică

$$B_c = 94 + 71f^{3/2}, \text{ unde } f \text{ este în kHz} \quad (2.15)$$

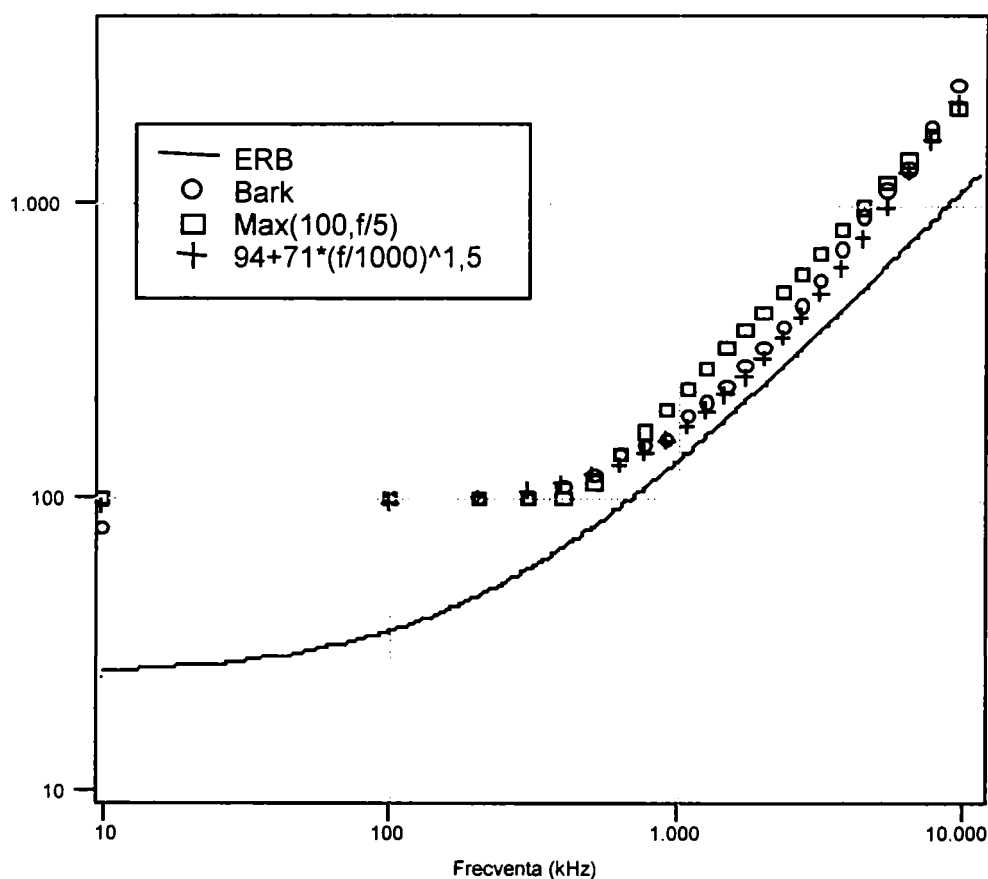


Figura 2.10 Lățimea benzilor critice și echivalente funcție de frecvență

Scala ERB se definește ca numărul de ERB-i sub fiecare frecvență [SMI99]:

$$\text{ERB}(f) = 21,4 \log_{10}(0,00437f + 1) \quad (2.16)$$

unde  $f$  este în Hz.

### 2.2.5 Mascarea auditivă

Mascarea auditivă descrie efectul prin care un semnal de intensitate redusă (mascatul) poate deveni inaudibil dacă apare simultan cu un semnal de intensitate mai mare (mascatorul).

Acest fenomen poate fi exploatat în codarea audio și vocală prin modelarea potrivită a zgomotului în procesul de codare. Această modelare a zgomotului permite o codare de calitate chiar și cu un raport semnal/zgomot redus. Fenomenul de mascare depinde de distribuția spectrală a mascatorului și mascatului și de variația lor în timp.

În interiorul benzilor critice intensitatea tonurilor individuale este însumată de ureche. Sistemul auditiv uman poate fi descris ca un grup de filtre trece-bandă, constând din filtre trece-bandă care se suprapun, având lățimea benzii de 100 Hz pentru semnale sub 500 Hz și până la 5.000 Hz pentru semnale de înaltă frecvență (vezi tab. 2.2). Se iau în considerare 26 de benzi critice, până la 24.000 Hz. În figura 2.11 se face o comparație între scara liniară de frecvență (0 – 24.000 Hz) și împărțirea în benzi critice (scara Bark). Scara liniară de frecvență este împărțită în 32 de subbenzi, având fiecare o lățime de 750 Hz. Această împărțire în subbenzi este folosită la codarea MPEG (vezi cap. 3). Conceptul de benzi critice trebuie privit cu atenție, deoarece benzile critice nu sunt entități fixe ci sunt definite de fenomenele acustice care ating urechea.

Efectul de mascare este cu atât mai accentuat cu cât două sunete de tărie comparabilă sunt mai apropiate ca frecvență. Dintre două sunete de tărie comparabilă, cel mai grav îl maschează în diferite grade pe cel mai înalt și nu invers. Acest lucru se întâmplă mai ales când sunetul grav conține armonici intense, dintre care una este apropiată sau coincide cu frecvența sunetului mai înalt.

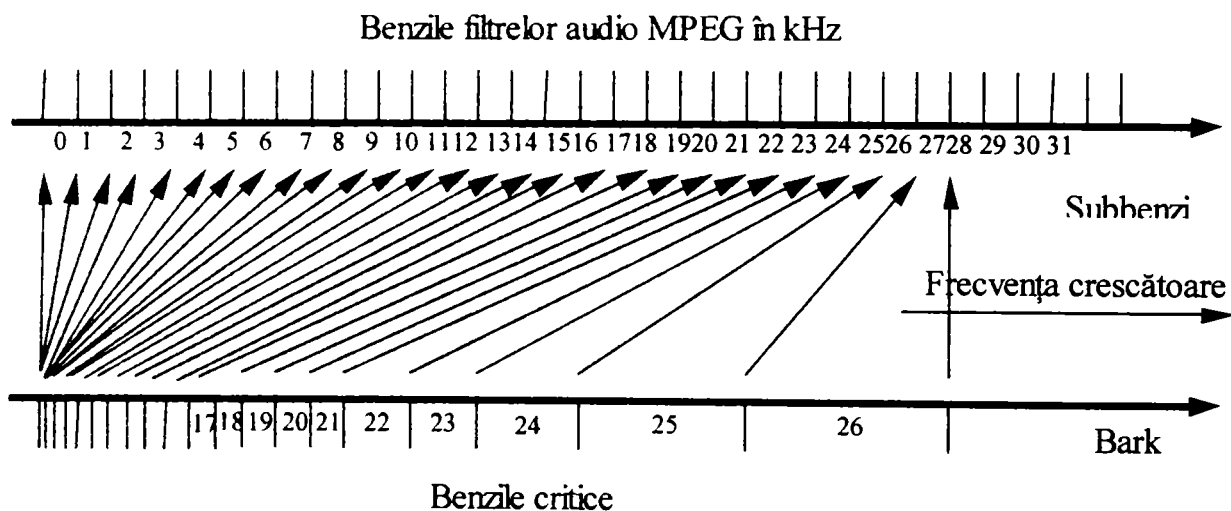


Figura 2.11 Benzile filtrelor audio MPEG și distribuția benzilor critice

### 2.2.5.1 Mascare simultană

Mascarea simultană este un fenomen ce ține de domeniul frecvență: un semnal de nivel mic, de exemplu un zgomot, poate deveni inaudibil (mascat) prin apariția simultană a unui semnal mai puternic, dacă mascatorul și mascatul au frecvențe suficient de apropiate. Fără mascator, un semnal este neaudibil dacă nivelul de presiune sonoră (SPL) se află sub pragul de audibilitate, care depinde de frecvență și are o gamă dinamică de aproape 80 dB (vezi curba inferioară din fig. 2.12). În această figură mascatorul este un zgomot cu frecvența centrală de 1.000 Hz și o lățime de 90 Hz. În prezența acestui mascator se poate măsura un nou prag de mascare sub care nici un semnal nu este audibil. Pragul de mascare depinde de structura, nivelul (SPL) și diferența de frecvență dintre semnale. Tot din fig. 2.12 se poate observa că dacă presiunea sonoră a unui sunet sinusoidal cu o frecvență apropiată de frecvența centrală a unui zgomot este cu 5 dB mai mică decât nivelul zgomotului, sunetul devine inaudibil. În figura 2.13 sunt scoase în evidență și alte detalii despre pragul de mascare: în interiorul benzilor critice, tonurile aflate sub prag (în zona hașurată) sunt mascate. Distanța dintre nivelul mascatorului și pragul de mascare se numește raport semnal/mascare (*signal-to-mask ratio* SMR). Valoarea maximă a acestui raport se află la marginea din stânga a benzii critice (punctul A), iar valoarea

minimă, rezultată în urma experimentelor, este de aprox. 5,5 dB. Să presupunem cuantizarea unui semnal audio cu  $m$  biți. În interiorul unei benzi critice zgomotul de cuantizare nu va fi sesizabil atât timp cât raportul semnal/zgomot (SNR) va fi mai mare decât SMR. Zgomotul și contribuțiile semnalului în afara respectivei benzi critice vor fi de asemenea mascate dacă nivelul lor (SPL) se află sub pragul de mascare.

Considerând  $SNR(m)$  ca fiind raportul semnal/zgomot rezultat în urma cuantizării pe  $m$  biți, distorsiunea perceptibilă într-o anumită subbandă se măsoară prin raportul zgomot/mascare NMR (*noise-to-mask ratio*)

$$NMR(m) = SMR - SNR(m) \text{ (dB)} \quad (2.17)$$

Raportul zgomot/mascare  $NMR(m)$  descrie raportul semnal/mascare și raportul semnal/zgomot care se poate obține pentru o cuantizare pe  $m$  biți. Valoarea NMR reprezintă de asemenea diferența (în dB) dintre nivelul zgomotului de cuantizare și pragul de audibilitate al distorsiunii într-o anumită subbandă.

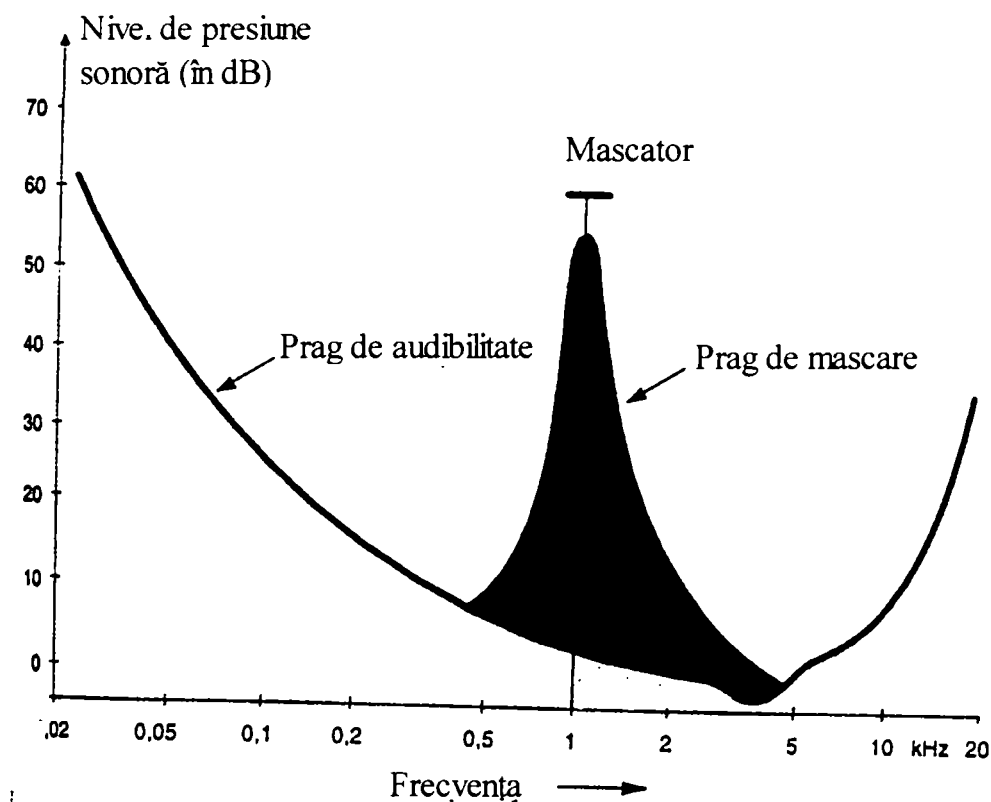


Figura 2.12 Pragul de audibilitate și pragul de mascare. Evenimentele acustice aflate în zona neagră nu sunt audibile.

Până acum a fost descrisă mascarea cu un singur mascator în interiorul unei benzi critice. Dacă semnalul sursă constă din mai mulți mascatori simultani, trebuie să avem în vedere că fiecare are propriul prag de mascare, iar pragul global de mascare se poate obține prin însumarea pragurilor individuale (vezi cap. 4).

Mascarea auditivă poate fi foarte utilă în codarea sursei; semnalul mascat poate fi zgomot de cuantizare sau componente ale semnalului sursă. În primul caz codorul trebuie să țină zgomotul de cuantizare sub pragul de mascare. În al doilea caz componentele semnalului sursă aflate sub pragul de mascare nu trebuie să fie codate și transmise. O codare eficientă a sursei va elimina toate componentele irelevante pentru ureche, iar dacă și zgomotul poate fi mascat codorul este *transparent perceptual*.



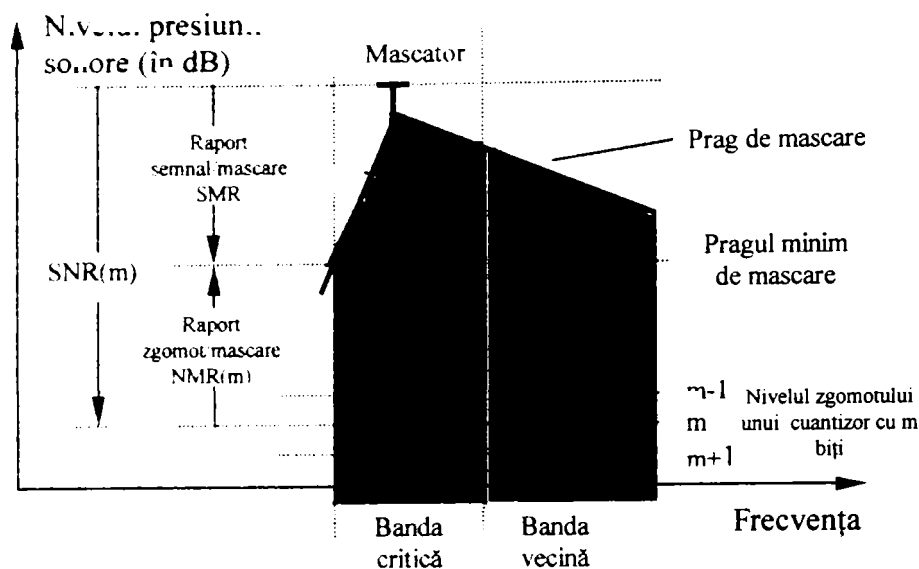


Figura 2.13 Pragul de mascare și raportul semnal/mascare (SMR) [NOL97]

Noțiunile de bandă critică și mascare simultană în contextul codării audio au condus la apariția unei terminologii specifice, ilustrate în fig. 2.13, unde s-a luat în considerare apariția unui singur ton mascator în mijlocul unei benzi critice. Toate nivelurile din figură sunt date în dB. Apariția unui ton de mascare de un anumit nivel generează o excitație de-a lungul membranei basilare, care este modelată de o funcție de împrăștiere și un *prag de mascare* corespunzător. Pentru banda luată în discuție, *pragul minim de mascare* reprezintă minimumul funcției de împrăștiere în interiorul benzii. Presupunând că mascatorul este cuantizat folosind un cuantizor scalar uniform pe  $m$  biți, zgomotul poate apărea la nivelul  $m$ . Raportul semnal/mascare SMR și raportul zgomot/mascare NMR reprezintă distanțele logaritmice dintre pragul minim de mascare și mascator, respectiv nivelul de zgomot.

Mascarea simultană poate apărea când mascatorul și mascatul sunt prezente în același timp și sunt cvasi staționare. Mascarea maximă se obține de obicei dacă ambele semnale au aceeași frecvență centrală. Dacă un semnal de tip zgomot maschează un semnal tonal, nivelul de mascare este aproape independent de frecvența mascatorului. Dacă un semnal tonal maschează un semnal zgomot, nivelul de mascare depinde de frecvența mascatorului. El poate fi estimat prin formula **(15,5+z/Bark) dB**, unde  $z$  este banda critică a mascatorului. Mai mult, la niveluri mari ale semnalului, efectele neliniare reduc pragul de mascare lângă mascator. În cazul combinării mai multor semnale, pragul de mascare este o funcție neliniară de pragurile de mascare ale fiecărui semnal în parte [BAU95].

Deși spectrul audio poate conține diferite scenarii complexe de mascare simultană, în cazul codării audio se iau în considerare trei tipuri de mascare simultană: zgomot-maschează-ton (NMT – *noise-masking-tone*), ton-maschează-zgomot (TMN – *tone-masking-noise*) și zgomot-maschează-zgomot (NMN – *noise-masking-noise*), ale căror caracteristici esențiale vor fi prezentate în continuare.

1) *Zgomot-maschează-ton*: În scenariul NMT, fig. 2.14.a, un zgomot de bandă îngustă (de ex. cu lățimea de 1 Bark) maschează un ton aflat în aceeași bandă critică, având un nivel mai mic decât zgomotul mascator. Numeroase studii privind zgomote aleatoare și stimuli tonali puri au apărut încă din anii 1930. S-a observat că raportul semnal/mascare SMR minim se obține atunci când frecvența tonului mascat se află aproape de frecvența centrală a mascatorului. De exemplu, în fig. 2.14.a un zgomot mascator cu frecvența centrală 410 Hz și intensitatea de 80 dB maschează un ton cu frecvența de 410 Hz, rezultând la pragul de detecție un SMR minim de 4dB. Puterea de mascare scade (SMR crește) pentru tonuri cu frecvențe peste sau sub frecvența centrală conform unei funcții de împrăștiere care depinde de nivel și frecvență, descrisă mai jos.

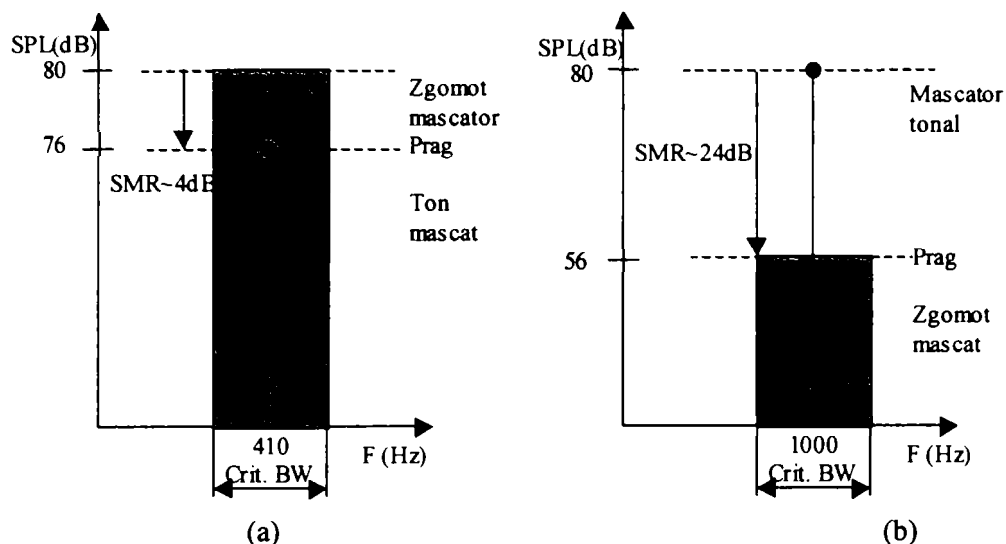


Figura 2.14 Asimetria mascării simultane: (a) zgomot-maschează-ton NMT, (b) ton-maschează-zgomot TMN

2) *Ton-maschează-zgomot*: În cazul TMN (fig. 2.14.b) un ton pur având frecvența centrală a unei benzi critice maschează un zgomot aflat în interiorul respectivei benzi critice având intensitatea sub un anumit prag. Pragul de detectare al unui zgomot mascat de un ton pur, respectiv minimul raportului SMR, apare când frecvența mascatorului este apropiată de frecvența centrală a mascatului și este de aproximativ 21 – 28 dB [SCH79]. În fig. 2.14.b un zgomot cu o lățime de bandă de 1 Bark și frecvența centrală de 1 kHz este mascat de un ton cu frecvența de 1kHz și intensitatea de 80 dB. Raportul SMR la pragul de mascare este de 24 dB. Ca și în cazul anterior, puterea de mascare descrește pentru zgomote a căror bandă au frecvența centrală peste sau sub frecvența centrală a zgomotului care a produs SMR minim.

3) *Zgomot-maschează-zgomot*: scenariu NMN în care un zgomot de bandă îngustă maschează un alt zgomot de bandă îngustă este mai greu de caracterizat decât cele două cazuri anterioare, datorită influenței reciproce a fazelor dintre componentele mascatorului și mascatului. S-au obținut valori de aproximativ 26 dB pentru pragul SMR [PAI00].

4) *Asimetria mascării*: Exemplele de mascare NMT și TMN din fig.2.14 demonstrează clar asimetria puterii de mascare dintre un mascator de tip zgomot și unul tonal. În ciuda faptului că ambii mascatori au avut un nivel de 80 dB, pragul SMR corespunzător a fost diferit cu aproximativ 20 dB. Cunoașterea acestei asimetrii se exploatează în modelarea distorsiunilor de codare pentru a nu fi detectate de sistemul auditiv uman. Pentru fiecare interval de analiză temporală, modelul codecului perceptual trebuie să identifice în tot spectrul de frecvență componentele de tip zgomot sau tonal atât din semnalul audio cât și din distorsiunile de codare. Împreună cu efectul de împrăștiere al mascării (descriș mai jos), proprietățile NMT, TMN și NMN se pot folosi pentru a construi un prag global de mascare. Deși metodele curente de estimare a pragului de mascare s-au dovedit eficiente, o aprofundare a înțelegerii asimetriei poate duce la îmbunătățirea modelelor perceptuale. În particular, s-a demonstrat [HAL97] că asimetria mascării se poate explica prin raportul dintre lățimile de bandă ale mascatorului și mascatului și nu numai prin proprietățile absolute ale mascatorului. Aceasta înseamnă că schemele standard de estimare a puterii de mascare, bazate pe energie, folosite în codecurile perceptuale sunt valabile numai atât timp cât lățimea de bandă a mascatorului depășește lățimea benzii mascatului. În cazul în care lățimea benzii mascatului depășește pe cea a mascatorului, calculul mascării va ține cont de măsura înfășurătoarei.

5) *Împrăștierea mascării*: Efectele mascării simultane nu se limitează doar la o singură bandă critică. De exemplu, un mascator având frecvența centrală într-o anumită bandă critică, poate produce efecte predictibile la detecția pragurilor și în alte benzi critice. Acest efect, numit *împrăștierea mascării*, este adesea modelat în aplicații de codare printr-o funcție de împrăștiere aproximativ triunghiulară care are pante de +25 și -10 dB/Bark. O expresie analitică a acestei funcții este dată în [SCH79]:

$$SF_{dB}(x) = 15,81 + 7,5(x + 0,474) - 17,5\sqrt{1 + (x + 0,474)^2} \text{ dB} \quad (2.18)$$

unde  $x$  este exprimat în Barks și  $SF_{dB}(x)$  în dB. După ce s-a făcut analiza benzilor critice și s-a ținut cont de împrăștierea mascării, se calculează pragurile de mascare:

$$TH_N = E_T - 14,5 - B \quad (2.19)$$

$$TH_T = E_N - K \quad (2.20)$$

unde

$TH_N$  și  $TH_T$  pragurile de mascare cauzate de mascatori de tip zgomot (N), respectiv tonal (T)  
 $E_N$  și  $E_T$  energia mascatorilor de tip zgomot (N), respectiv tonal (T) în cadrul benzii critice  
 $B$  numărul benzii critice

În funcție de algoritm, parametrul  $K$  are valori între 3 și 5dB. Evident, pragurile calculate cu relațiile (2.19) și (2.20) țin cont numai de contribuția mascatorilor individuali de tip zgomot sau tonal. În realitate, fiecare cadru conține mai mulți mascatori de ambele tipuri. După identificare, aceste praguri individuale sunt combinate pentru a forma un prag global de mascare. Pragul global de mascare conține o estimare a nivelului la care zgomotul de cuantizare devine sesizabil. În consecință, pragul global de mascare mai este numit și “distorsiunea abia sesizabilă” – JND (*just noticeable distortion*). Metodologia standard în codarea perceptuală parcurge trei etape: clasificarea semnalelor de mascare (zgomot sau tonale), calculul pragurilor corespunzătoare și apoi modelarea spectrului zgomotului astfel să fie sub JND (§2.2.9, ISO/IEC MPEG Model 1 și 2). De remarcat că și pragul absolut de audibilitate ( $Tq$ ) se ia în considerare la modelarea zgomotului, în general pragul de distorsiune permis fiind dat de MAX (JND,  $Tq$ ).

### 2.2.5.2 Mascare temporală

Așa după cum se poate observa din fig. 2.15 fenomenul de mascare se poate extinde și în afara intervalului de timp în care este prezent stimulul (mascatorul). Mascarea temporală poate apărea atunci când mascatorul și mascatul sunt prezente la momente diferite. În funcție de durata mascatorului, timpul de cădere al pragului poate fi între 5 ms (dacă mascatorul este un impuls Gaussian cu durata mai mică de 0,05 ms) și peste 150 ms (dacă mascatorul este un zgomot roz cu o durată de 1 s). Acest efect se numește *post-mascare*.

Semnalele slabe aflate chiar în fața unor semnale puternice pot fi de asemenea mascate. Intervalul de timp în care poate apărea acest efect de *pre-mascare* este de obicei sub 5 ms. Dacă mascatul este chiar deasupra pragului de mascare, el nu este perceput ca un semnal înaintea mascatorului, ci ca o modificare a mascatorului. Efectul pre-mascării diferă mult de la ascultător la ascultător.

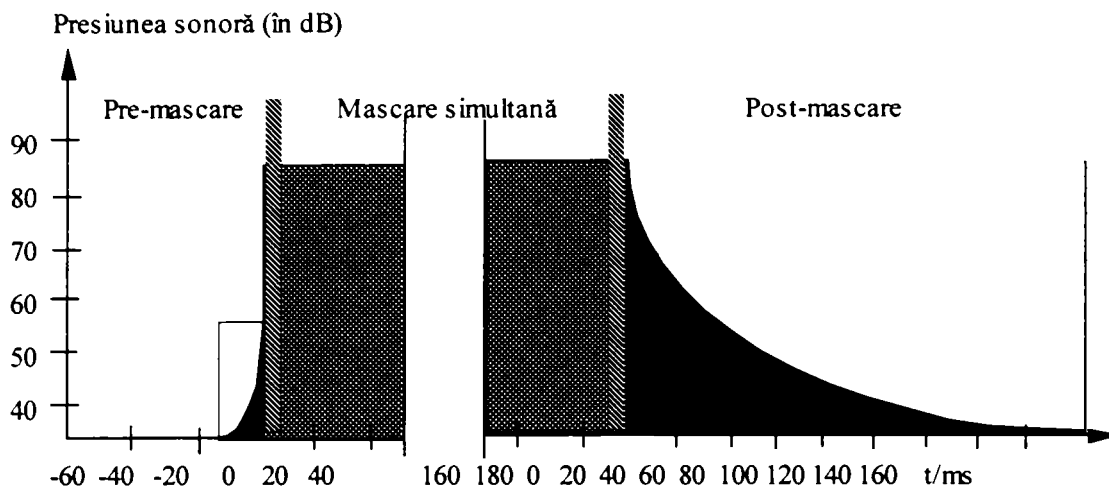


Figura 2.15 Mascarea temporală. Evenimentele acustice care apar în zonele hașurate nu sunt audibile [NOL97]

În continuare vor fi prezentate câteva proprietăți ale mascării temporale care trebuie luate în considerare în cazul codurilor audio perceptuale. Dintre cele două feluri de mascare temporală, postmăscarea este mai bine înțeleasă. În cazul în care mascatorul și mascatul au aceeași frecvență, studiile experimentale au arătat [JES82] că dimensiunea efectului de postmăscare depinde într-un mod predictibil de frecvența, intensitatea și durata mascatorului, întârzierea apariției mascatului față de momentul încetării mascatorului. De asemenea, fenomenul de postmăscare depinde de relația dintre frecvențele mascatorului și mascatului [MOO78]. Deși fenomenul de pre-măscare a făcut și el obiectul multor studii, el este mai puțin bine înțeles. Așa după cum se poate observa din fig. 2.15, fenomenul de pre-măscare are o durată mult mai scurtă. Nu s-a putut stabili o valoare exactă pentru durata maximă de persistență a fenomenului de mascare, aceste măsurări fiind în general subiective și depinzând în mod semnificativ de experiența subiecților. În cazul codării perceptuale trebuie avut în vedere că semnalele audio tranzitorii (atacul unui instrument de percuție) crează în timp zone de pre- și postmăscare în care ascultătorul nu va sesiza semnalele cu un nivel sub pragul de audibilitate produs de mascator. Fenomenul de mascare temporală a fost utilizat în diferiți algoritmi de codare audio (de ex. [BRA94]), iar fenomenul de pre-măscare în particular a fost folosit pentru a compensa distorsiunile de tip preecou (§4.2.2).

### 2.2.6 Entropia perceptuală

În timp ce lucra la laboratoarele Bell, Johnston a combinat noțiunile de mascare psihoacustică cu principiile de cuantizare a semnalelor pentru a defini entropia perceptuală (EP), ca o măsură a informației perceptuale relevante conținută într-o înregistrare audio. Exprimată în biți/eșantion, EP reprezintă limita teoretică de compresie a unui anumit semnal. Majoritatea materialelor audio pot fi transparent comprimate la calitate CD cu o EP de 2,1 bit/eșantion [PAI00]. Procesul de estimare al EP se realizează astfel: semnalul este mai întâi eșantionat într-o fereastră și transformat în domeniul frecvență. Se calculează pragul de mascare după principii perceptuale și apoi se determină numărul de biți necesar pentru a cuantiza spectrul fără a introduce zgomot perceptibil. EP se poate măsura construind histograma EP pentru mai multe cadre și alegând valoarea din cazul cel mai defavorabil.

Transformarea în domeniul frecvență se face cu o fereastră Hann urmată de o transformare FFT cu 2.048 de puncte. Pragurile de mascare se obțin din analiza benzilor critice (cu împrăștiere), determinând natura semnalului (tonal sau zgomot), aplicând regulile de prag în funcție de calitatea semnalului și ținând cont de pragul de audibilitate absolut. Componentele reale și imaginare ale transformării sunt convertite mai întâi în componente spectrale de putere:

$$P(\omega) = \text{Re}^2(\omega) + \text{Im}^2(\omega) \quad (2.21)$$

apoi se formează un spectru Bark discret însumând energia din fiecare bandă critică (Tabel 2.1)

$$B_i = \sum_{\omega = b_l}^{b_h} P(\omega) \quad (2.22)$$

unde limitele sumei sunt marginile benzii critice. Indicele  $i$  depinde de rata de eșantionare, pentru semnale de calitate CD  $i \in \{1, 25\}$ . O operație de convoluție se aplică apoi funcției de împrăștiere (2.18) și spectrului discret Bark

$$C_i = B_i * SF_i \quad (2.23)$$

pentru a ține cont de împrăștierea mascării. Estimarea naturii tonale sau zgomot a  $C_i$  se obține prin măsurarea netezimii spectrale SFM (*spectral flatness measure*):

$$SFM = \frac{\mu_g}{\mu_a} \quad (2.24)$$

unde  $\mu_g$  și  $\mu_a$  sunt mediile geometrică, respectiv aritmetică a componentelor densității spectrale de putere PSD (*power spectral density*) din fiecare bandă. SFM ia valori între 0 și 1. Valorile apropiate de 1 corespund unui spectru plat într-o anumită bandă, indicând o bandă decorelată (zgomotoasă). Valorile apropiate de 0 corespund unui spectru îngust într-o anumită bandă. Din SFM se poate defini un “coeficient de tonalitate”  $\alpha$  măsurat în dB:

$$\alpha = \min\left(\frac{SFM_{dB}}{-60}, 1\right) \quad (2.25)$$

care se folosește pentru a pondera relațiile pragului de mascare (2.19) și (2.20) (cu  $K = 5,5$ ) astfel încât să formeze un offset pentru fiecare bandă:

$$Q_i = \alpha(14,5 + i) + (1 - \alpha)5,5 \text{ (în dB)} \quad (2.26)$$

Un set de estimări ale pragului de distorsiune JND în domeniul frecvență se obține din scăderea offseturilor din componentele spectrale Bark:

$$T_i = 10^{\log_{10}(C_i) - Q_i/10} \quad (2.27)$$

Aceste estimări sunt scalate cu un factor de corecție pentru a simula deconvoluția de funcția de împrăștiere, apoi fiecare  $T_i$  se verifică față de pragul absolut de audibilitate și se înlocuiește cu  $\max(T_i, T_q(i))$ . Valorile obținute pentru EP sunt calibrate, la fel ca și calibrarea nivelului SPL descrisă la începutul paragrafului 2.2, considerând pragul minim absolut egal cu energia unui semnal de 4 kHz cu o amplitudine de  $\pm 1$  bit. Cu alte cuvinte se presupune că nivelul de redare al sistemului este astfel configurat încât amplitudinea minimă a semnalului să fie asociată cu nivelul pragului minim de audibilitate. Aplicând principiile cuantizării uniforme semnalului și setului de valori JND, se poate estima numărul minim de biți necesari pentru o codare transparentă. Valoarea EP în biți/eșantion este dată de:

$$EP = \sum_{i=1}^{25} \sum_{\omega=bl_i}^{bh_i} \log_2 \left( 2 \left\lceil n_{\text{int}} \left( \frac{\text{Re}(\omega)}{\sqrt{6T_i/k_i}} \right) \right\rceil + 1 \right) + \log_2 \left( 2 \left\lceil n_{\text{int}} \left( \frac{\text{Im}(\omega)}{\sqrt{6T_i/k_i}} \right) \right\rceil + 1 \right) \text{ (bit/eșantion)} \quad (2.28)$$

unde

- $i$             indicele benzii critice
- $bl_i$  și  $bh_i$  marginea inferioară și superioară a benzii  $i$
- $k_i$            numărul componentelor transformării în banda  $i$
- $T_i$            pragul de mascare în banda  $i$  (2.27)
- $n_{\text{int}}$         rotunjire la cel mai apropiat întreg

Dacă apare zero în argumentul logaritmului, rezultatul va fi zero. Pragurile de mascare folosite în calculul de mai sus al EP sunt baza algoritmului de codare prin transformări descris în paragraful 3.2.3. Modelul psihoacustic nr. 2 (§2.2.9), folosit de ISO/IEC MPEG-1 în codoarele “.mp3” este strâns legat de procedeul de calcul al EP. De la apariția în 1988, schema de calcul a EP a evoluat, în special prin îmbunătățirea estimării naturii tonale prin măsurarea SFM [RAU93]. Măsurarea netezimii spectrale SFM este limitată atât în timp cât și în frecvență. În timp, se analizează o singură estimare spectrală, care, în frecvență, adună mai multe linii spectrale. Noile scheme de măsură a naturii tonale (de ex. “măsura haosului” [ISO/IEC IS11172-3, 1992]) iau în considerare predictibilitatea în timp a componentelor de frecvență individuale, în funcție de amplitudine și fază. Valoarea prezisă pentru fiecare componentă se compară cu valoarea reală, iar distanța euclidiană se reține ca o măsură a predictibilității. Componentele spectrale cu un grad mare de predictibilitate sunt considerate tonale, iar cele impredictibile sunt considerate zgomot. Coeficientul de tonalitate se calculează cu (2.25).

### 2.2.7 Intensitatea sunetului

Intensitatea sunetului este cantitatea de energie ce trece într-o secundă printr-o suprafață de 1 mp, perpendiculară pe direcția de propagare. Intensitatea se măsoară în  $\text{W}/\text{m}^2$  și are dimensiunea unei puteri specifice. Relația dintre intensitate și presiune este:

$$I = \frac{P^2}{\rho c} \quad (2.29)$$

unde:  $I$  este intensitatea sunetului

$P$  – presiunea sonoră instantanee [ $\text{N}/\text{m}^2$ ]

$\rho$  – densitatea mediului (aerului) [ $\text{g}/\text{cm}^3$ ]

$c$  – viteza de propagare a sunetului în mediu (aer) [ $\text{cm}/\text{s}$ ]

Nivelul de intensitate sonoră a unui sunet de intensitate  $I$  se calculează în funcție de un sunet de o intensitate minimă audibilă (situat la pragul senzației auditive), luat ca nivel de referință  $I_{\text{ref}}$ , cu relația:

$$N = 10 \log \frac{I}{I_{\text{ref}}} \quad [\text{dB}] \quad (2.30)$$

unde  $I_{\text{ref}} = 10^{-16} \text{ W}/\text{cm}^2 = 10^{-12} \text{ W}/\text{m}^2$ , numit și nivel 0 acustic.

Tăria sunetului este caracterul senzației auditive legat de presiunea acustică a unui sunet, care permite ordonarea sunetelor pe o scară de la “slabe” la “puternice”. Nivelul de tărie al unui sunet este nivelul de intensitate sonoră al unei vibrații acustice de 1.000 Hz, apreciat de de un ascultător otologic normal ca având aceeași tărie cu cea a sunetului considerat. Nivelul de tărie

este o noțiune subiectivă și se măsoară în foni. Începând din anul 1956 s-a convenit să se utilizeze și o altă unitate de tărie mai mare, sonul, definită ca unitatea de tărie egală cu tăria unui sunet pur cu frecvența de 1.000 Hz și un nivel de 40 de foni (40 dB) deasupra pragului de audibilitate. [BER71]

În ceea ce privește senzația de tărie a sunetului, pragul diferențial este exprimat prin numărul de dB cu care trebuie să crească sau să scadă intensitatea unui sunet pentru a putea afirma că noul sunet este mai puternic sau mai slab decât cel inițial.

### 2.2.8 Audiția binaurală și localizarea sunetului

La un om cu auzul normal, la audiția directă binaurală, sunetul acționează separat și diferit asupra celor două urechi ale ascultătorului, care transmit semnalul simultan și independent, pe căi diferite, la centrii auditivi din creier unde se formează senzația auditivă.

O însușire importantă a omului constă în posibilitatea localizării în spațiu a unei surse sonore, prin determinarea cu precizie a direcției sursei și aprecierea distanței de la sursă la ascultător. Această însușire deosebită are la bază audiția binaurală și audiția inteligentă și îi permite omului ca atunci când ascultă direct un ansamblu de surse sonore dispuse într-o anumită ordine în spațiu, ca de exemplu o orchestră, să stabilească cu exactitate *câmpul sonor* complex creat de aceste surse, putând localiza în orice moment fiecare instrument fără a-l privi. De asemenea, poate determina mișcarea unei surse sonore în spațiu.

Audiția direcțională se poate împărți în trei planuri: planul lateral de la stânga la dreapta, planul față-spate sau median și planul vertical.

Pentru a înțelege cum sunt percepute pozițiile surselor sonore în planul lateral (atât în față cât și în spatele capului) trebuie să facem distincție între sunetele continue, repetitive (sunete pure) și sunetele tranzitorii sau zgomotele. Aceasta deoarece putem vorbi de o diferență de fază în grade între două semnale dacă sunt continue, repetitive și au o frecvență fundamentală, dar acest concept este irelevant în cazul unor semnale complexe, neperiodice, de scurtă durată. Diferența de fază între cele două urechi în cazul auzului direcțional se traduce de fapt prin diferența de timp a sosirilor la cele două urechi a unui același front de undă.

Din punct de vedere acustic un obiect devine un obstacol efectiv în calea sunetului atunci când lungimea de undă a sunetului este mică în raport cu dimensiunile obiectului. Deoarece lungimea de undă a sunetelor audibile este cuprinsă între 18 m și câțiva cm, se poate observa că pentru frecvențe joase capul este prea mic pentru a acționa ca o barieră, dar în domeniul frecvențelor înalte (peste 3.000 Hz) diferența de nivel între cele două urechi este cauza definirii direcției deoarece, prin faptul că lungimea de undă devine comparabilă cu dimensiunile capului, fenomenul de difracție și de atenuare a undei este mult mai evident. Pentru asemenea sunete capul ascultătorului constituie un obstacol efectiv. La frecvențe înalte, diferențele de intensitate sonoră la cele două urechi pot fi de ordinul 20 – 30 dB. Pentru semnale sinusoidale s-a observat că precizia maximă de localizare a sursei de sunet se obține pentru frecvențe mai mari de 5.000 Hz. În cazul unor frecvențe mai mici decât aceasta va exista o abatere în determinarea direcției sursei fictive, imaginea sonoră situându-se lateral, la un unghi mai mic decât cel din realitate. Devierea față de poziția reală va fi cu atât mai mare cu cât frecvența va fi mai mică.

În cazul unui același sunet emis de două surse diferite, Haas (1951), studiind efectul ecourilor asupra percepției direcționale, a arătat că direcția percepută a sunetului se îndreaptă spre sursa mai avansată în timp și că fenomenul poate fi sesizat până la diferențe de timp de aproximativ 50 ms. În cazul a două surse ambele urechi vor auzi ambele surse și astfel situația nu va mai fi aceeași ca în cazul unei singure surse în care unda sonoră ajunge la urechi cu o mică diferență de timp. Haas a arătat că pentru diferențe de până la 50 ms sunetele de la cele două surse sunt amestecate de creier, apărând ca o singură sursă percepută ca având direcția sursei al cărei sunet a ajuns primul. Peste 50 ms creierul sesizează sunetele distincte, cel de-al doilea

apărând ca ecou al primului. În cazul sunetelor scurte și nerepetitive efectul dispare mult mai repede decât în cazul sunetelor complexe, fiind observat până la aproximativ 5 ms.

În realitate sunetele sunt complexe, fiind formate din componente cu frecvențe și intensități diferite. Componenta spectrală a unui semnal complex conține atât componente de frecvențe joase cât și componente de frecvențe înalte, localizarea sursei făcându-se atât pe baza diferenței de timp cât și pe baza diferenței de intensitate.

În figura 2.16 este prezentată contribuția diferenței de timp ( $\Delta t$ ) și a diferenței de intensitate ( $\Delta I$ ) în localizarea surselor sonore, în funcție de frecvență. De aici rezultă că la localizarea surselor sonore cu frecvențe până la 800 Hz are importanță numai diferența de timp (diferența de fază). În domeniul frecvențelor medii (800 – 2.000 Hz) se folosesc ambele metode, iar la frecvențe peste 2.000 Hz contează numai diferența de intensitate. Datorită diferențelor de timp și a diferențelor de intensitate a componentelor spectrale ale semnalelor percepute de cele două urechi, rezultă că se poate localiza orice sursă de semnal complex, însă se produce o modificare a timbrului sunetelor complexe care excită cele două urechi, rezultând senzația unui sunet complex cu timbru diferit de cel real.

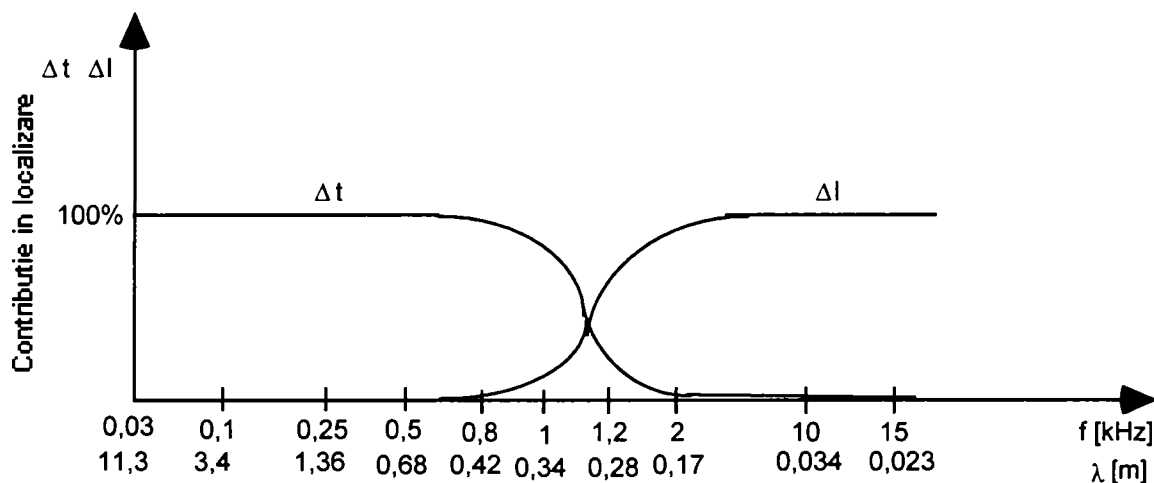


Figura 2.16 Contribuția diferențelor de timp și de intensitate în localizarea imaginii sonore [GRI 91]

Diferența dintre timbrul sunetului perceput de ascultător și timbrul real al acestuia este cu atât mai mare cu cât unghiul de incidență format de direcția sursei cu planul median al capului este mai mare. Rezultă că, pentru a nu altera sensibil timbrul semnalului complex perceput, se recomandă ca acest unghi să nu depășească  $60^\circ$ , ceea ce este suficient pentru cele mai complexe surse sonore întâlnite în practică.

Experiența arată că stabilirea direcției surselor de sunete pure este mult mai dificilă decât aprecierea direcției sunetelor complexe sau zgomote, fiind mai precisă pentru sunetele pure cu frecvențe peste 5.000 Hz [BER71].

## 2.2.9 Modelul psihoacustic

Modelul psihoacustic analizează semnalul audio și calculează cantitatea disponibilă de zgomot de mascare în funcție de frecvență. Capacitatea de mascare a componentei unui semnal dat depinde de poziția frecvenței, tăria și tonalitatea. Codorul folosește aceste informații pentru a decide modul optim de reprezentare a semnalului audio de intrare cu numărul de biți disponibil. Standardul MPEG Audio conține două exemple de implementare a modelului psihoacustic. Modelul 1 este mai puțin complex decât modelul 2, făcând câteva compromisuri pentru simplificarea calculelor. Ambele modele pot fi folosite la cele trei niveluri de codare MPEG, totuși numai modelul 2 include modificări specifice nivelului 3.



Implementarea modelului psihoacustic are o mare libertate, precizia fiind dependentă de gradul de compresie și aplicația dorită. În continuare se va face o trecere în revistă a pașilor necesari pentru calculul modelelor psihoacustice, cu evidențierea diferențelor dintre ele.

### 1. Alinierea în timp a datelor audio

La fiecare cadru se face o evaluare psihoacustică. Datele audio trimise modelului psihoacustic trebuie să fie concurente cu datele audio ce trebuie codate. Modelul psihoacustic trebuie să ia în considerare atât întârzierea datelor prin banca de filtre cât și offsetul cu care datele relevante sunt centrate în cadrul ferestrei psihoacustice de analiză. De exemplu, utilizarea modelului 1 pentru Nivelul-1 introduce o întârziere de 256 de eșantioane la trecerea prin bancul de filtre și un offset de 64 de puncte pentru a centra cele 384 de eșantioane ale cadrului de nivel-1 în fereastra de analiză cu 512 puncte  $(512 - 384)/2 = 64$ ). Rezultă un offset net de 320 de puncte pentru a alinia în timp datele modelului psihoacustic cu ieșirile bancului de filtre.

### 2. Conversia datelor audio în domeniul frecvență

Modelul psihoacustic folosește o transformare timp-frecvență, separată și independentă de bancul de filtre polifază, pentru că are nevoie de o rezoluție mai bună în frecvență pentru calculul pragurilor de mascare. Ambele modele psihoacustice folosesc în scopul evaluării structurii spectrale a semnalului audio o transformare Fourier rapidă (FFT – *Fast Fourier Transform*). FFT constituie un algoritm de calcul eficient, care face să corespundă unui șir de  $N$  eșantioane provenind din semnalul analogic analizat un nou șir de numere complexe (*linii spectrale*) de aceeași lungime  $N$ . De regulă  $N$  este o putere a lui 2, de exemplu  $N = 2^{10} = 1.024$  [STA99].

*Modelul psihoacustic 1* folosește o fereastră de analiză cu 512 eșantioane pentru Nivel-1 și 1024 de eșantioane pentru Nivelurile 2 și 3. Deoarece un cadru de Nivel-1 are numai 384 de eșantioane, fereastra cu 512 eșantioane este suficientă, reducând volumul de calcul. Nivelurile 2 și 3 folosesc un cadru cu 1.152 de eșantioane, astfel că fereastra cu 1.024 de eșantioane nu acoperă complet numărul de eșantioane dintr-un cadru. Deși cazul ideal ar fi ca fereastra de analiză să acopere toate eșantioanele ce trebuie codate, totuși dimensiunea de 1.024 de eșantioane reprezintă un compromis rezonabil. Eșantioanele care rămân în afara ferestrei de analiză nu au în general un impact major asupra evaluării psihoacustice.

*Modelul psihoacustic 2* folosește o fereastră cu 1.024 de eșantioane pentru toate nivelurile. Pentru Nivelul 1, modelul centrează cele 384 de eșantioane audio în fereastra psihoacustică, în modul descris mai sus. Pentru nivelurile 2 și 3 modelul efectuează două calcule de 1.024 de puncte pentru fiecare cadru: mai întâi sunt centrate primele 576 de eșantioane din cadru și apoi cealaltă jumătate. Modelul combină rezultatele celor două calcule folosind rapoartele semnal-mascare cele mai mari din fiecare subbandă, ceea ce are ca efect alegerea pragului de zgomot de mascare cel mai redus din fiecare subbandă.

### 3. Prelucrarea valorilor spectrale în grupări legate de lățimea benzilor critice

Pentru simplificarea calculelor psihoacustice, ambele modele prelucrează valorile frecvențelor în cuante perceptuale.

### 4. Separarea valorilor spectrale în componente tonale și netonale

Ambele modele identifică și separă componentele tonale de cele de tip zgomot din semnalul audio, datorită caracteristicilor diferite de mascare.

*Modelul psihoacustic 1* identifică componentele tonale bazate pe vârfurile locale din spectrul de putere audio. Maximele locale care depășesc componentele vecine cu cel puțin 7 dB sunt clasificate tonale. Setul “tonal”  $S_T$  se definește astfel:

$$S_T = \left\{ P(k) \left| \begin{array}{l} P(k) > P(k \pm 1), \\ P(k) > P(k \pm \Delta_k) + 7dB \end{array} \right. \right\} \quad (2.31)$$

unde

$$\Delta_k \in \begin{cases} 2 & 2 < k < 63 & (0,17 - 5,5\text{kHz}) \\ [2,3] & 63 \leq k < 127 & (5,5 - 11\text{kHz}) \\ [2,6] & 127 \leq k \leq 256 & (11 - 20\text{kHz}) \end{cases} \quad (2.32)$$

Mascatorii tonali  $P_{TM}(k)$  sunt calculați din vârfurile spectrale aflate în  $S_T$ :

$$P_{TM}(k) = 10 \log_{10} \sum_{j=-1}^1 10^{0,1P(k+j)} \text{ (dB)} \quad (2.33)$$

Cu alte cuvinte, pentru fiecare vecinătate a unui maxim, se combină energia a trei componente spectrale adiacente, având vârful în centru, pentru a forma un singur mascator tonal.

După prelucrarea componentelor tonale, modelul 1 însușește valorile spectrale rămase în afara vecinătății  $\pm \Delta_k$  a mascatorului tonal într-o singură componentă netonală în fiecare bandă critică.

$$P_{NM}(\bar{k}) = 10 \log_{10} \sum_j 10^{0,1P(j)} \text{ (dB)}, \quad \forall P(j) \notin \{P_{TM}(k, k \pm 1, k \pm \Delta_k)\} \quad (2.34)$$

unde  $\bar{k}$  este media geometrică liniilor spectrale din banda critică respectivă.

$$\bar{k} = \left( \prod_{j=l}^u j \right)^{1/(l-u+1)} \quad (2.35)$$

unde  $l$  și  $u$  sunt liniile spectrale de la marginea de jos, respectiv de sus a benzii critice. S-a considerat că energia spectrală reziduală dintr-o bandă critică, care nu a fost asociată cu un mascator tonal, se asociază cu un mascator netonal (zgomot).

*Modelul psihoacustic 2* nu separă componentele tonale și netonale, ci calculează un index de tonalitate funcție de frecvență. Modelul 2 folosește acest index pentru a interpola între zgomotul de mascare pentru tonuri pure sau zgomote. Indexul de tonalitate se bazează pe o măsură a predictibilității. Modelul 2 folosește datele din două ferestre de analiză anterioare pentru a prezice, prin extrapolare liniară, valorile componente din fereastra curentă. Componentele tonale sunt mai predictibile și au un indice de tonalitate mai mare. Acest proces se bazează pe mai multe date, fiind mai precis decât separarea componentelor tonale și netonale din modelul 1.

##### 5. Aplicarea funcției de împrăștiere

Capacitatea de mascare a unui semnal se întinde și asupra benzilor critice învecinate. Modelul determină pragurile de mascare prin aplicarea asupra componentelor semnalului a unei mascări determinate empiric, în cazul modelului 1, sau a unei funcții de împrăștiere, în cazul modelului 2.

##### 6. Alegerea marginii inferioare a valorilor de prag

Ambele modele includ un prag de mascare absolut determinat empiric, numit pragul de liniște. Acest prag reprezintă marginea inferioară de audibilitate a sunetului.

- Găsirea pragului de mascare pentru fiecare subbandă. Ambele modele psihoacustice calculează pragurile de mascare cu o rezoluție de frecvență mai mare decât a bancului de filtre polifazate. Ambele modele calculează o valoare de prag obținută din pragurile de mascare ale frecvențelor din subbanda respectivă.

*Modelul psihoacustic 1* selectează pragul de mascare minim din fiecare subbandă. Acest concept este valabil pentru frecvențe joase, unde subbenzile sunt apropiate ca lățime de benzile critice, dar poate fi imprecis pentru frecvențe înalte, unde într-o bandă critică sunt cuprinse mai multe subbenzi. Imprecizia crește pentru că modelul 1 concentrează toate componentele netonale dintr-o bandă critică într-o singură valoare, la o singură frecvență. De fapt modelul 1 transformă componentele netonale într-o formă de componentă tonală. O subbandă din interiorul unei benzi critice, aflată la distanță de componenta netonală concentrată, nu va avea parte de o evaluare corectă a mascării netonale. Acesta este un compromis pentru a reduce volumul de calcul.

*Modelul psihoacustic 2* alege pragul de mascare minim din subbandă, dacă lățimea subbenzii este mai mare decât a benzii critice corespunzătoare și folosește o medie a pragurilor de mascare, când lățimea subbenzii este apropiată de lățimea benzii critice. Precizia modelului 2 nu scade la subbenzile de la frecvențe înalte pentru că nu concentrează componentele netonale.

### 7. Calculul raportului semnal/mascare

Modelul psihoacustic calculează raportul semnal/mascare ca un raport al energiei semnalului din interiorul subbenzii (sau pentru nivel-3, un grup de benzi) cu pragul minim de mascare din subbandă. Modelul transmite această valoare secțiunii de alocare a biților din codor.

## 2.3 Concluzii

Conform DEX, "audibil" poate fi simplu definit "care poate fi auzit".

Există două tipuri de praguri de audibilitate:

- pragul de audibilitate (absolut), definit ca fiind energia necesară unui ton pur pentru a putea fi detectat de un ascultător într-un mediu lipsit de alte zgomote;
- pragul diferențial (relativ), exprimat prin cantitatea cu care trebuie să varieze (în plus sau în minus) un stimul pentru a determina o modificare *distinctă* a sunetului inițial.

Sistemul auditiv efectuează o analiză spectrografică a stimulilor auditivi. Membrana basilară poate fi considerată ca un grup de filtre cu ieșirile ordonate tonotopic, astfel încât se realizează o transformare frecvență-loc.

O însușire importantă a omului constă în posibilitatea localizării în spațiu a unei surse sonore, prin determinarea cu precizie a direcției sursei și aprecierea distanței de la sursă la ascultător. Această însușire deosebită are la bază audiția binaurală și audiția inteligentă și îi permite omului ca atunci când ascultă direct un ansamblu de surse sonore dispuse într-o anumită ordine în spațiu, ca de exemplu o orchestră, să stabilească cu exactitate *câmpul sonor* complex creat de aceste surse, putând localiza în orice moment fiecare instrument fără a-l privi. De asemenea, poate determina mișcarea unei surse sonore în spațiu.

Codoarele audio perceptuale realizează compresia datelor exploatând faptul că într-un semnal există informație nerelevantă din punct de vedere acustic, care nu poate fi detectată nici de cel mai sensibil și antrenat ascultător. Informația nerelevantă este identificată în timpul analizei semnalului prin introducerea în codor a mai multor principii psihoacustice: pragul de audibilitate, analiza frecvenței în benzi critice, mascarea simultană, împrăștierea mascării de-a lungul membranei basilare și mascarea temporală. Combinarea acestor noțiuni de psihoacustică cu proprietățile de bază ale cuantizării semnalului a condus la teoria entropiei perceptuale, o estimare cantitativă a limitelor fundamentale ale compresiei transparente a semnalelor audio.

# CAPITOLUL 3

## COMPRESIA SEMNALELOR AUDIO

Algoritmii de codare audio sau de compresie audio se folosesc pentru a obține reprezentări digitale compacte ale semnalelor audio de înaltă fidelitate (bandă largă), cu scopul de a fi transmise sau stocate eficient. Obiectivul central al codării audio este reprezentarea semnalului cu un număr minim de biți și reproducerea transparentă a semnalului, adică generarea unui semnal audio de ieșire care să nu poată fi distins de semnalul original de la intrare. În acest capitol vor fi prezentați algoritmi utilizați pentru codarea audio transparentă.

Prima generație de aplicații audio digitale (CD și DAT) au adus o serie de avantaje fără precedent în lumea analogică, dar cu prețul unor debite binare foarte mari. A doua generație de aplicații multimedia și în special sistemele de transmisie fără fir, sunt supuse unor limitări ale lățimii de bandă și a costurilor, incompatibile cu debitele mari de date. Pe de altă parte, datorită succesului primei generații, utilizatorii așteaptă o redare de “calitate CD” de la orice sistem digital. De aceea noile sisteme audio multimedia cu transmitere în rețea sau fără fir trebuie să reducă debitul de date fără a compromite calitatea redării.

Deși procesul reducerii cantității de date necesare pentru a reprezenta un semnal audio este considerat în general o *compresie de date*, descrierea mai exactă ar fi de *reducere a datelor*. Compresia implică un proces reversibil (prin expandarea materialului comprimat se obține originalul), dar majoritatea strategiilor de reducere a datelor sunt cu *pierderi*, ceea ce înseamnă că anumite date au fost eliminate și sunt irecuperabile. Un sistem fără pierderi (și acestea sunt puține) este cel în care anumite date din original nu au fost codate, dar pot fi recreate perfect la redare - adevărata *compresie a datelor*.

Reducerea datelor nu este specifică numai domeniului audio digital, metodele de reducere a zgomotului utilizate în sistemele Dolby și dbx fiind metode de reducere a datelor în domeniul analogic.

Reducerea datelor se folosește în camerele digitale, DVD, disc laser, samplare, stații de lucru digitale, sisteme de înregistrare multipistă pe harddisk, rețele de radio și televiziune, sisteme de telefonie și multe alte aplicații.

### 3.1 Compresia semnalelor audio

Cei patru parametri fundamentali ai codării sunt: **calitatea semnalului**, **debitul binar**, **întârzierea la prelucrare** și **complexitatea implementării**. Scopul principal al codării este scăderea debitului binar, menținând însă un anumit nivel de calitate. În general, nici timpul de prelucrare și nici complexitatea implementării nu trebuie să depășească un anumit nivel.

Pentru a avea o idee asupra eficienței acestor sisteme trebuie menționat că metodele de codare fără pierderi (cel puțin în audio) nu pot să depășească un raport de 2:1, pe când cele cu pierderi trec ușor de 4:1, ajungând la 12:1 și chiar 20:1. De regulă, cu cât raportul de reducere a datelor este mai mare, cu atât deteriorările și artefactele devin mai sesizabile, deși ultima generație de MPEG2, cu un raport de reducere de 12:1, sună foarte bine, dacă nu chiar “transparent”.

Există patru metode de reducere a datelor plecând de la un semnal liniar PCM: reducerea ratei de eșantionare, reducerea rezoluției de cuantizare (număr de biți), eliminarea datelor redundante și eliminarea datelor irelevante. Primele două metode permit reduceri modeste dacă

se dorește păstrarea unei calități acceptabile, cele mai eficiente și avansate sisteme bazându-se pe ultimele două metode.

Una din metodele cele mai simple de reducere a datelor constă în acceptarea unei benzi de frecvențe audio reduse și deci utilizarea unei frecvențe de eșantionare mai mici. Această metodă a fost adoptată de diferite procesoare de efecte și de sistemele digitale de emisie radio și TV. Aceasta este o tehnică cu pierderi pentru că frecvențele înalte, odată eliminate, nu mai pot fi rediate. Evident, cu cât scade frecvența de eșantionare, cu atât mai mult se vor reduce datele, dar cu reducerea corespunzătoare a calității audio. Am realizat o evaluare obiectivă a efectului scăderii frecvenței de eșantionare asupra calității audio percepute în §7.2.

Posturile de radio ce emit în MF folosesc sisteme analogice ce permit transmisia unei benzi audio de 15 kHz, astfel că frecvența de eșantionare de 32 kHz, folosită de formatele digitale NICAM și DAB este perfect acceptabilă și reprezintă o economie de 40% a debitului de date față de semnalul inițial de 48 kHz. În cazul înregistrărilor muzicale însă, frecvența de eșantionare trebuie aleasă cu mare grijă pentru a nu afecta structura armonică a materialului sonor original.

O altă tehnică este reducerea numărului nivelurilor de cuantizare, respectiv a numărului de biți necesar pentru descrierea fiecărui eșantion, dar această tehnică are ca efect creșterea zgomotului de cuantizare și scăderea gamei dinamice. Și aceasta este o tehnică cu pierderi pentru că anumite detalii sonore de nivel scăzut nu pot fi cuantizate într-un sistem cu rezoluție redusă. Am realizat o evaluare obiectivă a efectului reducerii nivelurilor de cuantizare asupra calității audio percepute în §7.3.

Deși această lucrare se referă la semnale audio, aceleași principii și performanțe ale codării se aplică și transmisiei de date. Diferențele unanim recunoscute ar fi că o transmisie de date trebuie să fie în general fără pierderi, implicând o calitate perfectă a semnalului, iar timpul de prelucrare, întârzierea produsă în procesul de codare, este mai puțin importantă decât pentru semnale audio.

### 3.1.1 Debite binare

Tabelul 3.1 definește banda de frecvențe și debitele binare în comunicațiile audio. Frecvența de eșantionare respectă în fiecare caz teorema lui Nyquist, fiind cel puțin dublul frecvenței maxime.

Debitul binar se măsoară în general în biți pe secundă. Numărul de biți pe secundă reprezintă produsul frecvenței de eșantionare (măsurată în Hz) și numărul mediu de biți pe eșantion folosiți de sistemul de cuantizare al codorului.

Tabel 3.1 Valori standard pentru semnale audio

Format audio	Frecvența de eșantionare	Banda de frecvență
Telefonie	8 kHz	200 – 3.400 Hz
Teleconferința	16 kHz	50 – 7.000 Hz
Compact disc (CD)	44,1 kHz	20 – 20.000 Hz
Digital audio tape (DAT)	48 kHz	20 - 20.000 Hz
Minidisc (MD)		

De exemplu, conform frecvenței de eșantionare din tabelul 3.1, la 8 biți pe eșantion în cazul telefoniei, rezultă un debit de 64 kbit/s și la 16 biți pe eșantion pentru fiecare canal audio (stânga și dreapta) în cazul formatului CD stereo, rezultă un debit de 1.410 kbit/s.

Există și sisteme care folosesc un debit binar variabil, asigurând o calitate audio constantă. Codarea cu debit binar variabil se recomandă în special la transmisii ce utilizează pachete de date.

### 3.1.2 Calitatea semnalului

Datorită faptului că beneficiarul și judecătorul final al unui sistem de compresie audio este urechea umană (ascultătorul), calitatea semnalului audio este cel mai bine descrisă de criterii subiective. Calitatea semnalului audio este evaluată pe o scară cu 5 trepte [ITU93] (vezi cap. 4).

### 3.1.3 Timpul de prelucrare

Timpul de prelucrare în procesul de codare este suma timpilor necesari în fazele de codare și decodare ale procesului. La codare, întârzierea se produce în procesul de analiză a semnalului pentru reducerea redundanței, de exemplu prin metodele de prelucrare a blocurilor de transformări liniare sau cuantizare vectorială. La decodare, întârzierile sunt produse de transformările inverse aplicate blocurilor.

O altă sursă de întârzieri în telecomunicații este transmiterea în rețea a semnalelor digitale. Aplicațiile în timp real, precum telefonía sau teleconferința necesită cele mai mici întârzieri, în funcție de calitatea dorită a semnalului și, respectiv, algoritmul de compresie folosit. În comunicațiile într-o singură direcție, precum emisia, întârzierile sunt mai puțin importante. În aplicațiile de stocare, timpul de codare este total irelevant, fiind mai importantă calitatea serviciului și timpul de decodare.

### 3.1.4 Complexitatea

Complexitatea se măsoară atât prin capacitatea de prelucrare aritmetică cerută de algoritm (măsurată în milioane de instrucții pe secundă - mips) cât și memoria necesară (kbytes de ROM sau RAM). În general, utilizarea Mips (sau Mflops) ca o măsură a complexității se face în cazul calculatoarelor sau procesoarelor de semnal (DSP). În cazul circuitelor integrate specializate (ASIC - *application-specific integrated circuits*) se folosesc alte unități de măsură a complexității, precum numărul de tranzistori sau porți.

Complexitatea este un parametru important al performanței sistemului de codare din două motive: necesitatea de reducere a costurilor și reducerea puterii disipate, în special la dispozitivele portabile.

În aplicații precum emisia, este importantă simplificarea decodării, complexitatea codării fiind neesențială.

Complexitatea unui cip disponibil (în putere de calcul și memorie) a crescut exponențial în ultimii ani, ceea ce permite utilizarea unor algoritmi tot mai sofisticăți pentru codarea semnalelor audio.

## 3.2 Tehnici de codare audio

Pentru a crea un echilibru între timpul de prelucrare și complexitatea prelucrării se folosesc mai multe sisteme de codare. Totuși, compresia semnalului se bazează pe două principii de bază:

- a) - eliminarea redundanțelor din semnalul sursă
- b) - adaptarea sistemului de cuantizare la proprietățile mecanismului uman de percepție auditivă.

În cazul datelor, compresia se bazează numai pe eliminarea redundanțelor.

Semnalele vocale au un model universal bine pus la punct care permite codarea liniară predictivă în domeniul timp. Pentru codarea semnalelor audio de bandă largă se folosește codarea în domeniul frecvență utilizând un model de alocare neuniformă a biților.

La codarea semnalelor audio sistemul de cuantizare poate folosi fenomenul perceptual numit “formarea zgomotului” (NS - *noise-shaping*) pentru a obține reducerea debitului binar.

Prin fenomenul de mascare un stimul puternic (semnalul dorit) acoperă complet un semnal mai slab (zgomot de cuantizare) într-o vecinătate spectrală sau spațio-temporală. Cu alte cuvinte, distorsiuni de cuantizare cu un nivel matematic semnificativ au un efect redus sau chiar nu au nici un efect asupra semnalului audio perceput. Cele mai mari progrese s-au realizat în compresia audio de bandă largă. Spre deosebire de semnalul vocal, semnalele audio de bandă largă nu au un model universal sau un cadru pentru reducerea redundanței. Pentru compresia acestor semnale se folosesc mai întâi tehnicile clasice pentru reducerea cât mai mult a redundanței, dar importanța cea mai mare o are codarea perceptuală. Se definește un prag de zgomot perceptibil, funcție de frecvență, și vor fi codate numai componentele semnalului care depășesc acest prag (vezi cap. 4). Codorul audio perceptual poate realiza un raport de compresie de peste 20:1, un exemplu fiind compresia unui semnal stereo de pe CD cu debitul de 1.410 kbit/s la un debit de aprox. 64 kbit/s fără o degradare semnificativă a calității subiective a semnalului. Calitatea subiectivă a semnalului audio codat este însă dependentă și de semnalul de intrare. De remarcat că există limite fundamentale în codarea semnalului, determinate de entropia perceptuală (§2.2.6), debitul binar cel mai scăzut posibil pentru un anumit nivel de calitate. Entropia perceptuală este funcție de semnalul de intrare și de mecanismul uman de percepție auditivă.

### 3.2.1 Codarea predictivă

O metodă fundamentală, care elimină redundanța dintr-un semnal audio, este codarea predictivă. Ideea de bază folosită în tehnicile predictive constă în înlăturarea redundanței mutuale dintre două eșantioane consecutive și codarea doar a informației noi. Codarea predictivă a fost utilizată cu succes în procesul de codare a semnalelor vocale și imaginilor. Predictorii de la emițător și receptor pot fi făcuți oricât de inteligenți dorim, cu condiția să lucreze în sincronism. Aceasta înseamnă că ei funcționează determinist conform cu același set de reguli și că au acces la aceeași bază de date.

Codarea predictivă poate fi explicată prin descrierea succintă a tehnicii de modulație diferențială a impulsurilor în cod DPCM (*Differential Pulse Code Modulation*), a cărei schemă bloc este prezentată în figura 3.1. Din semnalul de intrare, codat PCM, se extrage ieșirea predictorului și diferența se recuantizează, valoarea numerică fiind divizată cu mărimea noului pas. Diferența recuantizată devine semnalul de ieșire a codorului. Pentru a reduce această diferență, eroarea de recuantizare trebuie trimisă înapoi predictorului. Diferența recuantizată este adusă la valoarea numerică inițială (plus eroarea de recuantizare) prin înmulțirea cu pasul folosit în recuantizator. Aceasta este funcția cuantizorului invers. Diferența cuantizată invers se aplică la ieșirea predictorului, astfel încât următoarea intrare în predictor reprezintă semnalul de intrare curent plus eroarea de recuantizare.

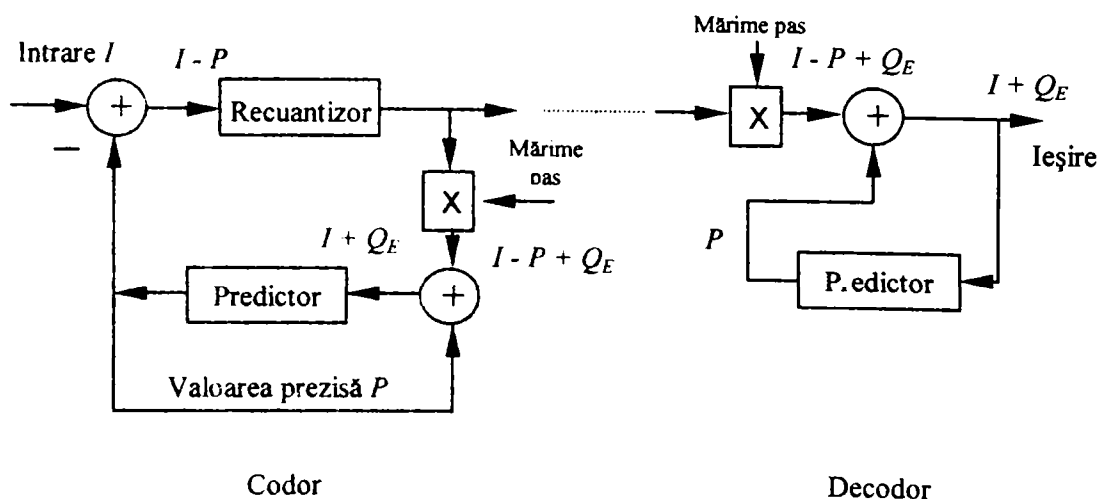


Figura 3.1 Codor predictiv

În cazul DPCM predicția se bazează pe ieșirea cuantizorului (eșantioane cuantizate) și nu pe intrarea acestuia (eșantioane necuantizate). De aici rezultă că predictorul se va găsi într-o buclă de reacție de la ieșirea la intrarea cuantizorului, astfel încât eroarea de cuantizare la un anumit moment dat să fie transmisă la intrarea cuantizorului pentru calculul erorii eșantionului următor. Acest lucru are un efect stabilizator care previne acumularea de erori în semnalul reconstruit [VLA97].

Decodorul conține același predictor și cuantizor invers (bucla de predicție) ca și codorul. Cuantizarea inversă redă diferențelor valoarea numerică corectă. Predictorul este comandat de ieșirea sistemului și va produce aceeași predicție ca și codorul. Diferența se adaugă la ieșirea predictorului pentru a recrea valoarea originală a eșantionului, care devine următoarea intrare a predictorului.

Deoarece semnalele audio sunt în mod fundamental nestaționare, adaptarea codorului și decodorului la caracteristicile semnalului poate aduce anumite avantaje în procesul de codare. O schemă PCM diferențială poate avea un cuantizor adaptiv sau un predictor adaptiv sau amândouă, obținându-se o modulație diferențială adaptivă a impulsurilor în cod (ADPCM - *Adaptive Differential Pulse Code Modulation*). Principiul de bază al adaptării este de a avea în codor (uneori chiar și în decodor) calcularea unei estimări locale pentru unele aspecte statistice ale semnalului audio ca energia, spectrul sau înălțimea. În cazul codorului DPCM adaptarea cuantizorului asigură îmbunătățirea cea mai semnificativă a calității audio la un debit dat.

O reducere suplimentară a datelor se poate obține folosind pași de recuantizare de mărime variabilă. Dacă semnalul diferență este mare, mărimea pasului va crește pentru a preveni saturația. Dacă diferența este mică, pasul se reduce pentru a crește rezoluția. Pentru a evita transmiterea mărimii pasului împreună cu diferența, recuantizorul poate fi făcut adaptiv. În acest caz mărimea pasului se bazează pe mărimea ieșirilor anterioare ale recuantizorului. Dacă ieșirea crește spre valoarea maximă de cod, crește și mărimea pasului pentru a evita saturația.

În figura 3.2 este prezentat modul în care un cuantizor adaptiv poate fi inclus într-un codor predictiv. Evident același pas trebuie aplicat cuantizorului invers din codor și din decodor.

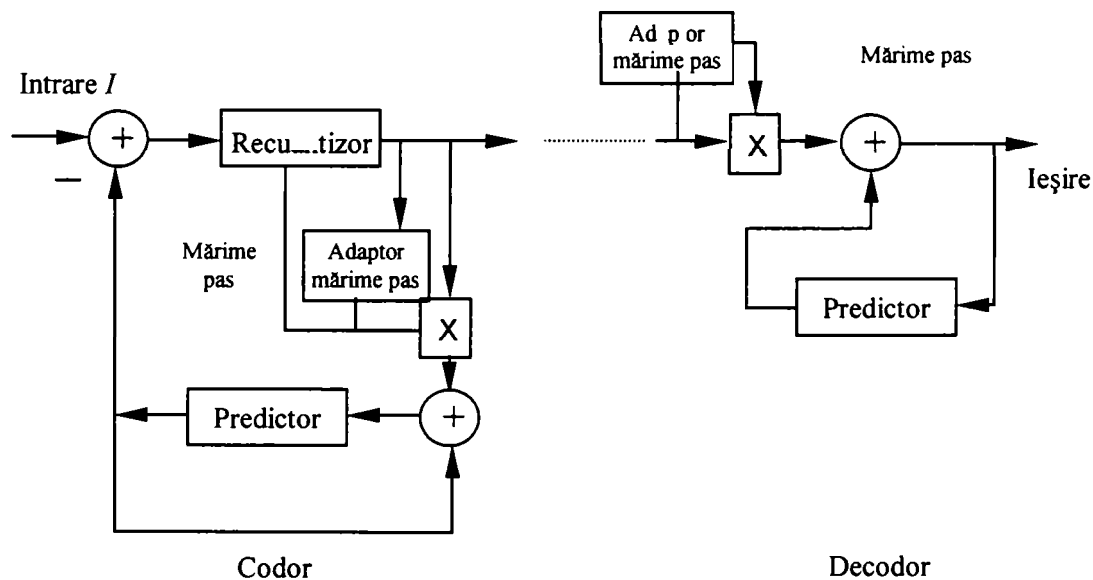


Figura 3.2 Codor predictiv cu cuantizor adaptiv

Există două metode de a distribui informația de adaptare între codor și decodor. În cazul adaptării înainte (*forward-adaptive*), valoarea estimării se calculează din eșantioanele de intrare de la emițător. Semnalul de intrare trebuie să fie memorat, iar valoarea estimării (sau un echivalent codat) trebuie să fie transmisă la receptor. În cazul adaptării înapoi (*backward-adaptive*), valoarea estimării se calculează din eșantioanele cuantizate transmise independent de



emițător și de receptor. Rezultă că în acest caz nu este necesară nici o informație colaterală despre adaptare.

Codoarele predictive acționează pentru găsirea redundanțelor din semnalul audio. De exemplu nota fundamentală a unui instrument muzical va fi un semnal sinusoidal care își modifică lent amplitudinea. Predictorul va învăța repede înălțimea sunetului și-l va elimina din fluxul de date, care va conține numai armonicile. Acestea au un nivel (intensitate) mai mic decât fundamentală și sunt reprezentate de mai puține date. La apariția unui semnal tranzitoriu, predictorul va fi nepregătit și va apărea o diferență foarte mare. Cuantizorul adaptiv va reacționa prin mărirea pasului, pentru a preveni saturația.

Codoarele DPCM au avantajul că lucrează cu forma de undă în domeniul timp și au nevoie de o istorie relativ scurtă a semnalului pentru a funcționa. De aceea procesul de codare/decodare produce întârzieri relativ mici. Un alt avantaj este că datele diferențiale sunt mai puțin sensibile la erori decât cele PCM, semnalele diferență reprezentând doar o mică parte din amplitudinea finală a semnalului. De asemenea, la același debit de codare și același cuantizor, tehnica DPCM produce o creștere de aproximativ 6 dB a raportului semnal/zgomot față de un semnal PCM. Cu un cuantizor adaptiv se poate obține un câștig mai mare de 5 dB a raportului semnal/zgomot față de sistemul DPCM cu cuantizor fix [FEH93].

Un prim pas în direcția reducerii debitului de date audio a fost tehnica de codare aproape instantanee NICAM (*near-instantaneously companded audio multiplex*). De exemplu BBC a folosit formatul de codare NICAM 728 pentru transmiterea sunetului în rețelele de televiziune și formatul de codare NICAM 3 pentru legăturile dintre studiourile de radio și emițătoarele MF stereo. Tehnica NICAM este un exemplu de codare PCM cu adaptare înainte a cuantizorului. De exemplu NICAM 728 folosește o rată de eșantionare de 32 kHz cu o cuantizare inițială pe 14 biți, urmată de o condensare a informației pe 10 biți în blocuri de 1 ms. Trei biți de informație laterală sunt folosiți pentru a codifica factorul de scală. Debitul binar net pentru un canal audio este  $10 \times 32 + 3 = 323$  kbit/s. O marjă de siguranță pentru detecția erorilor și sincronizare cresc debitul la 364 kbit/s pentru un canal mono sau 728 kbit/s pentru modul stereo (de unde și denumirea de NICAM 728). NICAM rezolvă problema gamei dinamice pentru codarea audio, dar nu este o schemă eficientă de compresie pentru că nu exploatează dependența statistică dintre eșantioane și nici efectele de mascare auditivă.

Tabelul 3.2 Niveluri de calitate audio disponibile în formatul CD-i

Calitate	Codare	Frecvență eșantionare	Număr biți/eșantion	Raport semnal/zgomot	Număr canale	Durata redare
CD-Audio	PCM	44,1 kHz	16	98 dB	1 stereo	1 oră
A	ADPCM	37,8 kHz	8	90 dB	2 stereo 4 mono	2 ore 4 ore
B	ADPCM	37,8 kHz	4	60 dB	4 stereo 8 mono	4 ore 8 ore
C	ADPCM	18,9 kHz	4	60 dB	8 stereo 16 mono	8 ore 16 ore
Informație	-	-	-	-	-	10 ore

Un alt exemplu de codare predictivă adaptivă este Discul Compact interactiv CD-i audio, unde informația audio poate fi stocată atât PCM, 16 biți, 44,1 kHz (calitate CD), cât și utilizând tehnica de codare ADPCM cu 3 niveluri de calitate audio, permițând un raport maxim de compresie de 1/16 (tab. 3.2), unde:

- A - calitate Hi-Fi (*High Fidelity* - înaltă fidelitate) cu banda audio de 18 kHz, similară cu a discurilor LP de vinil

- B - calitate Mid-Fi (*Middle Fidelity* - fidelitate medie) cu banda audio de 15 kHz, similară cu emisia radio MF

- C - calitate MA

În domeniul codării audio de calitate cu debit binar redus, între 64 și 128 kbit/s, codoarele bazate pe modelul sursei și predicția liniară (PL) nu au avut un succes deosebit. Cu excepția algoritmului Twin VQ, codecurile audio cu PL au rămas în general la nivel experimental. Cu ocazia ultimelor cercetări în vederea codării hibride vocale și audio cu debite foarte reduse, sub 16 kbit/s, tehnicile de codare audio cu PL au fost din nou luate în considerare. Cu ocazia testelor de ascultare din cadrul procesului de standardizare a MPEG-4 s-a observat că alegerea modelului de semnal depinde de materialul sursă. De exemplu, un codor CELP depășește calitativ un codor sinusoidal pentru semnale vocale, dar codorul sinusoidal depășește codorul CELP pentru muzică. De aceea, se dorește realizarea în viitor a unui algoritm hibrid care să combine părțile tari ale ambelor modele. Deși PL stă la baza algoritmilor vocali moderni, totuși în cazul codării audio i s-a acordat mai puțină atenție. Un motiv ar fi faptul că codoarele cu PL nu sunt potrivite pentru modelarea componentelor sinusoidale prezente în semnalele audio staționare. Aceste elemente creează vârfuri ascuțite în înfășurătoarea spectrală, care, în prezența zgomotului de cuantizare, pot duce la instabilitatea filtrului de sinteză PL. Un alt motiv pentru lipsa de interes este că sistemul sursă reprezentat prin cadrul de analiză-sinteză cu PL nu modelează toate mecanismele fizice ce generează semnalele audio. Corespondența dintre sistemul de analiză-sinteză cu PL și modelul de producere a vorbirii au stat la baza succesului predicției liniare în aplicații vocale. În cazul codării audio însă, succesul PL este dependent de natura semnalului. Totuși s-au realizat câțiva algoritmi cu PL care s-au aplicat cu succes în codarea audio de calitate CD și care vor fi prezentați în continuare. În plus, se va prezenta un nou codor bazat pe PL cu deviere de frecvență, care ar putea reduce complexitatea prin eliminarea modelului perceptual explicit.

#### A. Excitație multipuls

Singhal la Bell Labs [SIN90] a observat că excitarea multipuls cu o densitate suficientă de impulsuri se poate aplica pentru a corecta erorile înfășurătoarei PL introduse de lărgirea lății de bandă și cuantizare (fig.3.3). Acest algoritm folosește un filtru de sinteză LPC de ordinul 24 și optimizează poziția și amplitudinea impulsurilor pentru a minimiza erorile perceptuale de reconstrucție. Singhal a determinat că pentru a obține o calitate aproape transparentă este necesară o densitate de aproximativ un impuls la patru eşantioane de ieșire. Coeficienții spectrali sunt transformați în coeficienți “*inverse sine reflection*”, apoi codați diferențial și cuantizați folosind cuantizoare Max optimizate pdf. Se folosește de asemenea și codarea Huffman. Pozițiile impulsurilor sunt codate diferențial în funcție de poziția primului impuls. Amplitudinile impulsurilor sunt codate fracțional în funcție de impulsul cel mai mare și apoi cuantizate cu un cuantizor Max. Codorul audio MPLPC (*Multipulse Linear Prediction Coder*) propus a atins un raport semnal-zgomot de 35 – 40 dB la un debit de 128 kbit/s.

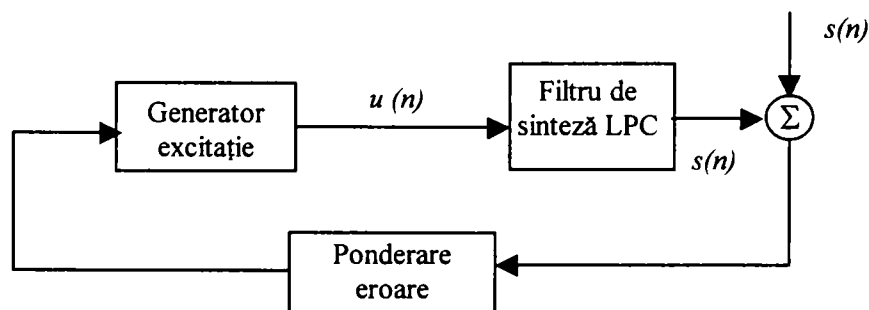


Figura 3.3 Model cu excitație multipuls

#### B. Codarea cu excitație wavelet discretă

O altă schemă perceptuală de codare audio bazată pe PL, cu un debit de 96 kbit/s, a fost propusă de Boland și Deriche [BOL98], având o calitate de ieșire comparabilă cu MPEG-1, Nivel 2 la 128 kbit/s. Schema propusă folosește o transformare DWT cu trei niveluri bazate pe

un filtru cu patru benzi egale. La fiecare nivel al transformării DWT, subbanda cea mai de jos a nivelului anterior se descompune în patru benzi egale. Această structură neuniformă pe zece benzi intenționează să simuleze într-o oarecare măsură benzile critice. Alocarea perceptuală a biților pentru coeficienții transformării s-a făcut folosind modelul psihoacustic 2 din MPEG-1.

### C. Predicție liniară cu deviație de frecvență

Dincolo de îmbunătățirile realizate prin folosirea a diverse modele de excitație, a apărut ideea devierii axei de frecvență înainte de analiza LP pentru a obține o rezoluție mai bună la anumite frecvențe. În contextul codării perceptuale, această deviere trebuie să fie conform scării Bark. Devierea axelor de frecvență pentru a obține o rezoluție FFT neuniformă a fost introdusă pentru prima oară de Oppenheim [OPP71], [OPP72], care a folosit pentru deviere o rețea de secțiuni de ordinul întâi înseriate, urmate de o transformare FFT standard. Această idee a fost mai târziu extinsă la ceea ce Strube [STR80] a numit “*warped linear prediction*” (WLP) și s-a utilizat apoi în codetul ADPCM [KRU88]. Din punct de vedere obiectiv, măsurarea erorii auditive a arătat o îmbunătățire considerabilă în cazul codării WLP, față de codarea LP convențională, folosindu-se același număr de biți alocați pentru reziduurile de predicție. Din punct de vedere subiectiv, algoritmul a atins o calitate transparentă pentru unele materiale, dar a avut probleme cu sunetele tranzitorii. Algoritmul a fost extins pentru a trata semnale stereofonice, folosind WLP pentru semnale complexe (CWLP). Calitatea obținută la un debit de 128 kbit/s, pentru un material eșantionat la 44,1 kHz, a fost sub calitatea CD.

### 3.2.2 Codarea în sub-benzi

Codarea în subbenzi ține cont de faptul că sunetele reale nu au o împrăștiere uniformă a energiei spectrale. Lungimea cuvântului pentru codarea audio PCM se bazează pe gama dinamică necesară, care în general este constantă cu frecvența, deși orice preaccentuare poate modifica situația. Când un semnal cu spectru neuniform este digitalizat PCM, gama dinamică completă va fi ocupată numai de componenta spectrală cea mai puternică, toate celelalte fiind codate cu o margine de codare excesivă. În forma cea mai simplă, codarea în subbenzi împarte semnalul audio în mai multe benzi de frecvență și codează fiecare bandă separat în funcție de nivelul din subbandă. În benzile cu o energie mai redusă va rezulta o amplitudine mai mică ce va putea fi codată printr-un cuvânt mai scurt. Astfel, în fiecare bandă vor rezulta eșantioane de lungime variabilă, suma tuturor lungimilor eșantioanelor fiind mai mică decât în cazul PCM, obținându-se o reducere a datelor. Codarea în subbenzi nu este folosită numai în domeniul digital, ea fiind folosită pe larg și de sistemul analogic de reducere a zgomotului Dolby.

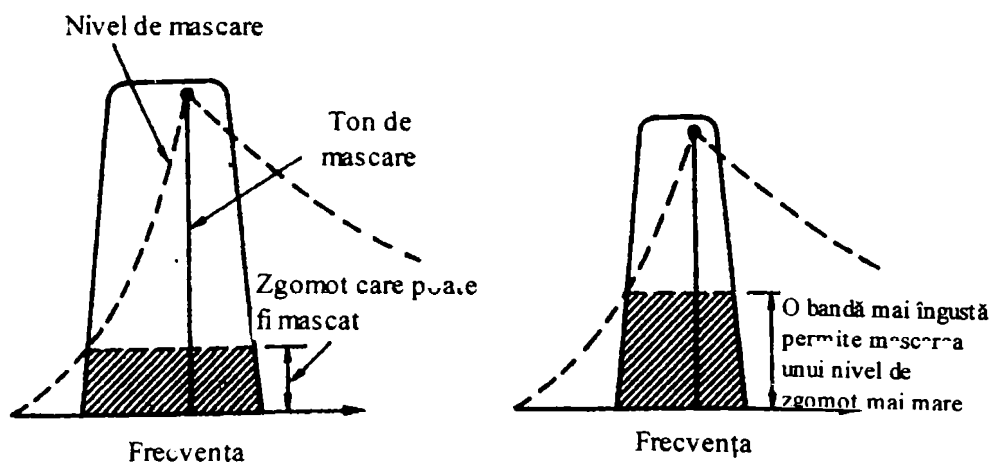


Figura 3.4 La codarea în subbenzi, cazul cel mai defavorabil apare când tonul de mascare se află în partea de sus a subbenzii. Cu cât banda este mai îngustă, cu atât nivelul zgomotului ce poate fi mascat este mai mare.

Numărul subbenzilor folosite depinde de tehnicile de reducere a datelor folosite în combinație cu codarea în subbenzi. Dacă se folosesc tehnici bazate pe mascarea auditivă este de preferat ca subbenzile să fie mai înguste decât benzile critice ale urechii. De exemplu, ISO/MPEG (§3.3.1) și PASC (§3.3.5) folosesc 32 de subbenzi. Figura 3.4 prezintă pragul de mascare pentru un ton de mascare aflat la marginea de sus a subbenzii. Se poate observa că dacă subbanda este mai îngustă, poate fi mascat mai mult zgomot de cuantizare. Totuși, un număr prea mare de subbenzi duce la creșterea complexității și a întârzierii la codare, crescând riscul de preecou la semnalele tranzitorii ce depășesc mascarea temporală.

Filtrele în oglindă în cuadratură QMF (*Quadrature Mirror Filters*) se folosesc numai în cazul benzilor de frecvență egale. S-ar putea crede că acesta este un dezavantaj, pentru că benzile critice ale urechii nu sunt uniforme. În realitate acest lucru devine o problemă numai în cazul debitelor binare foarte scăzute. Modelul de mascare al auzului trebuie să aibe în toate cazurile benzile critice corecte. Acest model poate fi suprapus pe benzi de orice lățime pentru a se determina mascarea și respectiv câștigul posibil prin codare. Câștigul obținut prin utilizarea benzilor cu lățime egală nu va fi la fel de mare ca în cazul benzilor critice, dar pentru multe aplicații creșterea complexității filtrelor nu se justifică față de reducerea obținută [WAT95].

Codoarele în subbenzi exploatează redundanța semnalului audio și irelevanțele psihoacustice în domeniul frecvență. În locul transformărilor unitare, aceste codoare se bazează pe reprezentări în domeniul frecvență obținute cu ajutorul unor bancuri de filtre trece bandă. Ieșirea fiecărui filtru este apoi eșantionată și codată. La receptor, semnalele sunt demultiplexate, decodate, demodulate și apoi însumate pentru a reconstrui semnalul. Metodele eficiente de cuantizare se bazează de obicei pe alocarea dinamică a biților controlată psihoacustic, astfel încât semnalul reconstruit să nu conțină zgomot de cuantizare sau alte artefacte.

În continuare vor fi prezentați algoritmi propuși de IRT (*Institut für Rundfunktechnik*), Philips Research Laboratories și CCETT, care, la început au fost cuprinși în programul european de standardizare a radioului digital DAB Eureka-147 și apoi au fost adunați într-un singur algoritm de către ISO/IEC, formând MUSICAM (*Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing*), algoritm inclus în partea audio a standardelor MPEG-1 și MPEG-2.

#### A. MASCAM

G. Theile a dezvoltat la IRT codorul în subbenzi MASCAM (*Masking Pattern Adapted Subband Coding*) bazat pe o structură arborigenă de filtre QMF care să simuleze benzile critice din structura sistemului auditiv. Codorul are 24 de subbenzi neuniforme, cu lățimea de 125 Hz sub 1 kHz, 250 Hz între 1 și 2 kHz, 500 Hz între 2 și 4 kHz, 1 kHz între 4 și 8 kHz și 2 kHz între 8 și 16 kHz. Filtrul prototip are 64 de pași. Secvențele obținute din subbenzi sunt prelucrate în blocuri de 2 ms. Un factor de scară se cuantizează pentru fiecare subbandă și se transmite pentru fiecare bloc. Alocarea biților în subbandă se face în urma unei analize psihoacustice simple. La început, codorul original lua în considerare numai mascarea simultană din interiorul unei benzi, apoi, la calculul alocării biților s-a luat în considerare și mascarea simultană între benzi și mascarea temporală. Codorul MASCAM are rezultate optime pentru o bandă de frecvență de 15 kHz la un debit binar de 80-100 kbit/s. În aceeași perioadă, un codor asemănător a creat și Philips, care utiliza 20 - 26 de benzi. Codorul Philips putea coda semnale de calitate CD cu un debit 110 kbit/s în varianta cu 26 de benzi și 180 kbit/s pentru 20 de benzi.

#### B. MUSICAM

Bazat pe codoarele dezvoltate de IRT și Philips, algoritmul MUSICAM a câștigat în 1990 competiția ISO/IEC pentru un nou standard de codare audio. El a fost baza standardelor audio MPEG-1 și MPEG-2, nivel 1 și 2. Față de predecesorii săi, MUSICAM (fig. 3.5) face câteva compromisuri între complexitate, întârziere și calitate. Folosind 32 de filtre polifazate egale în

locul filtrelor QMF, s-a obținut o reducere importantă a complexității (5 MFLOPS) și întârzierii (10,66 ms) față de codoarele IRT și Philips. Aceste îmbunătățiri s-au obținut utilizând un filtru suboptimal, cu benzi constante de 750 Hz pentru o rată de eșantionare de 48 kHz, care nu mai corespunde cu benzile critice. În ciuda lășimii excesive de bandă la frecvențe joase, se obține o codare de calitate datorată analizei psihoacustice îmbunătățite. Folosind o transformare FFT cu 1.024 de puncte în paralel cu un filtru polifazat se obține o rezoluție spectrală foarte bună de 46 Hz/linie la 48 kHz. Această structură paralelă permite o estimare îmbunătățită a pragurilor de mascare și a unui raport SMR mai precis în fiecare subbandă. Modelul psihoacustic folosit de algoritmul MUSICAM este același ca și în cazul standardului MPEG-1, descris la paragraful §2.2.9.

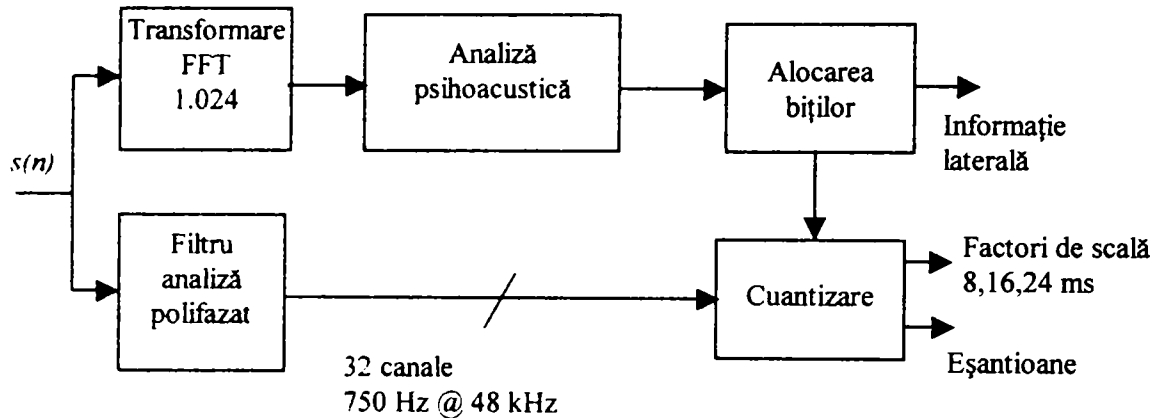


Figura 3.5 Codor MUSICAM

Algoritmul MUSICAM funcționează astfel: secvențele de ieșire din subbenzi sunt prelucrate în blocuri de 8 ms (12 eșantioane la 48 kHz), adică aproape de rezoluția temporală a sistemului auditiv (4 – 6 ms). Se extrag factorii de scală pentru fiecare bloc și se codează folosind 6 biți pe o scară dinamică de 120 dB. Ocazional, se exploatează redundanța temporală prin repetarea în 2 – 3 blocuri (16 – 24 ms) a factorilor de scală cu variație lentă dintr-o subbandă. Eșantioanele din subbenzi sunt cuantizate și codate în funcție de raportul SMR din subbandă determinat prin analiză psihoacustică. Alocarea biților în fiecare subbandă este transmisă ca informație laterală. Pe scara cu 5 trepte de evaluare subiectivă a calității, MUSICAM a obținut nota 4,6 (deviație standard 0,7) la 128 kbit/s și 4,3 (deviație standard 1,1) la 96 kbit/s pentru un canal mono. Totuși, în cazul unor semnale critice (triunghi, castaniete) la 96 kbit/s, se poate observa o scădere a calității, reflectată și prin deviația standard relativ mare (1,1).

### C. Descompuneri Wavelet

O altă clasă de codoare în subbenzi folosește, în locul băncilor de filtre, transformarea wavelet discretă (DWT – *Discrete Wavelet Transform*). Acestea oferă o flexibilitate mai bună decât procedurile MPEG, între selectivitatea temporală și cea frecvențială a analizei de semnal efectuate. Acest lucru este posibil, deoarece sistemele wavelet, spre deosebire de orice variantă de transformare DCT, pot oferi rezoluții temporale bune la frecvențe ridicate și respectiv rezoluție frecvențială mai bună la frecvențe joase [ASZ00].

În [PAI00] sunt prezentate câteva codoare bazate pe pachete wavelet (WP) realizate în anii '90, care au realizat codări transparente la debite binare sub 100 kbit/s. Ideea principală folosită de codoarele bazate pe DWT și DWPT este de a cuantiza și coda eficient coeficienții asociați cu fiecare stadiu al descompunerii wavelet folosind aceleași tehnici de formare a zgomotului ca la codoarele perceptuale în subbenzi decrișe anterior.

### 3.2.3 Codarea prin transformări

Algoritmii de codare prin transformări pentru audio de calitate folosesc transformări unitare pentru analiza timp/frecvență. În general, un semnal audio este reprezentat în domeniul timp, așa cum este el captat prin intermediul microfonului. Analiza Fourier permite reprezentarea unei forme de undă periodice printr-un set de componente armonice de o anumită amplitudine și fază. Teoretic, un semnal audio poate fi descompus în componentele sale spectrale, care pot fi apoi înregistrate sau transmise. Transformarea inversă permite recrearea precisă a semnalului original.

În realitate sunetele nu sunt periodice, conțin elemente tranzitorii pe care transformările nu le pot localiza precis în timp. Soluția este de a împărți forma de undă în segmente mai mici și de a le transforma separat. Timpul de calcul și întârzierile corespunzătoare vor fi reduse, dar pot apărea artefacte datorită trunchierii formei de undă prin ferestre de timp dreptunghiulare. O soluție este folosirea funcțiilor fereastră și suprapunerea segmentelor ca în fig. 3.6. Astfel fiecare eșantion de intrare apare în două transformări, dar cu pondere diferită în funcție de poziția în timp. Deși aparent, conform figurii, va fi generat un volum dublu de date, anumite transformări pot elimina redundanțele.



Figura 3.6 Codarea prin transformări se poate realiza practic numai prin blocuri scurte. Acestea se suprapun folosind funcții fereastră.

Transformarea de frecvență discretă DFT (*discrete frequency transform*) nu produce un spectru continuu, ci coeficienți la frecvențe discrete. Rezoluția de frecvență este egală cu numărul de eșantioane din fereastră. Dacă se folosesc ferestre suprapuse, va fi produs un număr dublu de coeficienți decât cel teoretic necesar. În plus DFT necesită un volum mare de calcul datorită reprezentării aritmetice complexe a amplitudinii și fazei. O alternativă ar fi folosirea transformării cosinus discrete DCT (*discrete cosine transform*), care are anumite avantaje în cazul utilizării ferestrelor suprapuse. În cazul transformării cosinus discrete modificate MDCT (*modified discrete cosine transform*) se folosesc ferestre cu o suprapunere de 50%, obținându-se un număr de coeficienți dublu decât cel necesar. Aceștia sunt subeșantionați (decimați) pentru a obține o transformare eșantionată critic, ceea ce duce la apariția unor produse de aliere în domeniul frecvență. Totuși, printr-o simplă modificare a transformării, produsele de aliere din a doua jumătate a unei ferestrei sunt egale dar de semn opus produselor de aliere din prima jumătate a ferestrei următoare, rezultând anularea lor la reconstrucția semnalului.

#### A. Codare optimă în domeniul frecvență (*Optimum Coding in the Frequency Domain – OCF-1, OCF-2, OCF-3*)

În 1987 Brandenburg a propus un algoritm denumit “Codare optimă în domeniul frecvență” pentru un debit de 132 kbit/s. Semnalul de intrare este împărțit în blocuri de 512 eșantioane și apoi transformat în domeniul frecvență folosind o DCT, apoi componentele sunt cuantizate și entropia codată. Se folosește un singur cuantizor pentru toate componentele transformării. Cuantizarea adaptivă și codarea entropiei sunt folosite împreună într-o procedură iterativă pentru a obține un debit binar fix. În bucla internă din figura 3.7 dimensiunea pasului de cuantizare este crescută iterativ, obținându-se un nou flux de date până se obține debitul binar dorit. Creșterea pasului la fiecare trecere duce la scăderea numărului de niveluri, respectiv la scăderea debitului binar.

În cea de a doua buclă se face o analiză perceptuală. Mai întâi se face o analiză pe benzi critice, apoi se aplică o funcție de mascare, care combină un prag de mascare fix de  $-6$  dB cu un prag de mascare între benzi, ceea ce conduce la estimarea unui JND pentru fiecare bandă critică. Dacă după cuantizarea din bucla interioară și codarea entropiei, distorsiunile depășesc JND în cel puțin o bandă critică, pasul de cuantizare se modifică numai pentru benzile critice în care apare această depășire. Această operație se repetă în bucla exterioră până la satisfacerea criteriului JND sau până la atingerea unui număr maxim de iterații. Componentele transformării sunt codate și apoi transmise receptorului, alături de informațiile laterale.

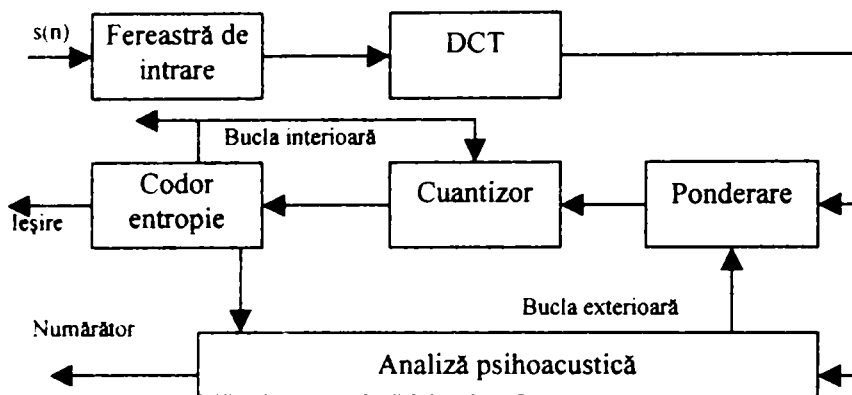


Figura 3.7 Codor OCF [PAI00]

În 1988, Brandenburg realizează OCF-2, cu calități audio subiective îmbunătățite pentru un debit de 110 kbit/s. Creșterea calității audio s-a obținut prin înlocuirea DCT cu MDCT și adăugarea unei bucle de detecție și compensare a preecoului. Cu ajutorul OCF-2 s-au obținut codări transparente pentru o mare varietate de materiale audio. O a treia variantă, OCF-3, a permis codări de calitate la debite de 64 kbit/s și o complexitate redusă.

### B. Codor cu transformare perceptuală (PXF)M

În timp ce Brandenburg dezvoltă algoritmul OCF, la AT&T Bell Labs, Johnston dezvoltă câteva codoare bazate pe transformări DFT, care au realizat codări transparente pentru semnale audio mono de calitate FM (fig. 3.8). Ideea ce stă la baza codorului cu transformare perceptuală (PXF)M este de a estima cantitatea de zgomot de cuantizare ce poate fi injectată inaudibil în fiecare subbandă folosind estimările entropiei perceptuale. Semnalul este transformat în domeniul frecvență folosind o fereastră cu o suprapunere de  $1/16$  și o transformare FFT cu 2.048 de puncte. Se estimează pragurile JND cu ajutorul entropiei perceptuale, descrisă la §2.2.6. Apoi, o buclă de cuantizare iterativă adaptează 128 de cuantizatori în subbenzi pentru a satisface pragurile JND până la atingerea debitului dorit. În 1989 Johnston extinde codorul PXFM pentru a lucra cu semnale stereofonice (SEPFM), obținând codarea transparentă a unui semnal stereofonic de calitate CD cu un debit binar de 192 kbit/s.

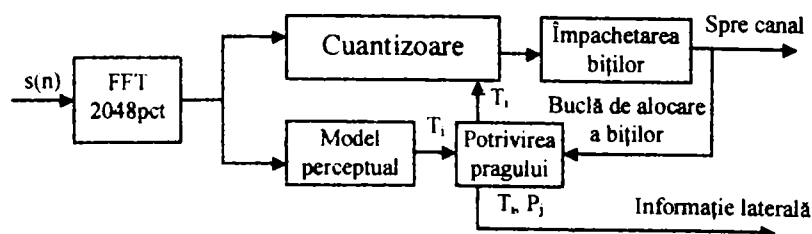


Figura 3.8 Codor PXFM [PAI00]

### C. Codor hibrid Brandenburg-Johnston

În 1990, Johnston și Brandenburg au colaborat la realizarea unui codor hibrid, care este atât un codor în subbenzi, cât și un codor prin transformări. Acest codor hibrid a fost creat în ideea de a mări rezoluția în timp și frecvență a algoritmilor OCF și PXXFM, construind o bancă de filtre mai asemănătoare sistemului auditiv. Semnalul audio este împărțit în subbenzi cu o lățime de patru octave folosind filtre QMF. Secvența de ieșire decimată din fiecare subbandă este urmată de una sau mai multe transformări pentru a obține rezoluția timp/frecvență dorită. Schemele de cuantizare și codare combină elemente atât din OCF cât și din PXXFM. Codorul hibrid, fără un mecanism special de control al preecoului a obținut o calitate egală sau chiar superioară cu OCF-3 la un debit de 64 kbit/s, dar cu o complexitate sporită. Această structură hibridă a fost adoptată în cadrul nivelului 3 din MPEG-1 și MPEG-2.

### D. ASPEC (Audio Spectro-Perceptual Entropy)

Codorul ASPEC este un algoritm flexibil, realizat în 1991, care încorporează cele mai bune caracteristici ale codoarelor de mai sus. El folosește o transformare DCT modificată, comutarea ferestrei ( $N = 256$  și  $N = 1.024$ ), un model perceptual și realizează o codare de calitate superioară celorlalte codoare la un debit de 64 kbit/s. Alocarea dinamică a biților și codarea Huffman a coeficienților cuantizați ai transformării are loc în cadrul a două bucle de iterație. ASPEC oferă diferite niveluri de calitate pentru debite cuprinse între 64 și 192 kbit/s pe canal, formând baza nivelului 3 al standardelor MPEG-1 și MPEG-2/BC-LSF [NOL97].

Alte tipuri de codoare perceptuale prin transformări, cum ar fi DPAC (*differential perceptual audio coder*), DCT și MDCT cu cuantizare vectorială, sunt prezentate în [PAI00].

## 3.2.4 Codoare sinusoidale

Deși modelele de semnale sinusoidale au fost aplicate cu succes în codarea vocală și în aplicații de sinteză muzicală, în codarea audio perceptuală ele au fost folosite doar de curând. În cadrul acestui paragraf vor fi prezentați algoritmi perceptuali de codare bazați pe modele de semnal sinusoidal pur sau hibrid. Standardul MPEG-4 a stabilit noi condiții pentru a obține o codare audio de calitate la debite foarte reduse, 6 – 24 kbit/s, debite rezervate până nu demult algoritmilor de codare vocală. Algoritmii sinusoidali care folosesc proprietățile perceptuale propuși recent pentru codare audio cu dbr sunt ASAC (*Analysis/Synthesis Audio Codec*), ASAC îmbunătățit și ASAC FM. Unii din acești algoritmi au fost adoptați în cadrul standardului MPEG-4 (vezi §3.3.1.D).

### A. Codorul ASAC (Analysis/Synthesis Audio Codec)

Codecul sinusoidal ASAC destinat codării semnalelor audio la debite între 6 și 24 kbit/s a fost realizat de Edler ș.a. la universitatea din Hanovra și propus pentru standardul MPEG-4 în 1995. ASAC împarte semnalul audio de intrare în cadre de analiză în care semnalul este considerat cvasistaționar. Parametrii de sinteză sinusoidali sunt extrași conform criteriilor perceptuale, cuantizați, codați și transmiși către decodor pentru sinteză. Algoritmii distribuie parametrii de sinteză în fluxul de bază sau îmbunătățit pentru a obține o calitate scalabilă pentru debite de 6 și 24 kbit/s. Schema ASAC constă (fig. 3.9) dintr-un bloc de preanaliză pentru alegerea ferestrei și extragerea înfășurătoarei, un bloc de estimare a parametrilor de analiză și sinteză, un model perceptual și un bloc de cuantizare și codare. Deși prezintă similarități cu codorul vocal sinusoidal și algoritmi de sinteză muzicală, codorul ASAC conține câteva tehnici noi. ASAC nu folosește eroarea medie pătratică, ci alege sinusoidale în ordinea descrescătoare a importanței perceptuale cu ajutorul unei bucle iterative de analiză/sinteză. Semnificația perceptuală a fiecărei componente sinusoidale este judecată în raport cu puterea de mascare a



semnalului sintetizat, determinată cu ajutorul unui model psihoacustic simplificat. Blocul iterativ de analiză și sinteză estimează câte un parametru și fiecare iterație identifică sinusoida cea mai semnificativă perceptual rămasă în sinteza reziduală,  $e_i(n) = s(n) - s_i(n)$  și o adaugă ieșirii sintetizate,  $s_i(n)$ . Semnificația perceptuală este evaluată prin compararea rezidului de sinteză față de pragul de mascare asociat ieșirii curente sintetizate și se alege sinusoida reziduală cea mai depărtată de prag. Bucla se repetă până la epuizarea bugetului de biți.

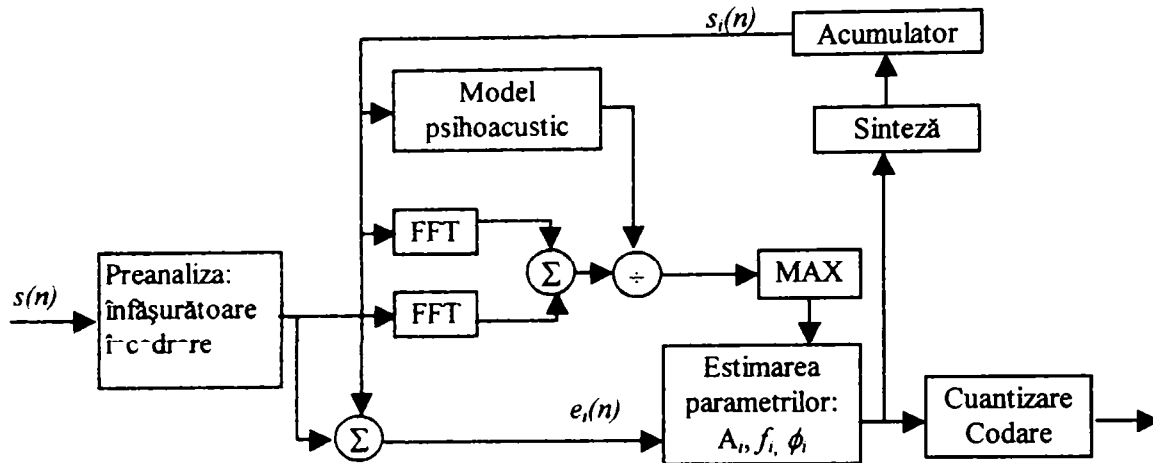


Figura 3.9 Codor ASAC

### B. Codorul HILN (Harmonic and Individual Lines Plus Noise)

Algoritmul ASAC a depășit algoritmi vocali la același debit cu ocazia testelor subiective cu anumite semnale de test, în special cu muzică cu un conținut spectral complex. Totuși, codecul ASAC original a dat greș în cazul altor semnale de test, cum ar fi semnalele vocale. Drept urmare, nucleul ASAC a fost inclus într-un algoritm ameliorat, care să adapteze mai bine modelul de semnal al codorului cu diversele caracteristici ale semnalelor de intrare. În cadrul cercetărilor pentru nucleul standardului MPEG-4, Purnhagen ș.a. de la Universitatea din Hanovra au dezvoltat, în colaborare cu Deutsche Telekom, un algoritm bazat pe obiecte. În acest concept, sinusoida armonice, sinusoida individuale și obiecte de tip zgomot se pot combina într-un model hibrid pentru a crea o reprezentare parametrică a semnalului. Algoritmul îmbunătățit, numit HILN (*linii armonice și individuale plus zgomot*), are o arhitectură similară cu algoritmul ASAC, cu anumite modificări (fig. 3.10). Blocul de analiză/sinteză a fost extins pentru a include o serie de stadii succesive de analiză pentru fiecare tip de obiect disponibil. În sistemul de analiză/sinteză ameliorat, se aplică mai întâi analiza armonică, urmată de analiza liniilor spectrale individuale, urmată de modelarea zgomotului rămas după cele două stadii.

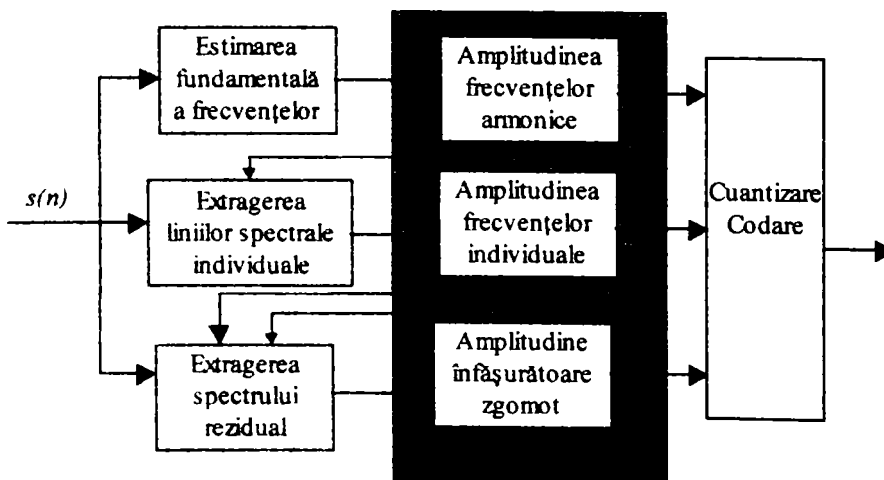


Figura 3.10 Codorul HILN

Rezultatele testelor de evaluare subiectivă la 6 kbit/s au demonstrat îmbunătățiri semnificative ale HILN față de ASAC, în special în cazul testelor critice ce au provocat artefactele în cazul ASAC. Totuși codoarele vocale CELP sunt mai eficiente în cazul codării vocii la debite reduse, iar codoarele “timp-frecvență” sunt mai eficiente pentru semnale audio generale la debite peste 32 kbit/s. Îmbunătățirile aduse HILN față de ASAC au convins comitetul de standardizare MPEG-4 să-l recomande ca algoritm audio parametric pentru debite reduse în proiectul standardului.

### C. Sinteza FM

Sinteza cu modulație de frecvență oferă anumite avantaje în plus față de metodele de codare armonice datorită abilității de a modela semnale cu multe armonice cu un număr relativ mic de parametri. De exemplu, în cazul celei mai simple sinteze FM, frecvența unei unde sinusoidale (purtătoare) este modulată de o altă sinusoidă (modulatoare) pentru a genera o formă de undă complexă, cu caracteristici spectrale ce depind de un index de modulare și de parametrii celor două unde sinusoidale. Expresia în funcție de timp a semnalului modulat în frecvență este:

$$s(t) = A \sin[2\pi f_c t + I \sin(2\pi f_m t)] \quad (3.1)$$

unde  $A$         amplitudinea  
 $f_c$         frecvența purtătoare  
 $f_m$         frecvența modulatoare  
 $I$          indexul de modulație  
 $t$          timpul

Seria Fourier asociată este:

$$s(t) = \sum_{k=-\infty}^{\infty} J_k(I) \sin(2\pi f_c t + 2\pi f_m t) \quad (3.2)$$

unde  $J_k(I)$  este funcția Bessel de gradul întâi. Cu ajutorul relației (3.2) se pot obține un mare număr de armonici (fig. 3.8) controlând numai trei parametri ai operatorului FM. Se poate observa că frecvențele fundamentale și armonicele sunt determinate de  $f_c$  și  $f_m$ , iar amplitudinea armonicilor este controlată de indexul de modulație  $I$ . Înfășurătoarea Bessel determină lățimea benzii spectrale FM.

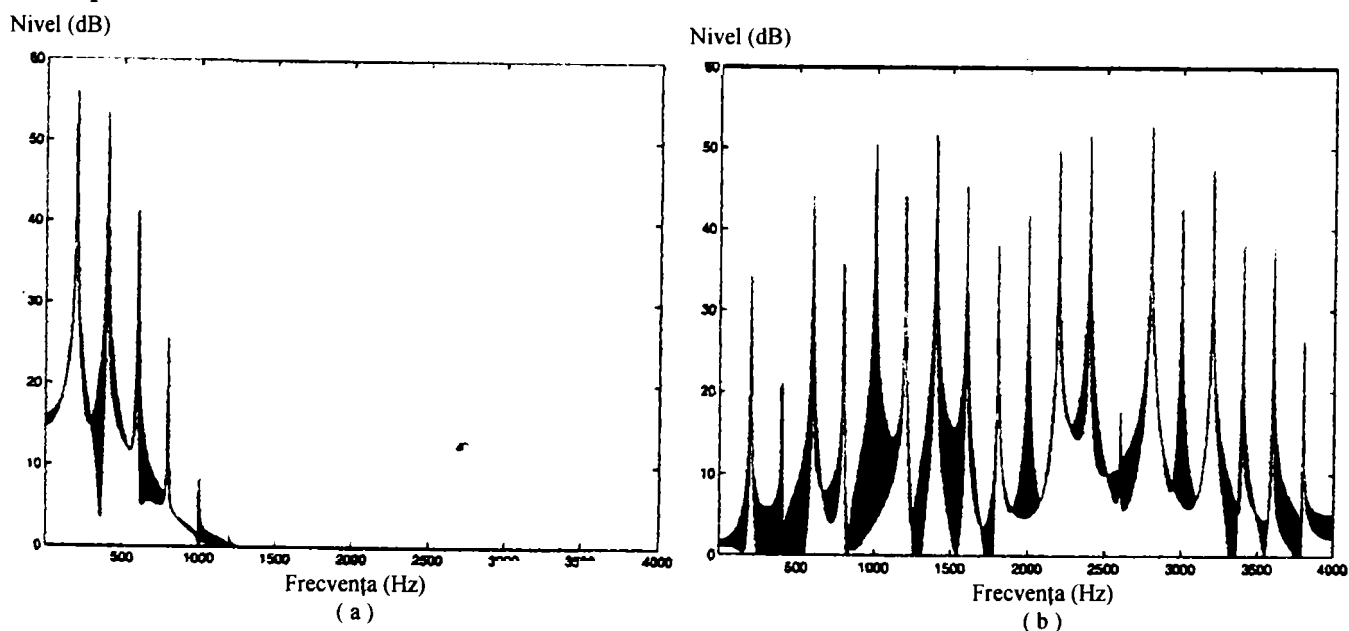


Figura 3.11 Spectrul armonic FM pentru  $f_c = f_m = 200\text{Hz}$ ,  $I = 1$  (a) și  $I = 15$  (b)

În fig. 3.11 este prezentat spectrul armonic FM pentru o purtătoare de 200 Hz și un indice de modulație 1 (a) și 15 (b). Deși ambele exemple au o structură armonică identică, înfășurătoarea amplitudinii și lățimea de bandă diferă în funcție de indicele de modulație  $I$ . Deci, problema principală pentru a beneficia efectiv de tehnica FM în modelarea semnalelor este precizia estimării parametrilor.

Winduratsna a propus o schemă de codare audio cu sinteză FM în care ieșirile, în paralel, a mai multor "operatori" FM sunt combinate pentru a modela sunetul unui instrument. Algoritmul funcționează astfel (fig. 3.12): blocul de preanaliză segmentează intrarea audio în cadre de analiză și extrage parametrii pentru un set de linii spectrale individuale, ca în [EDL96], apoi identifică structura armonică prin maximizarea unei funcții obiective [WIN98]. Fiind dată frecvența fundamentală  $f_0$ , estimată din preanaliză, bucla iterativă de extragere a parametrilor estimează parametrii operatorilor FM individuali și acumulează contribuțiile lor până când spectrul compus este asemănător cu originalul. Asemănarea perceptuală se consideră adecvată atunci când diferența dintre spectrul original și cel sintetizat se află sub pragul de mascare [BAU95]. Această schemă de codare reușește să reprezinte eficient sunetul unui singur instrument la debite între 2,1 și 4,8 kbit/s. Folosind o fereastră de analiză de 24 ms, s-a reușit codarea unui fragment critic de voce masculină prin sinteză FM, la un debit de 21,2 kbit/s, cu o calitate echivalentă cu cea obținută la codarea ASAC cu un debit de 45 kbit/s. În ciuda dificultăților estimate pentru codarea muzicii polifonice, sinteza FM va fi inclusă în noii algoritmi de codare audio bazați pe obiecte, precum HILN.

#### D. Codoare sinusoidale hibride

În vreme ce codoarele prin transformări și cele în subbenzi au ca scop o calitate transparentă la debite cuprinse între 32 și 128 kbit/s/canal, codoarele sinusoidale propuse s-au concentrat asupra aplicațiilor care necesită debite foarte reduse, între 2 și 16 kbit/s. Scopul acestor algoritmi nu a fost neapărat calitatea transparentă, cât posibilitatea de a prelucra o mare varietate de sunete la debite reduse fără a depinde de un anumit model de sursă. Algoritmii sinusoidali existenți (ASAC, HILN etc.) reprezintă porțiunile perceptual semnificative ale spectrului amplitudine din semnalul original, fără a se ocupa explicit de spectrul fază. În consecință, acești algoritmi nu realizează o codare perceptuală transparentă. În funcție de debitul disponibil, calitatea audio cea mai bună poate fi obținută prin diferite tehnici de codare. Astfel, algoritmul vocal CELP oferă performanțe maxime pentru o vorbire clară sub 16 kbit/s, tehnicile parametrice sinusoidale prelucrează semnale audio între 16 – 32 kbit/s, iar codecurile audio "timp-frecvență" funcționează bine la debite peste 32 kbit/s. Proiectanții sistemelor de codare cu o gamă foarte largă de debite binare trebuie să hotărască dacă vor înseria ieșirile mai multor codoare fundamentale diferite sau vor transmite simultan fluxuri independente de la diferite tipuri de codoare și la receptor să aleagă decodorul corespunzător.

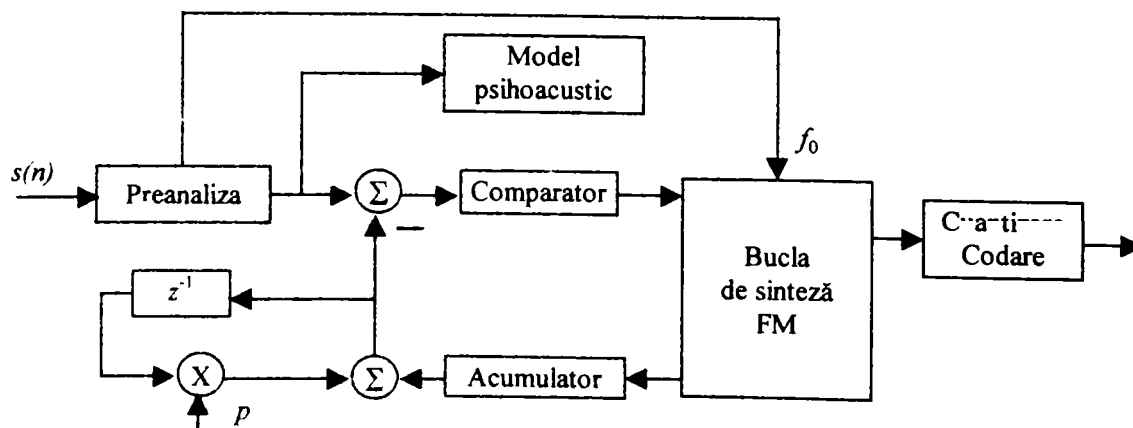


Figura 3.12 Schema de codare audio cu sinteză FM [WIN98]

Testele efectuate în cadrul standardului MPEG-4 au demonstrat că un codor hibrid în cascadă sinusoidal/timp-frecvență poate egala sau chiar depăși calitatea de ieșire la același debit și pentru anumite semnale critice de test, față de un codor prin transformări (timp-frecvență) singur [EDL96]. Este totuși dificil de spus care este algoritmul cel mai potrivit atunci când avem de-a face simultan cu semnale vocale și muzicale. În figura 3.13 se prezintă o schemă hibridă creată în scopul de a depăși transmiterea simultană (“simulcast”) CELP/parametric pentru semnale combinate vocal/muzical [EDL98]. Degradări ale calității semnalului au fost observate pentru semnale combinate, dominant muzicale. În viitor, structurile hibride vor beneficia de noi tehnici de discriminare între vorbă și muzică și de analiză și, eventual, descompunere a scenei auditive pentru o codare mai eficientă pe componente.

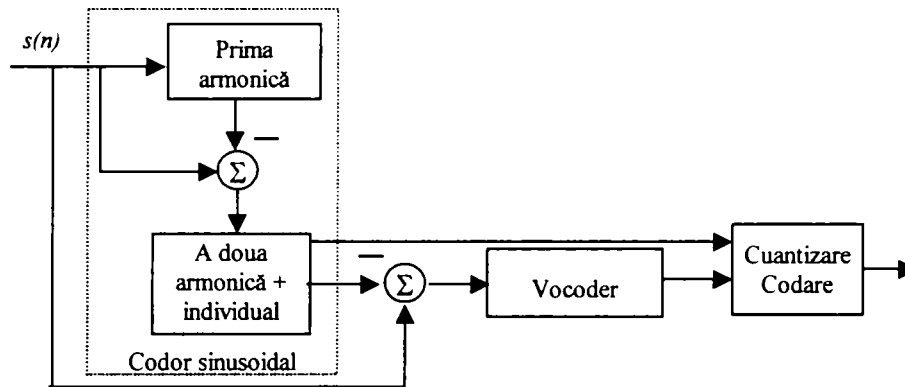


Figura 3.13 Schemă hibridă de codor sinusoidal/vocoder [EDL98]

### 3.3 Algoritmi și standarde de codare

Începând din anul 1988 Organizația Internațională de Standardizare ISO și Comisia Electrotehnică Internațională IEC au recunoscut importanța reducerii datelor în viitoarele produse audio-video, înființând grupul ISO/IEC/MPEG (*Moving Picture Experts Group*) și secțiunea sa MPEG/Audio cu scopul de a crea și evalua diferite scheme de codare, care să devină ulterior standarde internaționale.

Sistemul numit MUSICAM (*Masking pattern adapted Universal Sub-band Integrated Coding And Multiplexing*), care făcea parte din proiectul Eureka 147, a fost dezvoltat în colaborare de CCETT-Franța, IRT-Germania și Philips-Olanda, cu scopul de a fi utilizat în emisia digitală DAB.

În paralel s-a dezvoltat sistemul ASPEC (*Adaptive Spectral Perceptual Entropy Coding*), pornind de la câteva sisteme mai vechi, fiind destinat transmisiilor audio cu debit redus în ISDN. ASPEC a fost propus în colaborare de firmele AT&T Bell Labs, Thomson, Societatea Fraunhofer și CNET.

Aceste două sisteme au fost prezentate în iulie 1990, când la Radiodifuziunea suedeză a avut loc o serie completă de teste de ascultare critică. În urma acestor teste, grupul MPEG/Audio a combinat atributele sistemelor ASPEC și MUSICAM într-un standard având trei niveluri de complexitate și performanțe.

Performanțele unui codec depind de gradul de compresie și pentru a păstra calitatea audio la un raport de compresie mai ridicat trebuie folosite codoare mai complexe. Codarea ISO/MPEG permite la intrare rate de eșantionare de 32, 44,1 și 48 kHz și suportă la ieșire debite de 32, 48, 56, 64, 96, 112, 128, 192, 256 și 384 kbit/s. Transmisia poate fi mono, două canale (de ex. bilingvă), stereo și joint stereo (se ține cont de redundanța dintre cele două canale audio).

În tabelul 3.3 sunt prezentate câteva standarde de codare pentru semnale audio împărțite în două grupuri, telefonie și telefonie celulară și telefonie de bandă largă și audio.

Tabelul 3.3 Standarde de codare vocală și audio

Comitetul de standardizare	Standard	Anul	Algoritmul	Debit binar	Aplicație
CCITT	G.711	1972	Mu-Law și A-Law PCM	64 kbit/s	Telefonie
CCITT	G.721	1984	ADPCM	32 kbit/s	Telefonie
ETSI	GSM	1994	VSELP	8 kbit/s	Telefonie mobilă
CCITT	G.722	1986	ADPCM subbandă	32-64 kbit/s	Teleconferință
ISO	MPEG-1	1992	Musicam/ASPEC	128-384 kbit/s	Stocare audio (stereo)
ISO	MPEG-2	1996	-	320, 384 kbit/s	Stocare audio (cinci canale)
APT			apt-X100	64 - 384 kbit/s	Transmisie
Philips		1992	PASC	384 kbit/s	Stocare audio (DCC)
SONY		1992	ATRAC	292 kbit/s	Minidisc

### 3.3.1 Standardele de codare MPEG Audio

MPEG (*Moving Pictures Experts Group*) a fost creat de instituțiile de standardizare ISO/IEC în 1988 pentru a dezvolta standarde generale (pentru a fi utilizate în diferite aplicații) de reprezentare codată a filmelor și a benzii sonore asociate.

Din 1988 ISO/MPEG a efectuat standardizarea tehnicilor de compresie pentru video și audio. Scopul principal al MPEG a fost codarea video și audio pentru medii digitale de stocare (DSM - *Digital Storage Media*). Standardul de codare audio și-a găsit și alte aplicații, printre care:

- Digital Audio Broadcasting (EUREKA DAB, WorldSpace, ARIB, DRM)
- Transmisii ISDN pentru scopuri de contribuție și distribuție a programelor
- Arhivare și stocare pentru emisie
- Sunet pentru televiziunea digitală (DVB, Video CD, ARIB)
- Internet streaming (Microsoft Netshow, Apple Quicktime)
- Dispozitive audio portabile (mpman, mpnivel3, Rio, Lyra, YEPP etc.)
- Stocarea și schimbul de fișiere muzicale pe calculator.

Formatele de compresie audio cele mai folosite sunt MPEG-1/2 Audio Nivel 2 și 3, MPEG-2 AAC și Dolby AC-3.

#### A. ISO/IEC 11172-3 (MPEG-1) și ISO/IEC IS13818-3 (MPEG-2 BC)

MPEG-1 este numele primei faze a lucrărilor MPEG începute în 1988 și finalizate prin adoptarea standardului ISO/IEC IS 11172 la sfârșitul anului 1992. Partea de codare audio a MPEG-1 (ISO/IEC IS 11172-3) descrie un sistem general de codare proiectat să se potrivească mai multor aplicații. MPEG-1 audio constă din trei moduri de funcționare numite *niveluri (nivel)* de complexitate și performanțe crescânde de la nivelul-1 la nivelul-3. Nivelul-3 este modul cel mai complex, optimizat să furnizeze cea mai înaltă calitate la debite binare reduse (aprox. 128 kbit/s pentru un semnal stereo). Fiecare nivel succesiv îmbunătățește raportul de compresie, cu prețul creșterii complexității codoarelor și decodoarelor. Codoarele MPEG acceptă ca date de intrare fișiere PCM pe 16 biți cu diferite frecvențe de eșantionare. MPEG-1 definește compresia audio la 32 kHz, 44,1 kHz și 48 kHz. MPEG-2 extinde aceste frecvențe la 16 kHz, 22,05 kHz și 24 kHz. MPEG 2.5 este numele unei extensii a MPEG-1/2 Nivel-3, proprietate a institutului Fraunhofer, care lucrează cu frecvențe de eșantionare de 8 kHz, 11,05 kHz și 12 kHz.

Standardul MPEG-1 oferă patru moduri separate de codare: mono, stereo, dual independent mono (conținând de ex. versiunea audio în două limbi diferite) și joint stereo. MPEG audio nu lucrează cu un raport de compresie fix. Alegerea debitului binar este lăsată, în anumite limite, la

alegerea celui care implementează sau a utilizatorului codorului MPEG audio. Standardul definește o gamă de debite binare de la 32 kbit/s (pentru MPEG-1) sau 8 kbit/s (pentru MPEG-2 LSF - *Low Sampling Frequencies extension*) până la 320 kbit/s (respectiv 160 kbit/s pentru LSF). În cazul MPEG-1/2 nivel-3 codorul trebuie să suporte modificarea debitului binar de la un cadru audio la altul. Aceasta, împreună cu tehnologia rezervorului de biți, permit codarea cu debit binar variabil și codarea cu orice debit binar fix în limitele descrise de standard.

MPEG-2 (1994) extinde posibilitățile oferite de MPEG-1 pentru a suporta formatul 3/2 canale, adică stânga, dreapta, centru și stânga, dreapta spate. Primul standard MPEG-2 a fost compatibil cu MPEG-1, adică informația cu 3/2 canale transmisă de un codor MPEG-2 putea fi corect decodată și de un receptor MPEG-1. A doua variantă a standardului MPEG-2 sacrifică această compatibilitate pentru a elimina unele artefacte produse de zgomotul de cuantizare.

### 1. ISO nivel-1 - MUSICAM simplificat

Sistemul MUSICAM simplificat este un sistem de codare sub-bandă bazat pe schema bloc din fig. 3.5. O rețea de filtre polifazate în cuadratură împarte spectrul audio în 32 de benzi egale. Se folosesc cadre de câte 384 de eșantioane, cu câte 12 eșantioane în fiecare sub-bandă. Mărimea blocului se bazează pe fenomenul de pre-mascare (fig.3.14). Ca și la sistemul NICAM, eșantioanele din fiecare bloc dintr-o subbandă sunt comprimate în funcție de valoarea de vârf din blocul respectiv. Pentru fiecare sub-bandă se folosește un factor de scală pe 6 biți. Subbenzile se folosesc ca o analiză spectrală a semnalului de intrare pentru a determina alocarea biților. Datele de alocare a biților sunt multiplexate cu factorii de scală și eșantioanele recuantizate din sub-bandă pentru a forma cadrele mesajului codat. Codurile de alocare a biților sunt necesare pentru a permite decodorului să asambleze corect eșantioanele de lungime variabilă.

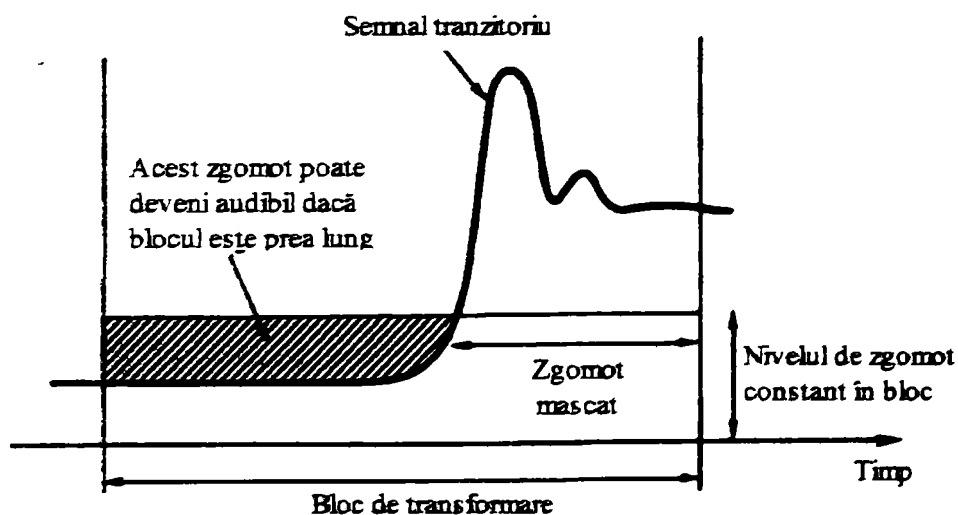


Figura 3.14 În cazul apariției unui semnal tranzitoriu la finalul unui bloc, zgomotul de cuantizare va fi prezent la începutul blocului, producând un preecou în care zgomotul este audibil înaintea semnalului tranzitoriu

### 2. ISO nivel-2 - MUSICAM

ISO nivel-2 este identic cu MUSICAM și este algoritmul cel mai răspândit în domeniul audio profesional. Această tehnică permite transmiterea bidirecțională a întregii benzi audio stereo folosind cele 2 canale B de 64 kbit/s ale unei linii ISDN. Principiul de funcționare a algoritmului MUSICAM este reducerea informațiilor redundante și irelevante din semnalul audio. Aceasta permite debite binare reduse fără o degradare perceptibilă a semnalului audio.

Codorul de nivel-2 folosește cadre cu 1.152 de eșantioane pe fiecare canal audio, formând câte trei grupe de câte 12 eșantioane din fiecare subbandă. Alocarea de bit are loc la nivelul unui astfel de bloc și sunt permise până la trei valori diferite ale factorului de scală (fiecare subgrupă poate avea propriul ei factor de scală, dacă este necesar pentru a evita distorsiuni audibile)

[ASZ00]. Codorul poate folosi același factor de scală pentru două sau chiar toate trei grupele, atunci când valorile sunt suficient de apropiate sau când codorul anticipează că mascarea temporală datorată sistemului auditiv uman va acoperi distorsiunea produsă prin folosirea unui singur factor de scală [NOL97].

### 3. ISO nivel-3

ISO-MPEG Audio nivel-3 (IS 11172-3 și IS 13818-3) a fost creat de Institutul Fraunhofer pentru Circuite Integrate (IIS - *Institut für Integrierte Schaltungen*) și Universitatea din Erlangen, Germania. Acesta este nivelul cel mai complex al standardului ISO și a devenit cel mai utilizat în transmisiile audio pe Internet și stocarea muzicii de către amatori. Se bazează pe sistemul ASPEC, cu unele modificări pentru un anumit grad de interoperabilitate cu nivelul 2. Codorul ASPEC original aplică direct o transformare MDCT eşantioanelor de intrare. La nivelul-3 aceasta a fost modificată pentru a utiliza o transformare hibridă care conține cele 32 de benzi din nivelurile 1 și 2. În nivelul-3, cele 32 de sub-benzi sunt prelucrate cu o transformare MDCT pe 12 benzi, obținându-se 384 de coeficienți de ieșire. Pentru a preveni fenomenul de pre-ecou la semnale tranzitorii se folosește o fereastră cu două dimensiuni. Comutarea ferestrei este realizată de modelul psihoacustic.

Pentru a beneficia de rezoluția foarte bună în frecvență se folosește un model perceptual foarte precis. De asemenea, se folosește cuantizarea neuniformă, împreună cu codarea Huffman, pentru reducerea redundanței, raportul de compresie tipic obținut fiind de 12:1.

Nivelul-3 suportă și codare/decodare cu debit binar variabil. Ultima etapă de codare prevede o procedură de codare entropică a datelor, o codare Huffman, în timp ce informațiile auxiliare (de descriere a nivelurilor de cuantizare, factori de scală etc.), sunt codate folosind o procedură de tip Run-Length. O metodă iterativă de analiză prin sinteză este utilizată pentru a menține zgomotul de cuantizare sub nivelul de mascare în timp ce procesul de scalare, cuantizare și codare entropică a eşantioanelor spectrale este reiterat în buclă. Decodarea cadrului MPEG obținut se face prin simpla inversare a pașilor de codare: extragerea parametrilor de codare, decodare Huffman, transformare MDCT inversă și trecerea prin filtrele de sinteză.

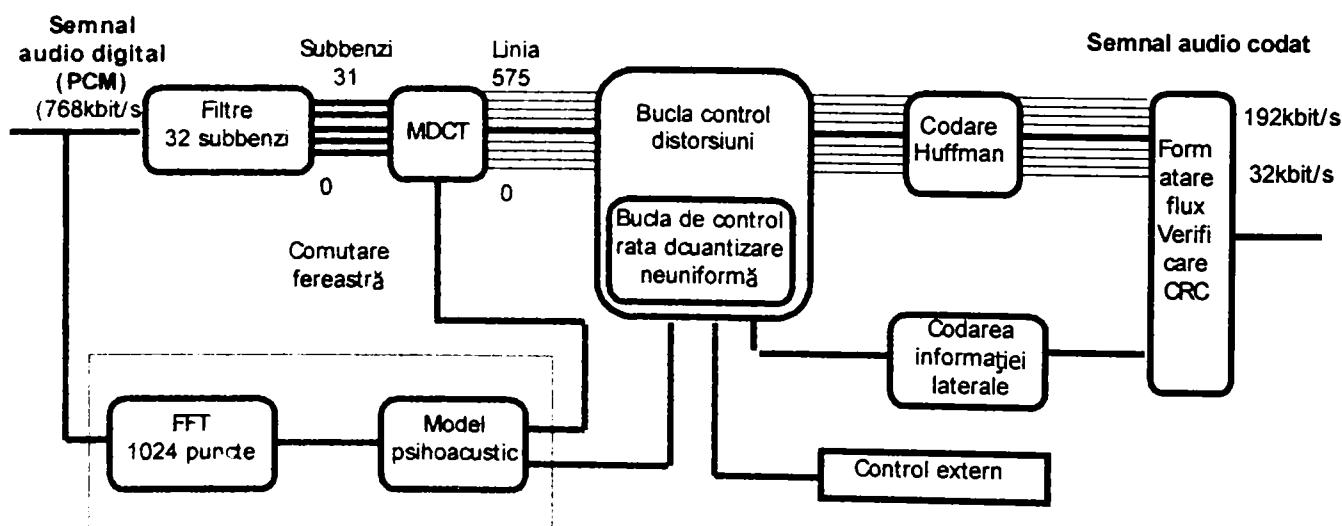


Figura 3.15 Schema bloc a unui codor MPEG-1 nivel-3

### B. ISO/IEC IS13818-7 (MPEG-2 NBC/AAC)

Standardele 11172-3 MPEG-1 și IS13818-3 MPEG-2 BC/LSF descriu algoritmi practici de codare de calitate a materialelor audio mono și stereofonice. La începutul anilor '90, a apărut necesitatea codării de calitate a sunetului multicanal la debite binare reduse. Standardul MPEG-2 BC/LSF putea coda sunet multicanal, dar condiția de compatibilitate cu MPEG-1, l-a făcut nepractic pentru codarea sunetului cu 5 canale la debite sub 640 kbit/s. De aceea MPEG a hotărât

realizarea unui sistem de codare avansat, necompatibil, care să realizeze o codare de calitate transparentă la un debit de 384 kbit/s pentru cinci canale. Astfel s-a ajuns la adoptarea standardului IS13818-7 MPEG-2 NBC/AAC (*Non-backward compatible/Advanced Audio Coding*), care prin algoritmi săi reușește o codare transparentă pe 5 canale la un debit de 320 kbit/s și reprezintă a doua generație de scheme de codare audio pentru codarea semnalelor stereo și multicanal. Deși asemănător cu predecesorii săi, algoritmul AAC conține câteva soluții noi de codare: formă adaptabilă a ferestrei bancului de filtre, predicția coeficientului spectral, formarea zgomotului temporal și scalabilitatea lățimii de bandă și a debitului binar [PAI00].

Figura 3.15 prezintă schema bloc a unui codor MPEG-2 AAC.

Obținerea unei calități comparabile la debite binare reduse este posibilă datorită următoarelor modificări față de Nivel-3:

- Rezoluție de frecvență ridicată: numărul liniilor de frecvență în AAC este de 1.024 față de 576 la nivel-3;

- Predicția spectrală;
- Codare joint stereo îmbunătățită;
- Codare Huffman îmbunătățită.

În AAC, codarea pe grupuri de câte patru linii de frecvență se aplică mai des. În plus, asignarea tabelor de cod Huffman la partițiile codorului poate fi mult mai flexibilă.

AAC conține și alte îmbunătățiri care permit obținerea unei calități foarte bune pentru clase de semnale foarte dificile:

- Comutarea blocurilor;
- TNS (*Temporal Noise Shaping*) pentru controlul preecoului.

Suma acestor mici îmbunătățiri permit AAC să obțină aceeași calitate ca și Nivel-3 la aprox. 70% din debitul binar.

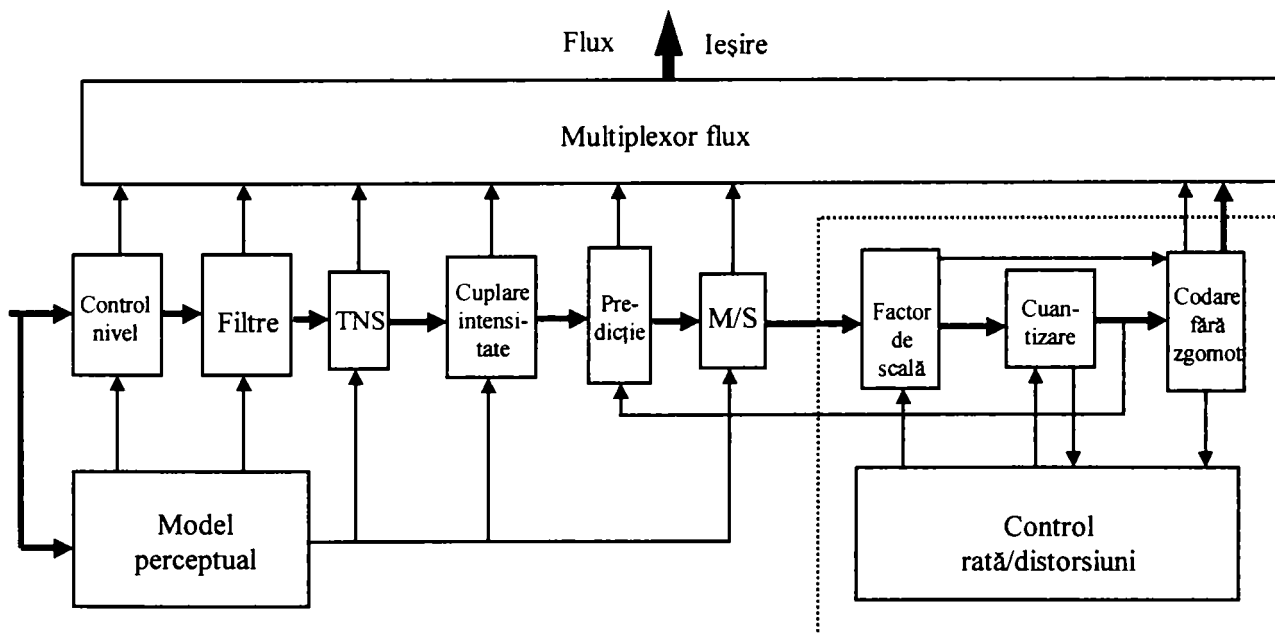


Figura 3.16 Schema bloc a unui codor MPEG-2 AAC

### C. MPEG-3

Destinat inițial codării video pentru televiziunea de înaltă definiție (HDTV), ca o fază următoare a standardelor MPEG, s-a constatat totuși că dispozitivele dezvoltate pentru codarea video în MPEG-2 conțin tot ce este necesar pentru HDTV, astfel că dezvoltarea MPEG-3 a fost inclusă în MPEG-2. Câteodată MPEG-1/2 nivel-3 (MP3) este numit din greșeală MPEG-3.



#### D. ISO/IEC 14496-3 (MPEG-4)

MPEG-4, finalizat în 1998-1999, este un standard major în lumea multimedia. Spre deosebire de MPEG-1 și MPEG-2, în MPEG-4 accentul se pune în special pe noi funcții. Cele mai importante aplicații ale MPEG-4 sunt terminale mobile și staționare, accesul în baze de date, comunicații și noi tipuri de servicii interactive. Acest standard facilitează creșterea interacțiunii și face legătura dintre tehnica de calcul, mass media electronică și telecomunicații.

MPEG-4 audio constă dintr-o familie de algoritmi de codare audio acoperind domeniul de la codarea vocală cu debit binar foarte redus (de la 2 kbit/s) până la codarea audio de înaltă calitate la 64 kbit/s pe canal și mai sus. Codarea audio generală pentru debite binare medii și înalte se face cu AAC. Caracteristicile care deosebesc MPEG-4 de predecesorii săi sunt scalabilitatea extinsă, reprezentări bazate pe obiecte, interactivitate cu utilizatorul și un larg set de unelte de codare care permit orice combinație între debitul binar, complexitate și calitate. Debitele foarte reduse se obțin prin folosirea reprezentărilor structurate pentru sinteză vocală și muzicală, precum “text-to-speech” și MIDI. Pentru debite binare mai mari și “audio natural” vocal și muzical, standardul furnizează instrumente de codare integrate care utilizează diferite modele de semnal, a căror alegere se face în funcție de debitul binar, lățimea de bandă, complexitatea și calitatea dorită.

1. *Instrumente de codare a sunetelor naturale:* prima versiune a MPEG-4 audio integrează un set de instrumente (fig. 3.17) pentru codarea sunetelor naturale cu debite binare cuprinse între 200 bps și 64 kbit/s pe canal. Pentru vocal și audio sunt integrați 3 algoritmi distincți, două codoare parametric pentru debite de 2 – 4 kbit/s și 8 kHz rată de eșantionare, precum și debite de 4 – 16 kbit/s și 8 sau 16 kHz eșantionare. Pentru o calitate vocală superioară, de bandă îngustă (eșantionare la 8 kHz) sau largă (eșantionare la 16 kHz) se folosește un codec CELP, care funcționează între 6 și 24 kbit/s. Pentru audio în general, la debite peste 16 kbit/s, se folosește un codor perceptual “timp-frecvență” și în special algoritmul MPEG-2 AAC.

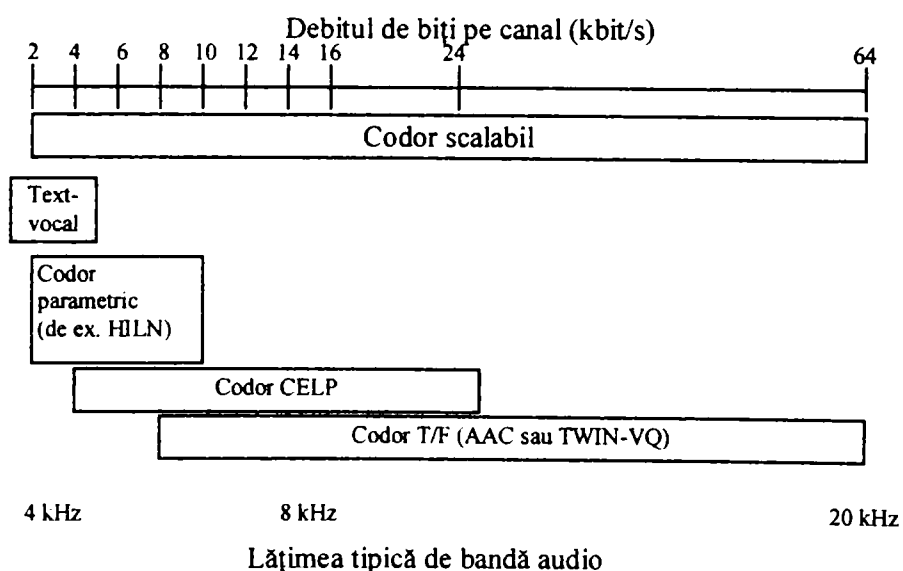


Figura 3.17 ISO/IEC MPEG-4 – unelte integrate pentru codare audio

2. *Instrumente de codare a sunetelor sintetice:* interfața text-vocal (TTSI – *text-to-speech-interface*) permite transmisii vocale sintetice cu debite între 200 și 1.200 bps, care pot fi reprezentate numai ca text sau text și parametri prozodici. MPEG-4 poate furniza și o sinteză muzicală generală, folosind un limbaj de descriere a orchestrei (SAOL – *structured audio orchestra language*). SAOL consideră muzica o colecție de instrumente și instrumentele ca o mică rețea de primitive de prelucrare a semnalului. Nu se specifică în mod expres o anumită tehnică de sinteză audio.

3. *Profile audio MPEG-4:* deși MPEG-4 audio are disponibile multe unelte de codare și prelucrare, considerentele de cost și complexitate fac ca într-un anumit sistem să nu fie toate

implementate. De aceea, versiunea 1 definește 4 profile audio, de complexități diferite, care să ajute în alegerea setului de instrumente potrivit aplicației. În ordinea crescătoare a debitului binar, ele sunt: profilul audio de sinteză, cu debit foarte redus, conține numai sinteză wavetable și interfața text-vocal (TTSI). Pentru prelucrarea audio naturală, profilul audio vocal pune la dispoziție un codor cu debit foarte redus și un codor vocal CELP. Profilul audio scalabil oferă un superset format din primele două profile, cu debite cuprinse între 6 și 24 kbit/s și lățimi de bandă între 3,5 și 9 kHz. Acest profil se recomandă pentru codare vocală și muzicală în aplicații pentru Internet streaming sau emisie digitală audio de bandă îngustă (NADIB – *narrow-band audio digital broadcasting*). În sfârșit, profilul principal este un superset al celorlalte profile și conține instrumente atât pentru semnale audio naturale cât și sintetice.

4. *MPEG-4 audio, versiunea 2*: versiunea 2, compatibilă cu versiunea 1, adaugă noi profile ce conțin un număr semnificativ de îmbunătățiri (parametrizarea proprietăților acustice ale scenei audio, modelarea proprietăților acustice ale camerei, sunet tridimensional etc.).

### E. MPEG-7

Spre deosebire de MPEG-1/2/4, MPEG-7, a cărui dezvoltare a început în 1996, nu definește algoritmi de compresie. MPEG-7, numit și “Interfață pentru descrierea conținutului multimedia”, standardizează reprezentarea conținutului pentru căutarea, filtrarea, managementul și procesarea informațiilor multimedia.

Standardul MPEG-7 este construit pe baza altor reprezentări standard: analogice, PCM, MPEG-1, -2, și -4. Una din funcționalitățile standardului este de a furniza referințe despre anumite porțiuni din reprezentările anterioare. De exemplu, un descriptor de contur folosit în MPEG-4 poate fi utilizat și în contextul lui MPEG-7 sau poate fi aplicat unui vector de mișcare utilizat în MPEG-1 sau MPEG-2.

Descriptorii din MPEG-7 nu depind de modul în care conținutul descris este codat sau stocat. O descriere MPEG-7 poate fi atașată unui film analogic sau unei fotografii pe hârtie. Deși descrierea MPEG-7 nu depinde de reprezentarea (codată) a materialului, standardul se bazează într-un fel pe MPEG-4, care permite codarea materialelor audio-vizuale ca obiecte cu o anumită relație în timp (sincronizare) și spațiu (pe ecran pentru video sau în cameră pentru audio). Utilizând codarea MPEG-4 se pot atașa descriptorii elementelor (obiectelor) din scenă, fie audio, fie vizuale. De asemenea, MPEG-7 va permite profunzimi (granularitate) diferite pentru descrierile sale, oferind posibilitatea de a avea diferite niveluri de discriminare.

Deoarece descrierea trebuie să aibă sens în contextul unei anumite aplicații, vor exista descrieri diferite în funcție de domeniul de utilizare și aplicație. Aceasta înseamnă că același material poate fi descris folosind diferite tipuri de caracteristici orientate spre domeniul de aplicație. De exemplu, în cazul unui material vizual, pentru un nivel redus de abstractizare, se va descrie forma, dimensiunea, textura, culoarea, mișcarea (traectoria) și poziția (locul din scenă în care se găsește obiectul), iar pentru audio se va da cheia, genul, tempo, modificări de tempo, poziția în imaginea sonoră. Nivelul cel mai înalt de abstractizare va da informații semantice: căutăm o scenă care să conțină în partea stângă un câine maro care latră, o minge albastră care cade în dreapta și sunet de trafic rutier în fundal.

Datele MPEG-7 pot fi asociate fizic cu materialul AV, în același flux de date sau în același sistem de stocare, dar pot exista și în altă parte, oriunde pe glob. Atunci când conținutul și descrierile nu se găsesc în același loc, se folosesc mecanisme bidirecționale de legătură între materialul AV și descrierile MPEG-7.

Există multe domenii și aplicații care ar putea beneficia de standardul MPEG-7. Câteva exemple de aplicații sunt:

- Biblioteci digitale (cataloge de imagini, dicționare muzicale, ... );
- Cataloge de servicii multimedia (de. ex. Pagini galbene);
- Selecția mediului de emisie (canal radio, canal TV, ...);
- Editare multimedia (știri personalizate etc.).

Aplicațiile potențiale sunt orientate către următoarele domenii:

- Educație;
- Jurnalism (de ex. căutarea discursurilor unui om politic folosind numele, vocea sau fața);
- Informații turistice;
- Servicii culturale (muzee de istorie, galerii de artă etc.);
- Distracție (căutare de jocuri, karaoke);
- Servicii de investigație (recunoașterea caracteristicilor umane, materiale judiciare);
- Sisteme de informații geografice;
- Supraveghere (control trafic, mijloace de transport, testare nedistructivă în medii agresive etc.);
- Aplicații biomedicale;
- Cumpărături (căutarea hainelor dorite);
- Arhitectură și design interior;
- Arhive de radio, video și film.

În încheiere trebuie remarcat că MPEG-7 nu va înlocui standardele anterioare (MPEG-1, MPEG-2 și MPEG-4) deoarece scopul său este de a furniza funcționalități complementare acestor standarde, care reprezintă informații despre conținut și nu conținutul propriu-zis (“biții despre biți”). Această funcționalitate este standardizarea descrierii conținutului multimedia.

### 3.3.2 Standardul G.722

Acest standard, dezvoltat de AT&T în S.U.A. și CNET în Franța și aprobat de CCITT (*Consultative Committee for International Telephony and Telegraphy*) în 1986, este destinat în primul rând codării vocale de bandă largă (50 – 7.000 Hz, 3 dB), dar se poate folosi și pentru codarea semnalelor audio. Sistemul se bazează pe tehnica de codare subbandă, cu două benzi la care se aplică algoritmul ADPCM (fig. 3.2).

Semnalul de intrare, codat PCM pe 14 biți, la 16 kHz, este împărțit, cu două filtre QMF identice (trece bandă), în două componente decimate (cu frecvența de eșantionare de 8 kHz), numite subbanda de jos și subbanda de sus (vezi fig. 3.18). Filtrele se suprapun și vor apărea efecte de alias datorită subeșantionării componentelor. Filtrele QMF de sinteză de la decodor asigură anularea produselor de alias. Totuși anumite erori de cuantizare din cele două subbenzi nu pot fi eliminate și de aceea se folosesc filtre QMF cu 24 de pași, cu o atenuare de bandă de 60 dB.

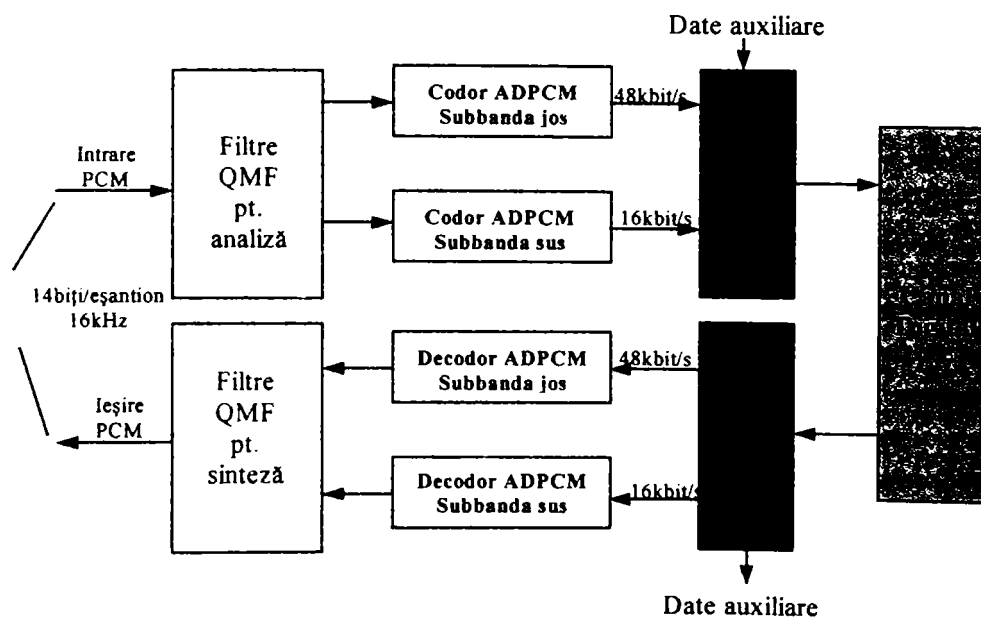


Figura 3.18 Structura unui codor vocal de bandă largă G.722

Codarea semnalelor din subbenzi se bazează pe o versiune modificată a codorului vocal ADPCM din standardul G.721, cu un debit de 32 kbit/s. Eșantioanele de intrare sunt prezise adaptiv, semnalul de eroare (diferență) fiind cuantizat și transmis. Codorul folosește o alocare fixă a biților, 6 biți/eșantion pentru subbanda de jos și 2 biți/eșantion pentru banda de sus. Tabelul 3.4 prezintă rezultatul testelor subiective de calitate efectuate pentru semnale vocale și audio, pentru două debite binare.

Tabel 3.4 Evaluarea calității subiective pentru codorul G.722

Debit binar kbit/s	Calitate vocală	Calitate audio
48	3,8	3,2
64	4,2	3,9

O versiune modificată a standardului G.722 folosește o frecvență de eșantionare de 32 kHz pentru a obține o bandă de frecvență audio de 15 kHz, rezultând un debit de 128 kbit/s pentru un canal mono.

Raportul de compresie al codorului audio G.722 este de 4:1. El oferă o bandă audio mono de 7,5 kHz pe un canal B de 64 kbit/s sau 15 kHz mono pe două canale B.

### 3.3.3 Algoritmul apt-X100

Sistemul apt-X100, creat de APT (*Audio Processing Technology*) din Belfast, Irlanda de Nord, se bazează pe tehnica ADPCM (*Adaptive Differential Pulse Code Modulation*). Acest algoritm de codare se bazează pe faptul că, în general, semnalul audio are o variație lentă în timp și care poate fi prezisă.

Codecul apt-X100 folosește tehnici de codare predictivă în patru subbenzi pentru a obține o compresie de 0,25 din debitul original. Subbenzile sunt obținute cu filtre în oglindă în cuadratură QMF și în fiecare subbandă are loc o codare predictivă continuă combinată cu o decodare continuă la receptor. Blocurile nu sunt folosite la codare ci numai pentru împachetarea valorilor diferență pentru transmisie. Blocul constă din 2.048 de biți și începe cu un semnal de sincronizare care permite decodorului să asambleze corect valorile diferențelor și să le atribuie subbenzii corespunzătoare.

Deoarece procesul se bazează pe forma de undă și nu pe conținutul spectral, nu este necesar un model de mascare auditivă mai precis și nici un număr mai mare de subbenzi. Drept urmare, apt-X100 poate funcționa cu diferite rate de eșantionare fără modificări, în timp ce pentru majoritatea codoarelor modificarea ratei de eșantionare înseamnă frecvențe diferite pentru sub-benzi și alți parametri de mascare. Un alt avantaj important al codării predictive este întârzierea redusă produsă de codec, sub 4 ms, ceea ce îl recomandă pentru aplicații live [WAT95]. Întârzierea pentru un ciclu complet de codare/decodare, prezentată în tabelul 3.5, se datorează în întregime utilizării filtrelor QMF și este invers proporțională cu frecvența de eșantionare.

Tabel 3.5 Întârzierea totală introdusă de algoritmul apt-X100

Rata de eșantionare (kHz)	Întârzierea (ms)
48	2,5
44,1	2,9
32	3,8
16	7,6

În funcție de aplicație, codecul poate permite o bandă audio de 7,5 kHz, ideală pentru știri și transmisii sportive, care folosește un canal B de 64 kbit/s, o bandă audio mono de 15 kHz, folosind ambele canale B de 64 kbit/s, sau pentru aplicații audio mai pretențioase, studiourile pot transmite și recepționa întreaga bandă audio stereo de 22,5 kHz folosind trei linii ISDN BRI (384 kbit/s).

### 3.3.4 Algoritmii Dolby AC-2 și AC-3

Dolby AC-2 și AC-3 sunt două tehnologii flexibile de compresie a datelor audio, destinate codificării unui canal audio, în cazul AC-2, sau mai multor canale audio într-un singur flux cu debit redus, pentru sunet surround cu 5.1 canale, în cazul AC-3.

1 – AC-2, AC-2A: AC-2 este o familie de algoritmi monocanal cu un debit între 128 și 192 kbit/s pentru un semnal de intrare cu o bandă de frecvență de 20 kHz, eșantionat la 44,1 sau 48 kHz. Sunt disponibile patru variante AC-2, care au în comun transformarea semnalului de intrare în domeniul frecvență printr-un filtru par TDAC cu o fereastră de analiză Kaiser-Bessel optimizată.

Două din cele patru variante AC-2 sunt destinate aplicațiilor cu complexitate și întârziere redusă, iar celelalte două pentru o calitate ridicată, implicând o complexitate și întârziere mai mare. Algoritmii AC-2A folosește un filtru TDAC cu comutare 128/512 puncte pentru îmbunătățirea calității la semnale tranzitorii. O caracteristică unică a AC-2 printre standardele audio de codare este modelul perceptual adaptabil înapoi, ceea ce înseamnă că alocarea biților nu se transmite explicit. Decodorul AC-2 extrage alocarea biților din înfășurătorul spectrală cuantizată folosind același model perceptual ca și codorul. Această structură duce la o reducere semnificativă a informației laterale și la o complexitate simetrică a codorului și decodorului, potrivită scopurilor inițiale ale AC-2 de mijloc de transport audio punct la punct. Un exemplu de sistem punct la punct care folosește algoritmul AC-2 cu întârziere redusă este codecul full-duplex DolbyFAX, care poate transmite simultan două canale în ambele direcții prin patru fluxuri B ISDN, asigurând colaborarea de la distanță dintre un studio de filmare și unul de televiziune. Legătura între două codecuri AC-2 se poate face și printr-un sistem de tip radioreleu în banda de 950 MHz (DSTL - *Digital Studio to Transmitter Link*).

Testele de evaluare subiectivă a calității efectuate în Australia și Franța au dat o valoare medie de 4,8 pe scara CCIR în 5 trepte, pentru un debit de 128 kbit/s pe canal. Rezultatele testelor de calitate și estimarea complexității implementării au condus CCIR TG10/2 să recomande AC-2 ca deplin corespunzător pentru aplicații de emisie. De aceea tehnologia AC-2 se folosește în sistemele de televiziune de înaltă definiție (HDTV) sau dispozitivele digitale de transmitere a semnalului între studio și emițător (DSTL) [BOS92].

Un avantaj al tehnologiei AC-2, care poate fi exploatat în aplicații multimedia este posibilitatea editării și prelucrării materialului audio codat. Datorită simetriei și complexității de calcul reduse a codecului, AC-2 poate înregistra/coda un canal simultan cu redarea/decodarea altui canal, fiind implementat pe un singur chip. În funcție de aplicație, se pot folosi rate de eșantionare între 16 și 48 kHz, obținându-se calitatea audio necesară aplicațiilor de studio.

2 – *Dolby AC-3/Dolby Digital/Dolby SR-D*: Formatul “surround” cu 5.1 canale, standardul de facto în redarea sunetului în cinematografe în anii '80, a devenit omniprezent în anii '90 în instalațiile casnice echipate cu sunet multicanal (de ex. Dolby ProLogic). Ca rezultat al acestor tendințe, a devenit evident că, pentru a cuceri marele public, aplicațiile viitoare ale codării audio perceptuale trebuie să poată prelucra sunet stereofonic sau multicanal. Deși algoritmi monocanal precum AC-2 pot funcționa în paralel pe mai multe canale independente, performanțele sistemului de codare se pot îmbunătăți prin tratarea simultană a canalelor pentru a exploata redundanțele și irelevanțele dintre canale.

Algoritmul Dolby Laboratories AC-3, denumit și “Dolby Digital” sau “SR-D” este următoarea generație de tehnologii de compresie Dolby, destinat aplicațiilor în sisteme audio multicanal. El îmbunătățește toate blocurile de bază ale algoritmului AC-2 – filtre, codarea înfășurătoarei spectrale, modelul perceptual și alocarea biților. Codorul poate prelucra 5.1 canale audio (stânga, centru, dreapta, stânga spate, dreapta spate și subwoofer), dar permite și compatibilitatea cu sistemele de redare mono sau stereofonice. Canalul “.1” este rezervat efectelor de frecvență joasă, banda de trecere fiind limitată la 120 Hz.

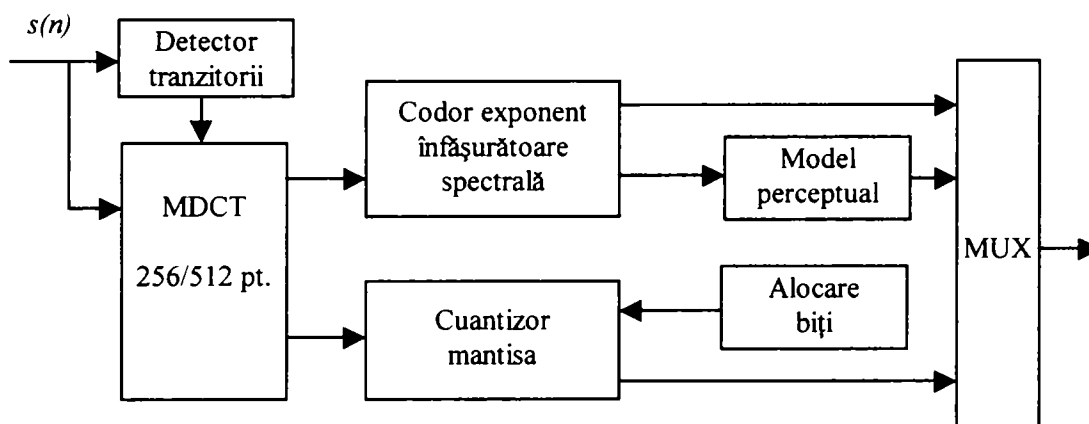


Figura 3.19 Codor AC-3

Principalele caracteristici ale algoritmului AC-3 sunt următoarele:

- rate de eșantionare: 32, 44,1 și 48 kHz;
- ieșire de înaltă calitate la un debit de 64 kbit/s/canal;
- filtre MDCT (TDAC), fereastră KBD;
- înfășurătoare spectrală reprezentată prin exponenți;
- model perceptual hibrid adaptabil înainte-înapoi;
- cuantizare uniformă a mantiselor;
- tratarea canalelor multiple ca un ansamblu;
- compatibilitate în jos a decodorului;
- debit binar: 32-640 kbit/s, variabil;
- întârziere aprox. 100 ms;
- alocarea parametrică a biților;
- posibilități de perfecționare a modelului perceptual;
- codare stereofonică în intensitate, în funcție de frecvență, LR, MS;
- sistem de control al dinamicii;
- decodare integrate în timp real.

Prima aplicație a algoritmului AC-3 a fost în cinematografie. Informațiile de tip “Dolby Digital” sau “SR-D” erau intercalate între fantele de antrenare ale peliculei de 35mm, pe o singură parte. Sistemul a fost instalat pentru prima dată în 1991, în trei cinematografe pentru filmul *Star Trek VI*, lansarea oficială făcându-se în 1992 cu *Întoarcerea lui Batman*. Până în 1997, mai mult de 900 de coloane sonore fuseseră codate cu AC-3. Astăzi, algoritmul AC-3 se utilizează în DVD, televiziunea prin cablu sau emisia directă prin satelit DBS (*Direct Broadcast Satellite*), decodoarele AC-3 sunt incluse în receptoare și amplificatoare de calitate. De asemenea, diferite organisme de standardizare internaționale (Digital Audio-Visual Council – DAVIC, Society of Cable and Telecommunication Engineers etc.) sau din S.U.A. (U.S. Advanced Television Systems Committee – ATSC) au inclus algoritmul AC-3 pentru componenta audio.

### 3.3.5 Tehnica de codare PASC (*Precision Adaptive Sub-band Coding*)

Tehnica de codare subbandă adaptivă de precizie PASC a fost realizată de Philips pentru caseta compactă digitală DCC (*Digital Compact Cassette*) și lansată în 1992. Ea este similară cu ISO/MPEG Nivel-1, sistem la care Philips a contribuit de asemenea. Sistemul de transport DCC funcționează cu un debit de 384 kbit/s, algoritmul PASC reducând debitul semnalului audio cu un raport de 4:1 în condițiile păstrării calității subiective a sunetului înregistrat.

În codorul PASC se utilizează 32 de sub-benzi egale produse cu filtre QMF având 512 coeficienți. Se pot folosi frecvențe de eșantionare de 32, 44,1 și 48 kHz, care conduc la sub-benzi cu lățimea de 500, 689 și 750 Hz și respectiv întârzieri de codare de 15, 12 și 10 ms. În fiecare sub-bandă se creează blocuri de câte 12 eșantioane cu o perioadă de 8 ms la 48 kHz.

Schema bloc a unui codor PASC este redată în fig. 3.20. Semnalul de intrare, care poate fi până la 18 biți PCM, este filtrat, rezultând eșantioane de 24 biți pe sub-benzi. Blocurile de eșantioane din fiecare sub-bandă sunt codate în virgulă mobilă. Se determină un factor de scală corespunzător valorii de vârf a celor 12 eșantioane dintr-un bloc. Eșantioanele sub-benzii sunt normalizate prin factorul de scală prin divizare. Factorul de scală reprezintă exponentul notației în virgulă mobilă. El este identic pentru toate eșantioanele codate ale sub-benzii și este înregistrat împreună cu eșantioanele codate. După normalizare eșantioanele reprezintă mantisa. În această fază ele sunt cuantizate în funcție de numărul de biți alocați sub-benzii (de ex. pentru 4 biți se obține o cuantizare pe 15 niveluri). Lungimea maximă a mantisei este de 15 biți, adică o rezoluție de 92 dB, dar această rezoluție poate fi plasată oriunde în gama +6 dB la -118 dB cu pași de 2 dB și un factor de scală pe 6 biți.

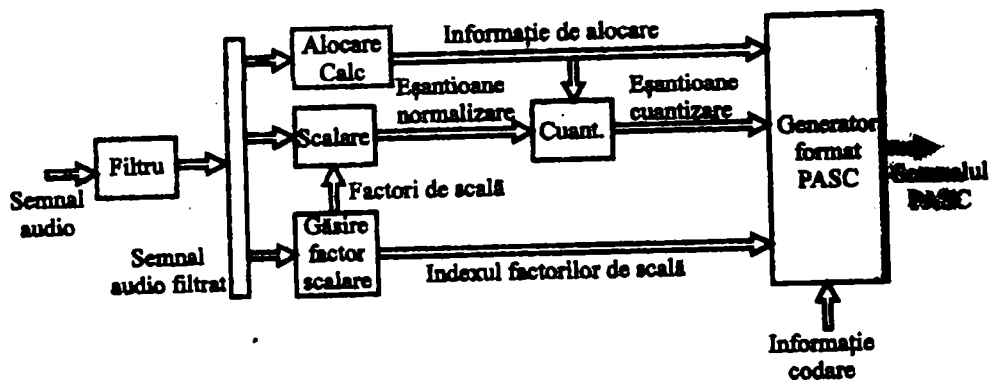


Figura 3.20 Schema bloc a unui codor PASC [OTE97]

Pentru stabilirea pragului de mascare, se calculează puterea semnalelor din subbandă pentru cadrul PASC corespunzător și apoi se determină contribuția mascării dintre sub-benzi sau a mascării din interiorul sub-benzii prin aplicarea pragului absolut de audibilitate. Pragul de mascare rezultat este comparat cu vârfurile de putere ale semnalelor din sub-bandă. Dacă vârfurile de putere dintr-o sub-bandă se află sub pragul de mascare înseamnă că informația audio este irelevantă și nu trebuie codificată. Dacă vârfurile de putere ale unei sub-benzii se află peste pragul de mascare, ele vor fi transferate ca un număr de mantisă a cărei lungime este proporțională cu distanța dintre vârful de putere și pragul de mascare. Deoarece pragul de mascare este recalculat pentru fiecare cadru PASC transferat, procesul se numește alocare adaptivă a capacității disponibile.

Principiile de mascare pot fi implementate în reprezentarea eșantioanelor sub-bandă variind lungimea mantisei. Rezoluția trebuie să fie doar atât de precisă cât să mențină zgomotul de cuantizare sub pragul de mascare [OTE97].

### 3.3.6 Algoritmul ATRAC (*Adaptive TTransform Acoustic Coding*)

Codarea acustică cu transformare adaptivă ATRAC, utilizată în minidisc, a fost dezvoltată de SONY și introdusă în 1992.

Codarea ATRAC este o combinație între codarea sub-bandă și codarea cu transformare sinus modificată (MDCT) prin care se realizează o împărțire neuniformă a axelor atât în frecvență cât și în timp, în concordanță cu principiile psihoacustice. Scopul este de a reduce debitul de date audio într-un raport de 5:1 (de la 1,4 Mbps la 292 kbit/s), folosind principii psihoacustice și păstrând un nivel de calitate ridicat. Procesul de codare este realizat în trei etape:

- divizarea în trei subbenzi a semnalului de intrare, cu ajutorul a două filtre oglindă în cuadratură QMF, în doi pași: separarea frecvențelor înalte (11,05 – 22,05 kHz) și separarea frecvențelor medii și joase (5,5025 – 11,025 kHz și 0 – 5,5025 kHz)
- compresia subbenzilor prin transformare sinus discretă modificată MDCT
- alocarea biților

Ca rezultat al codării ATRAC, datele de ieșire nu se pot descrie ca fiind date audio, ci date care descriu componente ale spectrului audio în lungimi de cuvinte și factori de scalare [OTE97].

### 3.3.7 Sistemul SDDS (*Sony Dynamic Digital Sound*)

SDDS reprezintă un sistem de sunet digital cinematic care conține opt module ATRAC independente pentru transportul informației unui program sonor multicanal folosit astăzi în cinematografe (stânga, centru-stânga, centru, centru-dreapta, dreapta, subwoofer, stânga-spate, dreapta-spate). Datele sistemului SDDS sunt înregistrate optic pe marginile peliculei, stânga și dreapta, în afara găurilor de antrenare, sub formă de matrice de puncte albe și negre, fiecare parte conținând patru canale. Fiecare canal asociat unui cadru de film conține 512 biți ATRAC, iar fiecare cadru optic conține o matrice de 52x192 biți. Datele SDDS nu înlocuiesc și nu se interferează cu sunetul analogic normal al filmului. Pentru a preveni erorile de citire datorate zgârieturilor pe film, particulelor de praf, defocalizare la citire sau copierea filmului, se folosește codul de corecție Reed-Solomon și informații redundante întârziate cu 18 cadre. În cazul unor erori digitale necorectabile se folosește informația de program analogic [PAI00].

### 3.3.8 Algoritmii Lucent Technologies PAC (*Perceptual Audio Coder*), EPAC (*Enhanced PAC*) și MPAC (*Multichannel PAC*)

Ca și codoarele MPEG, algoritmul Lucent PAC, folosit în sistemul Sirius Radio, este flexibil, putând suporta sunet monofonic, stereofonic și multicanal. De fapt, prin definiție, fluxul de date poate cuprinde până la 16 canale față, șapte spate (surround), șapte perechi de canale auxiliare și trei pentru frecvențe joase (LFE – *low frequency effects* sau *subwoofer*). În funcție de calitatea dorită, PAC suportă diferite debite binare. În continuare vor fi prezentate cele trei variante – PAC, EPAC și MPAC – insistând în special asupra noutăților care fac ca acești algoritmi să difere de cei prezentați anterior [PAI00].

1 – PAC: Sistemul PAC original asigură o codare stereofonică de calitate la 96 kbit/s. Ca și sistemele MPEG-1 nivel-3 și ATRAC, codorul PAC (fig. 3.21(a)) folosește un banc de filtre MDCT adaptabil la semnal pentru a analiza spectrul semnalului de intrare cu o rezoluție de frecvență potrivită. Pentru segmente staționare se folosește o fereastră lungă de 2.048 de puncte (1.024 de subbenzi) sau, în cazul segmentelor conținând sunete tranzitorii sau cu atac precis, se folosește o serie de ferestre scurte de 256 de puncte (128 de subbenzi). Spre deosebire de MPEG-1 și ATRAC, PAC se bazează numai pe transformarea MDCT, fără a incorpora analiza MDCT în structura unui banc de filtrare hibrid, realizând o simplificare a structurii bancului de filtre.



Pragurile de mascare sunt folosite pentru a alege unul din cei 128 de pași de cuantizare distribuiți exponențial în fiecare din cele 49 sau 14 benzi (ca și în ATRAC BFU) corespunzătoare modurilor de înaltă sau joasă rezoluție. Benzile codorului sunt cuantizate folosind o buclă de control cu rată iterativă în care pragurile sunt potrivite pentru a satisface simultan constrângerile legate de debit și criteriul de egală tărie (*equal loudness*) care încearcă să modeleze zgomotul de cuantizare astfel încât tăria sa absolută să fie constantă în raport cu pragul de mascare. Bucla de control a debitului permite debite binare variabile în timp, asemănător cu “rezervorul de biți” de la MPEG-1 nivel-3. Redundanțele statistice rămase sunt eliminate din fluxul de eşantioane spectrale cuantizate folosind opt tabele structurate Huffman multidimensionale, înainte de formarea fluxului de date final.

2 – *EPAC*: Pentru a îmbunătăți calitatea algoritmului PAC la debite reduse, Sinha și Johnston au introdus o nouă schemă de filtre adaptive MDCT/WP (fig. 3.21(b)), care a permis codarea aproape transparentă a unui material audio de calitate CD cu un debit de 64 kbit/s pentru o pereche de canale stereo. Unicitatea algoritmului EPAC constă în comutarea între două filtre diferite față de utilizarea unui filtru hibrid sau mai multe filtre neuniforme succesive. În cazul testelor de ascultare critice, realizate cu 12 ascultători, cu semnalele de test castaniete și triunghi, EPAC a depășit PAC la un debit de 64 kbit/s stereo cu o medie de 0,4 - 0,6 puncte pe scara de calitate subiectivă în 5 trepte.

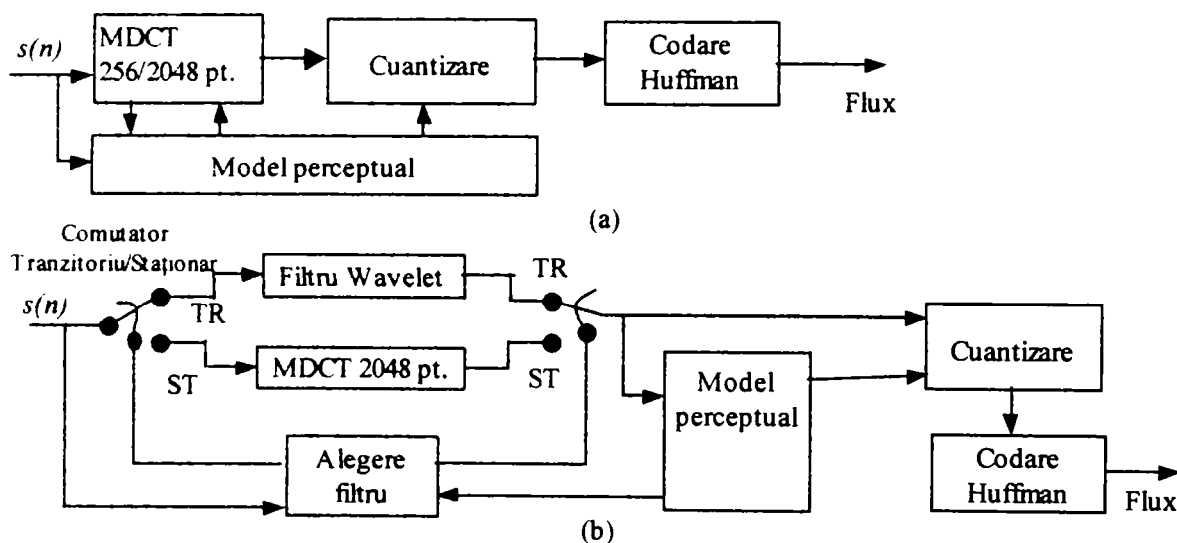


Figura 3.21 Schema bloc a codorului Lucent Technologies: (a) PAC și (b) EPAC

3 - *MPAC*: Ca și în cazul sistemelor MPEG, AC-3 și SDDS, și algoritmul PAC a fost extins pentru a prelucra semnale stereofonice sau multicanal. Algoritmul PAC stereo calculează pragurile individuale de mascare pentru semnalele stânga, dreapta, sumă și diferență ( $L$ ,  $R$ ,  $M = L + R$ ,  $S = L - R$ ) folosind o versiune mono a modelului perceptual modificată pentru a lua în considerare diferența binaurală de mascare *BLMD* (*Binaural Level Masking Difference*). Apoi, perechile  $L$ ,  $R$  sau  $M$ ,  $S$  sunt codate cu metode PAC mono. Totuși, pentru a minimiza debitul total, în bucla de control a debitului s-a introdus o procedură de comutare LR/MS, astfel încât în același cadru benzile pot fi codate diferit (LR sau MS). Într-un test recent al ISO se consideră că algoritmul MPAC dă cea mai bună calitate pentru cinci canale la un debit de 320 kbit/s (ISO94).

4 – *Aplicații*: Versiunile sistemului stereofonic PAC de 128 și 160 kbit/s se află în curs de standardizare în SUA, în cadrul proiectului de radio digital DAR (*Digital Audio Radio*). De asemenea, existența unui decodor JAVA PAC face ca algoritmul PAC să fie tot mai mult utilizat de furnizorii de programe audio în Internet. În sfârșit, algoritmul MPAC poate fi utilizat în cinematografie și televiziune.

### 3.4 Concluzii

Aplicațiile actuale curente ale codării audio sunt emisia și transmisia digitală, DVD-ul, televiziunea de înaltă definiție HDTV, cinematografia cu sunet spațial multicanal, și audio la comandă prin rețele de date globale, precum Internetul. Codarea audio a permis miniaturizarea mediilor de stocare audio digitale, precum minidiscul sau caseta digitală compactă DCC. Formatul audio “.mp3”, care indică fișierele audio comprimate folosind algoritmul de nivel 3 din standardul MPEG-1, a devenit principalul format pentru schimbul de informații multimedia prin Internet și, în ultimul timp, a fost integrat în diferite dispozitive de redare audio portabile.

Gradul de sofisticare și performanțele înalte atinse de standardele din tabelul 3.6 reflectă maturizarea rapidă a algoritmilor de codare audio în ultimii zece ani. Acum, accentul se pune pe realizarea unor algoritmi cu debite cât mai reduse, cu complexitate și întârziere redusă. Pe lângă scopurile inițiale de compresie transparentă la debite binare reduse (sub 64 kbit/s/canal) cu complexitate redusă, întârziere minimă și sensibilitate redusă la erori, o nouă direcție de cercetare în domeniul codării audio este dezvoltarea unor algoritmi care oferă scalabilitate. Algoritmii scalabili vor fi folosiți pentru transmisii audio prin intermediul rețelelor cu pachete comutate care compun Internetul sau telefonía mobilă. O altă direcție o reprezintă convergența dintre algoritmii de codare audio cu dbr și codoarele vocale, care folosesc tot mai mult mecanisme ce elimină irelevanțele perceptuale. Perfecționări continue se aduc de asemenea și blocurilor constructive ale codoarelor perceptuale, precum noi bănci de filtre pentru reducerea întârzierilor de codare și a pre-ecoului și noi tehnici de analiză psihoacustică a semnalului.

Tabel 3.6 Standarde și aplicații pentru codoare audio

Algoritm	Rata de eșantionare (kHz)	Canale	Debit binar (kbit/s)	Aplicații
APT-X100	44,1	1	176,4	Cinema
ATRAC	44,1	2	256/canal	MiniDisc
Lucent PAC	44,1	1 – 5.1	128/stereo	DBA: 128/160kbit/s
Dolby AC-2	44,1	2	256/canal	DBA
Dolby AC-3	44,1	1 – 5.1	32 – 384	Cinema, HDTV
MPEG I, LI-III	32, 44,1, 48	1, 2	32 – 448	“MP3”: LIII DBA: LII@256 kbit/s DBS: LI@224 kbit/s DCC: LI@384 kbit/s
MPEG-2	32, 44,1, 48, 16, 22, 24	1 – 5.1	32 – 640	Cinema
MPEG-2/AAC		1 – 96	8 – 64/canal	Internet/www
MPEG-4		1 -	200bit/s – 64kbit/s/canal	General

În radiodifuziune, fie ea analogică sau digitală, cele mai uzuale metode de codare audio sunt cele prezentate în tabelul 3.7, unde am calculat câteva mărimi specifice în funcție de algoritm și debitul binar utilizat, în comparație cu formatul PCM necomprimat. De asemenea, am făcut o clasificare în funcție de criterii subiective de calitate și am calculat capacitatea totală de stocare a unor fișiere audio în format stereo pentru câteva dimensiuni uzuale de harddiscuri. Pentru a avea o idee asupra numărului aproximativ de piese muzicale ce pot fi stocate, în diferite scopuri, în funcție de calitate, pe harddiscuri de diferite dimensiuni, am calculat numărul de piese cu o durată medie de 3 minute.

În domeniul codării audio de calitate cu debit binar redus, între 64 și 128 kbit/s, codoarele bazate pe modelul sursei și predicția liniară (PL) nu au avut un succes deosebit. Cu excepția algoritmului Twin VQ, codecurile audio cu PL au rămas în general la nivel experimental. Cu ocazia ultimelor cercetări în vederea codării hibride vocale și audio cu debite foarte reduse, sub

16 kbit/s, tehnicile de codare audio cu PL au fost din nou luate în considerare. Cu ocazia testelor de ascultare din cadrul procesului de standardizare a MPEG-4 s-a observat că alegerea modelului de semnal depinde de materialul sursă. De exemplu, un codor CELP depășește calitativ un codor sinusoidal pentru semnale vocale, dar codorul sinusoidal depășește codorul CELP pentru muzică. De aceea, se dorește realizarea în viitor a unui algoritm hibrid care să combine părțile tari ale ambelor modele. Deși PL stă la baza algoritmilor vocali moderni, totuși în cazul codării audio i s-a acordat mai puțină atenție, ele nefiind potrivite pentru modelarea componentelor sinusoidale prezente în semnalele audio staționare. Corespondența dintre sistemul de analiză-sinteză cu PL și modelul vocal au stat la baza succesului predicției liniare în aplicații vocale. În cazul codării audio însă, succesul PL este dependent de natura semnalului.

În final, se fac eforturi deosebite pentru realizarea unor metode cât mai eficiente de măsurare subiectivă și obiectivă a calității audio percepute în vederea evaluării codoarelor. De fapt, după o fază competițională și apoi o colaborare dintre mai multe colective, ITU-R a adoptat standardul BS.1387, care prezintă o metodă obiectivă de măsurare a calității audio, destinată selecției, evaluării și întreținerii codecurilor audio perceptuale (vezi cap. 6).

**Tabel 3.7** Compresia audio digitală – caracteristici practice ale unor algoritmi utilizați în radiodifuziune

<b>Calitate CD</b>	<b>Compresie</b>	<b>Debit binar</b>	<b>Dimensiunea fișierului în kB/minut</b>		
PCM Wav	1:1	-	10.336 Stereo 5.168 Mono		
MPEG Nivel-2	3:1	384	2.810 Stereo 1.405 Mono		
<b>Calitate aproape CD</b>	<b>Compresie</b>	<b>Debit binar</b>	<b>Dimensiunea fișierului în kB/minut</b>		
ADPCM	4:1	-	2.600 Stereo 1.300 Mono		
MPEG1 nivel-2	6:1	256	1.873 Stereo 936,5 Mono		
MPEG1 nivel-3	10:1	128	934,6 Stereo 467,3 Mono		
<b>Calitate FM</b>	<b>Compresie</b>	<b>Debit binar</b>	<b>Dimensiunea fișierului în kB/minut</b>		
MPEG1 nivel-2	8:1	192	1.404 Stereo 702 Mono		
MPEG1 nivel-3	12:1	112	818 Stereo 409 Mono		
<b>Timp total de stocare audio stereo</b>	<b>6 GB HDD</b>	<b>10 GB HDD</b>	<b>20 GB HDD</b>	<b>40 GB HDD</b>	<b>60 GB HDD</b>
PCM WAV	10h 08m 41s	16h 54m 30s	33h 47m	67h 35m	101h 23m
MPEG1 n-2, 3:1	37h 18m 20s	62h 10m 34s	124h 19m	248h 38m	372h 57m
ADPCM	40h 20m 26s	67h 14m 03s	134h 21m	268h 43m	403h 04m
MPEG1 n-2, 6:1	55h 58m 48s	93h 18m 01s	186h 30m	373h 01m	559h 31m
MPEG1 n-3, 10:1	112h 11m 41s	186h 59m 28s	373h 46m	747h 33m	1.121h 20m
MPEG1 n-2, 8:1	74h 42m 05s	124h 30m 26s	248h 48m	497h 37m	746h 26m
MPEG1 n-3, 12:1	128h 11m 04s	213h 38m 26s	427h 03m	854h 07m	1.124h 20m
<b>Număr piese de 3 minute stereo</b>	<b>6 GB HDD</b>	<b>10 GB HDD</b>	<b>20 GB HDD</b>	<b>40 GB HDD</b>	<b>60 GB HDD</b>
PCM WAV	202	338	675	1.351	2.027
MPEG1 n-2, 3:1	746	1.243	2.486	4.972	7.459
ADPCM	806	1.344	2.687	5.374	8.061
MPEG1 n-2, 6:1	1.119	1.866	3.730	7.460	11.190
MPEG1 n-3, 10:1	2.243	3.739	7.475	14.951	22.426
MPEG1 n-2, 8:1	1.494	2.490	4.976	9.952	14.928
MPEG1 n-3, 12:1	2.563	4.272	8.541	17.082	22.486

# CAPITOLUL 4

## METODE DE EVALUARE A CALITĂȚII AUDIO

### 4.1 Introducere

Reprezentările digitale ale formelor de undă analogice au ca efect apariția unor distorsiuni, numite *erori de codare* sau *zgomot de codare*. Aceste distorsiuni pot fi evaluate prin criterii subiective și criterii obiective. Criteriile obiective pot fi simple, precum raportul semnal/zgomot, ca o măsură a *similarității formei de undă* dintre sursă și semnalul reconstruit sau criterii obiective complexe, ca măsură a similarității perceptuale, care țin cont de caracteristicile percepției auditive umane.

Transmiterea și stocarea digitală a semnalelor audio depinde din ce în ce mai mult de algoritmi de compresie care țin cont de proprietățile sistemului auditiv uman. Scopul unui astfel de algoritm este de a controla distribuția spectrotemporală a distorsiunilor de codare rezultate și de a le menține sub pragul de audibilitate. Distorsiunile devenite astfel neauzibile sunt totuși fizic prezente, calitatea acestor codare perceptuale neputând fi evaluată precis prin metode convenționale de măsură a nivelului general de distorsiuni. Un exemplu pentru a ilustra această limitare [THI00] este adăugarea unui zgomot cu o structură spectrală adaptată semnalului audio și care nu este sesizabil deși raportul semnal/zgomot scade cu 13 dB. De aceea calitatea audio a unui codor perceptual este evaluată de obicei folosind criterii subiective. Deoarece testele subiective sunt scumpe sau dificil de realizat, este necesară o metodă de măsurare obiectivă care să modeleze procesele senzoriale și cognitive ce conduc la evaluările subiective (vezi cap. 6).

Scheme de măsurare a calității obiective ce utilizează proprietățile sistemului auditiv uman există din 1979 și au fost aplicate în special pentru codecurile vocale. Prima metodă de măsurare perceptuală aplicată unui codec audio de bandă largă a fost introdusă în 1987.

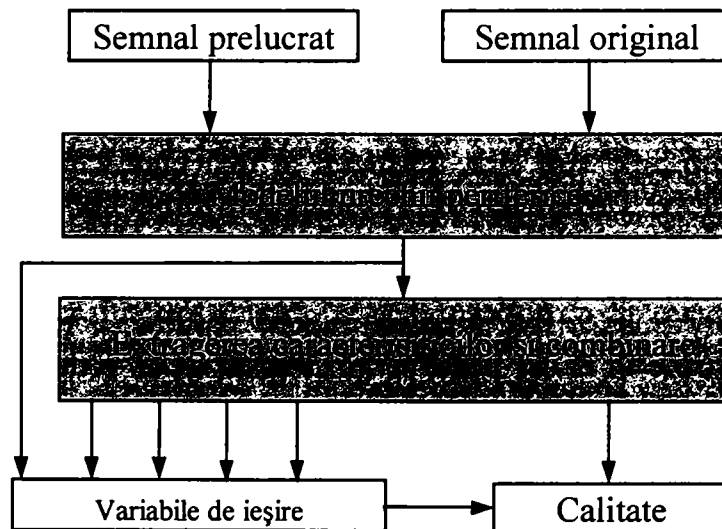


Figura 4.1 Schema bloc a schemei de măsurare

Măsurările obiective încearcă să modeleze percepția umană a semnalelor acustice. Metoda standard propusă, numită **PEAQ** (*Perceptual Evaluation of Audio Quality* - evaluarea perceptuală a calității audio), se bazează pe principii psihoacustice general acceptate. În figura 4.1 se prezintă o schemă bloc a metodei. În general, un semnal prelucrat printr-o anumită metodă este comparat cu semnalul original, aliniat în timp. Cadrele corespunzătoare ale semnalului original și prelucrat sunt transformate într-o reprezentare a membranei basilare și diferențele sunt

apoi analizate ca o funcție de frecvență și timp pe baza unui model cognitiv. Sunt extrase caracteristicile perceptuale relevante sub forma variabilelor de ieșire ale modelului **MOV** (*model output variables*), care se folosesc apoi pentru a calcula o măsură a calității.

Un anumit set de variabile este mapat pe o scală de calitate obiectivă. Această mapare s-a realizat prin minimizarea diferenței distribuției măsurărilor obiective și distribuția de calitate subiectivă corespunzătoare unui anumit set de date.

Fiabilitatea metodelor de măsurare a calității audio obiective poate fi estimată numai prin compararea rezultatelor obținute prin teste de ascultare critică aplicate aceluiași semnale audio. Recomandarea ITU-R descrie în amănunt metodologia de testare bazată pe judecăți subiective pentru a obține așa numita "calitate audio de bază" a unui dispozitiv. În cazul PEAQ se dorește obținerea aceleiași calități audio de bază, dar prin metode de măsurare obiective.

## 4.2 Tipuri comune de artefacte

Artefactele produse de codarea audio diferă mult prin natura și audibilitatea lor de bine cunoscutele distorsiuni audio din sistemele analogice (wow&flutter, saturația benzii, intermodulație, neliniaritate etc.) și sunt greu de identificat într-un sistem audio format dintr-un lanț de diferite echipamente de prelucrare și codare/decodare a sunetului.

În continuare se prezintă câteva tipuri de artefacte de codare: preecoul, artefacte produse în codarea vocală, diferența de nivel datorată mascării binaurale, pierderea imaginii stereofonice, limitarea în frecvență (sus), artefacte de aliere, artefacte datorate codării succesive (în tandem) și cele datorate codării muzicii cu codoare vocale.

### 4.2.1 Pierderea lățimii de bandă

Dacă un codor nu reușește să codeze un bloc de date muzicale cu fidelitatea cerută (de ex. zgomotul admis în banda critică) în limitele debitului binar disponibil, anumite linii de frecvență vor fi anulate (șterse). De obicei limitările sunt în domeniul frecvențelor înalte. Dacă banda de frecvențe nu este constantă și se schimbă de la un cadru la altul (de ex. la fiecare 24 ms), efectul devine mai supărător decât în cazul unei reduceri constante a benzii de frecvențe.

### 4.2.2 Preecoul

Precourile sunt artefactele cele mai obișnuite și mai cunoscute în sistemele de codare perceptuală de înaltă rezoluție. Denumirea de *preecou* poate înșela într-un fel: artefactul principal de codare este un zgomot ce apare chiar înaintea fragmentului muzical care l-a creat. În schema bloc a decodatorului unui sistem perceptual de codare, liniile de frecvență reconstruite sunt combinate într-un filtru de sinteză. Filtrul constă dintr-o matrice de modulare și o fereastră de sinteză. Eroarea de cuantizare în codor poate fi considerată ca un semnal adăugat peste linia de frecvență originală. Lungimea (în timp) a unui astfel de semnal este egală cu lungimea ferestrei de sinteză. Astfel, erorile de reconstrucție sunt împrăștiate pe întreaga lungime a ferestrei. Dacă semnalul muzical conține o creștere bruscă a energiei semnalului (precum un atac de castaniete), crește de asemenea și eroarea de cuantizare. După cum s-a spus mai înainte, eroarea de cuantizare (zgomot) este împrăștiată pe toată lungimea ferestrei de sinteză. Dacă atacul apare în interiorul ferestrei de analiză, această eroare va preceda cauza care l-a creat. Dacă acest zgomot se extinde înapoi dincolo de perioada de premascare, el devine audibil și se numește preecou. Există anumite tehnici pentru a preveni preecourile audibile (codarea cu debit binar variabil, creșterea locală a debitului binar pentru a reduce local amplitudinea preecoului), dar acest tip de artefact rămâne unul dintre cele mai greu de evitat.

Dimensiunea tipică a blocurilor utilizate în codoarele audio poate varia între 400 – 2.048 de eșantioane per bloc. Pentru fiecare bloc se realizează o filtrare pe subbenzi, folosind o

transformare (MDCT, MLT, FFT, WT) și se calculează pragul de mascare folosind o transformare FFT cu o rezoluție mărită în frecvență. Motivul pentru care se prelucrează blocuri audio mari se bazează pe faptul că informația laterală necesară ocupă un spațiu mai mic, semnalele cvasistaționare putând fi codate mai eficient. Totuși, dimensiunea blocului influențează direct întârzierea totală a codării, fiind un parametru critic în cazul aplicațiilor în care se dorește o comunicare duplex.

Presupunem că pe durata blocului care se codează apare o componentă tranzitorie în domeniul timp (de ex. sunet de castaniete). Deoarece atacul acestui semnal are o cantitate substanțială de energie, modelul perceptual va alocă numai câțiva biți cuantizatoarelor din subbenzi pentru că un semnal tranzitoriu în domeniul timp se va împrăști în frecvență pe mai multe subbenzi și, în plus, zgomotul de cuantizare poate fi mascat în timpul sau după semnalul tranzitoriu, datorită mascării spectrale și temporale. La decodare, eșantioanele din subbenzi sunt recuantizate și zgomotul de cuantizare admis, care se presupune că va fi mascat complet, se va împrăști în timp pe întreg blocul, precedând astfel apariția semnalului propriu-zis. Zgomotul de cuantizare care precede semnalul tranzitoriu va produce artefacte audibile în domeniul timp, pentru că el “anunță” apariția semnalului tranzitoriu. Deoarece semnalele din subbenzi pot fi cuantizate mai brut, numai pe câțiva biți, eroarea de cuantizare din subbenzi nu poate fi considerată necorelată cu semnalul propriu-zis.

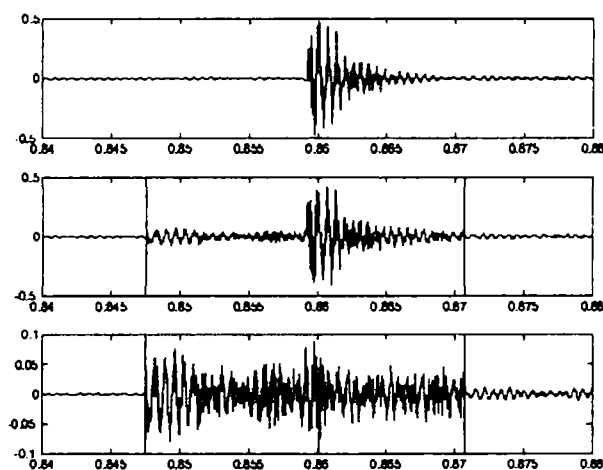


Figura 4.2 Semnalul original – castaniete (sus), după decodare (mijloc) și semnalul diferență (jos)

Dacă semnalul tranzitoriu (castaniete) apare la mijlocul blocului (fig. 4.2), eroarea de cuantizare corelată cu semnalul tranzitoriu va putea fi percepută înaintea atacului semnalului și se numește “preecou”, un artefact foarte asemănător cu “efectul de copiere” de la banda analogică.

În concluzie, eroarea de cuantizare care precede atacul poate produce artefacte audibile, pe când eroarea de cuantizare din timpul atacului va fi mascată de energia semnalului.

După cum s-a prezentat mai sus, în cazul blocurilor scurte informația laterală ocupă mai mult spațiu și nu permite exploatarea avantajelor unor semnale lent variabile. Dacă blocul este foarte scurt, efectul preecoului devine mai mic, iar artefactele pot fi mascate datorită premascării temporale. De aceea codoarele mai avansate folosesc o tehnică numită “comutarea adaptivă a blocurilor” [HER99]. O privire înainte asupra conținutului energetic al blocului următor va permite detectarea semnalelor tranzitorii și va comanda folosirea unor blocuri mai mici. Pentru semnale cvasistaționare se folosesc blocuri lungi (atât cât permite întârzierea maximă). În cazul codoarelor care folosesc transformări suprapuse, împreună cu micșorarea blocului se poate adapta și dimensiunea și forma ferestrei. Rezultatul va fi o schemă de codare mai puțin eficientă, deoarece o mare parte a debitului binar disponibil va fi folosit de informația laterală din blocurile mici.

### 4.2.3 Artefacte de alias

Fenomenul de alias (eroare de subeșantionare) este bine cunoscut din teorema eșantionării, care afirmă că frecvența de eșantionare trebuie să fie egală cu cel puțin dublul benzii de frecvență a semnalului eșantionat.

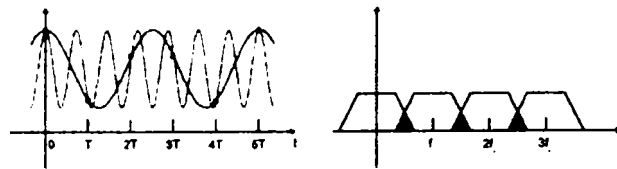


Figura 4.3 Fenomenul de alias în domeniul timp și frecvență

Codarea în subbenzi necesită un set de filtre trece bandă și, deși aceste filtre pot fi cu răspuns finit FIR, banda de trecere și atenuarea sunt limitate. Filtrele PQMF (*Polyphase Quadrature Mirror Filters*) au fost folosite în standardul MPEG-1 [BRA92] și ele pot fi derivate dintr-un filtru prototip, care în versiunea sa modulată formează o bandă compusă din filtre trece bandă egale. Deși aceste filtre se bazează pe anularea erorii de subeșantionare, nu se obține o reconstrucție perfectă. Totuși, în practică, datorită lungimii filtrelor cu 512 pași este puțin probabil ca artefactele să fie audibile în condiții normale.

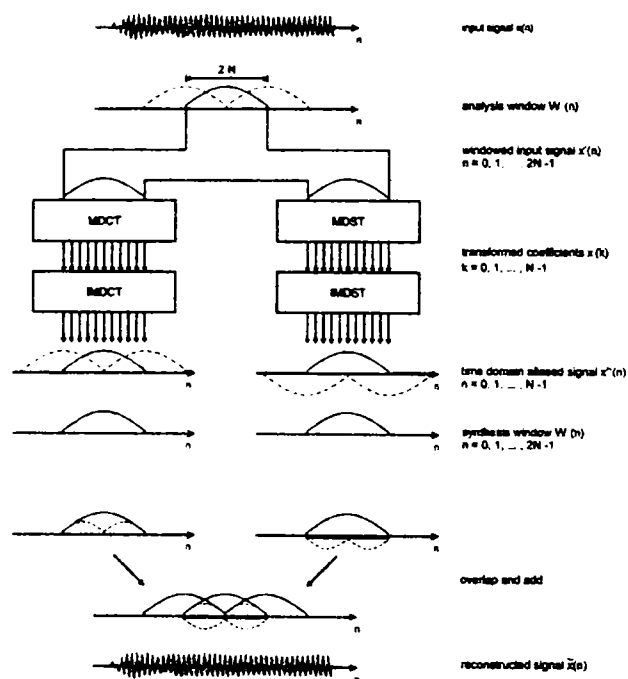


Figura 4.4 Principiul filtrelor TDAC

În figura 4.4 este prezentată o altă metodă, care a fost aleasă pentru filtrele cu anularea erorii de subeșantionare (alias) în domeniul timp (*TDAC – Time Domain Aliasing Cancellation*).

Semnalul în domeniul timp este vizualizat folosind o fereastră cu o lungime de două ori mai mare decât  $N$ , numărul subbenzilor, și cu o suprapunere de 50% între două blocuri succesive. Semnalul din fereastră este transformat prin intermediul unei transformări sinus discrete (DST) și a unei transformări cosinus discrete (DCT) unde, după transformările inverse corespunzătoare, semnalul reconstruit din domeniul timp va conține erori de subeșantionare. Componentele cu erori, care pentru o mai bună înțelegere sunt reprezentate (cu linie întreruptă) separate de semnal, sunt inversate în timp. Folosind o fereastră de sinteză și condiții de suprapunere, acești termeni eronați se vor anula, obținându-se o reconstrucție perfectă. Dar prelucrarea în subbenzi și o

cuantizare insuficientă pot influența procesul de anulare, putând apărea distorsiuni de alias audibile [WAT95].

Malvar [MAL95] a unificat conceptul de anulare a erorilor de subeșantionare din domeniul timp în cadrul transformărilor suprapuse (*LOT – lapped transforms*) și a demonstrat că un număr impar de MDCT poate fi considerat ca o transformare suprapusă modulată (*MLT – Modulated Lapped Transform*). MLT este larg folosită în codarea audio, în special în MPEG-1 și 2 pentru nivelul-3, în Dolby AC-3, în codarea audio avansată (*AAC – Advanced Audio Coding*) și în Windows Media Player.

#### 4.2.4 Artefacte în frecvență și variația lor în timp

Un codec audio perceptual include, prin definiție, un algoritm care modelează sistemul auditiv uman [BAU95], [BAU99]. Majoritatea codoarelor perceptuale estimează un prag de mascare pentru a determina nivelul cel mai mare de zgomot ce poate apărea în fiecare locație de frecvență fără a fi perceput de sistemul auditiv uman. Alte caracteristici ale sunetului, precum tăria și înălțimea pot face parte din model, dar ele sunt legate mai mult de interfața auditivă (urechea externă, medie și internă), fără a fi legate direct de procesele superioare din creier.

Se pare că procesele superioare din creier împart evenimentul sonor într-o colecție de articole independente (obiecte auditive) și le organizează în mai multe grupe (fluxuri) [NOL97]; de exemplu, o piesă muzicală pentru vioară și flaut (care cântă simultan) va conține obiecte audio (fiecare notă interpretată) și două fluxuri, unul pentru notele de vioară și celălalt pentru flaut.

După extragerea informațiilor despre înălțime, tărie și durată, ultima caracteristică ce trebuie interpretată este timbrul. Fluxul auditiv assemblează componentele timp-frecvență într-un singur obiect auditiv, având grijă de relațiile complicate timp-frecvență dintre părțile componente (atomi). De remarcat că componentele izolate pot crea un nou timbru și deci, prin simpla modificare a timbrului se pot crea noi obiecte perceptibile.

Caracteristicile timbrului depind de frecvența și structura temporală, dar relația dintre ele poate fi diferită. Anumite surse sonore își au caracteristicile principale în primii 4 – 5 kHz (vocea) și pot fi recunoscute în această bandă limitată (de ex. telefonul). Bineînțeles că un semnal vocal de calitate are nevoie de o bandă de 7 – 8 kHz, dar există și alte tipuri de sunete care necesită o bandă de frecvențe mult mai largă și care vor dispărea dacă se folosește o bandă limitată. Deoarece asignarea biților variază de la bloc la bloc, coeficienții spectrali pot apărea sau dispărea temporar. Rezultă modificări ale timbrului, variațiile de energie corespunzătoare frecvențelor înalte fiind cunoscute sub numele de “birdies” (păsărele, fig. 4.5).

O soluție pentru a evita acest tip de artefacte poate fi limitarea benzii de frecvență a semnalului înainte de codare. Deși limitarea benzii previne alocarea biților pentru frecvențe înalte, posibilitatea apariției acestui tip de artefacte este încă prezentă.

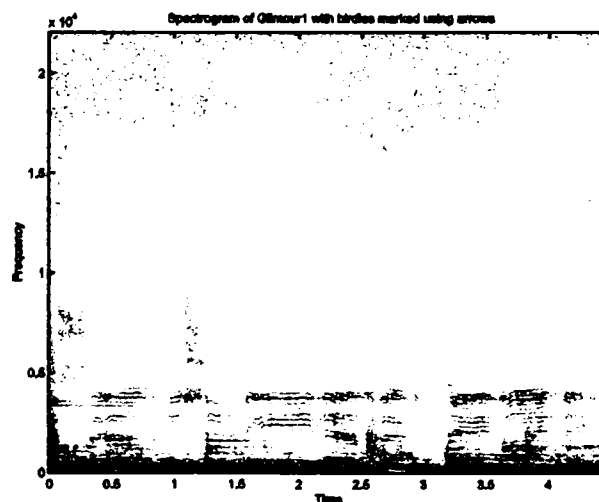


Figura 4.5 Spectrograma unui fragment sonor care indică existența artefactelor “birdies”



#### 4.2.5 Artefacte de reverberație vocală

În special la debite binare reduse și frecvențe joase de eșantionare apare o neconcordanță între rezoluția de timp a codorului și necesitățile structurii de timp a unor semnale. Acest efect este sesizabil la semnale vocale și la ascultarea în căști. Efectul este numit câteodată vorbire dublă pentru că o singură voce sună de parcă ar fi fost înregistrată de două ori și suprapusă.

Pentru a obține o codare cât mai eficientă, în codoarele audio se folosesc transformări de lungime mare sau filtre cu până la 2.048 de benzi.

Pentru majoritatea semnalelor staționare și/sau tonale, o rezoluție mare în frecvență îmbunătățește eficiența codorului datorită spectrului de intrare și de aceea pragul de mascare va avea vârfuri. Un banc de filtre cu mare rezoluție spectrală poate izola aceste vârfuri spectrale, le cuantizează separat, realizând astfel condiția ca zgomotul de cuantizare din vecinătatea unui vârf să rămână sub pragul de mascare. Din păcate principiul lui Heisenberg arată foarte clar că o rezoluție spectrală mare implică o rezoluție mică în timp. Semnalele nestaționare (castaniete) și cele vocale cu fricative și plozive pot necesita pe moment o rezoluție temporală mai mare. Rezultă că aceste semnale au un spectru mai larg, generând o curbă de mascare mai plată și necesitând o rezoluție spectrală mai redusă pentru a fi codate eficient.

Dacă codorul nu este adaptabil în ceea ce privește rezoluția timp/frecvență, atacul semnalelor și semnalele nestaționare vor produce un artefact numit “reverberație vocală”. Acest artefact devine evident pentru semnale vocale codate cu algoritmi ce folosesc transformări lungi (fig. 4.6).

O posibilitate de a evita această problemă este de a micșora temporar rezoluția spectrală a filtrului, de exemplu sub 1.024 de benzi, și de a mări rezoluția în timp (fig. 4.7).

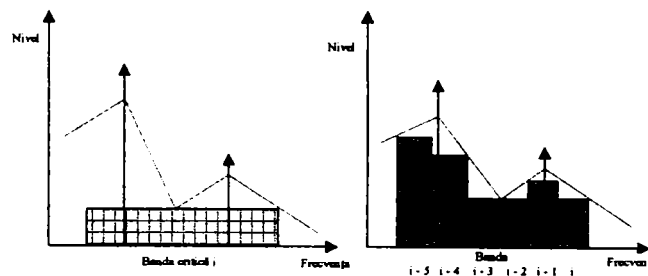


Figura 4.6 Un spectru cu vârfuri necesită o rezoluție bună în frecvență

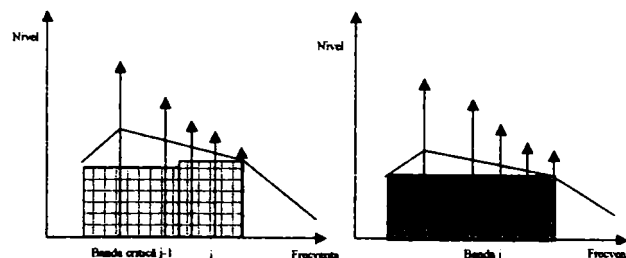


Figura 4.7 Pentru un semnal tranzitoriu cu un spectru continuu se poate folosi o rezoluție în frecvență mai scăzută

#### 4.2.6 Diferența binaurală de nivel de mascare

Diferența binaurală de nivel de mascare (BLMD – *Binaural Masking Level Difference*) este un fenomen care poate fi observat la frecvențe joase, unde mascatorul și proba au relații specifice în timp. Acest fenomen poate fi observat prin următorul experiment:

Un zgomot mascator având frecvența egală cu o bandă critică este prezentat identic ambelor urechi, în timp ce faza mascatului, proba sinusoidală, este modificată alternativ (cele două urechi

primesc semnalul în fază și în antifază). Efectul poate fi remarcat cel mai bine la frecvențe sub 500 Hz, deși el a fost observat chiar și între 2 – 3 kHz. În cazul în care atât mascatorul cât și mascatul sunt prezentate identic ambelor urechi, pragul de mascare este același ca și în cazul ascultării cu o singură ureche. În cazul de mai sus pragul de mascare pentru o probă sinusoidală este de aproximativ 5,5 dB.

În cazul în care proba se prezintă urechilor în antifază, diferența dintre semnalul în fază și în antifază este ușor de sesizat. Experimentul se poate face și cu mascatorul în fază și în antifază, pragul de mascare fiind mai ușor de observat, el crescând la aproximativ 20 dB.

Pentru a sesiza artefactele imaginii sonore, ascultătorul nu trebuie să se mai concentreze asupra artefactelor uzuale, cel puțin temporar, ci să-și construiască o imagine a semnalelor stereofonice și apoi să asculte artefactele și să le poziționeze în imagine. Nu contează tipul, ci numai sesizarea apariției unor “obiecte” sonore noi, care nu erau prezente în referință. Un exemplu poate fi obținut în cazul în care se dorește codarea unui CD comercial, unde canalele ascultate separat sună foarte aproape de original, dar ascultate împreună, semnalul stereofonic, sună prost, înfundat. Cauza acestui efect ar putea fi aplatizarea vârfurilor înfășurătoarei semnalului la frecvențe înalte, având ca efect neglijarea frecvențelor înalte de către sistemul auditiv, numai în cazul redării stereofonice, deoarece înfășurătoarele celor două canale nu sunt corelate nici între ele, nici cu înfășurătoarea frecvențelor joase din semnal.

#### 4.2.7 Pierderea imaginii stereofonice

Pentru codarea semnalelor audio stereofonice (sau multicanal) de înaltă calitate la debite binare reduse, se folosesc tehnici de codare combinate. Pe de o parte ele pun la dispoziție mecanismele care ne permit să ținem cont de efectele psihoacustice binaurale, pe de altă parte debitul binar necesar pentru codarea unui semnal stereofonic se reduce semnificativ față de debitul necesar în cazul codării separate a celor două canale. Tehnicile uzuale de codare stereofonică, ca și cele de înregistrare, sunt procedeul sumă/diferență M/S (*mid/side*) și de intensitate [BER71], [GRI91].

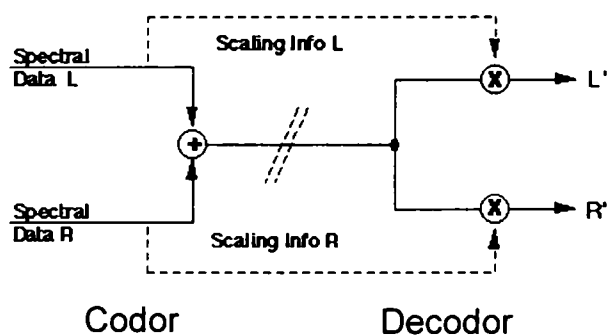


Figura 4.8 Principiul codării stereofonice în intensitate

Stereofonia în intensitate ia în considerare faptul că percepția componentelor de înaltă frecvență a sunetului (peste 4 kHz) se bazează mai mult pe analiza înfășurătoarei energie-timp decât pe forma de undă [HER94]. Astfel se presupune că este suficient să codăm înfășurătoarea unui astfel de semnal decât forma sa de undă. Aceasta se realizează prin transmiterea unui set comun de coeficienți spectrali (“semnal purtător”) pentru mai multe canale audio, în loc să transmitem pentru fiecare canal în parte (fig. 4.8). La decodor, semnalul purtător este scalat independent pentru fiecare canal pentru a reface înfășurătoarea originală (sau energia semnalului). Informațiile de scală sunt calculate și transmise o singură dată pentru fiecare grup de coeficienți spectrali (factor de scală al benzii). Imaginea stereofonică este recreată în decodor printr-o operație de panoramare a fiecărei benzi spectrale.

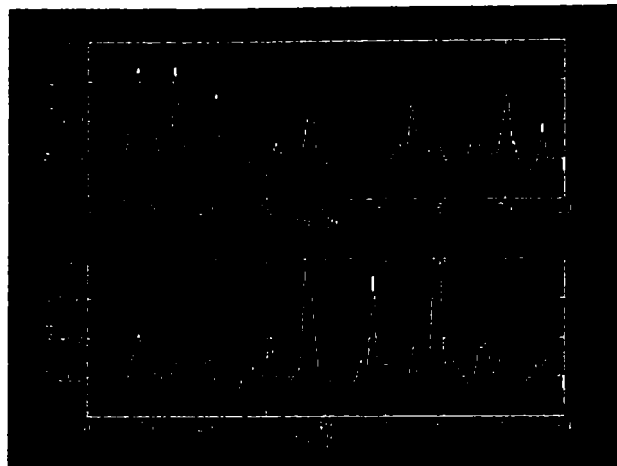


Figura 4.9 Fragment sonor "aplauze" – înfășurătoarea de frecvență a canalelor stâng și drept

Ca o consecință a procesului de codare/decodare stereofonică în intensitate toate semnalele de ieșire sunt versiuni scalate reconstruite după o singură purtătoare, adică au aceeași structură a anvelopei pe durata unui bloc de codare (de ex. 10 – 20 ms). Aceasta nu este o problemă pentru semnale staționare sau care au o structură similară a canalelor stereo codate în intensitate. În cazul semnalelor tranzitorii însă, cu înfășurători diferite în canale diferite, distribuția originală a înfășurătoarei originale între canalele codate nu se mai poate refăce (fig. 4.9, 4.10).

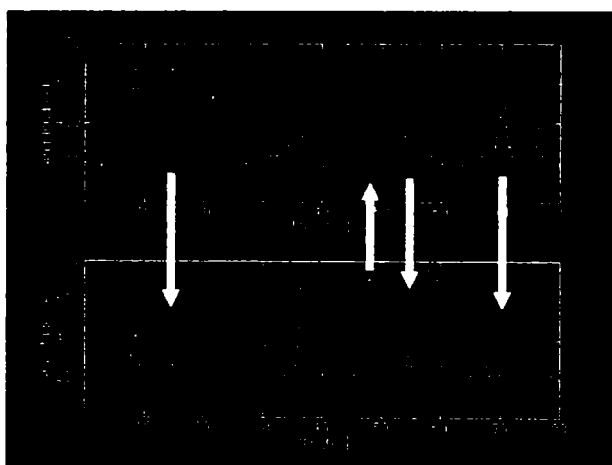


Figura 4.10 Fragment sonor "aplauze" – înfășurătoarea de frecvențe a canalelor stâng și drept după codare/decodare stereo în intensitate

#### 4.2.8 Codarea în tandem

În fiecare codor perceptual standard, reprezentarea spectrală a semnalului de intrare este ușor alterată de procesul de cuantizare. Cu cât debitul binar este mai redus, cu atât cuantizarea este mai grosolană pentru a putea reprezenta semnalul cu debitul disponibil. În acest fel se introduc distorsiuni prin creșterea zgomotului de cuantizare (zgomot de codare), care este format după criteriile modelului perceptual.

Odată cu dezvoltarea metodelor de codare audio cu dbr, compresia datelor audio apare în diferite stadii ale unui lanț de transmisie audio (contribuție, producție, distribuție, emisie etc.).

Deoarece prelucrarea/transmisia materialelor audio se face încă sub formă necompresată sau chiar analogică, avem de-a face cu cicluri repetate de decodare, prelucrare și recodare. La fel și în cazul modificării formatului de codare sau a debitului binar este necesar un ciclu de decodare/recodare.

Modul cel mai eficient de a evita acumularea erorilor este de a păstra semnalul audio cât mai mult posibil într-un format de codare. Se evită astfel noi procese de cuantizare, care introduc zgomot de cuantizare suplimentar. De altfel, în cazul unor operații precum copierea, nu trebuie

părăsit domeniul codat via PCM. În cazul în care un material audio trebuie convertit dintr-un algoritm de codare în altul, transcodarea (conversia în domeniul codat) poate îmbunătăți rezultatul unor codări succesive. Dacă totuși sunt necesare și alte prelucrări ale semnalului (modificarea nivelului, egalizare sau reverberație), întoarcerea în domeniul PCM este inevitabilă. Din păcate, metodele de prelucrare a sunetului în domeniul codat sunt încă destul de rare și în stadiu experimental [MAS00].

Dacă totuși este necesară decodarea/recodarea conținutului audio, este evident că va avea loc o degradare a calității semnalului. De aceea, pentru ca forma finală a semnalului audio codat să îndeplinească anumite condiții de calitate, calitatea fazelor intermediare trebuie să fie mai bună decât a celei finale. Astfel, pierderile de calitate datorate codării în tandem pot fi compensate prin creșterea calității de codare (implicit a debitului binar) în fazele intermediare.

#### 4.2.9 Audio prin codec vocal

Deși din punct de vedere istoric, tehnicile de codare în subbenzi au fost aplicate codării vocale, majoritatea codoarelor vocale se bazează pe modelul tractului vocal uman. Codarea liniară predictivă (LPC – *Linear Predictive Coding*) modelează acest filtru tract vocal utilizând un singur filtru liniar care are numai poli ( $p = 6 - 12$ ), care poate fi excitat de un tren de impulsuri (tonal) sau de un generator de zgomot (fig. 4.11).

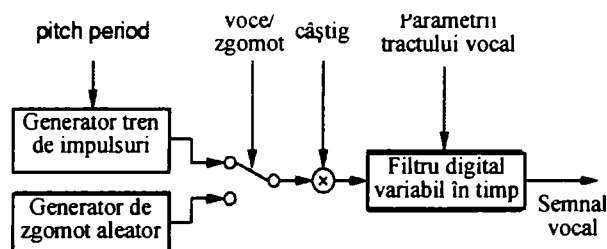


Figura 4.11 Model de sinteză vocală bazat pe codarea liniară predictivă [FEH87]

Este evident că acest sistem nu poate sintetiza simultan componente tonale sau de tip zgomot și deci semnalele audio de bandă largă aplicate acestor codoare vor suna nenatural.

Un alt sistem de codare vocală este CELP (*Code Excited Linear Prediction*) dezvoltat de AT&T Bell Laboratories (§3.2.1), cu care se pot efectua transmisii vocale de calitate acceptabilă cu un debit 2.400 bit/s [MOR95].

În general, codoarele vocale oferă o calitate limitată pentru semnalele audio de bandă largă pentru că:

- modelul lor se bazează pe generarea unui singur ton odată și de aceea nu pot reprezenta precis muzica polifonică
- toți parametri de codare sunt optimizați pentru semnale vocale
- forma zgomotului de cuantizare se bazează pe un criteriu perceptual foarte simplu (eroarea medie pătratică ponderată)
- codorul nu este adaptabil (timp-frecvență, dimensiune bloc etc.)

În general, se poate spune că, exceptând noile concepte bazate pe codarea parametrică folosite în HILN (§3.2.4 B.), MPEG-4 V.2 [EDL00] schemele de codare cu modele simple și folosind tehnici de sinteză și analiză, au rezultate slabe în comparație cu sistemele de codare în subbenzi.

### 4.3 Metode de măsurare a calității audio

Codecurile audio perceptuale sunt, în general, evaluate funcție de debitul binar, complexitate, întârziere, robustețe și calitatea audio. Codecurile perceptuale pot obține o calitate

transparentă pentru un domeniu foarte larg, dar dependent de semnal, al raportului semnal/zgomot (13 – 90 dB). De aceea, metodele clasice de măsurare a calității semnalului, raport semnal/zgomot și distorsiuni armonice, sunt total inadecvate. Pentru a măsura calitatea unui codec audio s-au dezvoltat în ultimii zece ani metode proprii de măsurare, care se pot clasifica în trei grupe: metode subiective (teste de ascultare critică), metode simple de măsurare obiectivă și tehnici de măsurare perceptuale.

#### 4.3.1 Metode subiective de evaluare a calității

##### A. Metodă subiectivă de evaluare a calității audio pentru degradări reduse

Până acum singura metodă disponibilă pentru a compara performanțele diferiților algoritmi de codare și a diferitelor codoare au fost testele de ascultare critică. ITU-R (Uniunea Internațională de Telecomunicații, secțiunea de Radiocomunicații) utilizând date de la anumiți broadcasteri și de la grupul MPEG audio a dezvoltat un set foarte complex de reguli pentru testele de ascultare critică. Scopul acestor teste este de a supune codoarele la condiții extreme, de exemplu de a găsi materialele cel mai greu de codat și de a evidenția performanțele codorului testat pe acest material. În majoritatea cazurilor, artefactele de codare devin audibile sau chiar supărătoare numai după o lungă perioadă de antrenament. Odată cu utilizarea echipamentelor bazate pe tehnologia de compresie audio, oricine poate deveni un ascultător expert în urma unor antrenamente intensive. Procedura de test bazată pe Recomandarea ITU-R BS.1116-1 necesită un număr relativ mare de subiecți, care să furnizeze o estimare demnă de încredere a calității audio. Principiul de bază al acestei metode subiective de evaluare a calității audio este “dublu orb cu stimul triplu și referință ascunsă”: ascultătorul poate alege între trei surse (“A”, “B” și “C”). Semnalul de referință cunoscut este întotdeauna pe sursa “A”. Semnalul de referință ascuns și semnalul testat sunt disponibile simultan, dar atribuite aleatoriu surselor “B” și “C”. Ascultătorul trebuie să evalueze diferențele dintre “B” și “A” și dintre “C” și “A” conform scării de diferențe subiective cu 5 trepte (tabelul 4.1). Diferența dintre una din sursele “B” sau “C” trebuie să fie imperceptibilă față de sursa “A”, iar cealaltă să prezinte anumite degradări ale calității audio.

Tabel 4.1 Scala ITU-R cu 5 trepte

5.0	Imperceptibil	0.0
4.9	Perceptibil, dar NEderanjant	-0.1
4.8		-0.2
4.7		-0.3
4.6		-0.4
4.5		-0.5
4.4		-0.6
4.3		-0.7
4.2		-0.8
4.1		-0.9
4.0		-1.0
3.9	Ușor deranjant	-1.1
3.8		-1.2
3.7		-1.3
3.6		-1.4
3.5		-1.5
3.4		-1.6
3.3		-1.7
3.2		-1.8
3.1		-1.9
3.0		-2.0
2.9	Deranjant	-2.1
2.8		-2.2
2.7		-2.3
2.6		-2.4
2.5		-2.5
2.4		-2.6
2.3		-2.7
2.2		-2.8
2.1		-2.9
2.0		-3.0
1.9	Foarte deranjant	-3.1
1.8		-3.2
1.7		-3.3
1.6		-3.4
1.5		-3.5
1.4		-3.6
1.3		-3.7
1.2		-3.8
1.1		-3.9
1.0		-4.0

Orice diferență percepută între referință și altă sursă trebuie interpretată ca o degradare. Pentru evaluarea calității se folosește un singur atribut, “Calitatea audio de bază” (*Basic Audio Quality* – BAQ). El se definește ca un atribut global care include toate diferențele detectate între semnalul de referință și semnalul testat.

Scara de notare cu 5 trepte folosită în procesul de evaluare a deteriorărilor audio se consideră continuă, cu cele cinci puncte de reper, și este conformă cu Recomandarea BS.562-3 a ITU-R.

Subiecții trebuie să noteze fragmentul în intervalul de 10 sec dintre stimuli, având la dispoziție un formular clar pentru notare. Se recomandă ca notarea să se facă cu o precizie de o zecimală.

Analiza rezultatelor testelor de ascultare critică se bazează în general pe gradul de diferență subiectivă SDG (*Subjective Difference Grade*), definit astfel:

$$SDG = Nota_{Semnal\ testat} - Nota_{Semnal\ referință} \quad (4.1)$$

În mod normal, valorile SDG trebuie să fie cuprinse între 0 și -4, unde 0 corespunde unor degradări imperceptibile și -4 unor degradări considerate foarte deranjante.

Pentru ca rezultatele unui astfel de test să fie relevante, se recomandă ca subiecții să fie persoane cu experiență în domeniul audio (acusticieni, proiectanți și constructori de instrumente, acordori de pian, experți în producția de muzică electronică, ingineri și tehnicieni de sunet, muzicieni etc.).

Conform recomandărilor ITU este necesar ca subiecții să poată identifica corect deteriorările produse sunetului de către codec. Distorsiunile suplimentare introduse de codarea perceptuală se deosebesc semnificativ de cele ale unui sistem analogic de redare a sunetului.

Procesul de instruire a subiecților are loc la începutul fiecărei sesiuni de evaluare:

- prezentarea experimentului - explicarea obiectivelor și a sistemelor examinate;
- familiarizarea cu materialul sonor folosit în timpul experimentului;
- descrierea atributului urmărit (calitatea audio - orice diferență între stimuli audio);
- explicarea verbală a categoriilor de deteriorări;
- descrierea metodei de testare;
- descrierea fazei de notare (scală cu 5 trepte de deteriorare);
- demonstrarea materialelor cele mai afectate;
- alegerea nivelului de volum cel mai confortabil.

**Tabelul 4.2** Categoriile de deteriorări sonore

Categoriile de artefacte	Descriere
Defecte de cuantizare	defecte asociate cu o rezoluție insuficientă, de ex. distorsiuni granulare sau modificarea nivelului de zgomot
Pierderi de frecvențe înalte	lipsa frecvențelor înalte
Exces de frecvențe înalte	exces de frecvențe înalte sau efecte asociate (sibilante sau făștit)
Efecte de modulație periodică	variații periodice - tremurat, pompat, agitat
Efecte de modulație neperiodică	efecte asociate cu tranzitorii - neclar, exploziv
Modificarea nivelului	modificarea nivelului sursei (de ex. aplauze)
Distorsiuni temporale	ecouri pre- și post-, sunet mănjit, neclar
Distorsiuni	armonice sau nearmonice
Sunete suplimentare	sunete parazite, fără legătură cu materialul
Efecte de corelare	intermodulație între canale
Calitatea imaginii	toate aspectele incluzând lărgimea imaginii, mișcare, stabilitate și efecte de fază

În tabelul 4.2 sunt enumerate câteva modificări nedorite ale semnalului sau *artefacte*, care pot apărea în urma codării perceptuale, pentru a servi subiecților ca referință în timpul instruirii și experimentului.

În ciuda metodologiei riguroase impuse de standardul ITU-R BS.1116, rezultatele testelor subiective sunt influențate de diferiți factori: context, locul de desfășurare și particularități individuale ale ascultătorului: fizice (acuitate) sau cognitive (preferințe). Nivelul de ascultare (SPL) și zgomotul înconjurător pot influența excitația și introduce efecte de mascare nedorite. Mai mult, modul de prezentare poate influența puternic calitatea percepută, deoarece atât difuzoarele, cât și încăperea de ascultare pot introduce propriile distorsiuni, creând o dependență a rezultatelor de locul de desfășurare a testului. Influența locului de desfășurare și dependența de subiecți a testelor de ascultare subiective, pot invalida compararea directă a rezultatelor obținute în teste independente pentru diferiți algoritmi. De aceea, compararea diferitelor codec-uri se recomandă să se facă într-un singur loc și cu aceeași subiecți.

Semnalele de test pot fi clasificate ca naturale și sintetice. Semnalele naturale de test sunt secvențe audio critice folosite în testele anterioare de evaluare a calității audio. Durata unui semnal de test natural trebuie să fie aceeași ca în testele de ascultare critică, adică de aproximativ 10 – 20 s. Semnalele trebuie să fie disponibile atât la locul de transmitere cât și la locul unde se efectuează măsurarea. Deci dispozitivul de măsurare trebuie să fie cu memorie. Este de dorit ca partea critică a semnalului de test, care produce majoritatea artefactelor, să fie limitată numai la un scurt fragment.

Semnalele sintetice sunt definite matematic și pot fi modificate după dorință. Aceste semnale pot fi generate independent la locul de transmitere și cel de măsurare, astfel că nu mai este necesară o memorie a dispozitivului de măsurare. Durata semnalelor sintetice de test trebuie să fie suficient de mare pentru a stresa codec-ul testat, care poate conține un buffer pentru semnalul audio codat. Luând în considerare lungimile acestor buffere și constantele de timp prezente în metoda de măsurare, durata fiecărui semnal de test poate fi mai mare de 500 ms. Datorită naturii acestor semnale, evaluarea lor subiectivă este foarte dificilă sau chiar imposibilă și de aceea pot fi limitate la așa durate scurte. De aceea această metodă de măsurare nu a fost validată prin evaluarea subiectivă a semnalelor sintetice.

### *B. Metodă subiectivă de evaluare a calității audio pentru transmisia prin Internet*

Unul din factorii principali care concură la răspândirea și acceptarea radioului prin Internet este calitatea audio a informației recepționate. Instrumentele de evaluare a calității audio a schemelor de compresie folosite pentru Internet Radio sunt testele de ascultare critică. Compararea diferitelor sisteme de codare este necesară din două motive:

- Producătorii sistemelor de codare pentru Internet Radio folosesc materiale demonstrative diferite pentru sisteme de codare diferite. Fiecare folosește propriile materiale de test alese astfel încât să scoată în evidență calitățile propriului sistem.

- La majoritatea expozițiilor profesionale, demonstrațiile sunt făcute cu materiale în limba engleză și nu se poate ști dacă în cazul altor limbi europene, pentru o anumită schemă de compresie, se păstrează calitatea audio și inteligibilitatea. De asemenea, nu se dă un răspuns la întrebarea dacă există diferențe de calitate între codarea unei voci feminine sau masculine sau între calitatea unui material vocal sau muzical sau codarea unei voci singure în comparație cu o voce cu fundal muzical, care reprezintă modul cel mai uzual de transmitere a informației astăzi [STO97].

În acest context, definirea unei metodologii de testare este destul de dificilă din următoarele motive:

1. Calitatea variază mult în timp datorită debitelor binare reduse și erorilor de transmisie. Din acest motiv, caracteristicile fragmentelor testate sunt foarte importante;

2. Nivelul de calitate este mai scăzut decât cel luat în considerare de metodele standardizate de testare;

3. Având în vedere că scopul investigației este aflarea gradului de satisfacție a utilizatorului, metodologia de testare nu trebuie să influențeze subiecții sau să-i facă prea critici.

În acest scop, grupul EBU B/AIM a propus o nouă metodă de testare numită MUSHRA (*MULTI Stimulus test with HIDDEN Reference and Anchors*), adică metodă de testare cu stimuli multipli, referințe ascunse și ancore [STO00]. Această metodă a fost propusă pentru a furniza rezultate utile și repetabile privind calitatea audio a semnalelor de calitate medie. Parametrul evaluat este “calitatea audio de bază”.

Spre deosebire de recomandarea BS.1116-1, care folosește principiul de testare “dublu orb cu stimul triplu și referință ascunsă”, în cazul metodei MUSHRA se folosește un principiu modificat “dublu orb, stimul multiplu, referințe ascunse și ancore ascunse”. Deoarece metoda MUSHRA a fost creată pentru a evalua modificări medii și mari ale semnalului audio, semnalele de referință nu trebuie să fie de înaltă calitate, critice, iar distanța perceptuală dintre semnalele de referință și cele de test va fi relativ mare. Pe de altă parte, distanța perceptuală dintre semnalele testate aparținând diverselor sisteme de codare, va fi relativ mică. De aceea, dacă fiecare sistem este comparat numai cu referința, diferențele dintre două sisteme vor fi prea mici pentru a fi relevante. În consecință, MUSHRA nu folosește numai o referință de calitate, ci și o comparație directă între două sisteme diferite. Ascultătorul poate comuta după dorință între semnalul de referință și oricare din sistemele testate. Deoarece ascultătorul poate compara direct semnalele codate, poate detecta mai ușor diferențele dintre semnale și le poate nota corespunzător. Se obține astfel o partajare mai precisă a sistemelor testate. De remarcat că notele date de ascultători sunt obținute atât prin compararea unui sistem de codare cu semnalul de referință, cât și cu celelalte semnale.

La desfășurarea testului se folosește un sistem de redare pe calculator. Într-o sesiune, ascultătorului i se prezintă o succesiune de fragmente audio, formată din semnalul de referință și semnalele prelucrate de codoarele testate. De exemplu, dacă sunt testate patru sisteme audio, ascultătorul are la dispoziție opt semnale pentru a le nota (o referință “cunoscută” + patru semnale codate + o referință “ascunsă” + 2 ancore “ascunse”). Ancorele sunt variante filtrate (3,5 kHz, 7 kHz) ale semnalului de referință. Ascultătorii încep sesiunea cu o trecere în revistă a semnalelor și o estimare primară a calității, urmată de un proces de clasificare și în final de acordare a notelor (tabelul 4.1).

Testele se concentrează numai asupra calității audio a sistemelor de compresie în condiții optime de transmisie. Nu se iau în considerare erorile de transmisie sau pierderile de pachete și se consideră un debit binar constant. Pentru a putea face o comparație echitabilă a sistemelor audio de Internet Radio se va face abstracție de variația în timp a caracteristicilor rețelei.

Pentru a scoate în evidență diferențele dintre sistemele de codare testate, materialele audio trebuie să fie suficient de critice pentru fiecare sistem testat în parte. Selectarea materialelor audio de test este o operație dificilă, dar foarte importantă, pentru că de ea depinde obiectivitatea evaluărilor obținute. Se recomandă utilizarea unor semnale “naturale”, de gen program radio. Nu se recomandă utilizarea unor semnale sintetice, prea critice, create special pentru a pune în dificultate sistemele de codare.

De asemenea, trebuie să se acorde atenție conținutului artistic sau intelectual al secvenței de program, care nu trebuie să fie nici prea atractiv, nici prea dezagreabil, pentru a nu distrage atenția ascultătorilor de la detectarea diferențelor. Alegerea trebuie să reflecte probabilitatea de difuzare a fiecărui tip de program în emisiunile posturilor de radio. Această condiție este destul de greu de îndeplinit, pentru că natura materialelor difuzate diferă de la o stație la alta, iar stilurile și preferințele muzicale evoluează în timp.

Numărul materialelor incluse într-un test poate varia, dar nu trebuie să fie prea mare, pentru a nu mări exagerat durata testului. Se recomandă ca numărul secvențelor de test să fie de 1,5 ori mai mare ca numărul sistemelor de codare testate, dar nu mai puțin de 5 secvențe pe sistem. Secvențele audio trebuie să aibă o durată cuprinsă între 10 s și 20 s. Toate sistemele de codare trebuie testate cu aceleași secvențe audio [STO00].



### 4.3.2 Tehnici simple de măsurare obiectivă

De nenumărate ori specialiștii au încercat să măsoare calitatea codorului prin mărimi ca raportul semnal/zgomot sau banda de frecvență a semnalului decodat. Deoarece modelul de bază al codorului perceptual se bazează pe îmbunătățirea calității subiective prin modelarea zgomotului de cuantizare în funcție de frecvență (și timp), conducând la obținerea unui raport semnal/zgomot mai redus decât cel obținut fără modelarea zgomotului, aceste măsurători sfidează scopurile codării perceptuale. Autorii care mai folosesc aceste măsurări nu fac decât să demonstreze că nu au înțeles ce fac. După cum se va arăta mai jos, a te baza pe banda de frecvență a semnalului codat nu reprezintă o bună înțelegere a subiectului.

O altă abordare o reprezintă studierea codecului din punct de vedere a rezultatelor obținute pentru diferite semnale de intrare (tranzitorii, semnale multi-tonale). Deși rezultatele acestor teste pot furniza expertului o mulțime de informații despre codec, este foarte periculos să te bazezi numai pe rezultatele acestor teste.

### 4.3.3 Tehnici de măsurare perceptuală

De 15 ani încoace au existat o mulțime de cercetări pentru a aplica modelarea psihoacustică la predicția calității unui codec și a audibilității anumitor artefacte. Deși tehnologia actuală nu este încă suficientă pentru a înlătura testele de ascultare critică, totuși tehnicile de măsurare perceptuale au progresat până în măsura în care pot completa testele de ascultare și în unele cazuri chiar le pot înlocui. Grupul de acțiune 10/4 din cadrul ITU-R lucrează de mai mulți ani la standardizarea tehnicilor de măsurare perceptuale și a realizat un sistem numit PEAQ (*Perceptual Evaluation of Audio Quality* - evaluarea perceptuală a calității audio). Recomandarea definește un sistem multimod bazat pe colaborarea cu principalele laboratoare care lucrează asupra tehnicilor de măsură perceptuale.

Prin definiție, calitatea sunetului unui codor perceptual poate fi evaluată numai de auzul uman. În consecință, un codor perceptual util poate fi proiectat numai pe baza unei cunoașteri aprofundate a mecanismului auditiv uman. Acuitatea urechii umane este surprinzătoare. Ea poate detecta distorsiuni foarte mici și suporta o gamă dinamică enormă (96 dB). Dacă urechea detectează un grad diferit de deteriorare a sunetului la două codecuri cu același debit binar cu ocazia unor teste de ascultare critică, putem spune care dintre ele este mai bun.

Sensibilitatea urechii la distorsiuni merită probabil mai multă atenție decât fidelitatea gamei dinamice sau a răspunsului în frecvență. Toate metodele de codare perceptuală se bazează pe înțelegerea fenomenului de mascare, deoarece el determină audibilitatea artefactelor. Membrana basilară a urechii joacă rolul unui analizor de spectru.

#### A. Concepte de măsurare perceptuale

În domeniul tehnicilor de măsură perceptuale se folosesc două concepte principale pentru estimarea distorsiunilor audibile: conceptul pragului de mascare și compararea reprezentărilor interne. În plus, anumite efecte sunt mai ușor de modelat folosind un spectru liniar decât reprezentarea membranei basilare. Acesta poate fi considerat al treilea concept și este denumit analiza spectrală a erorilor.

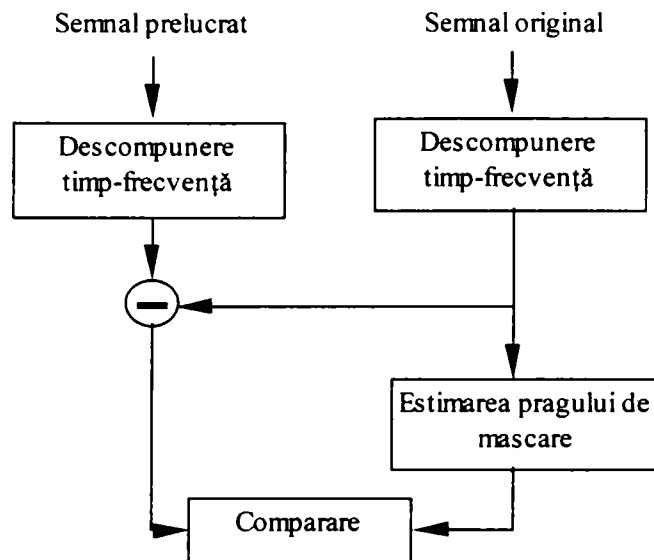


Figura 4.12 Conceptul pragului de mascare

### 1. Distanța de prag de mascare

Conceptul pragului de mascare (numit și evaluarea zgomotului) a fost utilizat în primele metode de măsurare perceptuale. În acest caz (fig. 4.12), semnalul de eroare, care este diferența dintre semnalul original și cel prelucrat, se compară cu pragul de mascare al semnalului original. La un anumit timp și frecvență, o eroare se consideră neaudibilă dacă are o amplitudine mai mică decât pragul de mascare predefinit. Avantajul principal al acestei metode este că parametrii modelului pot fi direct verificați folosind experimente de mascare.

### 2. Compararea reprezentărilor interne

Conceptul comparării reprezentărilor interne (sau compararea în domeniul cohlear), introdus în 1985 de Karjalainen, stă la baza majorității metodelor de măsură perceptuale actuale ([BEE92], [COL94]). El presupune modelarea semnalului de excitație a membranei basilare prin simularea transformărilor la care este supus semnalul în ureche. Măsurarea calității se face prin compararea formei de excitație a originalului și a semnalului prelucrat (fig. 4.13). Această metodă este mai apropiată de funcția fiziologică a sistemului auditiv decât conceptul de prag de mascare. De aceea ea reprezintă un punct de plecare mai bun pentru modelarea unor fenomene auditive mai complexe.

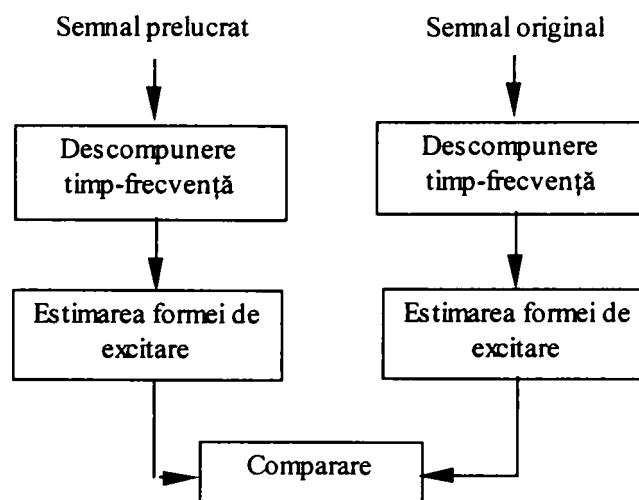


Figura 4.13 Compararea reprezentărilor interne

### 3. Analiza spectrală a erorilor

Anumite efecte, precum percepția frecvenței fundamentale și structura de armonici asociate, se modelează mai ușor folosind un spectru liniar, decât modelele de excitație ale membranei basilară. Evaluarea calității perceptuale obiective se îmbunătățește dacă modelele utilizează variabile bazate pe spectre liniare împreună cu celelalte două concepte.

### 4. Modelul urechii

După cum rezultă din fig. 4.1, modelul perceptual se împarte în două: modelul urechii periferice și un model pentru stadiile superioare de prelucrare. În acest context modelul urechii periferice conține toate etapele de prelucrare care transformă sunetul de intrare într-o reprezentare a membranei basilară.

### 5. Pragul absolut

Pragul absolut de audibilitate este cel mai redus în zona 2.000 - 3.000 Hz și crește odată cu creșterea, respectiv scăderea frecvenței. Această caracteristică se poate modela folosind funcția de transfer a urechii externe și medii combinată cu principiul zgomotului intern. Funcția de transfer a urechii externe și medii limitează banda de frecvență a semnalelor audio. Zgomotul intern este cauzat de circulația sângelui în cap și de activitatea nervoasă spontană.

### 6. Scala perceptuală de frecvență

Undele sonore sunt transformate în semnale electrice (neuronale) în cohlee conform unei transformări frecvență-loc. În funcție de frecvența semnalului de intrare, anumite porțiuni ale membranei basilară au o deplasare maximă. Celulele de păr, care sunt receptorii ce măsoară aceste deplasări, sunt egal distribuite pe întreaga membrană basilară. Fiecare celulă de păr reacționează la o regiune de frecvențe învecinate. Transformarea neliniară frecvență-loc împreună cu distribuția liniară a celulelor de păr de-a lungul membranei basilară produce percepția neliniară a frecvenței, așa numita înălțime. În funcție de experimentul psihoacustic, se pot găsi diferite funcții de transfer frecvență-înălțime. În cazul codării audio perceptuale și a măsurătorilor audio perceptuale se folosește de obicei scara Bark (§2.2.4.A). Această scară, care folosește conceptul de raport critic, a fost definită de Zwicker și Feldtkeller. Scara Bark împarte domeniul de frecvență 20 Hz – 15 kHz în 24 de secțiuni diferite (care nu se intersectează).

### 7. Excitarea

Celulele de păr din cohlea generează activitatea neuronală ca răspuns la vibrația membranei basilară indusă de unda sonoră. Fiecare celulă de păr reacționează la o anumită gamă de frecvențe ce poate fi descrisă prin caracteristica unui filtru. Panta acestor filtre se poate evidenția cel mai bine pe o scară perceptuală, ca cea descrisă mai sus. Forma filtrelor pe o asemenea scară este aproape independentă de frecvența centrală (fig. 4.14). Panta ascendentă a excitației este independentă de nivelul  $L$  al semnalului de intrare (aproximativ 27 dB/Bark). Panta descendentă, dependentă de energie și frecvență, este dată de aproximația lui Terhardt (4.2) și este mai accentuată pentru niveluri mici ale semnalului de intrare decât pentru niveluri mari (-5 până la -30 dB/Bark) [THI98]. Această caracteristică abruptă se datorează unui mecanism de reacție între două tipuri diferite de celule de păr și necesită un anumit timp pentru a se stabiliza. De aceea rezoluția în frecvență auditivă cea mai bună se obține pentru semnale staționare, la câteva milisecunde după stabilirea semnalului. Pentru semnale constând din mai multe componente, caracteristica de excitație se obține printr-o însumare neliniară.

$$\frac{S}{\text{dB/Bark}} = -24 - \frac{230\text{Hz}}{f_c} + 0,2 \cdot L / \text{dB} \quad (4.2)$$

unde  $L$  este energia locală și  $f_c$  este frecvența centrală a benzii de frecvență analizate.

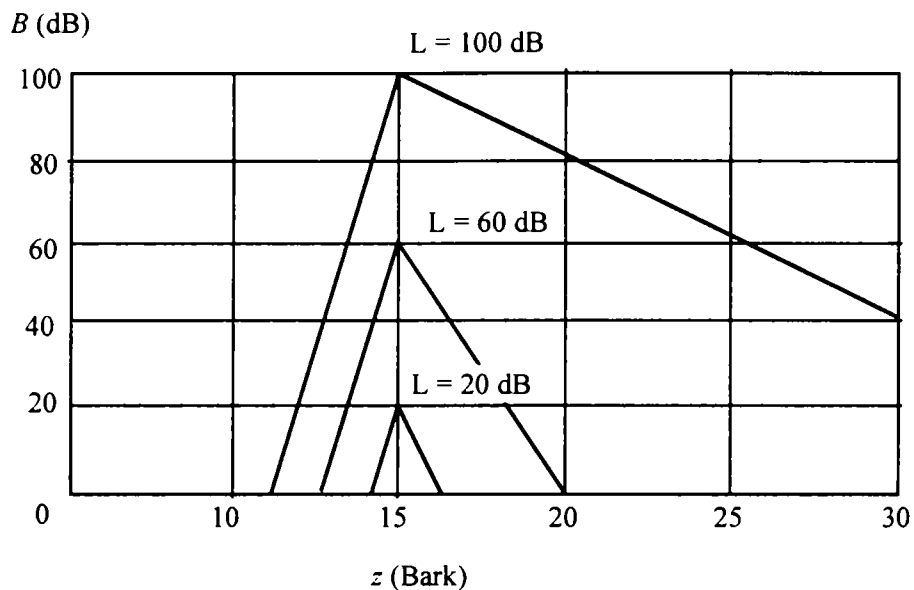


Figura 4.14 Dependența de nivel a excitației, după Terhardt (1979)

După expunerea la un semnal, celulele de păr și procesul neuronal au nevoie de un anumit timp pentru a reveni la sensibilitatea nominală. Durata procesului de revenire depinde de nivelul și durata semnalului și poate dura până la câteva milisecunde. Semnalele de nivel mare sunt prelucrate mai rapid decât cele de nivel mic pe traseul celule de păr – creier. De aceea, atacul unui semnal puternic poate masca un semnal precedent mai slab.

Un alt mod de abordare al modelului excitației se bazează pe scara ERB (§2.2.4.B, Moore, 1986), care folosește așa numitele filtre ROEX.

În contextul măsurării obiective a calității audio percepute, cele mai bune rezultate s-au obținut cu modelele bazate pe relațiile date de (Zwicker și Feldtkeller, 1967) și (Terhardt, 1979).

### 8. Discriminarea între semnale

Modelele de excitație ale semnalelor audio sunt prelucrate și stocate în creier. Se pot deosebi trei tipuri de memorie, care diferă prin gradul de detalii și durata de reținere a informației: memorie echoică, memorie de termen scurt și memorie de termen lung. Memoria echoică reține inițial informația audio până când trece în memoria de scurtă durată, unde este menținută printr-un proces de repetiție. Testele de ascultare critice care evaluează calitatea audio, definite în Recomandarea ITU-R BS.1116, depind de aceste două tipuri de memorie pentru a detecta micile diferențe dintre semnale. De aceea în testele de ascultare critică se pot face mai multe treceri între semnalul de referință și semnalul testat.

Deoarece măsurările pragului de audibilitate formează o distribuție cu o varianță măsurabilă, pragul poate fi definit ca nivelul semnalului cu probabilitatea 0,5 de a fi detectat. În jurul pragului, probabilitatea de detecție a diferențelor crește treptat de la 0 la 1.

Detecția pragului pentru modificări “abia sesizabile” ale nivelului semnalului (JNLD – *Just-Noticeable Level Difference*) depinde de presiunea sonoră a semnalului de intrare. De exemplu, la semnale de nivel mic diferențele de nivel trebuie să fie mai mari pentru a putea fi detectate (nivel: 20 dBSPL, JNLD: 0,75 dB). Pentru semnale de nivel mare sensibilitatea la diferențe mici este mai mare (nivel: 80 dBSPL, JNLD: 0,2 dB).

### 9. Mascarea

Un semnal care se poate auzi clar dacă este singur, poate fi complet mascat de un alt semnal simultan mai puternic. Acest fenomen psihoacustic stă la baza metodelor moderne de reducere a debitului binar și a fost prezentat pe larg în §2.2.5.

### 10. Tărie și mascare parțială

Tăria percepută a unui semnal audio nu depinde numai de nivelul de presiune acustică ci și de durata și structura sa temporală și spectrală. Tăria parțială a unui semnal audio este tăria percepută după ce a fost redusă de un mascator. Datorită auto-mascării tăria unui semnal complex este mai mică decât suma tăriei componentelor sale. Acest lucru este important în măsurătorile perceptuale, pentru că tăria parțială ține cont de scăderea tăriei percepute a unei distorsiuni audibile datorate unui mascator.

### 11. Claritate

Claritatea, una din valorile de bază ale senzației, este legată de timbru. Un sunet este perceput ca fiind clar dacă conține în special componente de înaltă frecvență. De exemplu, un sunet sinusoidal sau un zgomot care conține frecvențe înalte (peste 3 kHz) se spune că este precis. Totuși, structura detaliată în frecvență a semnalului audio nu are o mare influență asupra preciziei. Cercetările fundamentale privind claritatea sunetului au fost realizate de Von Bismarck în 1974. Alte cercetări au fost efectuate în 1984 de către W. Aures la Universitatea Tehnică din München și au condus la o funcție de ponderare ușor modificată față de cea definită de Bismarck. Ea are o contribuție mai mică la atribuirea clarității la benzile critice joase și foarte înalte și mai mare pentru benzile cuprinse între 14 și 20 Bark [THI98]. În plus, cercetările au arătat că claritatea semnalelor audio cu o mare variație a nivelului presiunii sonore și un conținut bogat de frecvențe înalte nu se poate baza numai pe tăria totală, ci pe o funcție de ponderare, care depinde de tăria totală.

### 12. Modelul cognitiv

Reprezentarea perceptuală furnizată de ureche se suprapune pe o reprezentare cognitivă existentă, greu de specificat. Totuși, se pot face anumite presupuneri rezonabile despre procesul cognitiv care presupune o evaluare a calității. De exemplu, dacă ascultătorii nu sunt obișnuiți cu semnalul prelucrat, opinia lor va fi influențată de cunoștințele lor generale. Deși mai multe persoane pot avea păreri diferite despre cum trebuie să sune un pian bun, există totuși caracteristici comune ale sunetului unui pian. De exemplu, dacă forma de undă este limitată în amplitudine, toți subiecții vor fi de acord asupra calității audio reduse.

Deoarece programele de calculator au puține cunoștințe generale, evaluările bazate pe calculator, care folosesc un model al sistemului auditiv uman, necesită semnalul original ca referință. Totuși, problema cunoștințelor generale insuficiente este numai parțial rezolvată prin utilizarea unei referințe. De exemplu, dacă semnalul original are caracteristicile unui zgomot, iar prelucrarea produce asupra semnalului de ieșire un efect estetic mai plăcut, cotarea calității poate fi superioară decât dacă ascultătorul ar fi comparat strict intrare cu ieșirea. Acest aspect al comportamentului ascultătorului este foarte greu de modelat fără a cunoaște reprezentarea ideală a semnalului audio pe care ascultătorul o are în minte.

Un alt fenomen, legat oarecum de idealul intern este organizarea perceptuală involuntară a intrării auditive. Dacă un semnal audio ce formează o entitate coerentă este ascultat fără anumite componente de timp și frecvență, semnalul rezultat va fi totuși o entitate coerentă. Dacă însă, semnalului i se adaugă o componentă timp-frecvență străină, rezultatul va fi sesizat ca două semnale diferite și va fi judecat supărător. Experimentele psihoacustice au arătat că dacă unui semnal audio i se adaugă o componentă timp-frecvență ce nu are nici o legătură cu structura sa de timp sau frecvență, imaginea auditivă se descompune în două evenimente sau fluxuri perceptuale diferite. Totuși, pentru a măsura dacă un codec produce noi componente timp-frecvență ușor deranjante sau foarte deranjante este necesar un analizor al *scenei auditive* care să descrie cum separă ascultătorii evenimentele auditive și le grupează în diferite obiecte. Un astfel de model al analizei scenei auditive nu face însă obiectul prezentei lucrări. În recomandarea ITU-R BS.1387-1 acest efect este modelat într-un mod simplificat prin tratarea asimetrică a perturbațiilor produse

prin adăugarea sau ștergerea unor elemente. Separarea distorsiunilor liniare și neliniare în fluxuri separate ușurează încorporarea acestui efect.

Învățarea este un alt fenomen cognitiv care joacă un rol important în evaluarea subiectivă a calității audio. O mică distorsiune necunoscută este mai greu de sesizat decât o mică distorsiune cunoscută. Acest efect se numește *mascare informațională*, când în urma unui antrenament adecvat pragul de sesizare a unui semnal complex mascat de un alt mascator complex poate scădea cu 40 dB.

Natura unei distorsiuni în raport cu liniaritatea influențează de asemenea percepția. În particular, distorsiunile liniare sunt mai puțin sesizabile decât cele neliniare. Separarea distorsiunilor neliniare de cele liniare se face printr-o filtrare adaptivă inversă a semnalului de ieșire. Metoda specificată în recomandarea ITU-R BS.1387-1 și descrisă în cap. 6, folosește o aproximație medie de scurtă durată a distorsiunii liniare pentru a face față codecurilor cu un răspuns în frecvență variabil în timp.

## B. Metode de măsurare perceptuală

### 1. DIX (*Disturbance Index*)

Metoda de măsurare perceptuală numită “Indexul perturbării” (DIX, Thiede și Kabot, 1996) se bazează pe un grup de filtre audio cu o rezoluție temporală mare, care permite o modelare mai precisă a efectelor temporale precum pre- și post-mascarea.

Frecvențele centrale ale filtrelor individuale sunt egal distribuite pe o scară perceptuală a înălțimii. Gama de frecvențe audio a fost acoperită mai întâi cu 80 de filtre în prima versiune DIX și redusă apoi la 40 de filtre, rezoluția în frecvență fiind de aprox. 0,6 Bark. Algoritmul filtrului este destul de rapid în comparație cu alte grupuri de filtre cu filtre individuale, dar totuși mult mai lent decât transformările FFT sau wavelet.

DIX adaptează dinamic nivelurile și spectrul dintre semnalul testat și de referință pentru a separa distorsiunile liniare de cele neliniare.

Prin compararea reprezentărilor interne ale semnalului testat și ale referinței se calculează mai mulți parametri, ca de exemplu tăria parțială a distorsiunilor neliniare, indicatori pentru cantitatea de distorsiuni liniare și măsurarea efectelor temporale și binaurale. Totuși, estimarea calității audio de bază se face utilizând numai doi parametri: tăria parțială a distorsiunilor neliniare și unul din indicatorii cantității de distorsiuni liniare.

### 2. NMR (*Noise-to-Masked-Ratio*)

Schema de măsură NMR (raportul zgomot/mascare – Brandenburg, 1987) evaluează diferențele de nivel dintre pragul de mascare și zgomot. Pentru a analiza structura spectrală a semnalului se folosește o transformare DFT cu o fereastră Hann de aproximativ 20 ms.

Pragul absolut este adaptat la rezoluția semnalului de intrare (de obicei 16 biți), dar nu cerințelor psihoacustice. De aceea NMR nu este sensibil la schimbare nivelului de redare. Rezoluția în înălțime este de aprox. 1 Bark. Deoarece puterea de calcul necesară este mică, NMR a fost implementat și ca sistem de măsură în timp real.

Modelul se utilizează din 1987 și în timp și-a demonstrat fiabilitatea.

Cele mai importante mărimi de ieșire ale NMR sunt procentul de cadre cu distorsiuni audibile, precum și valoarea totală și medie a NMR, care sunt două moduri diferite de mediere a distanței dintre eroarea energetică și pragul de mascare.

### 3. OASE (*Objective Audio Signal Evaluation*)

Schema de măsură OASE (Sporer, 1997) folosește pentru analiza semnalelor de intrare un grup de 241 de filtre. Frecvențele centrale sunt egal distanțate pe scala Bark la o distanță de 0,1 Bark. Filtrele se suprapun unele cu celelalte. Fiecare filtru este adaptat la frecvența de răspuns a unui punct din membrana basilară. Ca și în cazul NMR pantele depind de nivel. Filtrele cu

frecvență centrală joasă necesită un calcul la frecvența nominală de eșantionare, pe când cele cu frecvența centrală înaltă se pot calcula la frecvențe reduse. După filtre, se calculează un model al efectelor temporale ale sistemului auditiv uman. După acest pas este posibilă o reducere a frecvenței de eșantionare în toate benzile. Aceasta conduce la o rezoluție temporală a grupului de filtre de 0,66 ms la o frecvență de eșantionare de 48 kHz. Semnalele de ieșire pentru referință și semnalul testat sunt comparate cu o funcție de probabilitate a detecției. Această funcție folosește tăria semnalelor de intrare ca mărime de intrare pentru a calcula JNLD. Probabilitatea totală de detecție derivă din probabilitatea de detecție din fiecare bandă. Această operație se efectuează pentru ambele canale de intrare cât și pentru așa numitul canal central. Probabilitatea de detecție în canalul central pentru fiecare bandă reprezintă cazul cel mai defavorabil al probabilității de detecție pentru canalul stâng sau drept. De asemenea, pentru fiecare cadru de 0,66 ms se calculează suma pragurilor obținute în pașii de mai sus.

Pentru calculul mediei temporale a probabilității de detecție a pragului se folosesc mai multe metode:

- media temporală a probabilității de detecție;
- frecvența cadrelor cu o probabilitate de detecție peste 0,5;
- maximul probabilității de detecție căreia i s-a aplicat un filtru trece jos;
- numărul mediu de pași peste prag pentru cadre cu o probabilitate de detecție peste 0,5;
- numărul mediu de pași peste prag;
- numărul maxim de pași peste prag.

#### 4. PAQM (*Perceptual Audio Quality Measure*)

Principiul de bază al metodei de Măsurare Perceptuală a Calității Audio (PAQM – Beerends și Stemerding, 1992) este de a extrage reprezentările interne (reprezentări din interiorul capului subiectului) ale semnalului de referință și a celui degradat și de a mapa diferențele cu o mapare cognitivă a calității audio percepută subiectiv. Transformarea domeniului fizic extern în domeniul psiho-fizic intern se realizează prin patru operații:

- o mapare timp-frecvență cu ajutorul unei transformări DFT cu o fereastră Hann de 40 ms;
- distribuția pe frecvențe folosind scara Bark;
- împrăștiere timp-frecvență (convoluție neliniară);
- distribuție în intensitate (compresie).

Combinarea dintre împrăștiere și compresie permite modelarea comportamentului sistemului auditiv uman la și deasupra pragului de mascare. Optimizarea compresiei se realizează folosind rezultatele subiective obținute la prima evaluare MPEG a codecurilor audio [ISO90], [ISO91]. Diferența dintre reprezentările interne se exprimă ca perturbarea zgomotului. În ultimele versiuni PAQM trimise la ITU-R, au fost introduse două noi efecte cognitive, fluxul perceptual [BEE94] și mascarea informațională [BEE96].

S-a realizat și o variantă simplificată pentru măsurarea perceptuală a calității vocale PSQM (*Perceptual Speech Quality Measure*), care folosește modelul cognitiv prezentat în [BEE94], dar extins cu o funcție de ponderare pentru intervalele de liniște. La evaluarea calității vocale în telefonie, zgomotul produs în perioadele de liniște are mai puțină importanță decât zgomotul produs în timpul convorbirii. Datorită gradului înalt de corelare dintre rezultatele obiective și subiective obținute, metoda PSQM a fost standardizată sub Recomandarea ITU-T P.861 “Măsurarea calității obiective a codecurilor vocale în bandă telefonică (300 – 3.400 Hz)”.

#### 5. PERCEVAL (*PERCEptual EVALuation*)

PERCEVAL modelează caracteristicile de transfer ale urechii mijlocii și interne pentru a realiza o reprezentare internă a semnalului. Semnalul de intrare este descompus într-o reprezentare timp-frecvență folosind o transformare DFT. Semnalului de intrare i se aplică o fereastră Hann de aprox. 40 ms, cu o suprapunere de 50% între ferestrele succesive. Spectrul energetic este înmulțit cu o funcție dependentă de frecvență care modelează efectul canalului

auditiv și al urechii mijlocii. Valorile atenuate ale energiei spectrale sunt apoi mapate dintr-o scară de frecvențe într-o scară de înălțimi, care este mai liniară în raport cu proprietățile fizice ale urechii interne și față de efectele psiho-fizice observate. Componentele energetice transformate sunt apoi supuse unei operații de convoluție cu o funcție de împrăștiere pentru a simula dispersia energiei de-a lungul membranei basilară. În final, fiecărei componente de înălțime  $i$  se adaugă o energie dependentă de frecvență pentru a ține cont de pragul absolut de audibilitate. Conversia energiei în decibeli se face într-o reprezentare a membranei basilară a semnalului.

Ca instrument pentru estimarea calității audio, PERCEVAL calculează diferența dintre reprezentările semnalului de referință și semnalului testat. Folosind ipoteze rezonabile asupra nivelului perceptual și procesului cognitiv, se calculează un număr de variabile relevante perceptual și se combină pentru a da o măsură a calității obiective a semnalului testat.

#### 6. POM (*Perceptual Objective Measurement*)

Scopul metodei perceptuale obiective de măsură POM [COL94] este de a cuantifica un anumit grad de deteriorare care poate apărea între semnalul de referință și cel testat. Aceasta se realizează prin comparația reprezentării basilară interne pentru ambele semnale, indiferent de degradarea produsă. Reprezentarea basilară modelează diferitele procese prin care trece semnalul audio prin urechea umană.

POM folosește o transformare DFT cu o fereastră Hann de 40 ms și 50% suprapunere. Analiza basilară se face pe 620 de canale. Celelalte părți ale modelului auditiv sunt aproape identice cu cele folosite în PAQM și PERCEVAL.

Din acest model rezultă probabilitatea de a detecta o distorsiune între două semnale, precum și așa numita distanță basilară, care reprezintă distanța perceptuală între cele două excitații.

#### 7. *The toolbox approach*

Toolbox este o metodă cu trei pași pentru măsurarea distanței percepute între calitatea audio a unui semnal de test și a unui semnal de referință, furnizând o indicație asupra nivelului de calitate audio subiectivă a semnalului testat. Metoda se bazează pe modele perceptuale binecunoscute, care sunt folosite pentru a descrie reprezentarea perceptuală a diferențelor dintre cele două semnale audio. În plus, ea include o procedură de ponderare pentru calitatea audio percepută a unui semnal testat stereofonic, luând în considerare rezultatele obținute pentru canalele stâng și drept. Nu este necesară o corelare rigidă eșantion cu eșantion între semnalul de referință și cel testat.

Primul pas și principala funcție a toolbox este calculul tăriei specifice, folosind o transformare FFT cu 2.048 de puncte, cu o fereastră Hann de 40 ms. Fereastra se deplasează cu incrementele de 10 ms. Se adaugă efectele de mascare temporală, pre- și post-mascare. Din aceste valori de bază ale senzației se calculează alți parametri perceptuali (tăria integrată, tăria mascării parțiale, claritate și pre-ecou) ce vor fi folosiți în pașii următori.

Al doilea pas cuprinde proceduri de ponderare, care depind în principal de mărimea diferenței percepute și variația tăriei în timp.

Al treilea pas al metodei toolbox include generarea unui set de valori intermediare, care se bazează pe analiza statistică a valorilor obținute în pașii 1 și 2. Această analiză statistică include calculul valorilor medii, maxime, medie pătratică și deviația standard a valorilor medii. O sumă ponderată a acestor valori intermediare de ieșire se folosește pentru potrivirea finală a distanței perceptuale dintre semnalul testat și cel de referință. Dacă este necesar, valoarea unică de ieșire se poate adapta scării de diferență subiectivă, folosită de obicei în testele de ascultare critică, cu ajutorul unei funcții liniare sau polinomiale.



### 8. PEAQ (*Perceptual Evaluation of Audio Quality*)

Metoda de evaluare perceptuală a calității audio necesită la intrare un semnal de referință (semnalul original, neprelucrat) și semnalul codat. Ea evaluează diferențele dintre cele două semnale și le notează conform unor criterii psihoacustice. Diferențele individuale sunt combinate pentru a forma o estimare a gradului de diferență subiectivă, numit grad de diferență obiectivă (ODG – *Objective Difference Grade*). Schema bloc a metodei a fost prezentată în fig. 4.1.

Primul pas în acest proces este transformarea semnalelor de intrare în reprezentări ale modelului de excitație de-a lungul membranei basilare cu ajutorul unei transformări DFT și trecerea la o scară de înălțimi perceptuale. Modelele de excitație sunt apoi corectate luându-se în considerare pragul absolut de audibilitate și se aplică o funcție de împrăștiere, care reprezintă anumite efecte de mascare. Modelele de excitație rezultate sunt folosite apoi pentru a calcula raportul zgomot-mascare, distorsiuni liniare, diferențele de bandă de frecvențe, diferențele de modulație și curbele de mascare reale. Se calculează apoi probabilitatea de detecție pentru fiecare diferență individuală dintre semnalul de referință și cel codat.

Variabilele de ieșire ale modelului (MOV – model output variables) obținute sunt apoi combinate cu ajutorul unei rețele neuronale, pentru a calcula un grad de diferență subiectivă, deoarece relațiile dintre ele sunt foarte complicate și greu de modelat. Această rețea a fost antrenată și verificată cu rezultatele mai multor teste de ascultare. Rețeaua aproximează modul de judecată a unui ascultător expert mediu pentru diferențele individuale. O prezentare detaliată a acestei metode este făcută în capitolul 6.

Metoda PEAQ nu include o componentă statistică. Mărimea de ieșire este un singur număr care semnifică gradul de diferență [HOE03].

## 4.4 Aplicații posibile ale metodelor obiective de evaluare a calității audio

Comitetul ITU-R propune câteva posibile aplicații pentru aceste metode obiective de măsurare, prezentate în tabelul 4.3 [THI00].

**Tabelul 4.3** Aplicații propuse de ITU-R pentru metoda obiectivă de evaluare a calității audio

	Aplicații	Scurtă descriere	Categorie
1	Evaluarea implementărilor	Procedură pentru caracterizarea diferitelor implementări ale echipamentelor de procesare audio, în general codecuri audio	Diagnostic
2	Alinierea calității perceptuale	Procedură rapidă de testare a echipamentelor sau circuitelor înaintea intrării în serviciu	Operațional
3	Monitorizare on-line	Prelucrare continuă pentru monitorizarea transmisiilor audio	Operațional
4	Starea unui echipament sau conexiuni	Analiza detaliată a unui echipament sau a unui circuit	Diagnostic
5	Identificare unui codec	Procesul de identificare a tipului și implementării unui anumit codec	Diagnostic
6	Dezvoltarea unui codec	Analiza performanțelor unui codec în cele mai mici detalii	Dezvoltare
7	Proiectarea unei rețele	Optimizarea costurilor și performanțelor unei rețele de transmisie sub anumite condiții	Dezvoltare
8	Ajutor în evaluarea subiectivă	Procedeu de identificare a materialelor critice pentru a fi incluse în testele de ascultare subiective	Dezvoltare

Implementarea metodei obiective de măsurare poate fi făcută în timp real sau nu. De asemenea trebuie făcută o diferențiere între măsurările on-line și cele off-line. În cazul unei măsurări off-line procedeu de măsurare are acces complet la echipament sau conexiune, în timp ce măsurările on-line implică neîntreruperea programului în curs de desfășurare.

#### 4.4.1 Evaluarea implementărilor

Posturile de radio, operatorii de rețea și alți utilizatori au nevoie de evaluarea diferitelor implementări în echipamente a codecurilor audio, în vederea achiziționării sau a efectuării unor teste de acceptanță. Pentru acest tip de aplicații este necesară o precizie ridicată pentru a evalua mici deteriorări ale sunetului și a face un clasament corect al diferitelor implementări. Pentru utilizatori, o singură mărime de ieșire, precum ODG (*Objective Difference Grade*), este suficientă, dar pentru realizatorii de codecuri audio sunt disponibile și alte variabile de ieșire ale modelului.

#### 4.4.2 Alinierea calității perceptuale

Aceasta este o procedură rapidă, care are loc înaintea punerii în serviciu a unui echipament sau a unui circuit cu scopul de a verifica funcționalitatea și calitatea. Echipamentul de măsurare va fi operat de tehnicienii beneficiarului. Pot fi prezente orice tip de distorsiuni. Sunt necesare măsurători în timp real. Se pot folosi semnale de test sau anumite semnale audio preselectate. Mărimea care reprezintă calitatea audio, ODG, trebuie afișată de două ori pe secundă, sau în cazul unui semnal de test special, la sfârșitul acestuia.

#### 4.4.3 Monitorizarea on-line

Acesta este un proces continuu, care are loc în timpul unei transmisii audio. Programul nu trebuie întrerupt de procedura de măsurare, de aceea se va folosi ca semnal programul propriu-zis sau un fragment audio predefinit (de preferat un semnal de post sau un jingle). Sunt necesare măsurări în timp real. Mărimea care reprezintă calitatea audio, ODG, trebuie afișată de două ori pe secundă, sau în cazul unui semnal de test special, la sfârșitul acestuia. Afișarea altor variabile nu este necesară.

#### 4.4.4 Starea echipamentului sau a conexiunii

Pentru a asigura funcționarea unui echipament sau a unei conexiuni audio trebuie făcută o verificare periodică a calității. Spre deosebire de cele două aplicații anterioare, această aplicație necesită verificarea mai multor parametri tehnici. Sistemul de măsurare trebuie să furnizeze informații despre influența stării echipamentului sau conexiunii asupra calității audio percepute, prin afișarea, alături de ODG, a setului complet de MOV. Nu sunt necesare măsurări în timp real.

#### 4.4.5 Identificarea codecului

Pentru a identifica codecurile (diferiți algoritmi sau implementări diferite ale aceluiași algoritm) sistemul de măsurare trebuie să fie capabil să stocheze, să găsească și să compare modelele diferitelor caracteristici. Similitudinile dintre modele pot fi luate ca o măsură a asemănării dintre diferite implementări. Această procedură se folosește pentru a identifica tipul și implementarea unui anumit codec. Sistemul de măsurare trebuie să înregistreze numărul maxim de informații posibile.

#### 4.4.6 Dezvoltarea codecurilor

Pentru această aplicație, metoda de măsurare trebuie să caracterizeze performanțele codecului testat cât mai precis și mai detaliat posibil, în special în cazul micilor deteriorări. Este necesar accesul la setul complet de variabile MOV.

#### 4.4.7 Planificarea rețelelor

Planificarea rețelelor necesită evaluarea calității dorite în diferite puncte, în timpul procesului de proiectare. Se poate folosi o simulare software a componentelor rețelei, care să permită combinarea diferitelor stadii de procesare audio, pentru a examina diferite configurații în vederea optimizării calității audio. La final, se pot testa componentele reale în configurația aleasă. Planificarea rețelelor se face de către inginerii de sistem, care trebuie să regăsească informații detaliate asupra influenței caracteristicilor rețelei asupra calității audio. Clasificarea configurațiilor posibile ale rețelei se bazează pe un set de variabile MOV, corespunzător aplicației rețelei. Nu sunt necesare măsurări în timp real.

#### 4.4.8 Ajutor în evaluarea subiectivă

Metoda de evaluare obiectivă furnizează un instrument de selecție a materialelor audio critice ce vor fi folosite în testele de ascultare. Pentru analiza materialelor critice se poate folosi întreg setul de variabile MOV.

### 4.5 Concluzii

Testele de ascultare critică au o durată mare și sunt costisitoare. Pregătirea materialelor de test, antrenarea, ascultarea și analiza statistică a rezultatelor necesită efort și experiență. Realizarea unor astfel de teste cere, în general, resurse care depășesc posibilitățile unor organizații ce doresc monitorizarea unor activități online sau verificarea unor codec-uri destinate unor aplicații specifice. Rezultatele testelor obținute pentru o implementare a unui codec nu pot fi aplicate cu siguranță și altor implementări. Acesta a fost motivul pentru care ITU a dezvoltat o metodologie obiectivă de evaluare, numită PEAQ, concretizată în 1999 într-o primă formulare a recomandării BS-1387. Scopul ei este de a formaliza și automatiza procedurile subiective consumatoare de timp într-un instrument și o metodologie cu aplicare rapidă, care include experiența și cunoștințele acumulate în numeroase teste de ascultare critică, precum și stadiul actual al înțelegerii sistemului auditiv uman.

Această metodă nu va înlocui testele de ascultare, ci reprezintă un mod rapid de aproximare a rezultatelor. Mărimea sa de ieșire reprezintă o estimare a calității audio de bază, așa cum ar fi fost obținută printr-un test de ascultare critică.

Cu ajutorul acestei metode se poate monitoriza comportamentul unui codec în cadrul unui sistem și se pot realiza comparații între diferite implementări ale unei scheme de codare. De asemenea, proiectanții pot evalua rapid efectul modificărilor efectuate asupra unui algoritm de codare.

Pentru a scoate în evidență diferențele dintre sistemele de codare testate, materialele audio trebuie să fie suficient de critice pentru fiecare sistem testat în parte. Selectarea materialelor audio de test este o operație dificilă, dar foarte importantă, pentru că de ea depinde obiectivitatea evaluărilor obținute. Se recomandă utilizarea unor semnale "naturale", de gen program radio. Nu se recomandă utilizarea unor semnale sintetice, prea critice, create special pentru a pune în dificultate sistemele de codare.

De asemenea, trebuie să se acorde atenție conținutului artistic sau intelectual al secvenței de program, care nu trebuie să fie nici prea atractiv, nici prea dezagreabil, pentru a nu distra atenția ascultătorilor de la detectarea diferențelor. Alegerea trebuie să reflecte probabilitatea de difuzare a fiecărui tip de program în emisiunile posturilor de radio.

Un ascultător expert este antrenat pentru a identifica diferite artefacte produse de un codec. Deoarece un codec nou poate produce noi tipuri de artefacte trebuie avut în vedere că un

instrument, care nu a fost antrenat pentru aceste artefacte specifice, ar putea să nu le recunoască și deci să nu le noteze corespunzător.

În concluzie, metoda de evaluare obiectivă a calității audio descrisă în recomandarea ITU-R BS.1387, care va fi prezentată în capitolul 6, a fost propusă cu scopul de a crea un instrument rapid de evaluare a sistemelor ce includ codec-uri audio perceptuale și, de asemenea, de a veni în sprijinul realizatorilor de teste de ascultare. Deși metoda îndeplinește scopurile propuse, nu se recomandă ca ea să fie unicul criteriu la alegerea codec-ului cel mai potrivit pentru o anumită aplicație sau standard. Testele de ascultare critică, realizate conform recomandării ITU-R BS.1116, vor avea totuși ultimul cuvânt.



# CAPITOLUL 5

## INFLUENȚA CODĂRILOR SUCCESIVE ASUPRA CALITĂȚII

Datorită răspândirii echipamentelor ce folosesc tehnici de codare audio cu debit binar redus (dbr) în activitatea studiourilor de radiodifuziune, de la contribuție, prin studioul de producție până la distribuție și emisie, a apărut problema codărilor multiple, succesive sau în “tandem” cum mai sunt denumite. Într-un lanț digital de transmitere a materialelor audio pot exista combinații diferite de scheme perceptuale de codare succesive. S-a observat că în funcție de algoritmi și debitele binare folosite calitatea finală a sunetului poate fi serios afectată.

Deși sisteme care utilizează un algoritm de codare cu dbr funcționează bine când sunt testate singure, calitatea sunetului poate scădea semnificativ în cazul unor codări succesive. Performanțele audio ale codecurilor nu pot fi măsurate cu metode tradiționale (raport semnal/zgomot, distorsiuni armonice), astfel că pentru a evalua calitatea unor codări succesive se pot folosi tehnici simple de măsură obiective sau metode obiective perceptuale (vezi cap. 7).

Introducerea tehnicilor de codare cu dbr în studiourile de radiodifuziune aduce avantaje și probleme. Se pot face operații noi, care până acum erau imposibile, dar și operații obișnuite până acum au devenit dificil de realizat. Acesta este cazul echipamentelor ce folosesc tehnici de codare MPEG și trebuie integrate într-un sistem existent, cu echipamente convenționale. O altă problemă ar fi realizarea de echipamente care să poată aplica asupra semnalelor comprimate MPEG operații asociate în mod normal cu prelucrarea în domeniul liniar.

### 5.1 Clasificarea codărilor audio succesive

Există două domenii principale care beneficiază de tehnologiile de codare cu dbr: transmisia și stocarea. Mai multe scheme de codare au fost create și chiar standardizate. Standardul ISO/IEC MPEG audio definește o familie cu trei niveluri de algoritmi de codare. Nivelul 1 este recomandat aplicațiilor de amatori datorită complexității reduse. Un exemplu este tehnica de codare folosită de Philips pentru caseta digitală compactă (DCC), care este un subset al schemei de codare standard de nivel 1, cunoscută sub denumirea de PASC (§3.3.5). Nivelul 2 se folosește în aplicațiile profesionale și radioul digital (DAB - *Digital Audio Broadcast*). Nivelul 3 este destinat unor performanțe superioare chiar și în cazul aplicațiilor cu dbr, cum ar fi ISDN, rețele de satelit sau în ultimul timp stocarea pe calculator și transmiterea prin Internet. O descriere mai amănunțită a standardului MPEG audio și a celor trei niveluri de codare s-a făcut în capitolul 3. Alte scheme de codare sunt NICAM (§3.2.1) și Dolby AC-2 (§3.3.4) folosite pentru transmisii sau ATRAC, algoritmul de codare folosit de MiniDisc (§3.3.6).

Această trecere în revistă a tehnologiilor de codare audio disponibile ne permite să ne facem o idee asupra multitudinii de combinații (codări succesive) pe care le poate suferi în prezent și mai mult în viitor un material audio.

Deteriorarea calității sunetului datorată copiilor succesive este cunoscută de la înregistrările analogice. Fiecare nouă copie introducea zgomotul benzii și distorsiuni liniare sau neliniare. Până la un punct (unde nu apar erori) domeniul audio digital este ferit de asemenea artefacte specifice domeniului analogic. Datorită proceselor de cuantizare din sistemele de codare perceptuală, după un anumit număr de codări succesive pot apărea diferențe sesizabile.

În figura 5.1 se prezintă traseul unui semnal audio prin mai multe dispozitive de codare succesive.

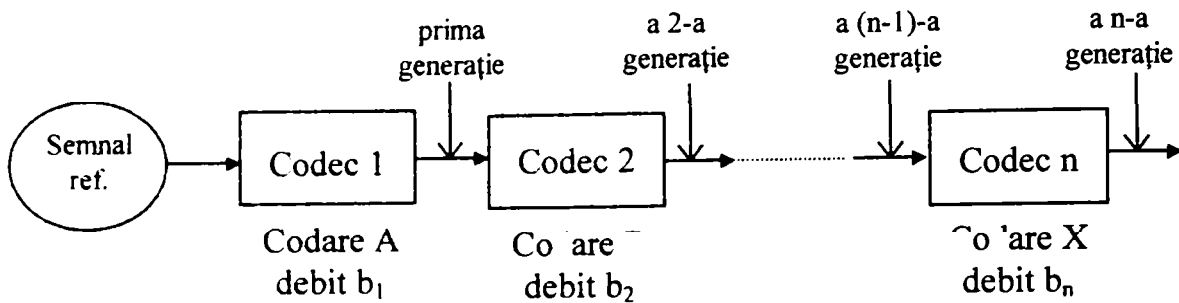


Figura 5.1 Codare audio succesivă

Fiecare codec folosește un algoritm de codare (notat cu A, B, ..., X) cu un anumit debit binar ( $b_1, b_2, \dots, b_n$ ). În cazul codărilor audio succesive se pot folosi unul sau mai multe coduri (cu algoritmi diferiți sau implementări diferite a aceluiași algoritmi) cu debite binare egale sau diferite, astfel încât se poate face următoarea clasificare:

**Clasa A:** Tandemuri omogene

Se aplică un singur algoritm

**A<sub>1</sub>:** la același debit binar (de ex. MD - MD)

**A<sub>2</sub>:** la debite diferite (de ex. rețea DAB cu MPEG nivel-2 contribuție la 180 kbit/s/c - distribuție la 180 kbit/s/c - emisie la 128 kbit/s/c)

**Clasa B:** Tandemuri heterogene

Se vor aplica diferiți algoritmi de codare

**B<sub>1</sub>:** la același debit binar (de ex. transmisie ISDN cu MPEG nivel-2 și nivel-3 la un debit de 64 kbit/s/c)

**B<sub>2</sub>:** La debite diferite (de ex. transmisie ISDN cu MPEG nivel-3 la 64 kbit/s/c - înregistrare pe harddisk cu nivel-2 la 128 kbit/s/c - stocare pe MD)

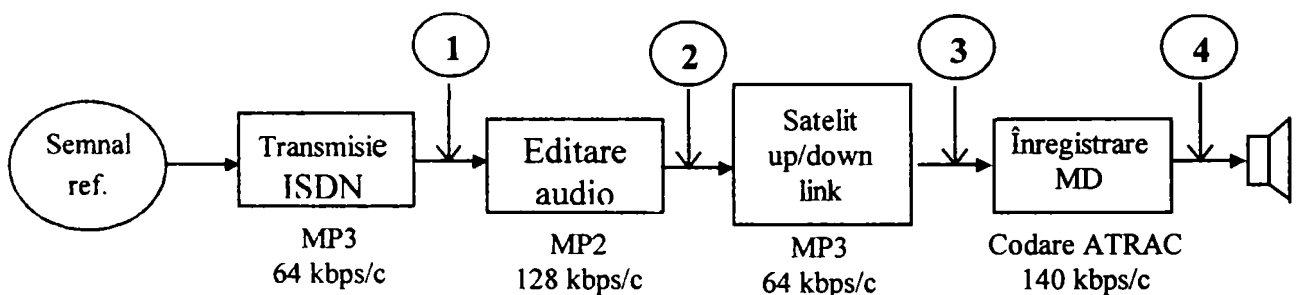


Figura 5.2 Traseul posibil al unui semnal audio într-un studiu de radiodifuziune

În multe studiouri de radiodifuziune semnalele audio sunt atât sub formă analogică cât și sub formă digitală (fig. 5.2), conversiile A/D și D/A, care se produc în diferite puncte (1 - 4) putând introduce diverse efecte nedorite. De asemenea, ar fi interesant de văzut dacă și ordinea aplicării algoritmilor de codare are vreun efect asupra calității sunetului.

ITU-R, prin grupul de lucru TG 10/2 creat în 1989 cu scopul de a selecta tehnicile de codare cu dbi adecvate aplicațiilor din domeniul radiodifuziunii, a propus un model de rețea (fig. 5.3), care să conțină cinci codări succesive la calitate de contribuție (180 kbit/s/c), urmate de trei codări succesive la calitate de distribuție (120 kbit/s/c) și o codare la calitate emisie (128 kbit/s/c). Debitul de 180 kbit/s/c și 120 kbit/s/c pentru aplicații de contribuție și distribuție au fost alese pentru a se putea încadra în canalele obișnuite de transmisie cu debite de  $N \times 64$  kbit/s.

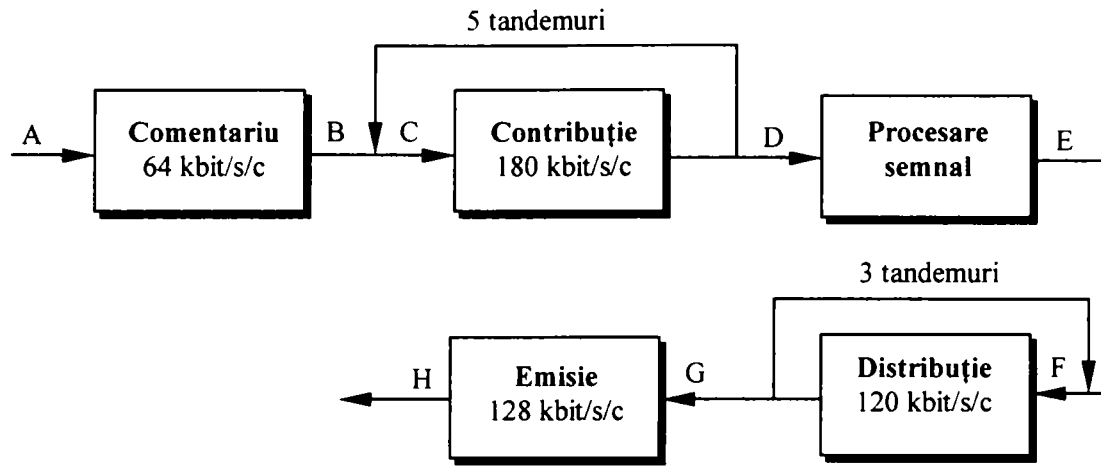


Figura 5.3 Modelul de referință al unei rețele de transmisie în radiodifuziune [BEA94]

Testele efectuate încă din 1992 pentru alegerea sistemelor de codare ce vor fi utilizate în cadrul radiodifuziunii digitale au condus la evaluarea calității obiective și subiective audio a materialelor codate o singură dată sau în cicluri succesive. Rezultatele acestor teste au dus la concluzia că nu se poate menține o calitate audio constantă pe tot parcursul lanțului de referință “5+3+1” ( $5 \times 180 \text{ kbit/s} + 3 \times 120 \text{ kbit/s} + 128 \text{ kbit/s}$ ). Recomandarea finală a ITU-R TG 10/2 definește algoritmul de codare ISO/MPEG nivel-2 ca fiind metoda recomandată pentru fazele de contribuție, distribuție și emisie, iar nivel-3 pentru comentariu.

## 5.2 Practici operaționale

Utilizarea tehnicilor de codare cu dbr în lanțul de transmisie a materialelor audio ridică noi probleme utilizatorilor. Întârzierile provocate de algoritmi de calitate provoacă probleme de monitorizare, numărul de codări succesive trebuind să fie cât mai redus posibil sau, la începutul lanțului, materialul să fie codat cu un debit cât mai mare. Editarea unui material, fără alte procesări, se recomandă să se facă sub formă codată, fără a decoda materialul în formă liniară PCM și apoi recoda.

Deși un codec poate fi “transparent” pentru un ciclu de compresie/ decompresie, după mai multe codări succesive pot apărea deteriorări perceptibile ale sunetului. Aceste deteriorări sunt în general dependente de complexitatea semnalului, un semnal sinusoidal nefiind în general suficient pentru a verifica calitatea audio perceptuală a lanțului.

Algoritmii de codare audio bazați pe mascarea auditivă încearcă să modeleze spectrul zgomotului de cuantizare astfel încât să se afle sub spectrul semnalului de intrare, adică zgomotul să fie mascat de semnalul util. Dacă raportul zgomot/mascare NMR nu conține o margine de codare suficientă, o simplă operație de egalizare poate modifica caracteristicile de mascare ale semnalului și zgomotul poate deveni audibil.

În aplicațiile de studiou, semnalele decodate de mai multe ori pot fi mixate și apoi recodate pentru transmisie. Acest lucru se poate întâmpla de mai multe ori pe parcursul producției unui program. Se poate întâmpla ca mixând două semnale audio decodate transparent, dar având diferite caracteristici de mascare a zgomotului, să rezulte un semnal în care zgomotul de cuantizare să devină audibil. Un algoritm de codare de calitate trebuie să fi mono-compatibil, adică mixarea canalelor stâng și drept ale unui semnal stereofonic nu trebuie să producă alte deteriorări ale semnalului rezultat în afară de pierderea imaginii stereofonice.

Deși, în cazul MPEG audio, sintaxa algoritmului este descrisă precis în standardul ISO 11172-3, realizatorul unui codor are o libertate destul de mare pentru implementarea algoritmului de codare. Implementarea trebuie făcută cu grijă și experiență pentru ca, pe lângă conformitatea cu standardul, să îndeplinească și cerințele de calitate ale unei rețele de radiodifuziune. Calitatea unei implementări conforme cu ISO 11172-3 poate fi afectată de mulți parametri.



Proiectarea modelului psihoacustic și strategia de alocare a biților la codor sunt deosebit de importante. Fără o bază teoretică solidă, implementarea unui codor ar putea funcționa bine pentru majoritatea semnalelor, dar să clacheze pentru semnale audio critice. Această dependență de semnal poate deveni evidentă într-un lanț de codări succesive. Proiectarea și optimizarea modelelor psihoacustice pentru a dispune de o “margină de codare” suficientă sau pentru a permite codări succesive fără deteriorări necesită multe teste subiective cu semnale critice. Implementarea blocului de filtre, a cuantizorului sau lungimea cuvântului din DSP pot afecta, de asemenea, performanțele codorului.

Interfața analogică este foarte importantă, calitatea conversiilor A/D și D/A fiind, în cazul codărilor succesive, mai importantă decât artefactele de codare. Precizia conversiei trebuie să fie suficientă pentru ca însumarea zgomotului analogic să nu reprezinte un factor limitator.

Utilizarea unui singur algoritm de-a lungul unui lanț de transmisie este recomandabilă, ducând la eliminarea transcodărilor. După cum s-a remarcat mai sus, concordanța cu standardul MPEG audio implică numai interoperabilitatea, fiind o condiție necesară dar nu și suficientă pentru un codec audio de calitate.

### 5.3 Cauzele deteriorării calității audio

Conversia din formatul MPEG în formă liniară și apoi în analogic introduce deteriorări ale semnalelor. Păstrarea lor cât mai mult în domeniul digital menține calitatea, cu condiția să putem efectua toate operațiile cu care ne-am obișnuit în domeniul analogic.

În continuare se va da o explicație deteriorărilor de calitate ce apar în cazul unor codări MPEG succesive, valabilă de altfel și în cazul altor scheme de codare ce au la bază principii psihoacustice, și câteva metode de reducere a lor.

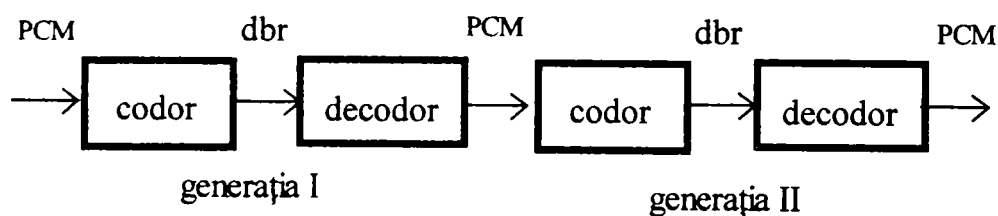


Figura 5.4 Două codări dbr succesive cu trecere intermediară prin domeniul liniar

Codarea repetată a unor semnale audio (fig.5.4) folosind algoritmi moderni de codare psihoacustică produc anumite deteriorări ale calității audio, deoarece semnalul de ieșire, în urma unui ciclu codare/decodare, nu mai este identic cu cel de intrare (eliminarea redundanțelor etc). Funcționarea unui codec, operațiile la care este supus un semnal audio, depinde de mai mulți parametri: structura pachetelor sau cadrelor impusă semnalului audio, transformarea timp-frecvență aplicată semnalului și numărul de biți alocați fiecărei părți din spectrul audio, determinată de modelul psihoacustic folosit. Pentru a afla cauzele acestei continue degradări se va analiza mai în detaliu codarea MPEG nivel 2 (fig.5.5).

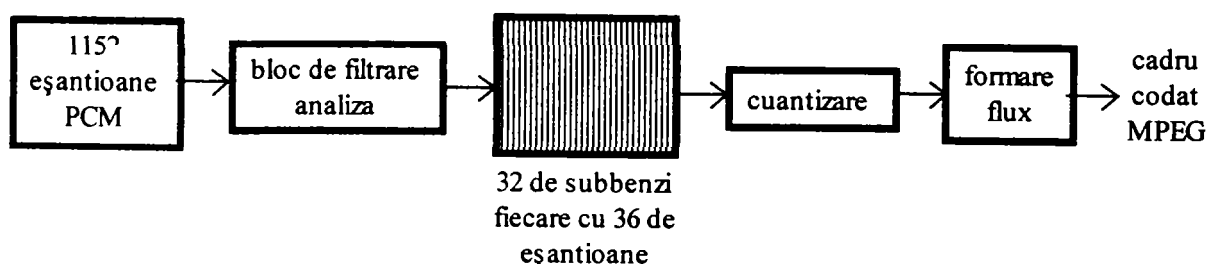


Figura 5.5 Schema bloc a codării MPEG nivel-2

### 5.3.1 Filtre

Specificațiile MPEG definesc un bloc de filtre subbandă care împarte semnalul audio în 32 de subbenzi de lățime egală. La o frecvență de eșantionare de 48 kHz, subbenzile au o lățime de 750 Hz. Filtrele subbandă au proprietăți de reconstrucție aproape perfecte, rezoluție de timp bună, dar o suprapunere (aliere) destul de mare cu subbenzile adiacente. Datorită acestor suprapuneri de frecvențe, pot apărea distorsiuni de aliere dacă eșantioanele din subbenzi nu sunt transmise cu suficientă precizie. Eșantioanele de intrare (în formatul PCM) sunt împărțite prin blocul de filtre în 32 de ieșiri paralele. Dacă același semnal este supus unei noi analize ulterioare, iar semnalul nu a fost sincronizat, blocul de filtre va crea un nou set de eșantioane pentru subbenzi. Blocul de filtre folosește un filtru prototip cu o anumită întârziere, ceea ce produce o întârziere totală de 481 de eșantioane pentru analiză și sinteză.

### 5.3.2 Cadre

În cazul MPEG nivel 2 un *cadru* al fluxului de informație reprezintă un bloc de 1.152 de eșantioane PCM (25 ms la 48 kHz). Blocul de eșantioane PCM este transformat pentru a forma 36 de seturi de eșantioane pe subbenzi (36 x 32 subbenzi = 1.152). Alocarea biților pentru fiecare subbandă se face conform modelului psihoacustic. Acesta definește numărul nivelurilor de cuantizare folosite în subbanda respectivă pe durata cadrului. Factorii de scală cu o rezoluție de – 2 dB descriu amplitudinea maximă a eșantioanelor din subbenzi pentru 1/3 din cadru (8 ms). Acești factori de scală se pot combina pentru a reprezenta 2/3 sau chiar un cadru întreg dacă au niveluri apropiate.

### 5.3.3 Factorii care afectează codarea succesivă

#### *Alinierea cadrelor*

Dacă începutul cadrului la noua codare se potrivește cu începutul cadrului din codarea anterioară, atunci cadrele sunt aliniate. În acest caz noile eșantioane pe subbenzi ar trebui să fie aproape identice cu cele din codarea anterioară sau cel puțin să se încadreze în rezoluția de reconstituire a blocului de filtre. Deci, aceleași eșantioane pe subbenzi vor fi disponibile pentru a fi recuantizate într-un cadru. Astfel ar fi posibil să se refacă identic fluxul binar original.

Dacă numai 32 de eșantioane dintr-un cadru sunt aliniate, atunci valorile eșantioanelor din subbenzi ar trebui să fie aceleași. Totuși setul de eșantioane din subbenzi care formează un cadru nu va fi același cu codarea anterioară. În acest caz factorii de scală pot fi diferiți, iar modelul psihoacustic va avea o intrare diferită și va produce o alocare diferită a biților.

Dacă decalarea cadrelor nu este multiplu de 32 de eșantioane, atunci va apărea un zgomot de eșantionare suplimentar pentru că cele 32 de eșantioane PCM sunt transformate în alte eșantioane pe subbenzi. Astfel semnalul codat va fi diferit, cu alți factori de scală și altă alocare a biților. Acesta este de fapt cazul real, pentru că nu se cunoaște sincronizarea cadrelor.

Deci, cunoscând alinierea cadrelor s-ar putea reconstitui (aproape perfect) eșantioanele din subbenzi pentru un cadru.

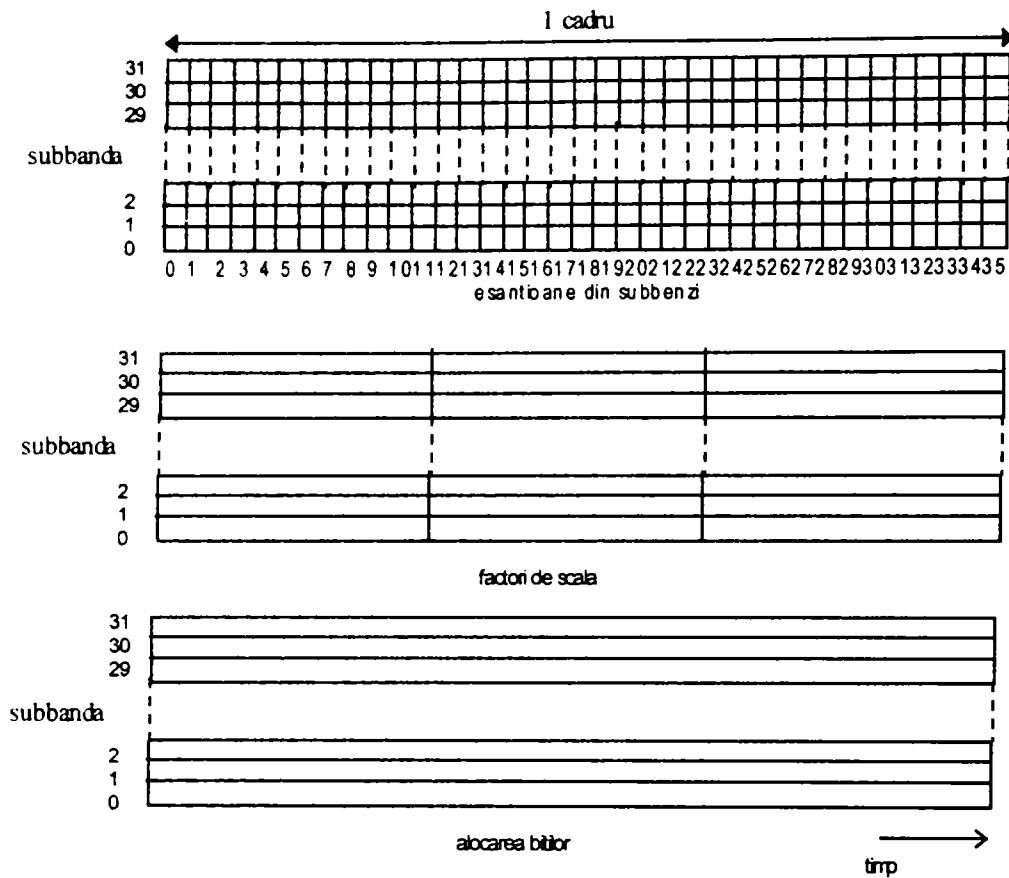


Figura 5.6 Structura cadrului pentru MPEG nivel-2 audio [MAS00]

### 5.3.4 Alocarea biților

Dacă alinierea cadrelor nu este perfectă atunci se va stabili o nouă alocare a biților, bazată pe un set diferit de eșantioane în subbenzi, ceea ce va duce la creșterea zgomotului de cuantizare și implicit la deteriorarea calității audio. În cazul ideal, de aliniere a cadrelor, descris mai sus, eșantioanele de cuantizare ar trebui să fie la fel ca în codarea anterioară. Din păcate, semnalul pe care dorim să-l codăm nu mai este identic cu cel aplicat codorului anterior datorită înlăturării informației de către procesul de cuantizare. Astfel modelul psihoacustic poate reacționa diferit la semnalul de intrare, producând o alocare diferită a biților.

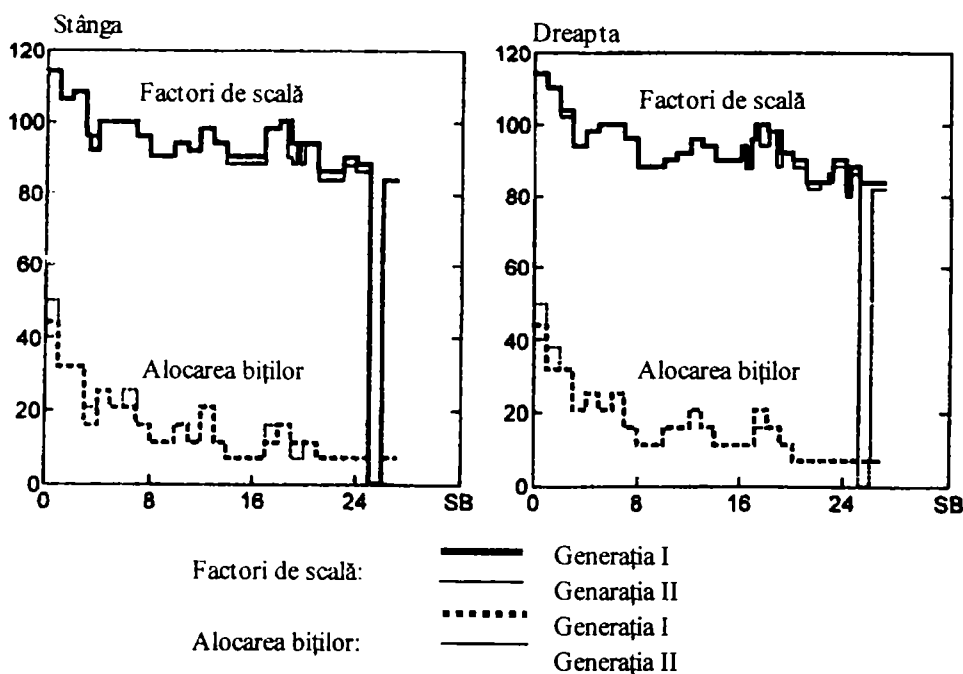


Figura 5.7 Refolosirea deciziilor de codare [MAS00]

Un alt caz în care alocarea biților este diferită apare atunci când se folosește un alt tip de codor, care utilizează un model psihoacustic diferit. Codorul va calcula o alocare complet diferită a biților. În concluzie, pentru a cuantiza identic un semnal trebuie să cunoaștem alocarea anterioară a biților.

### 5.3.5 Factori de scală

Cunoașterea alinierii cadrelor și alocarea biților sunt condiții necesare, dar nu suficiente pentru o recodare exactă. Datorită cuantizării eşantioanelor din subbenzi, unele valori cuantizate în al doilea codor pot avea o amplitudine mai mică decât eşantioanele necuantizate din primul codor. Astfel, al doilea codor va alege un alt set de factori de scală, ceea ce va introduce zgomot de cuantizare (vezi fig. 5.7).

De asemenea, reconstituirea blocului de filtre nu este perfectă: pot exista diferențe aritmetice în implementarea lor și anularea incompletă a componentelor de aliene. În concluzie, mici diferențe ale eşantioanelor din subbenzi conduc la factori de scală ușor diferiți.

Codorul curent ar trebui să aibă un algoritm diferit de combinare a factorilor de scală față de decodorul anterior. Dacă codorul curent ar trebui să aleagă o metodă diferită de combinare a factorilor de scală, s-ar introduce zgomot de cuantizare. O combinare diferită a factorilor ar putea însemna reprezentarea cu un număr diferit de biți a factorilor de scală, ceea ce ar putea duce la modificarea numărului de biți disponibili pentru cuantizarea eşantioanelor din subbenzi.

Deci, dacă am cunoaște și factorii de scală am avea toate informațiile necesare pentru o recodare audio identică.

În concluzie, având la dispoziție toate informațiile de mai sus, am putea recoda aproape identic un fișier audio codat anterior. Pot apărea totuși mici diferențe datorită modului de implementare a blocului de filtre.

La decodarea în format liniar PCM, fișierele audio din prima și a doua generație sunt de obicei identice, diferențe putând apărea doar de ordinul bitului cel mai puțin semnificativ (LSB).

## 5.4 Transcodarea prin modificarea debitului binar

Modificarea debitului binar al unui flux de date audio MPEG se poate face folosind un codor și un decodor. Totuși, așa după cum s-a arătat mai sus, fiecare trecere printr-un codec introduce modificări semnalului audio.

### 1. Transcodarea la un debit mai mare

Creșterea debitului unui fișier audio MPEG nivel-2 se poate face destul de simplu, deoarece nu avem nimic de câștigat prin modificarea fluxului de date. Creșterea numărului de biți alocați pentru a permite o rezoluție de cuantizare mai fină nu va fi de nici un ajutor, pentru că este puțin probabil ca recuantizarea valorilor eşantioanelor din subbenzi cu un cunatizor diferit să ducă la un rezultat mai apropiat de semnalul original. Cel mai apropiat pas de cuantizare al cuantizorului mai fin poate produce o valoare mai îndepărtată de eşantionul original decât cuantizorul actual, cu un pas mai mare. Aceasta ar duce la creșterea distorsiunilor din semnalul original.

Cele mai bune rezultate se pot obține menținând factorii de scală, alocarea biților și valorile eşantioanelor din subbenzi de la semnalul cu debit mai redus și umplând cu zerouri ultimele poziții pentru ca lungimea cadrului să corespundă noului debit.

## 2. *Transcodarea la un debit mai redus*

Scăderea debitului binar a unui flux MPEG nivel-2 este o operație mai complexă decât creșterea lui. Scopul acestei operații este reducerea numărului de biți folosiți pentru descrierea semnalului, astfel încât să se poată folosi un cuantizor mai brut. Dacă mărimea pasului de cuantizare se modifică semnificativ, atunci se vor introduce distorsiuni de cuantizare suplimentare. Mascarea psihoacustică va reduce audibilitatea zgomotului rezultat, dar, la fel ca în cazul primului proces de codare, trebuie avut în vedere nivelul de zgomot admisibil în fiecare subbandă. De aceea, reducerea debitului binar necesită o abordare mai atentă. O cale ar fi utilizarea alocării originale a biților ca reper pentru cantitatea de distorsiuni admisă în fiecare subbandă. De exemplu, alocarea biților pentru fiecare subbandă se va micșora cu un bit. Aceasta nu ține cont de variațiile raportului semnal/mascare produse de primul codor, unele subbenzi ar putea să nu suport reducerea alocării biților, iar altele să nu sufere nici un efect.

# CAPITOLUL 6

## METODĂ DE EVALUARE OBIECTIVĂ A CALITĂȚII AUDIO PERCEPUTE

Metoda de măsurare obiectivă a calității audio percepute care se dorește a fi implementată este conformă descrierii din Recomandarea ITU-R BS.1387-1 din 2001. Deoarece mai multe aspecte ale standardului sunt sub-documentate, în acest capitol s-au adus mai multe contribuții și interpretări alternative pentru o implementare mai clară și mai eficientă.

Deși s-a utilizat varianta ITU-R BS.1387-1 a standardului, care conține anumite corecții și clarificări față de varianta inițială BS.1387, au mai rămas totuși o serie de detalii neclarificate.

Standardul BS.1387-1 are două opțiuni: versiunea de bază și versiunea avansată. Versiunea de bază folosește un model al urechii bazat pe transformarea FFT, iar versiunea avansată folosește, pe lângă acest model, și unul bazat pe un grup de filtre. În ambele cazuri, variabilele de ieșire ale modelului sunt combinate folosind o rețea neuronală pentru a da o singură măsură, Gradul de diferență obiectivă, care măsoară degradarea intrării testate față de intrarea de referință.

În mod normal, un standard trebuie să specifice fără ambiguități pașii necesari pentru o implementare conformă. Anumite părți ale BS.1387-1 sunt insuficient specificate, nepermițând alegerea unor alternative plauzibile. Anumite variabile de ieșire sunt prea vag descrise, exemplul cel mai elocvent fiind al variabilei *EHS*.

În continuare, obiectul acestui capitol îl va face versiunea de bază a standardului BS.1387-1, cu scopul de a lămurii interpretările ambigui sau vagi ale anumitor secțiuni, în vederea unei implementări experimentale în MATCAD, care să permită o mai bună înțelegere a tehnicilor de evaluare obiectivă a calității audio.

### 6.1. Descrierea metodei

Metoda constă dintr-un model al urechii periferice, câteva etape intermediare (numite “preprocesarea formelor de excitație”), calculul variabilelor de ieșire ale modelului MOV (*Model Output Variables*) bazate pe principii psihoacustice și obținerea pe baza lor a unei singure valori reprezentând *calitatea audio de bază* a semnalului testat (fig.6.1).

Mărimile de intrare pentru calculul MOV sunt:

- mărimile de excitație pentru semnalele de referință și de test;
- mărimile de excitație adaptate spectral pentru semnalele de referință și de test;
- mărimile specifice tăriei pentru semnalele de referință și de test;
- mărimile de modulație pentru semnalele de referință și de test;
- semnalul de eroare calculat din diferența spectrală dintre semnalul testat și referință.

În cazul semnalelor stereofonice, calculele se efectuează independent și la fel pentru ambele canale.

În toate ecuațiile, indexul “Ref.” se referă la mărimile calculate pentru semnalul de referință, iar indexul “Test” se referă la mărimile calculate pentru semnalul testat. Indexul “*k*” se referă la variabile de frecvență discrete (de ex. benzi de frecvență) și “*n*” se referă la variabile de timp discrete (de ex. numărul cadrului sau eșantionului). Dacă valorile *k* sau *n* nu sunt definite explicit, calculele trebuie efectuate pentru toate valorile posibile ale lui *k* și *n*.

Metoda obiectivă de evaluare prezentată folosește un număr de 11 variabile de ieșire, calculate cu un model al urechii bazat pe transformarea FFT, care în final dau un singur parametru – *calitatea audio de bază*.

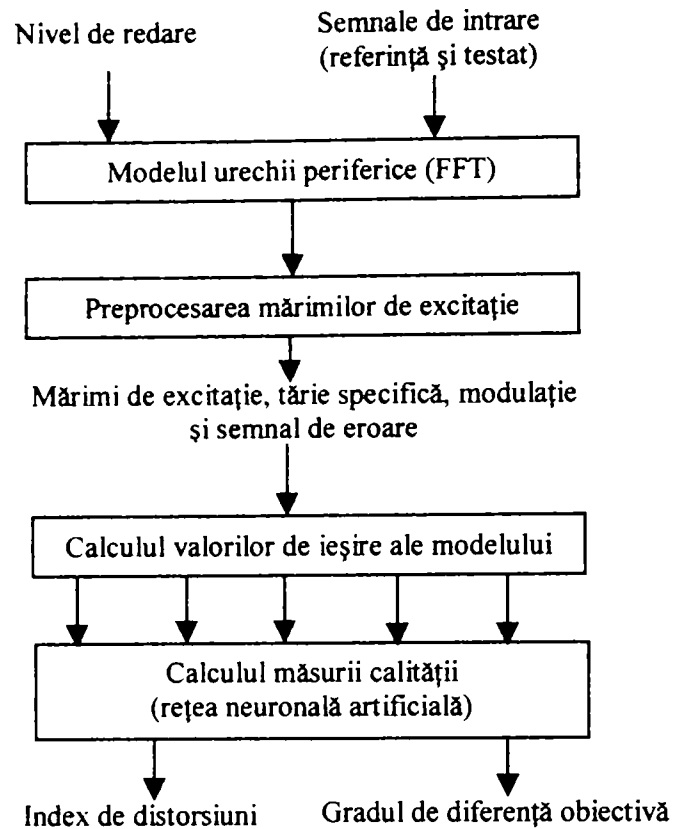


Figura 6.1 Schema bloc a metodei de măsurare

## 6.2. Modelul urechii periferice

### 6.2.1 Generalități

Semnalele de intrare (referință și test) în modelul urechii bazat pe transformarea FFT (fig. 6.2) sunt eșantionate cu frecvența  $F_s = 48$  kHz, aliniat în timp și împărțite în cadre de aproximativ 0,042 s, având o suprapunere de 50%. Fiecare cadru este transformat în domeniul frecvență folosind o fereastră Hann și o transformare FFT și apoi sunt scalate la nivelul de redare. Coeficienților spectrali li se aplică o funcție de ponderare, care ține cont de răspunsul în frecvență al urechii externe și mijlocii. Coeficienții spectrali ponderați sunt apoi grupați în benzi critice, având loc astfel o reprezentare în funcție de scala de înălțimi. Pentru a simula zgomotul intern al sistemului auditiv, se adaugă un offset dependent de frecvență. Pentru a modela filtrele spectrale auditive în domeniul frecvență se folosește o funcție de împrăștiere, care depinde de nivel. Urmează o împrăștiere în domeniul timp, care ține cont de efectele de mascare.

*Mărimile de excitație* obținute sunt folosite pentru a calcula *mărimile specifice tăriei și mărimile de mascare*. Mărimile dinainte de împrăștierea finală în domeniul timp (*unsmoothed excitation patterns*) sunt folosite pentru a calcula *mărimile de modulație*.

Pentru a modela semnalul de eroare, mărimile de ieșire ale urechii externe și medii pentru semnalul de referință și de test sunt combinate și grupate în benzi critice conform scării de înălțimi.

Aceste mărimi de ieșire, împreună cu mărimile de excitație sunt folosite pentru a calcula variabilele de ieșire ale modelului.

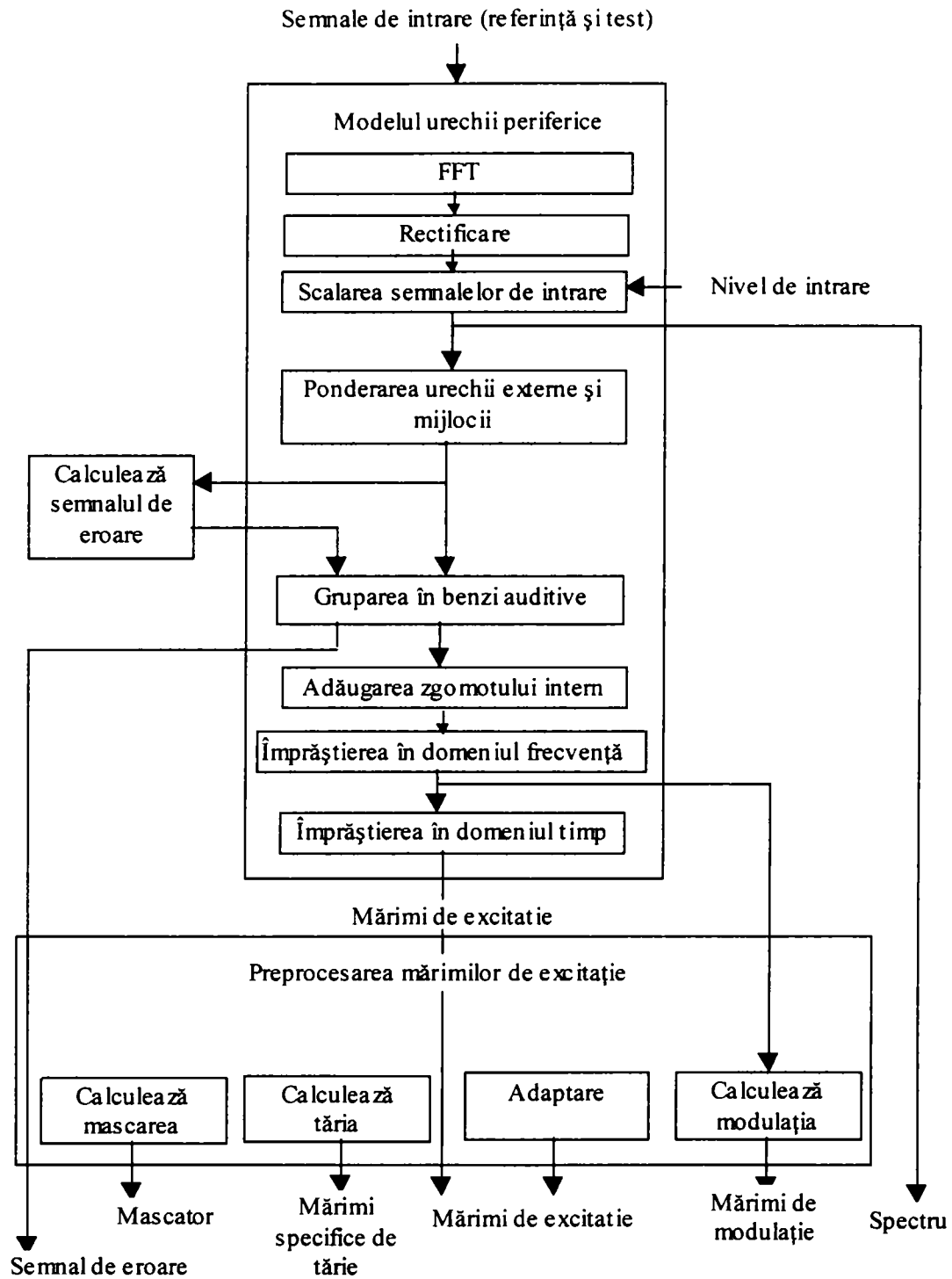


Figura 6.2 Modelul urechii periferice și preprocesarea mărimilor de excitație pentru partea bazată pe FFT

### 6.2.2 Prelucrarea în domeniul timp

La intrarea în modelul urechii bazat pe transformarea FFT, semnalele de test și de referință sunt împărțite în cadre cu lungimea  $N = 2.048$  de eșantioane cu o suprapunere de 1.024 de eșantioane:

$$t_n[k_t, n] = t[1024 \cdot n + k_t] \quad n = 0, 1, 2, \dots \quad k_t = 0 \dots 2047 \quad (6.1)$$

unde  $n$  este indicele cadrului de timp și  $k_t$  este un contor de timp în interiorul cadrului.



### 6.2.3 Transformarea FFT

Trecerea din domeniul timp în domeniul frecvență se face folosind o fereastră Hann:

$$h[k, N] = \begin{cases} \frac{1}{2} \left[ 1 - \cos\left(\frac{2\pi k}{N-1}\right) \right], & 0 \leq k \leq N-1 \\ 0, & \text{altfel} \end{cases} \quad (6.2)$$

Fereastra are valoarea zero la capetele intervalului, adică pentru  $k = 0$  și  $k = N-1$ , ceea ce înseamnă că fereastra va avea numai  $N-2$  termeni nenuli. Suma pătratelor valorilor ferestrei este:

$$\sum_{k=-\infty}^{\infty} h^2[k, N] = \frac{3}{8}(N-1) \quad (6.3)$$

Această valoare se folosește la scalarea ferestrei pentru a compensa pierderile de energie datorate formei ei ascuțite, astfel încât pentru c.c. sau zgomot alb, energia pe eșantion după aplicarea ferestrei să fie aceeași cu energia pe eșantion de dinainte.

Fereastra reală folosită în standard este o versiune scalată a ec. (6.2),

$$h_w[k] = \frac{1}{2} \sqrt{\frac{8}{3}} \left[ 1 - \cos\left(2\pi \frac{k}{N-1}\right) \right] \quad (6.4)$$

unde  $N = 2.048$

$$t_w[k_t, n] = h_w[k_t] \cdot t_n[k_t, n] \quad (6.5)$$

Datele obținute în fereastră sunt trecute în domeniul frecvență folosind o transformare Fourier discretă scalată:

$$F_f[k_f, n] = \frac{1}{2048} \sum_{k_t=0}^{2047} t_w[k_t, n] e^{-j \frac{2\pi}{2048} k_f k_t} \quad (6.6)$$

Valorile DFT sunt definite pentru  $0 \leq k \leq N-1$ . Totuși, numai valorile lui  $k$  cuprinse între 0 și  $N/2$ , corespunzătoare frecvențelor cuprinse între 0 și  $F_s/2$  (24 kHz) sunt necesare în continuare.

Anumiți factori de calitate perceptuală depind de nivelul real de presiune sonoră a semnalului testat. De aceea este necesară o operație de calibrare pentru transformarea nivelului de intrare al semnalului în tărie. Factorul de scalare pentru transformarea FFT se calculează cu ajutorul nivelului maxim de presiune sonoră  $L_p$  corespunzător unei sinusoide:

$$fac = \frac{10^{\frac{L_p}{20}}}{Norm} \quad (6.7)$$

unde  $L_p$  este presiunea sonoră (SPL) în dB a unei sinusoide de test la un nivel de 0dBFS. Dacă nu se cunoaște nivelul de presiune sonoră, se stabilește  $L_p$  la 92dB<sub>SPL</sub>, după cum se specifică în standard. Numitorul relației (6.7) notat cu  $Norm$  are forma:

$$Norm = \gamma(f_c) \sqrt{\frac{8}{3}} \frac{A_{\max}}{4} \frac{N-1}{N} \quad (6.8)$$

unde  $A_{\max} = 32.768$  – amplitudinea maximă a unei sinusoide pe 16 biți, iar  $\gamma(f_c)$  este un factor care variază între 0,84 și 1. Pentru frecvența de 1.019,5 Hz indicată de standard,  $\gamma(f_c) = 0,8497$ . Cu aceste valori obținem  $fac = 3,504$ .

În standard se afirma: “unde factorul de normalizare  $Norm$  se calculează considerând o sinusoidă cu frecvența de 1.019,5 Hz și nivel 0 dB drept semnal de intrare și calculând valoarea maximă absolută a coeficienților spectrali pe 10 cadre.” Această măsurare nu este necesară, deoarece factorului  $Norm$  se poate calcula analitic, cu rel. (6.8) și coeficienții de mai sus. De asemenea, standardul sugerează măsurarea valorii maxime a coeficienților spectrali pe 10 cadre. De fapt, combinația dintre frecvența de test și lungimea cadrului nu produce multe faze diferite. Avansul fazei pentru sinusul dintre două cadre adiacente (1.024 de eșantioane diferite) este  $1,499\pi$ . Neglijând inversările de semn, valorile sinusului se repetă la fiecare două cadre.

Forma finală a transformatei Fourier scalate va fi:

$$F[k_f, n] = fac \cdot F_f[k_f, n] \quad (6.9)$$

### 6.2.4 Urechea externă și mijlocie

Răspunsul în frecvență al urechii externe și mijlocii este modelat printr-o funcție de ponderare dependentă de frecvență:

$$W[k]/\text{dB} = -0,6 \cdot 3,64 \cdot \left(\frac{f[k]}{\text{kHz}}\right)^{-0,8} + 6,5 \cdot e^{-0,6 \cdot \left(\frac{f[k]}{\text{kHz}} - 3,3\right)^2} - 10^{-3} \cdot \left(\frac{f[k]}{\text{kHz}}\right)^{3,6} \quad (6.10)$$

unde:

$$f[k]/\text{Hz} = k \frac{F_s}{N} = k \cdot 23,4375 \quad 0 \leq k \leq N/2 \quad (6.11)$$

este reprezentarea frecvenței pe linia  $k$  care se aplică ieșirilor transformatei FFT (ec. (6.12)).

$$F_e[k_f, n] = |F[k_f, n]| \cdot 10^{\frac{w[k_f]}{20}} \quad (6.12)$$

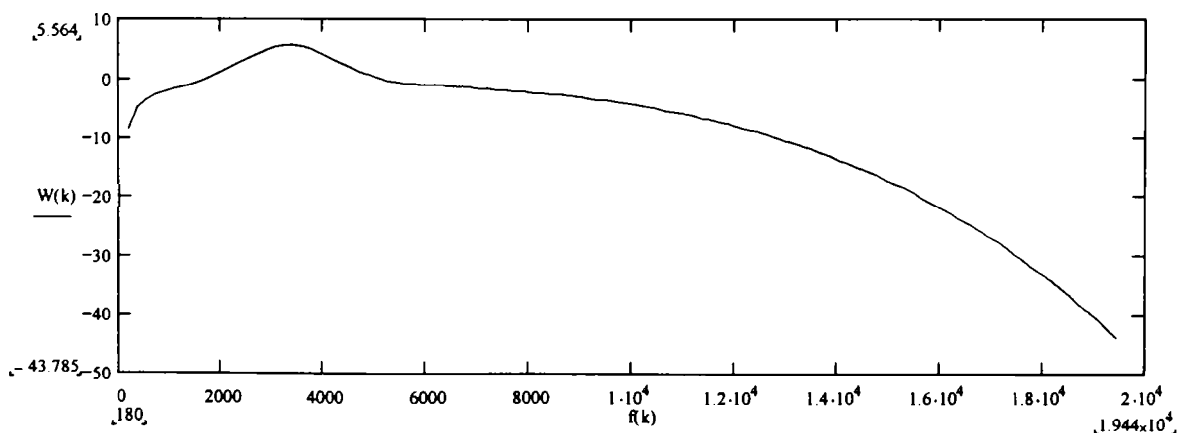


Figura 6.3 Răspunsul urechii externe și mijlocii

$F_e[k_f]$  se numește “leșirea ponderată a transformatei FFT pentru urechea externă”. Pentru  $f = 0$  răspunsul în dB este  $-\infty$ , la 1 kHz este  $-1,9$  dB, iar maximum, de  $+5,6$  dB, apare la 3,3 kHz.

Răspunsul ecuației (6.10), prezentat în figura 6.3, este similar cu cel dat de Terhardt [TER79]. Apare însă o diferență în primul termen, care controlează răspunsul la frecvențe joase. Terhardt folosește un factor egal cu 3,64, care în standard este folosit numai parțial, aplicându-se un coeficient de 0,6 ( $0,6 \times 3,64 = 2,184$ ). Cealaltă parte a coeficientului apare mai târziu, în §6.2.6, sub forma  $0,4 \times 3,64$ , ca o contribuție la zgomotul intern.

### 6.2.5 Gruparea în benzi critice

Pentru gruparea în benzi critice se folosește scara de înălțime auditivă, care se calculează cu relația aproximativă dată de Schroeder în 1979, (2.8).

$$z[\text{Bark}] = B(f) = 7 \cdot \text{ar sinh} \left( \frac{f / \text{Hz}}{650} \right) \quad (6.13)$$

Unitatea de măsură a înălțimii se numește Bark, deși această scară nu este o reprezentare exactă scării Bark așa cum a fost definită de Zwicker și Feldtkeller în 1967 (vezi §2.2.4).

Benzile de frecvență pot fi caracterizate printr-o frecvență inferioară  $f_i$ , o frecvență centrală  $f_c$  și o frecvență superioară  $f_s$ . Valorile corespunzătoare benzii pe scara Bark sunt date de relațiile:

$$\begin{aligned} z_i[k] &= z_l + k\Delta z \\ z_s[k] &= \begin{cases} z_l + (k+1)\Delta z, & k+1 \leq (z_s - z_l) / \Delta z \\ z_s, & \text{altfel} \end{cases} \\ z_c[k] &= \frac{z_s[k] - z_l[k]}{2} \end{aligned} \quad (6.14)$$

unde  $z_l = B(f_l)$  și  $z_s = B(f_s)$ .

Trecerea inversă din Bark în frecvență se face cu relația:

$$f[\text{Hz}] = B^{-1}(z) = 650 \sinh(z/7) \quad (6.15)$$

Folosind relația inversă a lui (6.14) obținem valorile corespunzătoare benzilor, în frecvență, valorile în Hz fiind date cu 3 zecimale în tabelul 6.1:

$$\begin{aligned} f_i[k] &= B^{-1}(z_i[k]) \\ f_c[k] &= B^{-1}(z_c[k]) \\ f_s[k] &= B^{-1}(z_s[k]) \end{aligned} \quad (6.16)$$

Domaniul de frecvență al filtrelor este între  $f_l = 80$  Hz și  $f_s = 18.000$  Hz. Lățimile și distanțele dintre benzile filtrelor au o rezoluție  $res = 0,25$  Bark, adică un număr  $Z = 109$  benzi de frecvență.

Tabelul 6.1 Benzile de frecvență folosite la modelul urechii bazat pe FFT

Grup	Frecv. inf./Hz	Frecv. centr./Hz	Frecv. sup./Hz	Lățimea benzii/Hz
$k$	$f_i[k]$	$f_c[k]$	$f_s[k]$	$f_b[k]$
0	80	91,708	103,445	23,445
1	103,445	115,216	127,023	23,577
2	127,023	138,87	150,762	23,739
3	150,762	162,702	174,694	23,932
4	174,694	186,742	198,849	24,155
5	198,849	211,019	223,257	24,408
6	223,257	235,566	247,95	24,693
7	247,95	260,413	272,959	25,009
8	272,959	285,593	298,317	25,358
9	298,317	311,136	324,055	25,738
10	324,055	337,077	350,207	26,151
11	350,207	363,448	376,805	26,598
12	376,805	390,282	403,884	27,079
13	403,884	417,614	431,478	27,594
14	431,478	445,479	459,622	28,145
15	459,622	473,912	488,353	28,731
16	488,353	502,95	517,707	29,354
17	517,707	532,629	547,721	30,014
18	547,721	562,988	578,434	30,713
19	578,434	594,065	609,885	31,451
20	609,885	625,899	642,114	32,229
21	642,114	658,533	675,161	33,048
22	675,161	692,006	709,071	33,909
23	709,071	726,362	743,884	34,814
24	743,884	761,644	779,647	35,763
25	779,647	797,898	816,404	36,757
26	816,404	835,17	854,203	37,799
27	854,203	873,508	893,091	38,888
28	893,091	912,959	933,119	40,028
29	933,119	953,576	974,336	41,218
30	974,336	995,408	1016,797	42,461
31	1016,797	1038,511	1060,555	42,461
32	1060,555	1082,938	1105,666	45,111
33	1105,666	1128,746	1152,187	46,521
34	1152,187	1175,995	1200,178	47,991
35	1200,178	1224,744	1249,7	49,522
36	1249,7	1275,055	1300,816	51,116
37	1300,816	1326,992	1353,592	52,776
38	1353,592	1380,623	1408,094	54,502
39	1408,094	1436,014	1464,392	56,298
40	1464,392	1493,237	1522,559	58,167
41	1522,559	1552,366	1582,668	60,109
42	1582,668	1613,474	1644,795	62,128
43	1644,795	1676,641	1709,021	64,226
44	1709,021	1741,946	1775,427	66,406

Grup	Frecv. inf./Hz	Frecv. centr./Hz	Frecv. sup./Hz	Lățimea benzii/Hz
$k$	$f_l[k]$	$f_c[k]$	$f_s[k]$	$f_b[k]$
45	1775,427	1809,474	1844,098	68,671
46	1844,098	1879,31	1915,121	71,023
47	1915,121	1951,543	1988,587	73,466
48	1988,587	2026,266	2064,59	76,003
49	2064,59	2103,573	2143,227	78,637
50	2143,227	2183,563	2224,597	81,371
51	2224,597	2266,34	2308,806	84,208
52	2308,806	2352,008	2395,959	87,154
53	2395,959	2440,675	2486,169	90,21
54	2486,169	2532,456	2579,551	93,382
55	2579,551	2627,468	2676,223	96,672
56	2676,223	2725,832	2776,309	100,086
57	2776,309	2827,672	2879,937	103,627
58	2879,937	2933,12	2987,238	107,302
59	2987,238	3042,309	3098,35	111,112
60	3098,35	3155,379	3213,415	115,065
61	3213,415	3272,475	3332,579	119,164
62	3332,579	3393,745	3455,993	123,415
63	3455,993	3519,344	3583,817	127,823
64	3583,817	3649,432	3716,212	132,395
65	3716,212	3784,176	3853,348	137,136
66	3853,348	3923,748	3995,399	142,051
67	3995,399	4068,324	4142,547	147,148
68	4142,547	4218,09	4294,979	152,432
69	4294,979	4373,237	4452,89	157,911
70	4452,89	4533,963	4616,482	163,592
71	4616,482	4700,473	4785,962	169,48
72	4785,962	4872,978	4961,548	175,585
73	4961,548	5051,7	5143,463	181,915
74	5143,463	5236,866	5331,939	188,476
75	5331,939	5428,712	5527,217	195,278
76	5527,217	5627,484	5729,545	202,329
77	5729,545	5833,434	5939,183	209,637
78	5939,183	6046,825	6156,396	217,214
79	6156,396	6267,931	6381,463	225,067
80	6381,463	6497,031	6614,671	233,208
81	6614,671	6734,42	6856,316	241,646
82	6856,316	6980,399	7106,708	250,392
83	7106,708	7235,284	7366,166	259,458
84	7366,166	7499,397	7635,02	268,854
85	7635,02	7773,077	7913,614	278,594
86	7913,614	8056,673	8202,302	288,688
87	8202,302	8350,547	8501,454	299,152
88	8501,454	8655,072	8811,45	309,996
89	8811,45	8970,639	9132,688	321,237
90	9132,688	9297,648	9465,574	332,887
91	9465,574	9636,52	9810,536	344,962

Grup	Frecv. inf./Hz	Frecv. centr./Hz	Frecv. sup./Hz	Lățimea benzii/Hz
$k$	$f_l[k]$	$f_c[k]$	$f_s[k]$	$f_b[k]$
92	9810,536	9987,683	10168,013	357,477
93	10168,013	10351,586	10538,46	370,447
94	10538,46	10728,695	10922,351	383,891
95	10922,351	11119,49	11320,175	397,824
96	11320,175	11524,47	11732,438	412,264
97	11732,438	11944,149	12159,67	427,231
98	12159,67	12379,066	12602,412	442,742
99	12602,412	12829,775	13061,229	458,817
100	13061,229	13296,85	13536,71	475,48
101	13536,71	13780,887	14029,458	492,748
102	14029,458	14282,503	14540,103	510,645
103	14540,103	14802,338	15069,295	529,192
104	15069,295	15341,057	15617,71	548,415
105	15617,71	15899,345	16186,049	568,339
106	16186,049	16477,914	16775,035	588,986
107	16775,035	17077,504	17385,42	610,385
108	17385,42	17690,045	18000	614,58

Trecerea de la frecvență la înălțime se face cu ajutorul algoritmului următor, unde  $F_{sp}[k_f]$  este reprezentarea energiei ieșirii ponderate a transformatei FFT pentru urechea externă:

$$F_{sp}[k_f, n] = |F_e[k_f, n]|^2 \quad (6.17)$$

sau reprezentarea energiei semnalului de eroare

$$F_{sp}[k_f, n] = |F_{noise}[k_f, n]|^2 \quad (6.18)$$

Vezi § 6.3.4 pentru calculul semnalului de eroare.

Mărimile de ieșire ale acestei etape de prelucrare sunt energiile grupurilor de frecvențe,  $P_e[k, n]$ .

### 6.2.5.1 Gruparea vectorului energie în benzile critice

$$\begin{aligned} \text{Indexul grupurilor de frecvențe} & \quad i := 0..108 \\ \text{Constantă pentru rezoluția frecvenței} & \quad F_{res} := \frac{48000}{2048} \\ f_{l_i} := FF_{i,0} & \quad f_{c_i} := FF_{i,2} \quad f_{s_i} := FF_{i,1} \end{aligned}$$

Calculul energiilor grupurilor de frecvențe  $P_e$  (MATCAD):

```

Peref := | Pe ← 0
         | for n ∈ 0..nn
         |   for i ∈ 0..108
         |     Pei,n ← 0
         |     for k ∈ 0..108
         |       Pei,n ← Pei,n + Fsprefk,n if [[(k - .5)·Fres ≥ fii] ∧ [(k + .5)·Fres ≤ fsi]]
         |       Pei,n ← Pei,n + Fsprefk,n ·  $\frac{(fs_i - fi_i)}{Fres}$  if [(k - .5)·Fres < fii ∧ (k + .5)·Fres > fsi]
         |       Pei,n ← Pei,n + Fsprefk,n ·  $\frac{[(k + .5)·Fres - fi_i]}{Fres}$  if [(k - .5)·Fres < fii ∧ (k + .5)·Fres > fii]
         |       Pei,n ← Pei,n + Fsprefk,n ·  $\frac{[fs_i - (k - .5)·Fres]}{Fres}$  if [(k - .5)·Fres < fsi ∧ (k + .5)·Fres > fsi]
         |       Pei,n ← Pei,n otherwise
         |     Pei,n ← max(Pei,n, .000000000000000)
         | Pe

```

```

Petest := | Pe ← 0
          | for n ∈ 0..nn
          |   for i ∈ 0..108
          |     Pei,n ← 0
          |     for k ∈ 0..108
          |       Pei,n ← Pei,n + Fsptestk,n if [[(k - .5)·Fres ≥ fii] ∧ [(k + .5)·Fres ≤ fsi]]
          |       Pei,n ← Pei,n + Fsptestk,n ·  $\frac{(fs_i - fi_i)}{Fres}$  if [(k - .5)·Fres < fii ∧ (k + .5)·Fres > fsi]
          |       Pei,n ← Pei,n + Fsptestk,n ·  $\frac{[(k + .5)·Fres - fi_i]}{Fres}$  if [(k - .5)·Fres < fii ∧ (k + .5)·Fres > fii]
          |       Pei,n ← Pei,n + Fsptestk,n ·  $\frac{[fs_i - (k - .5)·Fres]}{Fres}$  if [(k - .5)·Fres < fsi ∧ (k + .5)·Fres > fsi]
          |       Pei,n ← Pei,n otherwise
          |     Pei,n ← max(Pei,n, .000000000000000)
          | Pe

```

### 6.2.6 Adăugarea zgomotului intern

Energiilor din fiecare grup de frecvențe li se adugă un offset dependent de frecvență  $P_{Thres}$ , pentru a compensa zgomotul intern generat de ureche, modelat cu:

$$P_{Thres}[k] = 10^{0,4 - 0,364 \cdot \left(\frac{f_c[k]}{\text{kHz}}\right)^{-0,8}} \quad (6.19)$$

$$P_p[k, n] = P_e[k, n] + P_{Thres}[k] \quad (6.20)$$

Factorul  $0,4 \times 3,64 = 1,456$  este partea care lipsea din formula lui Terhard, la care ne-am referit în §6.2.4. Forma răspunsului este dată în fig.6.4. La 1 kHz, contribuția este de 1,46 dB. Mărimile de ieșire ale acestei etape de prelucrare sunt “înălțimile”,  $P_p[k,n]$ .

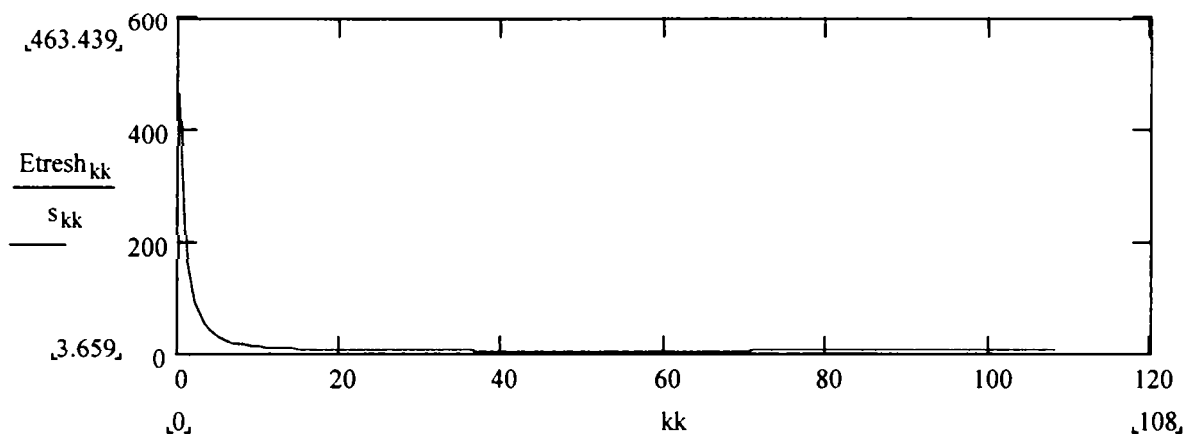


Figura 6.4 Contribuția zgomotului intern

### 6.2.7 Împrăștierea

Înălțimile  $P_p[k,n]$  sunt întinse în frecvență folosind o funcție de împrăștiere dependentă de nivel. Funcția de împrăștiere este o exponențială cu două părți. Panta ascendentă este întotdeauna de 27 dB/Bark, iar panta descendentă depinde de frecvență și energie (vezi rel. 4.2).

Pantele se calculează astfel:

$$\frac{S_u[k, L[k, n]]}{\text{dB/Bark}} = -24 - \frac{230 \text{ Hz}}{f_c[k]} + 0,2 \cdot L[k, n] / \text{dB} \quad (6.21)$$

$$S_l[k, L[k, n]] = 27 \frac{\text{dB}}{\text{Bark}} \quad (6.22)$$

cu:

$$L[k, n] = 10 \cdot \log_{10}(P_p[k, n])$$

Împrăștierea are loc independent pentru fiecare grup de frecvențe  $k$ :

$$E_2[k, n] = \frac{1}{\text{Norm}_{SP}[k]} \left( \sum_{j=0}^{Z-1} E_{line}[j, k, n]^{0,4} \right)^{\frac{1}{0,4}} \quad (6.23)$$

unde  $E_{line}$  este dată de:



$$E_{line}[j, k, n] = \left\{ \begin{array}{l} \frac{10^{\frac{l[j, n]}{10}} \cdot 10^{\frac{res \cdot (j-k) \cdot s_l[j, l[j, n]]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot s_l[j, l[j, n]]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot s_u[j, l[j, n]]}{10}}} \\ \frac{10^{\frac{l[j, n]}{10}} \cdot 10^{\frac{res \cdot (k-j) \cdot s_u[j, l[j, n]]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot s_l[j, l[j, n]]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot s_u[j, l[j, n]]}{10}}} \end{array} \right. \quad (6.24)$$

$Norm_{SP}[k]$  se calculează astfel:

$$Norm_{SP}[k] = \left( \sum_{j=0}^{Z-1} \tilde{E}_{line}[j, k]^{0.4} \right)^{\frac{1}{0.4}} \quad (6.25)$$

unde:

$$\tilde{E}_{line}[j, k] = \left\{ \begin{array}{l} \frac{10^{\frac{-res \cdot (j-k) \cdot s_l[j, 0]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot s_l[j, 0]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot s_u[j, 0]}{10}}} \quad k < j \\ \frac{10^{\frac{res \cdot (k-j) \cdot s_u[j, 0]}{10}}}{\sum_{\mu=0}^{j-1} 10^{\frac{-res \cdot (j-\mu) \cdot s_l[j, 0]}{10}} + \sum_{\mu=j}^{Z-1} 10^{\frac{res \cdot (\mu-j) \cdot s_u[j, 0]}{10}}} \quad k \geq j \end{array} \right. \quad (6.26)$$

și  $res$  reprezintă rezoluția scării înălțimii în Bark (0,25).

Mărimile obținute în această fază a prelucrării,  $E_2[k, n]$ , sunt numite “*unsmearred excitation patterns*” și folosite mai târziu pentru calculul mărimilor de modulație (“*unsmearred*” în timp).

### 6.2.8 Împrăștierea în domeniul timp

Până acum ne-am referit numai la prelucrarea unui singur cadru. Acum se introduce o împrăștiere în timp, care depinde de mai multe cadre. În acest scop se introduce un index de cadru  $n$  pentru a împrăști energia în domeniul Bark,  $E_2[k, n]$ . Cadrele sunt actualizate la fiecare 1.024 de eșantioane și au o rată de:

$$F_{ss} = \frac{F_s}{N/2} = 46,875 \text{ Hz} \quad (6.27)$$

Pentru a modela mascarea înainte, energiile din fiecare grup de frecvențe sunt întinse în timp cu filtre trece jos de ordinul întâi. Constantele de timp depind de frecvența centrală a fiecărui grup (vezi ec. (6.13) și tabelul 6.1) și se calculează cu:

$$\tau[k] = \tau_{\min} + \frac{100\text{Hz}}{f_c[k]} \cdot (\tau_{100} - \tau_{\min}) \begin{cases} \tau_{100} = 0,030 \text{ s} \\ \tau_{\min} = 0,008 \text{ s} \end{cases} \quad (6.28)$$

Constanta de timp  $\tau_{100}$  corespunde frecvenței de 100 Hz, care este numai cu puțin mai mare decât cea mai joasă frecvență centrală (91,708 Hz). Constanta de timp minimă corespunde celei mai înalte frecvențe centrale (17.690,045 Hz) și este foarte apropiată de  $\tau_{\min}$ .

Filtrele trece jos de ordinul întâi se calculează cu:

$$E_f[k, n] = a[k] \cdot E_f[k, n-1] + (1-a) \cdot E_2[k, n] \quad (6.29)$$

$$E[k, n] = \max(E_f(k, n), E_2(k, n)) \quad (6.30)$$

unde  $a[k]$  controlează constanta de timp a medierii pentru energiile descrescătoare și se calculează cu constantele de timp de mai sus:

$$a[k] = e^{-\frac{4}{187,5} \frac{1}{\tau[k]}} \quad (6.31)$$

Funcția  $\max(\cdot)$  din ec. (6.30) are semnificația că  $E[k, n]$  urmează creșteri instantanee de energie.

Condiția inițială de filtrare este  $E_f[k, 0] = 0$ . Presupunând că indexarea cadrelor se face de la  $n = 0$ , condiția inițială devine  $E_f[k, -1] = 0$ .

Mărimile obținute în această fază a prelucrării,  $E[k, n]$ , se numesc “mărimi de excitație”.

### 6.2.9 Pragul de mascare

Efectul de mascare se produce atunci când un semnal slab, dar distinct audibil, devine inaudibil odată cu apariția unui alt semnal mai tare (§2.2.5). Acest prag se calculează prin ponderarea mărimilor de excitație cu un parametru  $m_{dB}[k]$  (măsurat în dB), care reprezintă cantitatea cu care pragul de mascare se află sub energia (Bark) împrăștiată timp-frecvență. Parametrul  $m_{dB}[k]$  este dat ca o funcție liniară pe scara Bark:

$$m_{dB}[k] = \begin{cases} 3,0 & k \cdot res \leq 12 \\ 0,25 \cdot k \cdot res & k \cdot res > 12 \end{cases} \quad (6.32)$$

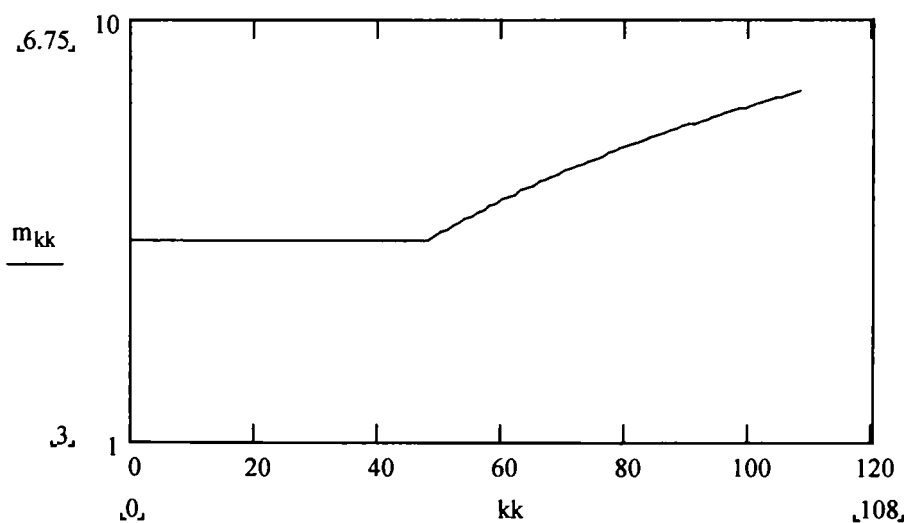


Figura 6.5 Offsetul de frecvență funcție de banda critică

Punctul de inflexiune în caracteristica offsetului pragului de mascare apare la  $z_l + 12$  pe scala Bark ( $z_l$  a fost definit la §6.2.5), corespunzător frecvenței de 1.987 Hz. Această funcție discretă poate fi extinsă ca o funcție continuă pe scara Bark (fig.6.5):

$$m_{dB}(z) = \begin{cases} 3 & z \leq z_l + 12, \\ 0,25 \cdot (z - z_l) & z > z_l + 12 \end{cases} \quad (6.32a)$$

Pragul de mascare, în unități de energie, este:

$$M[k, n] = \frac{E[k, n]}{10^{\frac{m_{dB}[k]}{10}}} = g_m[k] \cdot E[k, n] \quad (6.33)$$

Factorul de ponderare  $g_m[k]$  poate fi precalculat.

Mărimile obținute în această fază a prelucrării,  $M[k, n]$ , se numesc “mărimi de mascare”.

### 6.3. Preprocesarea mărimilor de excitație

Mărimile obținute în secțiunea anterioară sunt prelucrate mai departe și vom avea în vedere două semnale de intrare și ieșirile lor corespunzătoare – semnalul de referință și semnalul testat, notate cu indicii *Ref* și respectiv *Test*. Cazul semnalelor binaurale va fi discutat mai târziu în legătură cu variabilele de ieșire ale modelului. În această secțiune se vor considera cele două canale, stânga – dreapta, ca două semnale separate, de referință și, respectiv, test, la care indexul  $k$  reprezintă banda de frecvență și indexul  $n$  reprezintă numărul cadrului. Pentru modelul urechii bazat pe transformarea FFT, avem variabilele  $StepSize = 1.024$  și  $Z = 109$ . Pentru celelalte variabile, standardul BS.1387 afirmă: “toate variabilele și filtrele recursive sunt inițializate la zero, dacă nu se specifică altfel.” Condiția “dacă nu se specifică altfel” lasă de înțeles că ar putea exista unele operații de filtrare care nu sunt inițializate de la zero. În standard însă, nu există nici un exemplu de inițializare diferită de zero.

#### 6.3.1 Adaptarea nivelului și mărimilor

Intrările sunt mărimile de excitație (împrăștiate în frecvență și curățate în timp):  $E_{Ref}[k, n]$  și  $E_{Test}[k, n]$ . Acestea sunt funcții de frecvență și timp. Interpretarea următoarelor operații este dată în [TH100].

Pentru a compensa diferențele de nivel și distorsiunile liniare dintre semnalul testat și cel de referință, nivelurile medii ale celor două semnale sunt adaptate.

În primul pas, energiile din fiecare bandă de frecvențe sunt curățate cu un filtru trece jos de ordinul întâi. Constantele de timp depind de frecvențele centrale ale benzilor și se aleg astfel:

$$\tau[k] = \tau_{\min} + \frac{100 \text{ Hz}}{f_c[k]} \cdot (\tau_{100} - \tau_{\min}) \quad \left. \begin{array}{l} \tau_{100} = 0,050 \text{ s} \\ \tau_{\min} = 0,008 \text{ s} \end{array} \right\} \quad (6.34)$$

Împrăștierea în timp are loc independent pentru fiecare bandă de frecvențe și separat pentru semnalele de referință și test. Condițiile inițiale de filtrare sunt zero. Filtrele trece jos de ordinul întâi se calculează astfel:

$$P_{Ref}[k, n] = a[k] \cdot P_{Ref}[k, n-1] + (1 - a[k]) \cdot E_{Ref}[k, n] \quad (6.35)$$

$$P_{Test}[k, n] = a[k] \cdot P_{Test}[k, n-1] + (1 - a[k]) \cdot E_{Test}[k, n] \quad (6.36)$$

unde  $E_{Test}$  și  $E_{Ref}$  sunt mărimile de excitație care trebuie adaptate și  $a[k]$  se calculează cu ajutorul constantelor de timp:

$$a[k] = e^{-\frac{1024}{48000 \cdot \tau[k]}} \quad (6.37)$$

### 6.3.1.1 Adaptarea nivelului

Din mărimile de intrare filtrate pe benzile de frecvență  $P_{Test}$  și  $P_{Ref}$  se calculează un factor de corecție momentan  $LevCorr$ :

$$LevCorr[n] = \left( \frac{\sum_{k=0}^{Z-1} \sqrt{P_{Test}[k, n] \cdot P_{Ref}[k, n]}}{\sum_{k=0}^{Z-1} P_{Test}[k, n]} \right)^2 \quad (6.38)$$

De remarcat că numitorul este întotdeauna pozitiv pentru că s-a impus un prag energetic minim ( $E_{min} = 1 \times 10^{-12}$ ) și s-a adăugat un termen corespunzător zgomotului intern al urechii.

Dacă factorul de corecție este mai mare ca 1, semnalul de referință se divide cu factorul de corecție, iar dacă este mai mic ca 1 semnalul de test se înmulțește cu factorul de corecție.

$$E_{L,Ref}[k, n] = \begin{cases} E_{Ref}[k, n] / LevCorr[n] & LevCorr[n] > 1 \\ E_{Ref}[k, n] & LevCorr[n] \leq 1 \end{cases} \quad (6.39)$$

$$E_{L,Test}[k, n] = \begin{cases} E_{Test}[k, n] & LevCorr[n] > 1 \\ E_{Test}[k, n] \cdot LevCorr[n] & LevCorr[n] \leq 1 \end{cases} \quad (6.40)$$

### 6.3.1.2 Adaptarea modelului

Factorii de corecție pentru fiecare canal sunt calculați prin compararea înfășurătoarei temporale a ieșirilor pentru semnalele de test și referință:

$$R[k, n] = \frac{\sum_{i=0}^n a[k]^i \cdot E_{L,Test}[k, n-i] \cdot E_{L,Ref}[k, n-i]}{\sum_{i=0}^n a[k]^i \cdot E_{L,Ref}[k, n-i] \cdot E_{L,Ref}[k, n-i]} \quad (6.41)$$

Valorile lui  $a[k]$  sunt calculate cu ec. (6.37) și constantele de timp din ec. (6.34), cu aceleași condiții inițiale de zero. Dacă  $R[k, n]$  este mai mare ca 1, factorul de corecție pentru semnalul de test devine  $R[k, n]^{-1}$ , iar factorul de corecție pentru semnalul de referință devine 1. Altfel, factorul de corecție pentru semnalul de referință devine  $R[k, n]$  și factorul de corecție pentru semnalul de test devine 1.

$$R_{Test}[k, n] = \frac{1}{R[k, n]}, R_{Ref}[k, n] = 1 \quad |R[k, n] \geq 1 \quad (6.42)$$

$$R_{Test}[k, n] = 1, R_{Ref}[k, n] = R[k, n] \quad |R[k, n] < 1$$

- Dacă numitorul din relația (6.41) este 0 (și  $R[k, n]$  ar fi nedefinit) și numărătorul este mai mare ca 0,  $R_{Test}[k, n] = 0$  și  $R_{Ref}[k, n] = 1$ .

- Dacă numărătorul relației (6.41) este și el 0, rapoartele  $R_{Test}[k, n]$  și  $R_{Ref}[k, n]$  sunt copiate din banda de frecvențe anterioară ( $R_{Test}[k, n] = R_{Test}[k-1, n]$  și  $R_{Ref}[k, n] = R_{Ref}[k-1, n]$ ).

- Dacă nu există o bandă anterioară ( $k = 0$ ) atunci rapoartele  $R_{Test}[k, n]$  și  $R_{Ref}[k, n]$  sunt egale cu 1.

De fapt, condițiile de mai sus nu sunt necesare, pentru că termenii sunt întotdeauna pozitivi deoarece s-a impus un prag energetic minim ( $E_{min} = 1 \times 10^{-12}$ ) și s-a adăugat un termen corespunzător zgomotului intern al urechii.

Factorii de corecție sunt mediați pe  $M$  benzi și curățați în timp (ec. 6.43) folosind aceleași constante de timp ca mai sus (ec. (6.34) și (6.37)). Lățimea ferestrei de frecvență  $M$  este 8.

$$PattCorr_{Test}[k, n] = a \cdot PattCorr_{Test}[k, n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Test}[k+i, n] \quad (6.43)$$

$$PattCorr_{Ref}[k, n] = a \cdot PattCorr_{Ref}[k, n-1] + (1-a) \cdot \frac{1}{M} \cdot \sum_{i=-M_1}^{M_2} R_{Ref}[k+i, n]$$

unde

$$M_1 = M_2 = \frac{M-1}{2} \quad \text{pentru } M \text{ impar}$$

$$M_1 = \frac{M}{2} - 1, M_2 = \frac{M}{2} \quad \text{pentru } M \text{ par}$$

Intervalul nominal pe care se aplică aceste corecții este între  $k - M_1$  și  $k + M_2$ . La capetele de jos și de sus ale scalei de frecvență, unde fereastra ar depăși gama benzilor de frecvențe, lățimea ferestrei de frecvență se reduce corespunzător:

$$M_1[k] = \min(M_1, k), \quad M_2[k] = \min(M_2, z - k - 1), \quad M = M_1 + M_2 + 1 \quad (6.44)$$

Mărimile de intrare adaptate în nivel sunt ponderate cu factorii de corecție corespunzători  $PattCorr_{Test/Ref}[k, n]$  pentru a obține *mărimile spectrale adaptate*.

$$E_{P,Ref}[k, n] = E_{L,Ref}[k, n] \cdot PattCorr_{Ref}[k, n] \quad (6.45)$$

$$E_{P,Test}[k, n] = E_{L,Test}[k, n] \cdot PattCorr_{Test}[k, n] \quad (6.46)$$

Mărimile spectrale adaptate sunt mărimile finale de ieșire ale acestei etape de prelucrare.

### 6.3.2 Modulația

Cu ajutorul mărimilor de excitație “unsmeared” (împrăștiate în frecvență, dar nu și în timp),  $E_2[k, n]$ , care sunt mărimi de intrare pentru această etapă, se calculează mediile și diferențele medii într-un domeniu aproximativ de tărie (prin ridicarea excitației la puterea 0,3). Tăria medie este:

$$\begin{aligned}\bar{E}_{\text{Ref}}[k, n] &= a[k] \cdot \bar{E}_{\text{Ref}}[k, n-1] + (1-a[k]) \cdot E_{2\text{Ref}}[k, n]^{0,3}, \\ \bar{E}_{\text{Test}}[k, n] &= a[k] \cdot \bar{E}_{\text{Test}}[k, n-1] + (1-a[k]) \cdot E_{2\text{Test}}[k, n]^{0,3}.\end{aligned}\quad (6.47)$$

Diferențele medii sunt:

$$\begin{aligned}\bar{E}_{\text{derRef}}[k, n] &= a[k] \cdot \bar{E}_{\text{derRef}}[k, n-1] + (1-a[k]) \cdot \frac{48000}{1024} \cdot |E_{2\text{Ref}}[k, n]^{0,3} - E_{2\text{Ref}}[k, n-1]^{0,3}|, \\ \bar{E}_{\text{derTest}}[k, n] &= a[k] \cdot \bar{E}_{\text{derTest}}[k, n-1] + (1-a[k]) \cdot \frac{48000}{1024} \cdot |E_{2\text{Test}}[k, n]^{0,3} - E_{2\text{Test}}[k, n-1]^{0,3}|.\end{aligned}\quad (6.48)$$

Valorile pentru  $a[k]$  sunt calculate cu relația (6.37), constantele de timp cu relația (6.34) și aceleași condiții de zero inițial.

Această valoare și valoarea absolută a derivatei sale temporale sunt întinse în timp.

Cu valorile de tărie rezultate,  $\bar{E}_{\text{der}}$  și  $\bar{E}$ , se calculează o măsură pentru modulația înfășurătoarei pentru ieșirea fiecărui filtru:

$$\begin{aligned}Mod_{\text{Ref}}[k, n] &= \frac{\bar{E}_{\text{derRef}}[k, n]}{1 + \bar{E}_{\text{Ref}}[k, n]/0,3} \\ Mod_{\text{Test}}[k, n] &= \frac{\bar{E}_{\text{derTest}}[k, n]}{1 + \bar{E}_{\text{Test}}[k, n]/0,3}\end{aligned}\quad (6.49)$$

Acești parametri de modulație, precum și valoarea medie a tăriei semnalului de referință ( $\bar{E}_{\text{Ref}}[k, n]$ ), vor fi folosite mai departe la calculul diferenței de modulație.

### 6.3.3 Tăria

Mărimile tăriei specifice semnalului testat și cel de referință sunt folosite mai târziu pentru a alege cadrele ce vor fi incluse în variabilele de ieșire referitoare la tăria zgomotului, și se calculează cu relația:

$$\begin{aligned}N_{\text{Ref}}[k, n] &= const \cdot \left( \frac{1}{s[k]} \cdot \frac{E_{\text{Thres}}[k]}{E_0} \right)^{0,23} \cdot \left[ \left( 1 - s[k] + \frac{s[k] \cdot E_{\text{Ref}}[k, n]}{E_{\text{Thres}}[k]} \right)^{0,23} - 1 \right] \\ N_{\text{Test}}[k, n] &= const \cdot \left( \frac{1}{s[k]} \cdot \frac{E_{\text{Thres}}[k]}{E_0} \right)^{0,23} \cdot \left[ \left( 1 - s[k] + \frac{s[k] \cdot E_{\text{Test}}[k, n]}{E_{\text{Thres}}[k]} \right)^{0,23} - 1 \right]\end{aligned}\quad (6.50)$$

dată de Zwicker și Feldkeller în 1967.

Unitatea de măsură a tăriei este son-ul, definit ca tăria obținută cu un sunet sinusoidal cu frecvența de 1 kHz și un nivel de 40 dB<sub>SPL</sub> (§2.2.7). De aceea în relația de mai sus  $E_0 = 10^4$  (40 dB față de 0 dB<sub>SPL</sub>), iar  $const = 1,07664$ .

Indexul pragului  $s$  și excitația la prag  $E_{\text{Thres}}$  se calculează cu relațiile:

$$s[k] = 10^{\frac{1}{10} \left( -2 - 2,05 \cdot \text{am} \left( \frac{f[k]}{4 \text{ kHz}} \right) - 0,75 \cdot \text{am} \left( \left( \frac{f[k]}{1.600 \text{ Hz}} \right)^2 \right) \right)}\quad (6.51)$$

$$E_{Thres}[k] = 10^{0,364 \left(\frac{f}{1 \text{ kHz}}\right)^{-0,8}} \quad (6.52)$$

Indexul de prag este raportul dintre intensitatea unui ton de test abia audibil și intensitatea zgomotului intern în interiorul benzii critice. Forma folosită este atribuită lui Kapust [THI99]. Pragul de excitare reprezintă partea de joasă frecvență de la filtrarea urechii externe și mijlocii și termenii reprezentând zgomotul intern care au apărut anterior.

Tăria totală a semnalului testat și a celui de referință se calculează ca suma tăriilor specifice pentru toate benzile de frecvență care au valori mai mari decât 0.

$$\begin{aligned} N_{totalRef}[n] &= \frac{24}{109} \cdot \sum_{k=0}^{108} \max(N_{Ref}[k, n], 0) \\ N_{totalTest}[n] &= \frac{24}{109} \cdot \sum_{k=0}^{108} \max(N_{Test}[k, n], 0) \end{aligned} \quad (6.53)$$

Factorul 24 reprezintă numărul total de barks dintr-un semnal audio, iar 24/109 reprezintă lățimea fiecărei benzi de frecvență.

NOTA – Datorită modelelor diferite ale urechii periferice, tăria calculată aici nu este identică cu cea definită în standardul ISO 532 (Acustică - Metodă de calcul nivelului de tărie 1975). Un sunet sinusoidal cu frecvența de 1 kHz și nivelul de 40 dB SPL va avea tăria de 0,584 son. Amplitudinea sunetului sinusoidal de 1 kHz a fost stabilită la  $A_{max} 10^{Lp/20} / 10^{40/20}$ , unde  $Lp$  este nivelul de presiune sonoră de calibrare (92 dB) și  $A_{max}$  este amplitudinea maximă a sinusoidei de calibrare. Crescând energia sinusoidei de 10 ori (amplitudinea  $\times \sqrt{10}$ ) ar trebui să obținem o creștere a tăriei de 1 son. În realitate tăria crește numai cu 0,770 son. Aceasta se datorează faptului că exponentul (0,23) a fost ales pentru a uniformiza zgomotul (zgomot cu aceeași energie în fiecare bandă critică).

### 6.3.4 Calculul semnalului de eroare

Semnalul de eroare se calculează în domeniul frecvență luând în considerare diferența dintre mărimile filtrate ale spectrului urechii externe și mijlocii ale semnalului de referință și testat (vezi §6.2.4).

$$F_{noise}[k_f, n] = \|F_{eRef}[k_f, n] - F_{eTest}[k_f, n]\| \quad (6.54)$$

$F_{noise}$  este transferat în domeniul înălțime folosind algoritmul descris în §6.2.5.

Mărimea de ieșire obținută cu acest algoritim,  $P_{noise}[n, k]$ , se numește “*Noise pattern*”.

## 6.4. Calculul variabilelor de ieșire ale modelului

### 6.4.1 Generalități

Ieșirile obținute din pașii anteriori sunt în general funcții de timp și frecvență ale semnalului de referință și de test. În continuare, funcțiile sunt transformate în funcții de timp și apoi mediate pentru a obține o singură valoare - variabila de ieșire a modelului (MOV).

Metoda folosește 11 variabile de ieșire ale modelului prezentate în tabelul 6.2.

Pentru semnalele binaurale, majoritatea variabilelor se calculează separat pentru fiecare canal și apoi valorile corespunzătoare sunt mediate. În cazul variabilelor *MFPD* și *ADB*, canalele sunt combinate frecvență cu frecvență înainte de medierea în timp.

Figura 6.6 prezintă intrările și ieșirile din secțiunea anterioară, care sunt folosite în continuare în această secțiune.

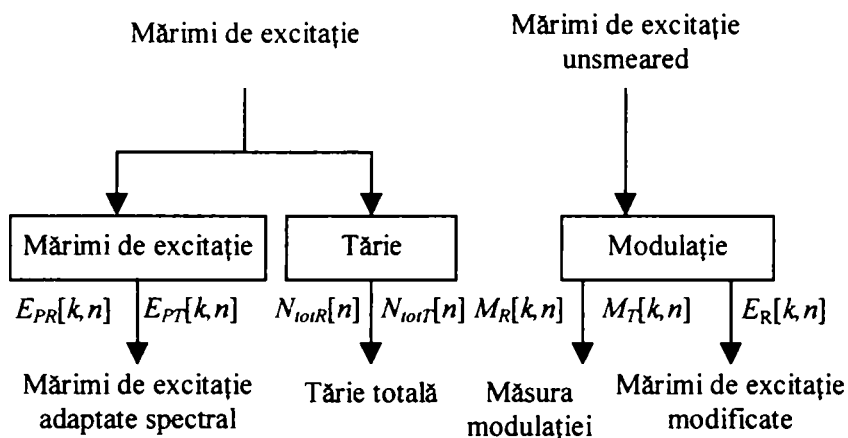


Figura 6.6 Intrările și ieșirile din secțiunea de prelucrare

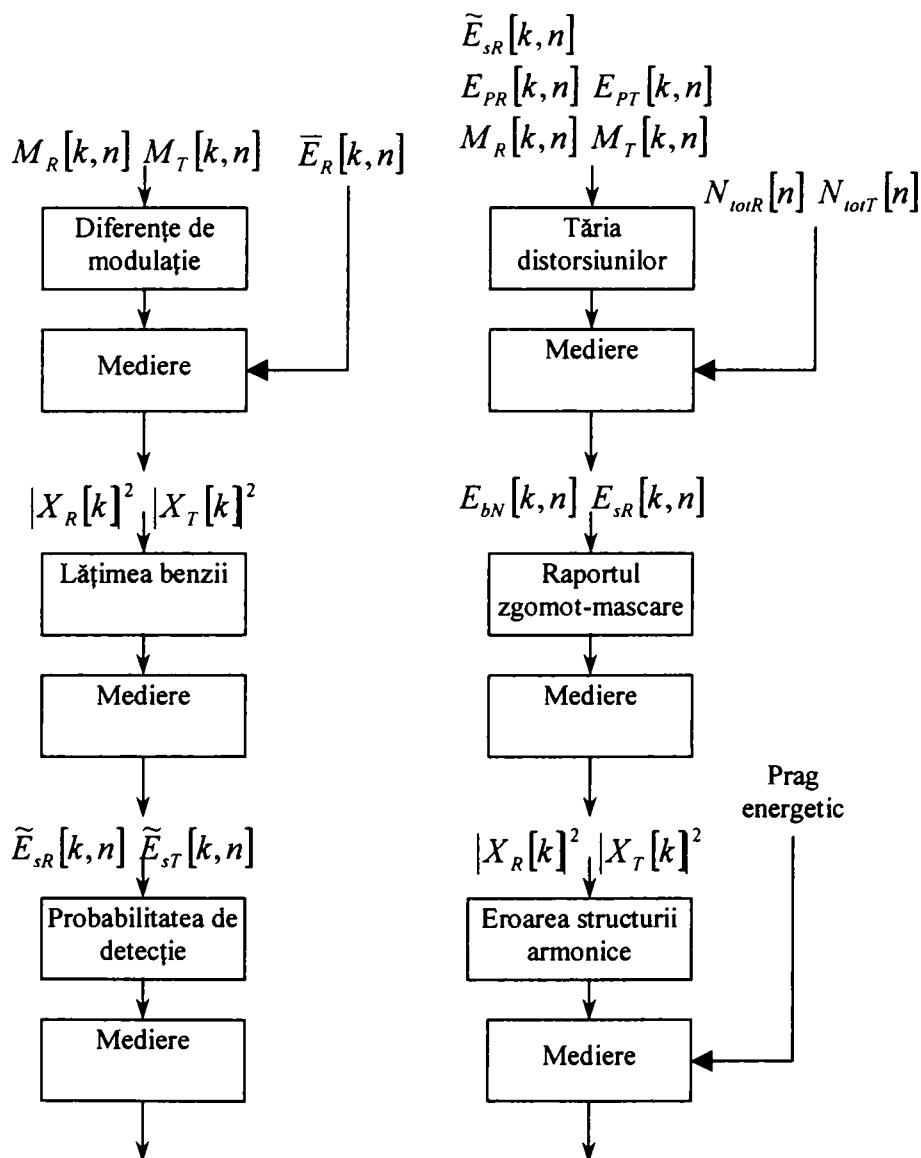


Figura 6.7 Intrările pentru calculul variabilelor de ieșire ale modelului

Variabilele de ieșire ale modelului folosesc mărimile de ieșire ale secțiunii anterioare și câteodată și alte semnale. Figura 6.7 prezintă relațiile dintre ele.



Parametrii ce vor fi mediați în timp pentru a deveni variabile de ieșire ale modelului sunt definiți într-un anumit domeniu, valabil pentru toate variabilele.

**Tabel 6.2** Variabilele de ieșire ale modelului

Variabila de ieșire a modelului	Model	Descriere
<i>BandwidthRef</i>	FFT	Lățimea benzii semnalului de referință
<i>BandwidthTest</i>	FFT	Lățimea benzii semnalului de test
<i>TotalNMR</i>	FFT	Raportul zgomot-mascare
<i>WinModDiff1</i>	FFT	Diferența de modulație în fereastră
<i>ADB</i>	FFT	Distorsiunea medie în bloc
<i>EHS</i>	FFT	Structura armonică a erorii
<i>AvgModDiff1</i>	FFT	Diferența medie de modulație
<i>AvgModDiff2</i>	FFT	Diferența medie de modulație
<i>RmsNoiseLoud</i>	FFT	Tăria distorsiunii
<i>MFPD</i>	FFT	Probabilitatea maximă de detecție filtrată

Cadrelle de nivel mic de la începutul și sfârșitul secvenței de intrare sunt identificate. Se definește începutul și sfârșitul datelor ca prima locație, de la începutul sau de la sfârșitul fișierului, în care suma valorilor absolute pe cinci eșantioane succesive depășește 200, într-unul din canalele audio corespunzătoare. Cadrele care sunt complet în afara acestui domeniu vor fi ignorate.

Se presupune că valoarea pragului este valabilă pentru date pe 16 biți. Pragul pentru un semnal cu amplitudinea maximă  $A_{\max}$  este:

$$A_{thr} = 200 \frac{A_{\max}}{32768} \quad (6.55)$$

Implementarea cere o fereastră alunecătoare de  $L = 5$  eșantioane. Testul cere de fapt ca mărimea medie a eșantioanelor dintr-o fereastră să fie mare de  $A_{thr}/L$  (adică mai mare de 40 pentru un semnal pe 16 biți).

În concluzie, limita de start a datelor corespunde cu primul eșantion din primul grup de 5 eșantioane care satisfac criteriul. Similar pentru sfârșitul fișierului, limita de sfârșit a datelor corespunde cu ultimul eșantion din ultimul grup de 5 eșantioane care satisfac criteriul. Din punct de vedere al cadrelor, cadrul de start este primul cadru care conține eșantionul de start al datelor, iar cadrul de sfârșit este ultimul cadru care conține eșantionul de sfârșit al datelor.

#### 6.4.2 Diferența de modulație

Înfășurătoarele temporale pentru fiecare bandă de frecvență sunt combinate în mai multe variabile de ieșire ale modelului. Diferențele de modulație a înfășurătoarelor temporale a semnalelor de referință și de test sunt măsurate prin calculul diferenței locale de modulație pentru fiecare canal (ec. (6.56)), unde  $Mod_{Test}$  și  $Mod_{Ref}$  se obțin prin aplicarea relației (6.49) semnalului de referință  $R_{Test}$ .

Diferența instantanee de modulație este:

$$ModDiff[k, n] = w \cdot \frac{|Mod_{Test}[k, n] - Mod_{Ref}[k, n]|}{offset + Mod_{Ref}[k, n]} \quad (6.56)$$

$$\begin{cases} w = 1,0 & |Mod_{Test}[k, n] > Mod_{Ref}[k, n] \\ w = negWt & |Mod_{Test}[k, n] < Mod_{Ref}[k, n] \end{cases}$$

Diferența momentană de modulație se calculează ca o medie a diferențelor locale de modulație pentru toate canalele (ec. (6.57)).

$$ModDiff[n] = \frac{100}{109} \sum_{k=0}^{108} ModDiff[k, n] \quad (6.57)$$

Pragul de liniște este luat în considerare prin intermediul unui factor de ponderare (ec. (6.58)) calculat din mărimile de excitație modificate pentru semnalul de referință (ec. (6.48)) și funcția de zgomot intern definită cu relația (6.13).

$$TempWt[n] = \sum_{k=0}^{108} \frac{\bar{E}_{Ref}[k, n]}{\bar{E}_{Ref}[k, n] + levWt \cdot E_{Thres}[k]^{0,3}} \quad (6.58)$$

Medierea temporală a diferențelor momentane de modulație  $ModDiff[n]$  folosind factorii de ponderare  $TempWt[n]$  este descrisă în §6.5.2. Valorile constantelor  $negWt$ ,  $offset$  și  $levWt$  sunt date în tabelul 6.3.

**Tabel 6.3** Variabilele de ieșire ale modelului care estimează diferența de modulație

MOV ( $X_{xx} = Win/Avg/Rms$ )	negWt	offset	levWt
$X_{xx}ModDiff1$	1	1	100
$X_{xx}ModDiff2$	0,1	0,01	100

#### 6.4.2.1 WinModDiff1

Variabila de ieșire a modelului  $WinModDiff1$  reprezintă media diferenței de modulație calculată din modelul urechii bazat pe transformarea FFT. Pentru medierea temporală vezi §5.2.3 și tabelul 6.3 pentru constante. Factorul de ponderare temporală din ec. (6.58) nu se aplică pentru această variabilă.

Variabila finală se obține prin medierea într-o fereastră alunecătoare cu  $L = 4$  (85 ms).

$$WinModDiff = \sqrt{\frac{1}{N - L + 1} \cdot \sum_{n=L-1}^{N-1} \left( \frac{1}{L} \cdot \sum_{i=0}^{L-1} \sqrt{ModDiff[n-i]} \right)^4} \quad (6.59)$$

#### 6.4.2.2 AvgModDiff1 și Avg ModDiff2

Variabilele  $AvgModDiff1$  și  $Avg ModDiff2$  sunt media liniară a diferenței de modulație calculată din modelul bazat pe FFT. Diferența dintre  $Avg ModDiff2$  și  $AvgModDiff1$  este

alegerea diferită a constantelor. Pentru medierea temporală vezi §5.2.1 și tabelul 6.1 pentru constante.

Pentru  $AvgModDiff1$  diferența modulațiilor instantanee este dată de rel. (6.56), media pe benzile de frecvență este dată de rel. (6.57), iar forma finală a variabilei este dată printr-o mediere temporală:

$$AvgModDiff1 = \frac{\sum_{n=0}^{108} TempWt[n] \cdot ModDiff1[n]}{\sum_{n=0}^{108} TempWt[n]} \quad (6.60)$$

Pentru  $AvgModDiff2$  diferența modulațiilor instantanee este dată de:

$$ModDiff2[k, n] = \begin{cases} \frac{Mod_{Test}[k, n] - Mod_{Ref}[k, n]}{0,01 + Mod_{Ref}[k, n]} & Mod_{Test}[k, n] \geq Mod_{Ref}[k, n] \\ 0,1 \cdot \frac{Mod_{Ref}[k, n] - Mod_{Test}[k, n]}{0,01 + Mod_{Ref}[k, n]} & Mod_{Test}[k, n] < Mod_{Ref}[k, n] \end{cases} \quad (6.61)$$

Media pe benzile de frecvență este:

$$ModDiff2[n] = \frac{100}{109} \sum_{k=0}^{108} ModDiff2[k, n] \quad (6.62)$$

Forma finală a variabilei este dată printr-o mediere temporală:

$$AvgModDiff2 = \frac{\sum_{n=0}^{108} TempWt[n] \cdot ModDiff2[n]}{\sum_{n=0}^{108} TempWt[n]} \quad (6.63)$$

### 6.4.3 Tăria zgomotului

Aceste variabile estimează tăria parțială a distorsiunilor adăugate în prezența semnalului de referință de mascare. Formula tăriei parțiale (ec. (6.63)) a fost proiectată pentru a da tăria specifică a zgomotului, conform Zwicker și Feldkeller, 1967, dacă nu este prezent nici un sunet de mascare și ceva asemănător raportului dintre zgomot și semnalul de mascare, dacă zgomotul este foarte mic în comparație cu semnalul de mascare.

Tăria parțială a zgomotului se calculează cu [THI99]:

$$NL[k, n] = \left( \frac{E_{Thres}[k]}{s_{Test}[k, n] E_0} \right)^{0,23} \cdot \left[ \left( 1 + \frac{\max(s_{Test}[k, n] \cdot E_{Test}[k, n] - s_{Ref}[k, n] \cdot E_{Ref}[k, n], 0)}{E_{Thres}[k] + s_{Ref}[k, n] \cdot E_{Ref}[k, n] \cdot \beta[k, n]} \right)^{0,23} - 1 \right] \quad (6.64)$$

unde  $E_0 = 1$ ,  $E_{Thres}[k]$  este funcția de zgomot intern definită cu (6.36) și  $s[k, n]$  se calculează cu:

$$\begin{aligned} s_{\text{Ref}}[k, n] &= \text{ThresFac}_0 \cdot \text{Mod}_{\text{Ref}}[k, n] + S_0 \\ s_{\text{Test}}[k, n] &= \text{ThresFac}_0 \cdot \text{Mod}_{\text{Test}}[k, n] + S_0 \end{aligned} \quad (6.65)$$

Dacă nu se specifică altfel, *mărimile de excitație adaptate spectral* (vezi §6.3.1) se folosesc ca intrări:  $E_{\text{Test}} = E_{p,\text{Test}}[k, n]$  și  $E_{\text{Ref}} = E_{p,\text{Ref}}[k, n]$ . Coeficientul  $\beta$ , care determină cantitatea de mascare, se calculează cu:

$$\beta[k, n] = \exp\left(-\alpha \cdot \frac{E_{\text{Test}}[k, n] - E_{\text{Ref}}[k, n]}{E_{\text{Ref}}[k, n]}\right) \quad (6.66)$$

Valorile parametrilor  $\alpha$ ,  $\text{ThresFac}_0$  și  $S_0$  depind de variabila pentru care se calculează.

Pentru calculul acestei variabile se folosește medierea cu întârziere (vezi §6.5.2.4.1). Valorile momentane ale tăriei zgomotului nu sunt luate în considerare decât la 50 ms după ce tăria totală pe un canal (stâng sau drept) a depășit o dată valoarea de  $N_{\text{Thres}} = 0,1$  son atât pentru semnalul testat cât și pentru cel de referință (vezi §6.5.2.4.2). În concluzie, întârzierea totală este dată de întârzierea de 0,5 s folosită la mediera întârziată și întârzierea specificată de pragul de tărie.

Condiția de test se poate scrie astfel (6.66a):

$$\begin{aligned} (N_{\text{totTest}}[n] \geq N_{\text{thres}}) \wedge (N_{\text{totRef}}[n] \geq N_{\text{thres}}) & \quad | \text{monaural} \\ ((N_{\text{totTest1}}[n] \geq N_{\text{thres}}) \wedge (N_{\text{totRef1}}[n] \geq N_{\text{thres}})) \vee ((N_{\text{totTest2}}[n] \geq N_{\text{thres}}) \wedge (N_{\text{totRef2}}[n] \geq N_{\text{thres}})) & \quad | \text{binaural} \end{aligned} \quad (6.66.a)$$

unde  $N_{\text{thres}} = 0,1$  son, iar indicii numerici indică numărul canalului. Offsetul cadrelor corespunzător unei întârzieri de 50 ms este:

$$N_{\text{off}} = \lceil \tau_{\text{off}} F_{ss} \rceil \quad (6.66b)$$

unde  $\tau_{\text{off}} = 50$  ms și  $\lceil \cdot \rceil$  indică o funcție limită, rezultând o întârziere de 3 cadre.

În medierea spectrală, valorile momentane sunt normalizate cu numărul filtrelor per bandă critică în loc de numărul total al filtrelor, de exemplu rezultatul medierii spectrale se multiplică cu 24.

Dacă tăria momentană a zgomotului se află sub prag,  $NL_{\text{min}}$  se face zero.

**Tabel 6.4** Variabilele de ieșire ale modelului care estimează tăria totală a zgomotului

MOV (Xxx=Win/Avg/Rms)	$\alpha$	$\text{ThresFac}_0$	$S_0$	$NL_{\text{min}}$
<i>XxxMissingComponents</i>	1,5	0,15	1	0
<i>XxxNoiseLoud</i>	1,5	0,15	0,5	0

### RmsNoiseLoud

Variabila de ieșire a modelului *RmsNoiseLoud* reprezintă media pătratică a tăriei zgomotului calculată din modelul FFT al urechii. Vezi §6.5.2.2 pentru medierea temporală și tabelul 6.4 pentru constante.

Medierea spectrală se face în felul următor:

$$\tilde{N}_L[n] = \frac{24}{109} \sum_{k=0}^{108} N_L[k, n] \quad (6.67)$$

Dacă valoarea momentană a tăriei este mai mică decât zero, atunci:

$$\tilde{N}_L[n] = \begin{cases} \tilde{N}_L[n] & \tilde{N}_L[n] \geq 0 \\ 0 & \tilde{N}_L[n] < 0 \end{cases} \quad (6.68)$$

Valoarea finală a variabilei este dată de medierea temporală a pătratului tăriei zgomotului:

$$RmsNoiseLoud = \sqrt{\frac{1}{109} \sum_{n=0}^{108} (\tilde{N}_L[n])^2} \quad (6.69)$$

#### 6.4.4 Lățimea benzii

Lățimea benzii se estimează pentru semnalele de referință și de test. Operațiile care se efectuează, pe ieșirile DFT în dB, sunt următoarele:

- Semnalul de test – se caută componenta cea mai mare peste 21,6 kHz și se numește nivel de prag;
- Semnalul de referință – se caută de la 21,6 kHz în jos prima valoare care depășește nivelul de prag cu 10 dB. Se înregistrează această frecvență ca lățimea de bandă a semnalului de referință;
- Semnalul de test – pornind de la frecvență ce reprezintă lățimea de bandă a semnalului de referință, se caută în jos prima valoare care depășește nivelul de prag cu 5 dB. Se înregistrează această frecvență ca lățimea de bandă a semnalului de test.

Dacă frecvența găsită pentru semnalul de referință nu este peste 8,1 kHz, lățimea de bandă pentru acest cadru nu se ia în considerare.

Aceste valori de ieșire ale modelului estimează lățimea de bandă medie a semnalului testat și a semnalului de referință în liniile FFT.

Pentru fiecare cadru se calculează lățimea de bandă locală  $Bw_{Ref}[n]$  și  $Bw_{Test}[n]$  cu secvența următoare (MATCAD):

Intrări

Nivelul ieșirilor FFT în dB

$$FLevelTest_{kf, n} := 10 \cdot \log(FFT_{kf, n}) \quad FLevelRef_{kf, n} := 10 \cdot \log(FFT_{kf, n})$$

Ieșiri  $Bw_{Ref}$ ,  $Bw_{Test}$

$$kx_{kk} := \text{round}\left(f_{c_{kk}} \cdot \frac{2048}{48000}\right)$$

$$kx1 := \text{round}\left(8109 \cdot \frac{2048}{48000}\right) \quad kx1 = 346 \quad kx2 := \text{round}\left(21586 \cdot \frac{2048}{48000}\right) \quad kx2 = 921$$

Pragul lățimii de bandă

$$ZeroThreshold := FLevelTest(921)$$

$$Bw_{Ref} := 0 \quad Bw_{Test} := 0$$

$$\text{ZeroThreshold}_n := \begin{cases} \text{for } k \in 921..1023 \\ \text{ZeroThreshold} \leftarrow \max(\text{ZeroThreshold}, \text{FLevelTest}_{kf,n}) \\ \text{ZeroThreshold} \end{cases}$$

### ***BandwidthRef* și *BandwidthTest***

Variabila *BandwidthRef* este media liniară a valorilor instantanee a lățimii de bandă a semnalului de referință pentru cadrul  $k$ ,  $BwRef[k]$ :

$$\text{BandwidthRef} = \frac{1}{109} \sum_{n=0}^{108} BwRef[n] \quad (6.70)$$

unde se însumează numai cadrele pentru care lățimea de bandă a semnalului de referință depășește 346 (8,1 kHz).

Variabila *BandwidthTest* este media liniară a valorilor instantanee a lățimii de bandă a semnalului de referință pentru cadrul  $k$ ,  $BwTest[k]$ :

$$\text{BandwidthTest} = \frac{1}{109} \sum_{n=0}^{108} BwTest[n] \quad (6.71)$$

unde se însumează numai cadrele pentru care lățimea de bandă a semnalului de referință depășește 346 (8,1 kHz).

Cadrele cu energie redusă de la începutul și sfârșitul fragmentului sunt ignorate (vezi §6.5.2.4.4). Vezi §6.5.2.1 pentru medierea temporală.

### **6.4.5 Raportul Zgomot-Mascare**

Valorile următoare ale modelului sunt calculate din valorile zgomotului și mascării.

Raportul Zgomot-Mascare (NMR) este raportul dintre zgomot și pragul de mascare. În acest caz zgomotul este diferența dintre mărimea semnalului de referință și de test (6.48). Mărimile de zgomot (Noise Patterns) din benzi sunt notate cu  $P_{noise}[n,k]$  ( $k$  fiind indexul benzii). Raportul Zgomot-Mascare în banda  $k$  este:

$$\text{NMR}[k,n] = \frac{P_{noise}[k,n]}{M[k,n]} \quad (6.72)$$

Valoarea locală *NMR* a cadrului curent  $n$  se definește astfel:

$$\text{NMR}_{local}[n] = 10 * \log_{10} \frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{noise}[k,n]}{M[k,n]} \quad (6.73)$$

Variabila de ieșire a modelului *Total NMR* reprezintă media liniară, în dB, a mediei raportului zgomot –mascare dintr-un cadru:

$$NMR_{\text{tot}} = 10 * \log_{10} \frac{1}{N} \sum_n \left( \frac{1}{Z} \sum_{k=0}^{Z-1} \frac{P_{\text{noise}}[k, n]}{M[k, n]} \right) \quad (6.74)$$

Cadrele cu energie scăzută de la începutul și sfârșitul fragmentului sunt ignorate (vezi §6.5.2.4.4).

#### 6.4.6 Relative Disturbed Frames

Variabila de ieșire a modelului *Relative Disturbed Frames* (prescurtat: *RelDistFrames*) reprezintă numărul de cadre cu:

$$RelDistFrames[n] = \max_{\forall k} \left( 10 \cdot \log \left( \frac{P_{\text{noise}}[k, n]}{M[k, n]} \right) \right) \geq 1.5 \text{ dB} \quad k \in [0, 108] \quad (6.75)$$

relativ la numărul total de cadre ale fragmentului. Se consideră distorsionat un cadru în care NMR depășește 1,5 dB. Variabila finală măsoară numărul de cadre pentru care NMR depășește 1,5 dB.

Cadrele cu energie scăzută de la începutul și sfârșitul fragmentului sunt ignorate (vezi §6.5.2.4.4).

#### 6.4.7 Probabilitatea detecției

Variabilele definite în această secțiune se bazează pe  $\tilde{E}[k, n]$  ( $k$  benzi,  $n$  cadre), care sunt mărimile de excitație  $E[k, n]$  exprimate în dB:

$$\begin{aligned} \tilde{E}_{\text{RefdB}}[k, n] &= 10 \cdot \log_{10}(E_{\text{Ref}}[k, n]) \\ \tilde{E}_{\text{TestdB}}[k, n] &= 10 \cdot \log_{10}(E_{\text{Test}}[k, n]) \end{aligned} \quad (6.76)$$

Pentru fiecare cadru  $n$ :

Următorii pași se fac independent pentru fiecare canal  $c$  (valorile lui  $c$  stânga și dreapta). Mărimile logaritmice de excitație sunt  $\tilde{E}_{\text{Ref}}[k, n]$  pentru semnalul de referință și  $\tilde{E}_{\text{Test}}[k, n]$  pentru semnalul testat.

Pentru fiecare bandă  $k$ :

- se calculează excitația medie asimetrică:

$$L[k, n] = \begin{cases} 0,3 \cdot \tilde{E}_{\text{RefdB}}[k, n] + 0,7 \cdot \tilde{E}_{\text{TestdB}}[k, n] & |\tilde{E}_{\text{Ref}}[k, n] > \tilde{E}_{\text{Test}}[k, n] \\ \tilde{E}_{\text{TestdB}}[k, n] & |\tilde{E}_{\text{Ref}}[k, n] \leq \tilde{E}_{\text{Test}}[k, n] \end{cases} \quad (6.77)$$

- se calculează mărimea efectivă a pasului de detecție  $s$ . Formula următoare este o aproximație a pragului de detecție a diferenței de nivel, măsurat de Zwicker și Fastl, 1990:

Dacă  $L[k, n] > 0$ :

$$s[k, n] = 5,95072 \cdot ((6,39468) / L[k, n])^{1,71332} + 9,01033 \times 10^{-11} \cdot L[k, n]^4 + 5,05622 \times 10^{-6} \cdot L[k, n]^3 - 0,00102438 \cdot L[k, n]^2 + 0,0550197 \cdot L[k, n] - 0,198719$$

$$\text{și dacă } L[k, n] \leq 0 \text{ atunci } s[k, n] = 1,0 \times 10^{-30} \quad (6.78)$$

- se calculează eroarea cu semn  $e$ :

$$e[k, n] = \tilde{E}_{\text{refdB}}[k, n] - \tilde{E}_{\text{TestdB}}[k, n] \quad (6.79)$$

- se calculează factorul de scală  $a$ :

$$a[k, n] = \frac{10^{\frac{\log_{10}(\log_{10}(2.0))}{b}}}{s[k, n]} \quad (6.80)$$

unde panta  $b$  este:

$$b = \begin{cases} 4 & |\tilde{E}_{\text{Ref}}[k, n]| > \tilde{E}_{\text{Test}}[k, n] \\ 6 & |\tilde{E}_{\text{Ref}}[k, n]| \leq \tilde{E}_{\text{Test}}[k, n] \end{cases} \quad (6.81)$$

Diferența de pantă penalizează mai mult adăugarea de componente audibile decât dispariția lor.

- se calculează probabilitatea detecției. Ecuația (6.80) setează factorul de scală  $a$  astfel încât dacă diferența de excitație  $e[k, n]$  este egală cu pragul de detecție  $s[k, n]$ , probabilitatea de detecție  $p_c[k, n]$  devine 0,5:

$$p_c[k, n] = 1 - 10^{(-a[k, n]e[k, n])^b} \quad (6.82)$$

- se calculează numărul total de pași peste prag:

$$q_c[k, n] = \frac{|INT(e[k, n])|}{s[k, n]} \quad (6.83)$$

Probabilitatea de detecție și numărul de pași peste prag se calculează pentru fiecare canal, în cazul unui semnal multicanal. Pentru fiecare frecvență și timp, probabilitatea totală de detecție și numărul total de pași peste prag se calculează cu ajutorul celei mari valori de pe canale:

- probabilitatea binaurală de detecție este:

$$p_{bin}[k, n] = \max(p_{st}[k, n], p_{dr}[k, n]) \quad (6.84)$$

- numărul de pași peste prag pentru canalul binaural este:

$$q_{bin}[k, n] = \max(q_{st}[k, n], q_{dr}[k, n]) \quad (6.85)$$

Probabilitatea de detecție totală a canalului  $c$  al cadrului  $n$  este:

$$P_c[n] = 1 - \prod_{\forall k} (1 - p_c[k, n]) \quad (6.86)$$

unde  $c$  poate fi *stânga*, *dreapta* sau *bin*. Numărul total de pași peste prag pentru canalul  $c$  al cadrului  $n$  este:



$$Q_c[n] = \sum_{\forall k} q_c[k, n] \quad (6.87)$$

#### 6.4.7.1 Probabilitatea maximă filtrată de detecție MFPD (*Maximum Filtered Probability of Detection*)

Se calculează o versiune mai fină a probabilității de detecție pentru fiecare canal  $c$ :

$$\tilde{P}_c[n] = (1 - c_0) \cdot P_c[n] + c_0 \cdot \tilde{P}_c[n - 1] \quad (6.88)$$

unde  $P_c[-1] = 0$ . Constanta  $c_0$  depinde de *StepSize*:

$$c_0 = 0,9^{\text{StepSize} / 1024} \quad (6.89)$$

$c_0$  reduce sensibilitatea la distorsiuni foarte scurte și corespunde unei constante de timp de 0,202 s.

Probabilitatea maximă filtrată de detecție (*MFPD*) se calculează astfel:

$$PM_c[n] = \max(PM_c[n - 1] \cdot c_1, \tilde{P}_c[n]) \quad (6.90)$$

unde  $PM_c[-1]$  este zero. Constanta  $c_1$  depinde de *StepSize*:

$$c_1 = 0,99^{\text{StepSize} / 1024} \quad (6.91)$$

$c_1$  modelează efectul prin care distorsiunile la începutul fragmentului audio sunt mai puțin severe decât la finalul fragmentului. De remarcat că această constantă este utilă pentru modelarea testelor de ascultare critică în care subiecților nu le este permis să selecteze secvențe mai scurte ale fragmentului. Pentru acest model, care se calibrează folosind date din testele de ascultare conform recomandării ITU-R BS.1116,  $c_1$  poate fi 1.0.

Variabila *MFPD* are valoarea lui  $PM_{bin}[n]$  pentru ultimul cadru:

$$MFPD = PM_{bin}[N - 1] \quad (6.92)$$

#### 6.4.7.2 Blocul mediu distorsionat ADB (*Average Distorted Block*)

Numărul cadrelor valabile cu probabilitatea de detecție a canalului binaural  $P_{bin}[n]$  peste 0,5 este contorizat ( $n_{distorted}$ ).

Pentru toate cadrele valabile se calculează numărul total de pași peste prag ai canalului binaural  $Q_{bin}[n]$ :

$$Q_{sum} = \sum_{\forall n} Q_{bin}[n] \quad (6.93)$$

Se calculează distorsiunea blocului mediu distorsionat ADB (în acest context termenul “bloc” este echivalent cu “cadru”):

- dacă  $n_{distorted} = 0$  atunci  $ADB = 0$  (distorsiunile nu sunt audibile);
- dacă  $n_{distorted} > 0$  și  $Q_{sum} > 0$  atunci  $ADB = \log_{10}(Q_{sum} / n_{distorted})$ ;
- dacă  $n_{distorted} > 0$  și  $Q_{sum} = 0$  atunci  $ADB = -0,5$

### 6.4.8 Structura armonică a erorii

Un semnal de referință conținând armonici puternice (de exemplu clarinet bas, harpsicord) are un spectru caracterizat de un număr de vârfuri separate prin văi adânci la distanțe regulate. În anumite condiții, semnalul de eroare poate moșteni această structură. De exemplu, zgomotul amestecat cu un astfel de semnal poate rămâne nemascat în văile spectrale ale semnalului. Spectrul de eroare rezultat va avea o structură similară cu spectrul original, dar deplasată în frecvență pentru a corespunde poziției văilor. Această structură poate conduce la o distorsiune cu calități tonale ce pot crește proeminența erorii.

Eroarea se definește ca diferența dintre logaritmul spectrului semnalului de referință și testat, ponderat fiecare cu răspunsul urechii externe și medii (vezi §6.2.4, ec.(6.10)).

$$F[k] = \log\left(|W[k]X_{test}[k]|^2\right) - \log\left(|W[k]X_{Ref}[k]|^2\right) = \log\left(\frac{|X_{test}[k]|^2}{|X_{Ref}[k]|^2}\right) \quad |0 \leq k \leq 1024 \quad (6.94)$$

Dacă răspunsul la o frecvență oarecare este zero pentru unul din semnale, valorile logaritmilor ar fi nedeterminate. De aceea se folosește un prag energetic (vezi §6.5.2.4.3), care elimină cadrele cu energie redusă.

Mărimea structurii armonice se obține prin identificarea și măsurarea celui mai mare vârf din spectrul funcției de autocorelare. Fiecare corelare se calculează ca și cosinusul unghiului dintre doi vectori conform formulei următoare, unde  $\mathbf{F}_0$  este vectorul de eroare și  $\mathbf{F}_t$  este același vector cu o întârziere  $t$ .

Se definește vectorul de eroare de lungime  $M$ ,  $\mathbf{F}_0$ , din  $F[k]$ :

$$\mathbf{F}_0 = [F[i], \dots, F[i+M-1]]^T \quad (6.95)$$

Se calculează autocorelarea normalizată:

$$C(t) = \frac{\tilde{\mathbf{F}}_0 \cdot \tilde{\mathbf{F}}_t}{|\tilde{\mathbf{F}}_0| \cdot |\tilde{\mathbf{F}}_t|} \quad (6.96)$$

Întârzierea maximă pentru a obține funcția de autocorelare este cea mai mare putere a lui 2 care este mai mică decât jumătate din numărul componente FFT corespunzătoare frecvenței de 18 kHz:

$$T_{\max} = 2^{\lfloor \log_2(N_F F_{\max} / F_s) \rfloor - 1} \quad (6.97)$$

De exemplu, la o frecvență de eșantionare de  $F_s = 48$  kHz și o fereastră FFT de  $N_F = 2.048$  de eșantioane, componenta FFT corespunzătoare frecvenței  $F_{\max} = 18$  kHz este  $(18/24) \times 1.024 = 768$ . Deci, întârzierea maximă va fi 384. Valoarea reală a întârzierii  $T_{\max}$  va fi 256, care este cea mai mare putere a lui 2 mai mică decât 384. Prima valoare a funcției de corelare se va obține aliniind  $\mathbf{F}_t[0]$  cu  $\mathbf{F}_0[0]$  și ultima aliniind  $\mathbf{F}_t[0]$  cu  $\mathbf{F}_0[255]$ .

În standardul BS.1387 se face următoarea afirmație:

”Lungimea corelării este egală cu maximul întârzierii (256 în exemplul de mai jos).”

Această afirmație este ambiguă. Ea se poate referi la numărul termenilor corelației sau la dimensiunea vectorilor folosiți pentru calculul corelației. În continuare vom presupune că atât numărul termenilor corelației ( $N_T$ ), cât și dimensiunea vectorilor ( $M$ ) sunt egale cu  $T_{\max}$ .

Corelația calculată cu (6.96) este de fapt  $C[t] = C(t,0)$ . Calculul acestei valori necesită calculul lui  $|\mathbf{F}_t|^2$ . Acest termen poate fi calculat recursiv:

$$|\mathbf{F}_t|^2 = \begin{cases} |\mathbf{F}_0|^2 & t = 0 \\ |\mathbf{F}_{t-1}|^2 + F[t + M - 1]^2 - F[t - 1]^2 & 1 \leq t \leq T_{\max} \end{cases} \quad (6.98)$$

Dacă cele două semnale sunt egale ( $F[k] = 0$  pentru toți  $k$ ), pot apărea probleme la normalizarea valorilor corelației. Dacă semnalele de referință și de test sunt egale, dar nenule, corelația se poate face egală cu 1.

Vectorul rezultat al corelațiilor este pus într-o fereastră Hann normalizată și, după extragerea componentei de c.c. prin scăderea valorii medii, se calculează un spectru de puteri cu o transformare FFT. Ca și în §6.2.3, se presupune că factorul de normalizare aplicat este  $\sqrt{8/3}$ .

Valorile întârzierii de corelare reprezintă valoarea diferenței de frecvență – întârzierea  $t$  corespunde unei diferențe de frecvență  $tF_s/N_T$  ( $t \times 23$  Hz). Corelația se calculează pentru întârzieri de la 1 la 256, corespunzătoare gamei de frecvență 23 – 6.000 Hz. În general, pentru un material audio tipic, spațiul armonic 50 – 2.000 Hz este cel mai interesant, căruia îi corespunde gama de întârzieri 2 – 85. Partea inferioară a acestei game corespunde vocilor, iar partea superioară instrumentelor cu sunete acute.

Discuția de mai sus sugerează că fereastra ar trebui să scoată în evidență partea inferioară a gamei de întârzieri. Aplicând o fereastră Hann ca cea descrisă în §6.2.3, ar fi scoase în evidență întârzierile din mijlocul domeniului (centrate în jurul a 3 kHz).

În continuare presupunem că fereastra are lungimea  $T_{\max}$ , adică un număr  $T_{\max}-2$  componente nenule. Fereastra de întârziere care se va folosi este:

$$H[t] = \begin{cases} \sqrt{\frac{8}{3}} \frac{1}{2} \left[ 1 - \cos\left(\frac{2\pi t}{N_T - 1}\right) \right] & 0 \leq t \leq T_{\max} - 1 \\ 0 & \text{altfel} \end{cases} \quad (6.99)$$

Corelația în fereastră se obține în felul următor:

$$C_w[m] = H[m] (C[m+1] - \bar{C}) \quad 0 \leq m \leq T_{\max} - 1 \quad (6.100)$$

unde

$$\bar{C} = \frac{1}{T_{\max}} \sum_{t=1}^{T_{\max}} C[t] \quad (6.101)$$

În continuare, se presupune că lungimea transformării FFT este  $N_T$ . Factorul de scalare nu se specifică, dar, ca la §6.2.3, se va folosi un factor de scalare egal cu inversul lungimii transformării, adică  $1/N_T$ .

Secvența de putere spectrală este:

$$S[k] = \left| \frac{1}{N_T} \sum_{t=0}^{N_T-1} C_w[t] e^{j2\pi kt / N_T} \right|^2 \quad (6.102)$$

Vârful maxim din spectru după prima vale identifică frecvența dominantă în funcția de autocorelare. Corelația puterii spectrale este simetrică, astfel încât căutarea maximumului se poate

opri la mijloc. Căutarea primei văi începe cu indexul 1 și se oprește când  $S[k] > S[k-1]$ . Căutarea maximului se face până la indexul  $N_T/2$  (inclusiv).

Valoarea medie a acestui maxim pe cadre înmulțită cu 1.000,0 reprezintă eroarea structurii armonice, variabila  $EHS$ .

$$EHS = \frac{100}{N} \sum_{n=0}^N EHS_{\max}(n) \quad (6.103)$$

## 6.5. Medierea

### 6.5.1 Medierea spectrală

Algoritmul următor se folosește la medierea valorilor locale în benzile de frecvență, dacă nu se specifică altfel în descrierea variabilelor (vezi § 6.4).

#### 6.5.1.1 Media liniară

Valoarea mediei liniare se calculează cu:

$$AvgS = \frac{1}{Z} \cdot \sum_{k=0}^{Z-1} S[k] \quad (6.104)$$

unde  $S$  reprezintă numele variabilei de ieșire a modelului, iar  $Z$  este numărul benzilor de frecvență.

### 6.5.2 Media temporală

Următorii algoritmi se folosesc la medierea valorilor momentane în timp, dacă nu se specifică altfel în descrierea variabilelor (vezi § 6.4). Factorul de ponderare temporală (dacă se aplică) se notează cu  $W$  și  $Z$  este numărul benzilor de frecvență.

#### 6.5.2.1 Media liniară

Valoarea mediei liniare (prefix “Avg”) se calculează cu:

$$AvgX = \frac{1}{N} \cdot \sum_{n=0}^{N-1} X[n] \quad (6.105)$$

unde  $X$  reprezintă numele variabilei și  $N$  este numărul eșantioanelor de timp pentru care s-au calculat valorile momentane ale lui  $X$ .

Dacă se aplică o ponderare temporală (vezi §6.4.2), media liniară se calculează cu:

$$AvgX = \frac{\sum_{n=0}^{N-1} W[n] \cdot X[n]}{\sum_{n=0}^{N-1} W[n]} \quad (6.106)$$

### 6.5.2.2 Media pătratică

Valoarea mediei pătratice (prefix “*Rms*”) se calculează cu:

$$RmsX = \sqrt{\frac{1}{N} \cdot \sum_{n=0}^{N-1} X[n]^2} \quad (6.107)$$

unde  $X$  reprezintă numele variabilei și  $N$  este numărul eșantioanelor de timp pentru care s-au calculat valorile momentane ale lui  $X$ .

Dacă se aplică o ponderare temporală (vezi §6.4.2), media pătratică se calculează cu:

$$RmsX = \sqrt{Z} \cdot \sqrt{\frac{\sum_{n=0}^{N-1} W[n]^2 \cdot X[n]^2}{\sum_{n=0}^{N-1} W[n]^2}} \quad (6.108)$$

### 6.5.2.3 Media în fereastră

Valoarea mediei în fereastră (prefix “*Win*”) se calculează cu:

$$WinX = \sqrt{\frac{1}{N-L+1} \cdot \sum_{n=L-1}^{N-1} \left( \frac{1}{L} \cdot \sum_{i=0}^{L-1} \sqrt{X[n-i]} \right)^4} \quad (6.109)$$

unde  $X$  reprezintă numele variabilei,  $N$  este numărul eșantioanelor de timp pentru care s-au calculat valorile momentane ale lui  $X$  și  $L$  este lungimea ferestrei în eșantioane de timp.

Dacă lungimea ferestrei este de aprox. 100 ms,  $L = 4$  pentru modelul urechii bazat pe FFT.

### 6.5.2.4 Alegerea cadrului

#### 6.5.2.4.1 Mediere cu întârziere

Pentru variabilele de ieșire ale modelului care folosesc acest criteriu, valorile calculate în primele 0,5 s ale măsurării nu sunt luate în considerare la medierea temporală. Astfel, numărul de cadre care vor fi omise este:

$$N_{del} = \lceil \tau_{del} F_{ss} \rceil \quad (6.110)$$

unde  $\tau_{del} = 0,5$  s și  $F_{ss} = 46,875$  Hz (vezi ec. (6.27)). Deci, indexul cadrelor  $n$  va include numai cadrele care apar după întârzierea inițială și numărul total de eșantioane  $N$ , care se folosește la mediere, va conține numai aceste valori. Interpretând acest criteriu din standard, putem spune că numărul cadrelor omise trebuie să dea o întârziere de cel puțin 0,5 secunde.

*Medierea cu întârziere* se folosește pentru următoarele variabile:

*WinModDiff1, AvgModDiff1, AvgModDiff2, RmsNoiseLoud.*

#### 6.5.2.4.2 Pragul de tărie

Pentru variabilele care folosesc acest criteriu, toate valorile momentane calculate până la 50ms după ce tăria generală a unui canal audio a atins o dată valoarea de prag  $N_{Thres}$  son nu se iau

în considerare la medierea temporală. Pragul de tărie se folosește numai pentru variabilele descrise la §6.4.3.

#### 6.5.2.4.3 Pragul energetic

În calculul variabilei *EHS* nu sunt incluse cadrele cu energie redusă. În standardul BS.1387 se enunță următoarea condiție:

“Dacă energia celei mai recente jumătăți de cadru din cele 2048 de eșantioane este mai mică de 8000<sup>1</sup>, în canalul mono sau în fiecare din canalele stâng sau drept din fișierele de referință sau testat, cadrul este ignorat.”

Pragul pentru un semnal cu amplitudinea maximă  $A_{\max}$  este:

$$A_{thr}^2 = 8000 \left( \frac{A_{\max}}{32768} \right)^2 \quad (6.111)$$

Calculul energiei pentru un anumit cadru (cu  $n$  eșantioane) se face astfel:

$$A^2 = \sum_{n=1024}^{2047} x^2[n] \quad (6.112)$$

Cadrele au o suprapunere de 50% și se evaluează numai o jumătate de cadru care conține date noi. Comparația se face astfel încât energia medie a unei jumătăți de cadru să fie mai mică decât  $A_{thr}^2 / 1024$  (adică 7,8 pentru date pe 16 biți, corespunzător unei valori RMS de 2,8). Acum se poate da o formulare mai clară enunțului de mai sus:

“Pentru un semnal monaural, un cadru este ignorat dacă energia celei mai recente jumătăți de cadru este mai mică decât 8.000, atât pentru semnalul de referință, cât și pentru cel de test. Pentru un semnal binaural, cadrul este ignorat dacă energia celei mai recente jumătăți de cadru este mai mică de 8.000 atât pentru ambele canale ale semnalului de referință, cât și pentru ambele canale ale semnalului testat.”

Acest criteriu se aplică numai variabilelor descrise la §6.4.8. În concluzie, pentru calculul variabilei *EHS*, un cadru este ignorat dacă:

$$\begin{aligned} & (A_{Test}^2 < A_{thr}^2) \wedge (A_{Ref}^2 < A_{thr}^2) && \text{monaural} \\ & (A_{Testst}^2 < A_{thr}^2) \wedge (A_{Refst}^2 < A_{thr}^2) \wedge (A_{Testdr}^2 < A_{thr}^2) \wedge (A_{Refdr}^2 < A_{thr}^2) && \text{binaural} \end{aligned} \quad (6.113)$$

Aplicarea acestui criteriu previne analiza cadrelor cu energie foarte mică, adică calculul logaritmilor pentru cadre cu valoarea zero. Pentru semnale monaurale, valoarea instantanee trebuie să fie evaluată numai când cel puțin un semnal conține eșantioane nenule. Pentru semnale binaurale, cel puțin unul din cele patru semnale (2 semnale și 2 canale) este nenul.

#### 6.5.2.4.4 Date limită

Dacă fișierul prelucrat conține zgomot înainte sau după datele propriu-zise, eroarea relativă poate fi foarte mare pentru că nivelul de referință se apropie de  $-\infty$ . Pentru ca această eroare să nu fie considerată un artefact, ea poate fi ignorată prin introducerea unui criteriu de rejecție a datelor aflate sub o anumită limită.

<sup>1</sup> Acest număr se referă la datele de intrare pe 16 biți, cu valori între  $-32768$  și  $32767$ , ca la Compact Disc

La deschiderea fișierelor, mai întâi se identifică începutul și finalul efectiv al secvenței sonore din fișierul de referință. Începutul și sfârșitul datelor se definește ca prima locație, scanând de la începutul sau de la sfârșitul fișierului, în care suma valorilor absolute pe cinci eșantioane succesive depășește 200 într-unul din canalele audio. Cadrele aflate în afara acestui domeniu sunt ignorate.

Acest criteriu se aplică tuturor variabilelor.

### 6.5.3 Medierea pe canale audio

Dacă nu se indică altfel, în cazul semnalelor stereofonice, variabilele corespunzătoare canalelor stâng și drept, după medierea temporală, sunt mediate liniar.

## 6.6. Estimarea calității audio de bază percepute

Calitatea audio de bază percepută se estimează printr-un singur număr, rezultat prin combinarea variabilelor obținute în secțiunile anterioare folosind o rețea neuronală, care ne dă gradul de diferență obiectivă ce măsoară degradarea semnalului testat față de semnalul de referință. Rețeaua neuronală a fost antrenată astfel încât rezultatele obținute să fie conforme cu scala de apreciere subiectivă a calității audio, prezentată în tabelul 6.5, care măsoară diferența dintre aprecierea dată semnalului testat și aprecierea semnalului de referință.

Tabel 6.5 Scala de diferențe subiective

Grad de diferență	Descrierea deosebirilor
0	Imperceptibilă
-1	Perceptibile, dar nederanjante
-2	Ușor deranjante
-3	Deranjante
-4	Foarte deranjante

Înainte de prelucrarea finală, variabilele de ieșire ale modelului sunt translatate și scalate.

$$M'_v[i] = \frac{M_v[i] - a_{\min}[i]}{a_{\max}[i] - a_{\min}[i]} \quad (6.114)$$

Tabel 6.6 Parametrii de scalare și translație ai variabilelor

Index (i)	Variabila de ieșire a modelului x(i)	Valoare min. $a_{\min}(i)$	Valoare max. $a_{\max}(i)$
0	<i>BandwidthRef</i>	393,916656	921,0
1	<i>BandwidthTest</i>	361,965332	881,131226
2	<i>Total NMR</i>	-24,045116	16,212030
3	<i>WinModDiff1</i>	1,110661	107,137772
4	<i>ADB</i>	-0,206623	2,886017
5	<i>EHS</i>	0,074318	13,933351
6	<i>AvgModDiff1</i>	1,113683	63,257874
7	<i>AvgModDiff2</i>	0,950345	1145,018555
8	<i>RmsNoiseLoud</i>	0,029985	14,819740
9	<i>MFPD</i>	0,000101	1,0
10	<i>RelDistFrames</i>	0,0	1,0

Tabelul 6.6 prezintă parametrii de scalare și translație pentru fiecare variabilă. Acești parametri permit formarea unei imagini asupra gamei normale de valori ale variabilelor.

### 6.6.1 Rețeaua neuronală artificială

Funcția de activare a rețelei neuronale este un sigmoid asimetric:

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} = \frac{1}{2} + \frac{1}{2} \tanh\left(\frac{x}{2}\right) \quad (6.115)$$

Rețeaua folosește I intrări și J noduri în stratul ascuns. Aplicația se definește printr-un set de factori de scală de intrare  $a_{\min}[i]$ ,  $a_{\max}[i]$ , un set de ponderi de intrare  $w_x[i]$ , un set de ponderi de ieșire  $w_y[j]$  și o pereche de factori de scală de ieșire  $b_{\min}$  și  $b_{\max}$ . Intrările sunt aplicate unui *index de distorsiuni*

$$ID = w_y[J] + \sum_{j=0}^{J-1} \left( w_y[j] \cdot \text{sig} \left( w_x[I, j] + \sum_{i=0}^{I-1} w_x[i, j] \cdot \frac{x[i] - a_{\min}[i]}{a_{\max}[i] - a_{\min}[i]} \right) \right) \quad (6.116)$$

care este legat direct de *calitatea audio de bază percepută* estimată pe o scală de diferență obiectivă (*ODG – objective difference grade*). Relația dintre indexul de distorsiuni și diferența obiectivă este dată de:

$$ODG = b_{\min} + (b_{\max} - b_{\min}) \cdot \text{sig}(ID) \quad (6.117)$$

unde  $b_{\min} = -3,98$  și  $b_{\max} = 0,22$ .

*ODG* este valoarea de ieșire a metodei obiective de evaluare, corespunzătoare *SDG* (§4.3.1) din domeniul subiectiv. Rezoluția *ODG* este limitată la o zecimală. Totuși, chiar și așa, diferențe de o zecime între două evaluări nu sunt semnificative. Această observație este valabilă și pentru testele subiective.

După cum se poate observa din figura 6.6, *ODG* poate lua și valori pozitive. Aceste valori pot apărea datorită modelului cognitiv folosit pentru a face corespondența dintre valorile variabilelor de ieșire ale modelului și rezultatele testelor subiective de ascultare critică. În cazul testelor subiective, *SDG* poate avea o valoare pozitivă dacă o persoană a evaluat greșit semnalul de referință și cel testat.

Indicele de distorsiune *ID* are aceeași semnificație ca și *ODG*. Totuși, *ID* și *ODG* pot fi comparați numai cantitativ, nu și calitativ. În figura 6.6 sunt prezentate două curbe, care reprezintă relația dintre calitatea audio și valoarea *ID* (a) și relația dintre calitatea audio și valoarea *ODG* (b). În figura 6.6.a se poate observa că indicele de distorsiune *ID* prezintă o saturație mai mică decât caracteristica *ODG*. Mai mult, domeniul de valori este diferit.

În general, se recomandă ca pentru evaluarea calității audio să se ia în considerare numai valorile  $ODG \geq -3,6$ . În acest domeniu *ODG* se corelează foarte bine cu evaluările subiective. Pentru intervalul  $-3,6 \leq ODG \leq -4$  se poate utiliza *ID*.

De remarcat că nu se recomandă compararea valorilor *ODG* obținute într-o evaluare cu valorile *ID* obținute la o altă evaluare.



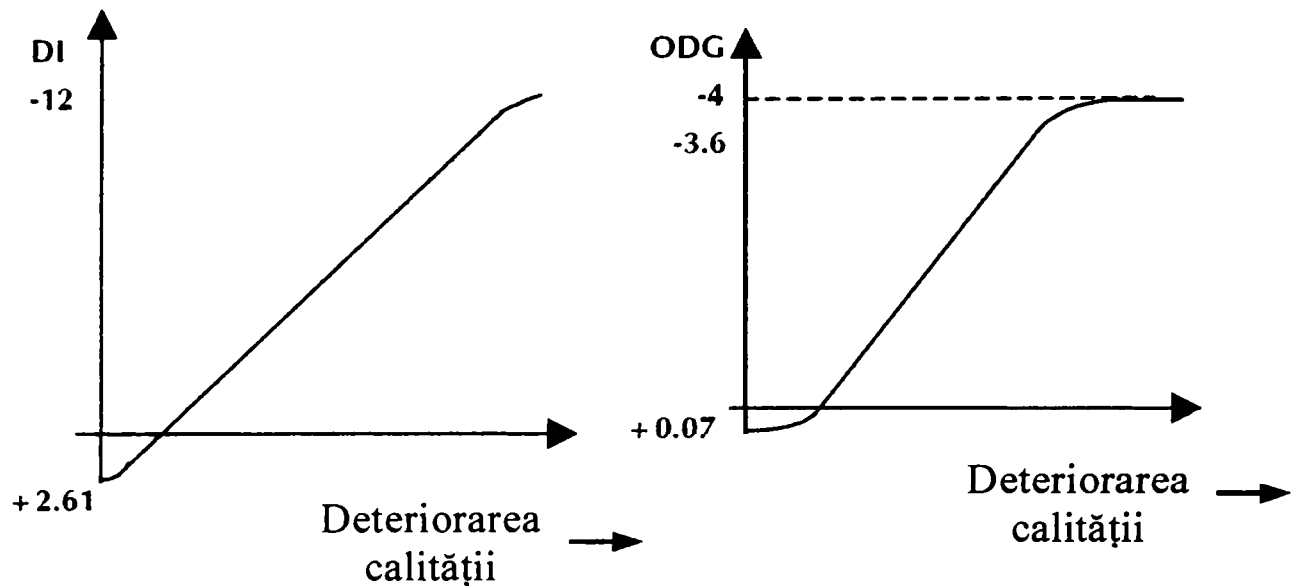


Figura 6.6 Comparație între indexul de distorsiuni ID și gradul de diferență obiectivă ODG

În final, ar trebui specificat modul de tratare a semnalelor de la începutul și sfârșitul fișierelor, deoarece poate afecta rezultatul final. Standardul nu tratează în mod explicit această problemă, care poate afecta conformitatea implementării. Cadrele de la începutul și sfârșitul fișierelor care au un nivel foarte redus sunt excluse folosind criteriul de prag (§6.4.1). Considerăm că fișierele de referință și de test încep cu eșantionul zero și au lungimea (în eșantioane)  $N_{Ref}$ , respectiv  $N_{Test}$ . În continuare, considerăm că eșantioanele dinaintea eșantionului zero și după ultimul eșantion sunt nule. Primul eșantion din primul cadru se va alinia cu eșantionul  $n_{off}$ , pentru care avem următoarele variante:

- Primul cadru începe la eșantionul zero,  $n_{off} = 0$ . Aceasta înseamnă că eșantioanele aflate la începutul semnalului apar cu ponderi mici datorită folosirii ferestrei Hann în faza de analiză FFT.

- Decalăm cadrul cu o jumătate de cadru,  $n_{off} = -N_F/2$ . Acum primul cadru va conține numai zero în prima jumătate și date în a doua jumătate. Aceleași date vor apărea și în prima jumătate a celui de-al doilea cadru, deoarece cadrele sunt suprapuse 50%.

Pentru finalul datelor se iau în considerare aceleași variante, dar apare problema că semnalele pot fi de lungimi diferite. O soluție ar fi de a considera lungimea semnalului ca maximul dintre lungimile semnalelor de referință și de test.

$$N_s = \max(N_{sRef}, N_{sTest}) \quad (6.118)$$

În acest caz trebuie considerat momentul în care se oprește prelucrarea semnalelor.

- Ultimul cadru este cel care conține date pe mai mult de jumătate din cadru. Dacă s-ar mai adăuga un cadru, el ar conține numai eșantioane care au apărut deja în cadrul precedent.

- Ultimul cadru este cel care conține cel puțin un eșantion de semnal.

Luând în considerare condițiile de mai sus, numărul de cadre ce va fi prelucrat este:

$$N_p = \left\lceil \frac{N_s - n_{off} - n + 1}{N_F / 2} \right\rceil \quad \text{Cel puțin } n \text{ eșantioane în ultimul cadru} \quad (6.119)$$

Opțiunea 1 pentru începutul datelor este consistentă cu opțiunea 1 pentru sfârșitul datelor, iar opțiunea 2 pentru începutul datelor este consistentă cu opțiunea 2 pentru sfârșitul datelor.

Metoda folosește modelul urechii bazat pe transformarea FFT și utilizează 11 variabile de ieșire: *BandwidthRef*, *BandwidthTest*, *Total NMR*, *WinModDiff1*, *ADB*, *EHS*, *AvgModDiff1*, *AvgModDiff2*, *RmsNoiseLoud*, *MFPD* și *RelDistFrames*. Pe baza acestor variabile se obține un index de calitate unic folosind o rețea neuronală conform descrierii din §6.6.1 cu trei noduri în stratul ascuns. Parametrii aplicației sunt dați în tabelele 6.6 – 6.9.

Tabel 6.7 Variabile de ieșire

Variabila	Scop
<i>WinModDiff1</i>	Schimbări ale modulației
<i>AvgModDiff1</i>	
<i>AvgModDiff2</i>	
<i>RmsNoiseLoud</i>	Tăria distorsiunii
<i>BandwidthRef</i>	Distorsiuni liniare (răspuns în frecvență etc.)
<i>BandwidthTest</i>	
<i>RelDistFrames</i>	Frecvența distorsiunilor audibile
<i>Total NMR</i>	Raportul Zgomot-Mascare
<i>MFPD</i>	Probabilitatea detecției
<i>ADB</i>	
<i>EHS</i>	Structura armonică a erorii

Tabel 6.8 Ponderi pentru nodul de ieșire

nod 1 ( $w_y[0]$ )	nod 2 ( $w_y[1]$ )	nod 3 ( $w_y[2]$ )	bias ( $w_y[3]$ )
-3,817048	4,107138	4,629582	-0,307594

Tabel 6.9 Ponderi pentru nodurile de intrare

index (i)	Variabila ( $x[i]$ )	nod 1 ( $w_x[i,0]$ )	nod 2 ( $w_x[i,1]$ )	nod 3 ( $w_x[i,2]$ )
0	<i>BandwidthRef</i>	-0,502657	0,436333	1,219602
1	<i>BandwidthTest</i>	4,307481	3,246017	1,123743
2	<i>Total NMR</i>	4,984241	-2,211189	-0,192096
3	<i>WinModDiff1</i>	0,051056	-1,762424	4,331315
4	<i>ADB</i>	2,321580	1,789971	-0,754560
5	<i>EHS</i>	-5,303901	-3,452257	-10,814982
6	<i>AvgModDiff1</i>	2,730991	-6,111805	1,519223
7	<i>AvgModDiff2</i>	0,624950	-1,331523	-5,955151
8	<i>RmsNoiseLoud</i>	3,102889	0,871260	-5,922878
9	<i>MFPD</i>	-1,051468	-0,939882	-0,142913
10	<i>RelDistFrames</i>	-1,804679	-0,503610	-0,620456
11	<i>bias</i>	-2,518254	0,654841	-2,207228

## 6.7. Conformitatea implementării

Pentru a verifica corectitudinea implementării metodei, au fost alese un număr de 16 fragmente de test, care să producă o gamă cât mai largă de valori pentru variabilele de ieșire și indexul de distorsiuni (ID).

Fragmentele de test sunt disponibile de la ITU-R sub formă de fișiere .wav eșantionate la 48 kHz, 16-bit PCM. Semnalele de referință și de test furnizate de ITU sunt adaptate ca nivel și timp și nu necesită compensări ulterioare. Algoritmii de măsurare trebuie reglat la un nivel de ascultare de 92 dB SPL.

Tabel 6.10 Fragmentele de test și valorile corespunzătoare pentru ID și ODG

Item	ID	ODG
acodsna.wav	1,304	-0,676
bcodtri.wav	1,949	-0,304
ccodsax.wav	0,048	-1,829
ecodsmg.wav	1,731	-0,412
fcodsb1.wav	0,677	-1,195
fcodtr1.wav	1,419	-0,598
fcodtr2.wav	-0,045	-1,927
fcodtr3.wav	-0,715	-2,601
gcodcla.wav	1,781	-0,386
icodsna.wav	-3,029	-3,786
kcodsme.wav	3,093	0,038
lcodhrp.wav	1,041	-0,876
lcodpip.wav	1,973	-0,293
mcodcla.wav	-0,436	-2,331
ncodsfe.wav	3,135	0,045
scodclv.wav	1,689	-0,345

Pentru a fi conform Recomandării, valorile calculate ale indicelui de distorsiuni ID trebuie să se încadreze în valorile din tabelul 6.10.

## 6.8 Concluzii

În această secțiune a fost prezentată o metodă de măsurare obiectivă a calității audio percepute, bazată pe standardul BS.1387-1 al ITU. Deoarece specificațiile standardului sunt în unele locuri insuficiente sau neclare, acestea au fost completate cu ajutorul informațiilor suplimentare obținute din literatura de referință, aducându-se următoarele contribuții:

- Clarificarea modului de calcul al factorului de scalare a datelor de intrare în transformarea FFT.
- Explicarea contribuției coeficienților din funcția de ponderare care modelează răspunsul în frecvență al urechii externe și mijlocii.
- Explicarea formulelor de calcul a caracteristicilor benzilor de frecvență pentru trecerea de la scala de frecvență (Hz) la scala de înălțime auditivă (Bark) și invers.
- Precizarea condițiilor de inițializare a operațiilor de filtrare, în special în cazul factorilor de corecție ai modelului și eliminarea condiției “dacă nu se specifică altfel”.
- Clarificarea limitelor datelor de început și de sfârșit ale fișierului.
- Explicarea modului de aplicare a medierii întârziate.
- Explicarea aplicării pragului de tărie.
- Extinderea descrierii calculelor pentru structura armonică a erorii prin lămurirea ambiguităților privind scalarea, fereastra și transformarea. Lămurirea criteriului de prag energetic pentru semnale binaurale.

# CAPITOLUL 7

## REZULTATE EXPERIMENTALE

### 7.1 Evaluarea obiectivă a algoritmului ATRAC

Evaluarea schemelor de codare perceptuală audio de înaltă calitate și debit binar redus nu este întotdeauna relevantă folosind metode electroacustice tradiționale, de exemplu aplicând un semnal sinusoidal pentru a măsura raportul semnal-zgomot. Codarea perceptuală de tip ATRAC, prin natura ei, va acorda unui semnal cu un conținut spectral simplu mai mulți biți. Deci zgomotul de cuantizare sau riscul dispariției unor componente spectrale este mai redus. În plus, sistemele de codare care folosesc o dimensiune variabilă pentru blocul de transformare cosinus modificată (MDCT), ca în cazul ATRAC, permit o rezoluție mai bună a frecvenței cu semnale tranzitorii lente ca rezultat al unui bloc de transformare mai mare [BRA98], [SPE00].

Pentru a evalua performanțele unei scheme de codare pentru diferite tipuri de programe muzicale am creat câteva semnale de test în acest scop. Acestea includ forme de undă complexe cu conținut armonic derivat din studii de psihoacustică, forme de undă care să demonstreze performanțele sistemului pe întregul domeniu de frecvențe și forme de undă care să scoată în evidență deficiențele specifice schemei de codare supuse evaluării.

#### 7.1.1 Scopuri experimentale

Obiectivele experimentelor executate pentru evaluarea obiectivă a sistemului de codare ATRAC, implementat pe mai multe tipuri de minidiscuri, au fost următoarele:

- demonstrarea creșterii zgomotului de cuantizare a componentelor spectrale ca rezultat al codării ATRAC, folosind semnale de test pe 16 biți, cu frecvența de eșantionare de 44,1 kHz;
- să demonstreze alocarea biților în funcție de gama de frecvență a componentelor spectrale;
- să demonstreze situațiile potențiale în care toți biții să fie alocați și efectul asupra lățimii benzii de frecvență;
- să compare nivelul de pre-ecou prezent în înregistrările ATRAC ale unor semnale tranzitorii bruște cu alte sisteme digitale de codare audio de înaltă calitate și debit binar redus;
- să illustreze problemele potențiale, în domeniul timp, care pot apărea ca rezultat al codărilor ATRAC succesive (multigenerate).

#### 7.1.2 Metode experimentale

S-au avut în vedere următoarele aspecte:

- Semnalele sursă au fost generate digital folosind generatorul de semnal din programul SoundForge 5.0, în afară de semnalele tranzitorii, unde s-a folosit un CD de test (AES) cu înregistrări de percuție. Toate semnalele pe 16 biți și frecvența de eșantionare de 44,1 kHz.
- Conținutul spectral al fiecărui semnal generat digital a fost ales având în vedere înțelegerea teoretică a modului de funcționare a obiectului testat.
- Amplitudinea maximă a fiecărui fragment (numită de aici încolo 0 dB) a fost de 80% din amplitudinea maximă digitală disponibilă pe 16 biți.

- S-au folosit semnale sinusoidale pentru testele bazate pe teoria mascării perceptuale și semnale dreptunghiulare pentru a demonstra situațiile în care sunt utilizați toți biții, rezultând un zgomot de cuantizare maxim ca efect al comprimării.

- Transferul pentru înregistrarea fișierelor de test s-a făcut exclusiv digital, pe intrarea optică a minidiscului Sharp MD-7222, ca în fig. 7.1, prin interfața digitală S/PDIF (*Sony/Philips Digital Interface*), eliminându-se posibilitatea ca circuitele analogice să afecteze rezultatele testelor.

- Semnalele de test înregistrate pe MD au fost transferate înapoi în calculator folosindu-se ieșirea optică digitală S/PDIF a unui minidisc SONY MDS-JB920 și salvate din nou ca fișiere de sunet (.WAV) pe 16 biți, 44,1 kHz.

- Fișierele de sunet originale și cele codate ATRAC au fost supuse unei analize FFT folosind programul SoundForge 5.0.

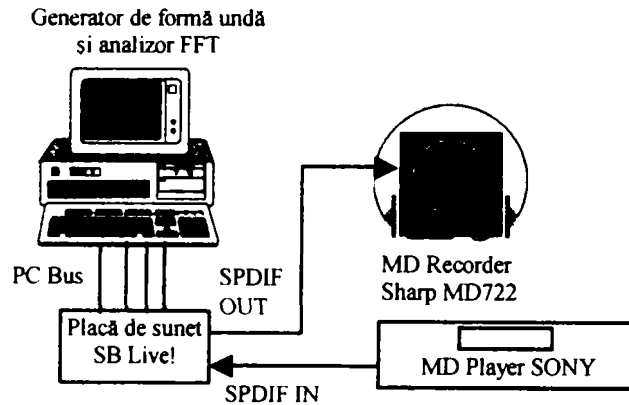


Figura 7.1 Conectarea aparatului pentru transferul și analiza fișierelor audio

În cazul analizei efectelor codărilor ATRAC succesive, transferul fișierelor s-a făcut în două moduri: înregistrare și redare analogică cu minidiscul portabil Sharp MD-7222, pentru a vedea efectul conversiei repetate A-D și D-A (fig. 7.2.a) și transfer digital pentru înregistrare și redare folosind numai minidiscul staționar Sony MDS JB-920 (fig. 7.2.b).

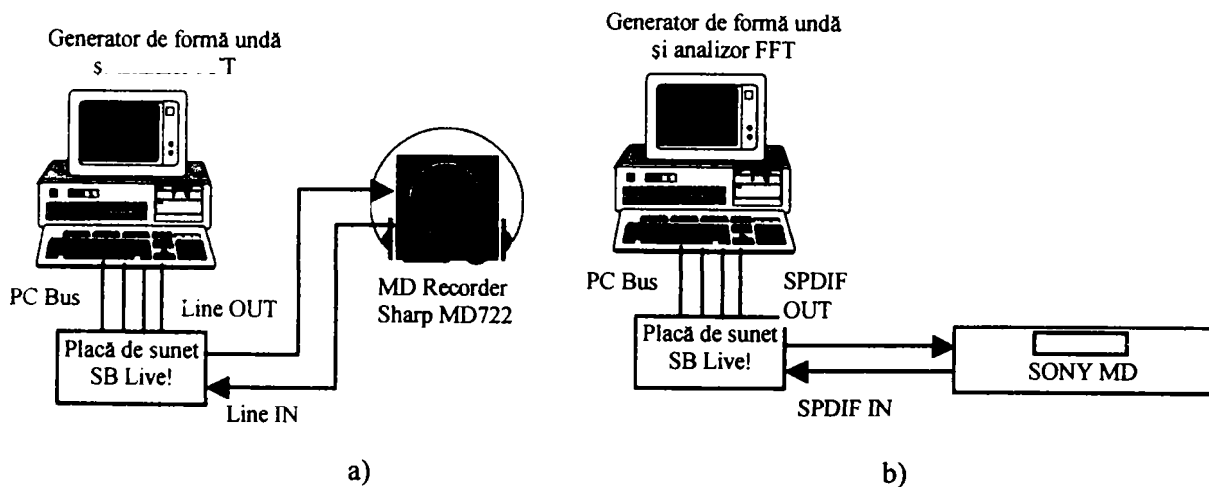


Figura 7.2 Conectarea aparatului pentru transferul analogic (a) și digital (b) al fișierelor audio

### 7.1.3 Testarea caracteristicilor de bază

#### 1. Test 1

**Obiectiv** - să ilustreze creșterea zgomotului de cuantizare a componentelor mascate perceptual și variația sa cu frecvența și amplitudinea.

**Conținutul semnalelor** - pentru determinarea efectului codării ATRAC asupra componentei mascate se aplică semnal care constă dintr-o componentă de *maskare*,  $f_1$  și o componentă *mascată*,  $f_2$ . Având în vedere că banda critică este de aproximativ 1/4 de octavă, semnalul creat conține două componente, având frecvențele separate de o bandă critică, după următoarea formulă:

$$f_{\text{mix}} = f_1 + f_2 \text{ unde } f_2 = 5/4 f_1, \quad (7.1)$$

Frecvența  $f_2$  a fost atenuată conform pragului de mascare definit de Sundberg.

### Rezultate

La frecvențe joase și medii, codorul ATRAC nu prezintă creșteri măsurabile ale zgomotului spectral de cuantizare (fig.7.3 - 7.4). La frecvențe înalte totuși, apare o creștere semnificativă a zgomotului spectral de cuantizare, creștere de aproximativ +10 dB de la 10 kHz în sus (fig. 7.5).

Acest rezultat reflectă modul de lucru al algoritmului ATRAC, care alocă un număr mai mare de biți la frecvențe joase, respectiv la frecvențele în care urechea este sensibilă. Rezultatele sugerează că numărul fix de biți pentru frecvențe joase este întotdeauna mai mare decât numărul fix și variabil de biți la frecvențe înalte, indiferent de complexitatea semnalului.

De remarcat că sistemul testat poate înregistra frecvențe peste 19 kHz, în cazul semnalelor sinusoidale.

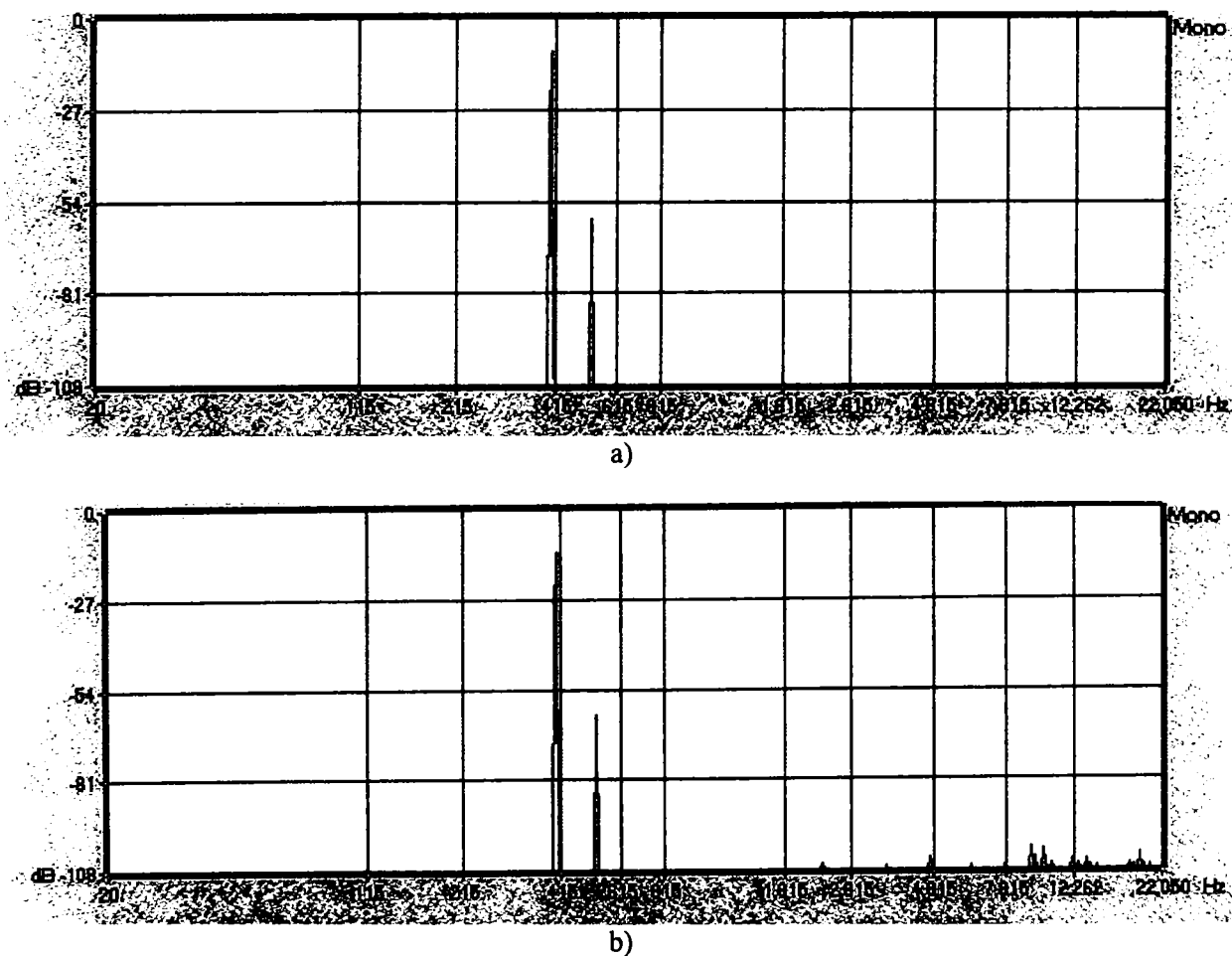
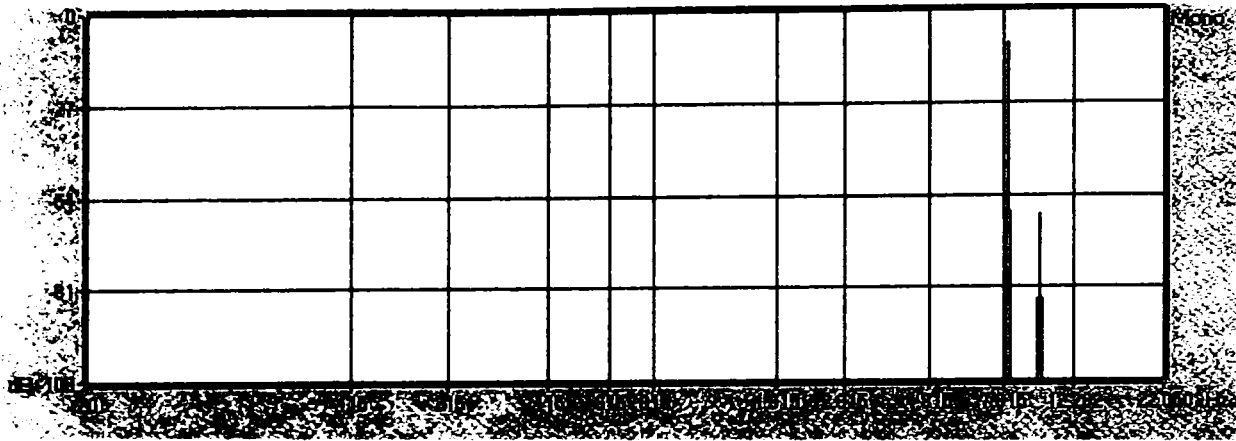
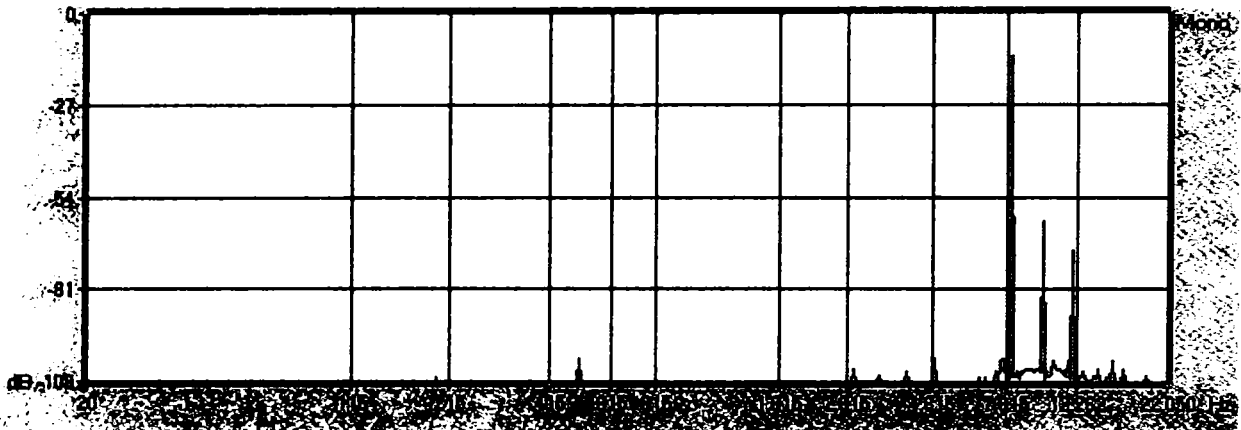


Figura 7.3 400 Hz (0 dB) + 520 Hz (-50 dB) 16 bit, 44,1 kHz, semnal PCM (a) și codat ATRAC (b)

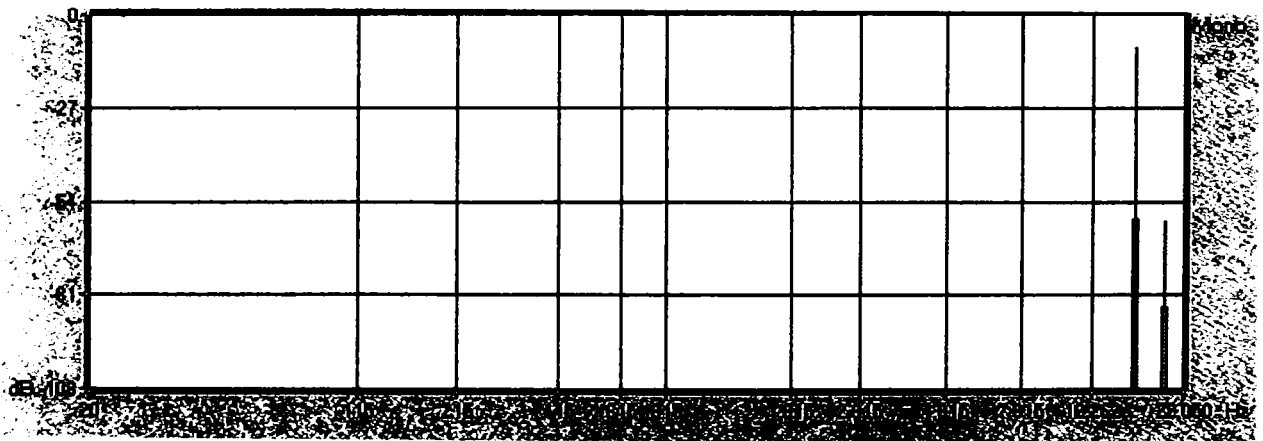


a)

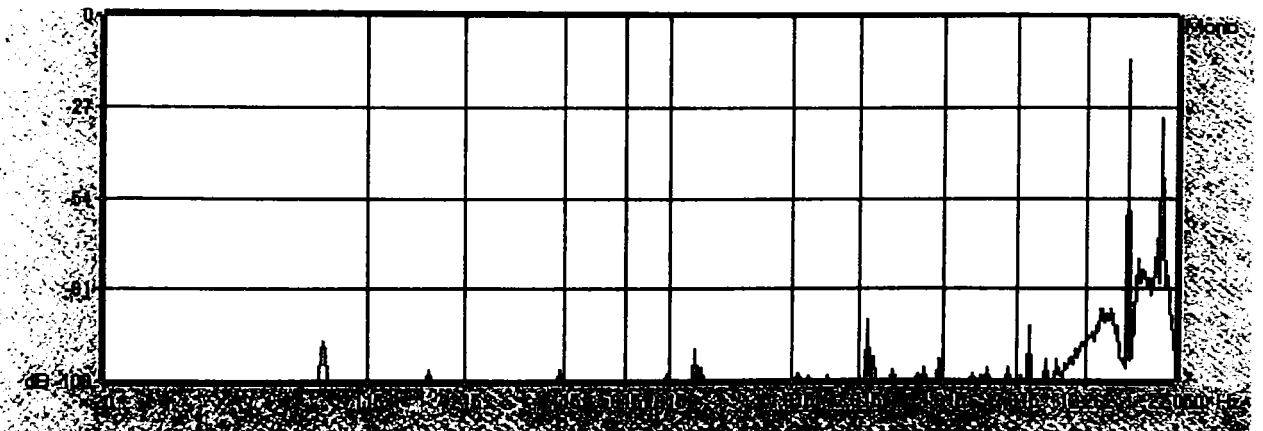


b)

Figura 7.4 8.000 Hz (0 dB) + 9.800 Hz (-50 dB) 16 bit, 44,1 kHz, semnal PCM (a) și codat ATRAC (b)



a)



b)

Figura 7.5 16.000 Hz (0 dB) + 19.500 Hz (-50 dB) 16 bit, 44,1 kHz, semnal PCM (a) și codat ATRAC (b)

## 2. Test 2

**Obiectiv** - să illustreze reducerea lărgimii de bandă și creșterea zgomotului spectral de cuantizare la codarea unui semnal bogat în armonice pe întreg spectrul audio.

**Conținutul semnalului** – 100 Hz dreptunghiular la 0 dB relativ (aproximativ -4 dB față de maximul amplitudinii digitale)

Un semnal dreptunghiular constă dintr-o frecvență fundamentală  $f_1$  și un șir infinit de armonici impare  $f_3 + f_5 + f_7 \dots + f_{2n+1}$ . Această abundență de componente spectrale se extinde pe toată gama de frecvențe a unei înregistrări cu frecvența de eșantionare de 44,1 kHz, cu niveluri de amplitudine considerabile în raport cu fundamentala.

Astfel, aplicarea unui semnal dreptunghiular poate da o indicație folositoare a creșterii zgomotului de cuantizare în urma codării ATRAC.

**Rezultate** - se poate observa o diferență între semnalul de referință și cel codat ATRAC: zgomotul de cuantizare crește cu aproximativ +50 dB față de semnalul de referință pe întreg spectrul de frecvențe și banda de frecvență a semnalului se micșorează cu 4 kHz în partea de sus a spectrului (fig. 7.6).

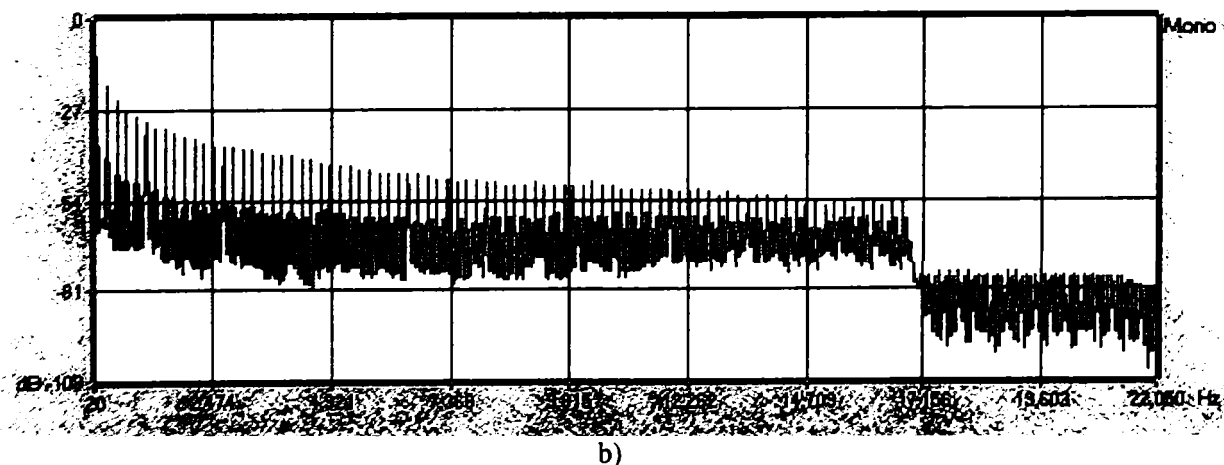
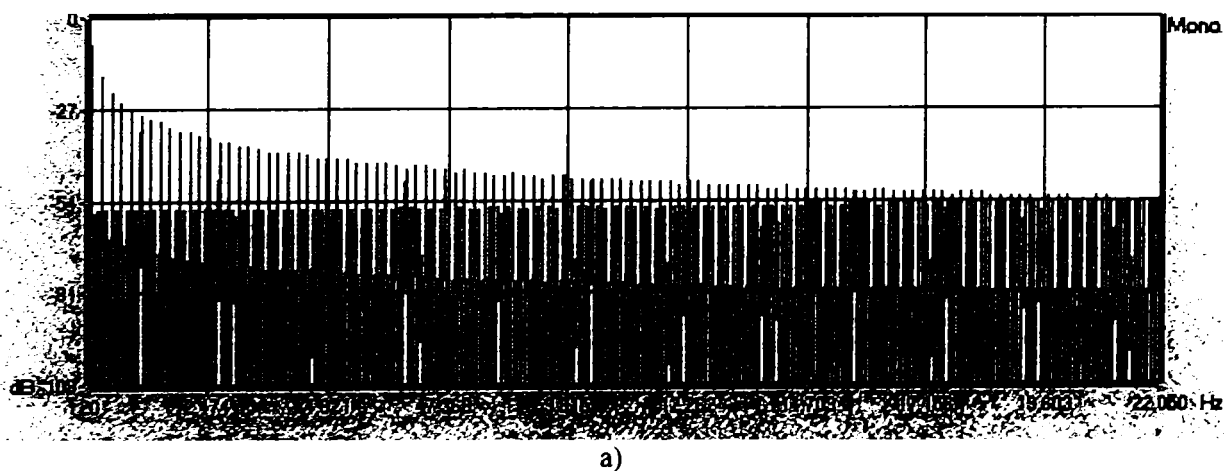


Figura 7.6 100 Hz (0 dB) semnal dreptunghiular PCM (a) și codat ATRAC (b)

Aceasta semnifică faptul că toți biții disponibili sunt folosiți pentru a reprezenta semnalul dreptunghiular, numărul de biți alocați frecvențelor înalte este sacrificat pentru a reda cât mai bine celelalte benzi.

În acest caz banda de frecvență a sistemului de înregistrare cu minidisc este mai redusă decât în cazul formatului PCM și reflectă importanța pe care o dă algoritmul ATRAC frecvențelor joase (reflectând sensibilitatea urechii).

Aplicând algoritmul de eșantionare din [TSU92], dacă un semnal dreptunghiular este considerat mai aproape de zgomot decât de un sunet pur (sinusoidal), tonalitatea T va fi mai



aproape de 0 decât de 1. Dacă raportul dintre biții fixi și variabili se ponderează cu  $1-T$ , cu cât  $T$  este mai mic cu atât mai mare va fi numărul de biți alocați.

### 3. Test 3

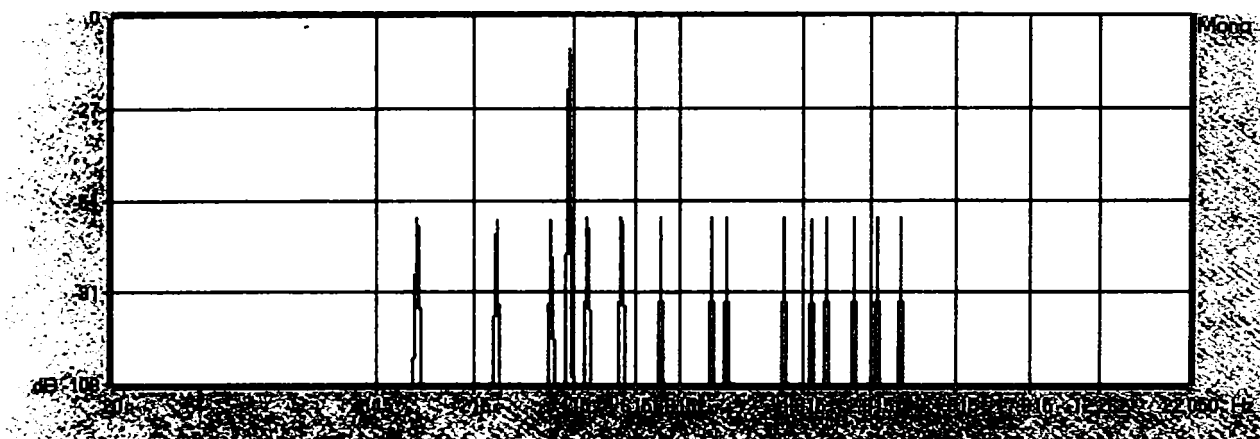
**Obiectiv** – punerea în evidență a mecanismului de alocare a biților într-o codare ATRAC în funcție de frecvențele subbenzilor și amplitudine.

#### Conținutul semnalelor:

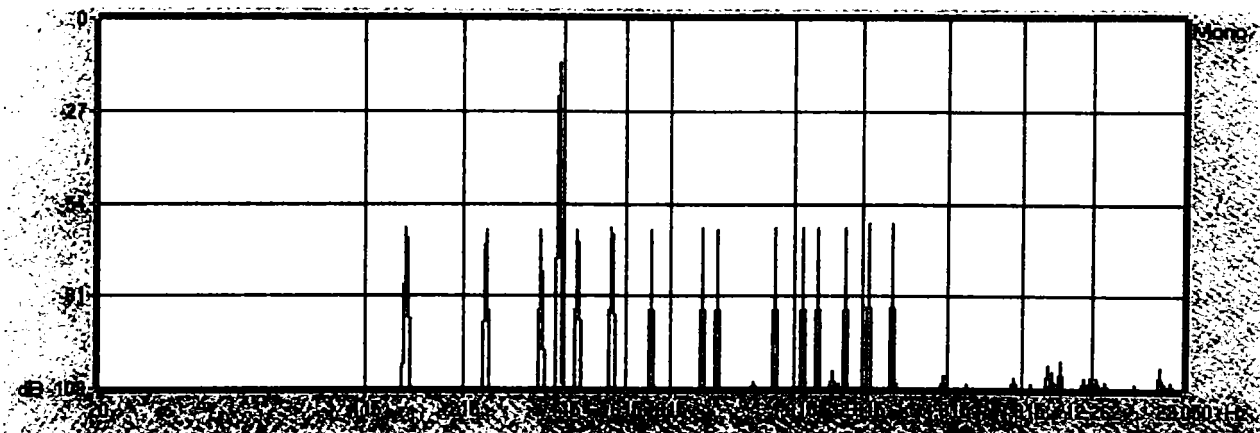
*Semnal 1:* Ton de mascare 400 Hz (0 dB) + 150 Hz, 250 Hz, 355 Hz, 450 Hz, 560 Hz, 720 Hz, 1 kHz, 1,1 kHz, 1,6 kHz, 1,9 kHz, 2,1 kHz, 2,5 kHz, 2,9 kHz, 3,4 kHz, toate la -60 dB

*Semnal 2:* Ton de mascare 400 Hz (0 dB) + 150 Hz, 250 Hz, 355 Hz, 450 Hz, 560 Hz, 720 Hz, 1 kHz, 1,1 kHz, 1,6 kHz, 1,9 kHz, 2,1 kHz, 2,5 kHz, 2,9 kHz, 3,4 kHz, toate la -40 dB

*Semnal 3:* Ton de mascare 400 Hz (0 dB) + (150 Hz, 250 Hz, 355 Hz, 450 Hz, 560 Hz, 720 Hz, 1 kHz, 1,1 kHz, 1,6 kHz, 1,9 kHz, 2,1 kHz, 2,5 kHz, 2,9 kHz, 3,4 kHz, 3,7 kHz, 4,4 kHz, 5,8 kHz, 6,4 kHz, 7,7 kHz, 9,5 kHz, 15,5 kHz toate la -40 dB) + (4,9 kHz, 5,8 kHz, 8,4 kHz, 12,6 kHz, 13,5 kHz, 17,7 kHz la -60 dB)

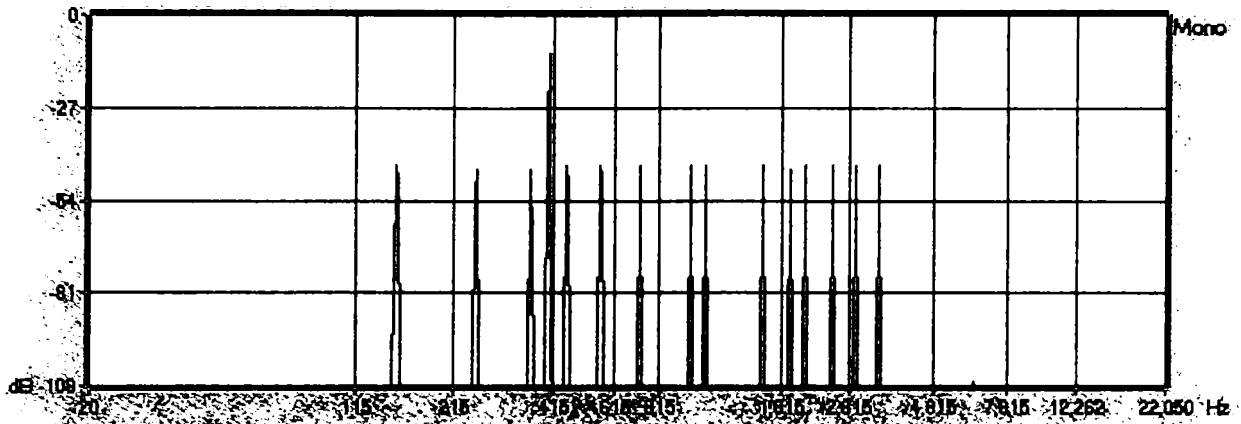


a)

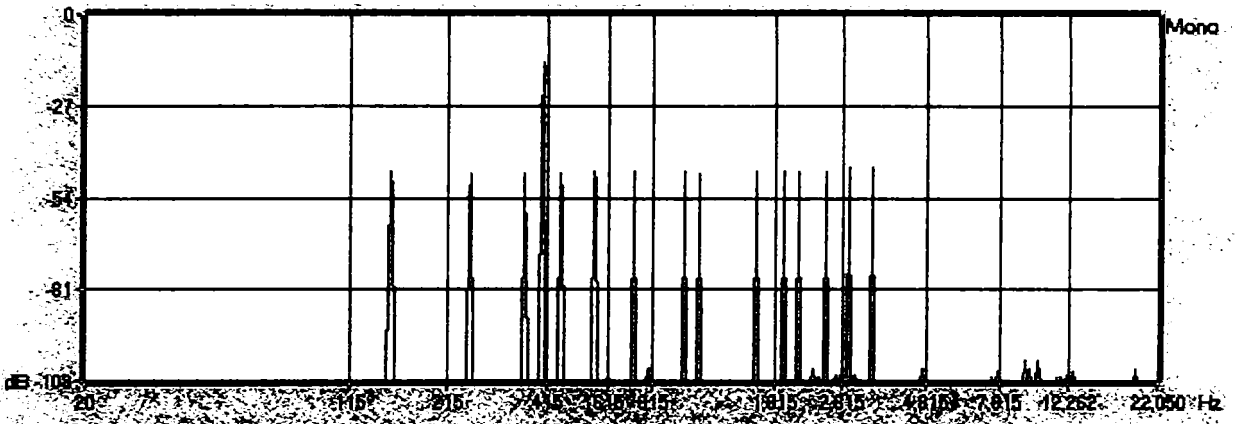


b)

Figura 7.7 Semnalul 1, 16 bit, 44,1 kHz, original PCM (a) și codat ATRAC (b)

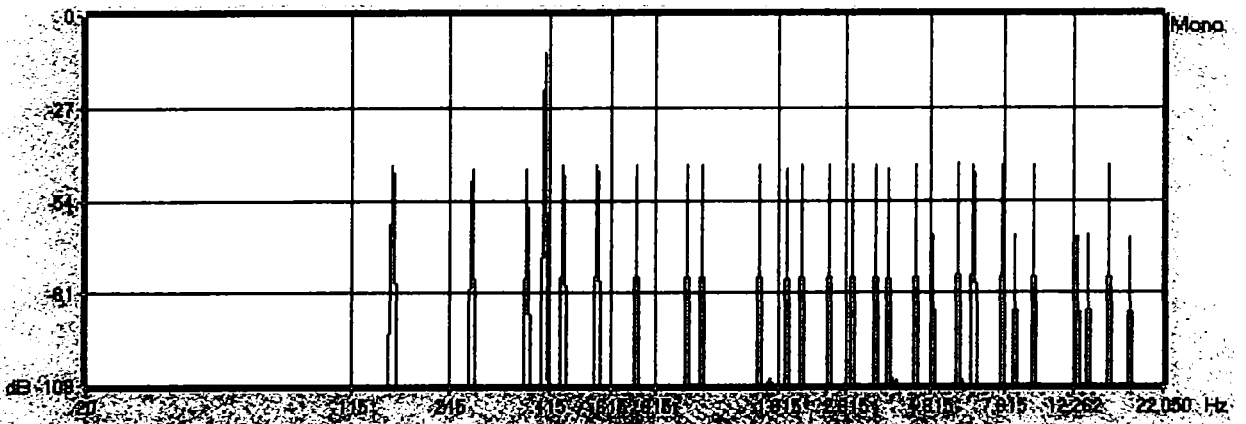


a)

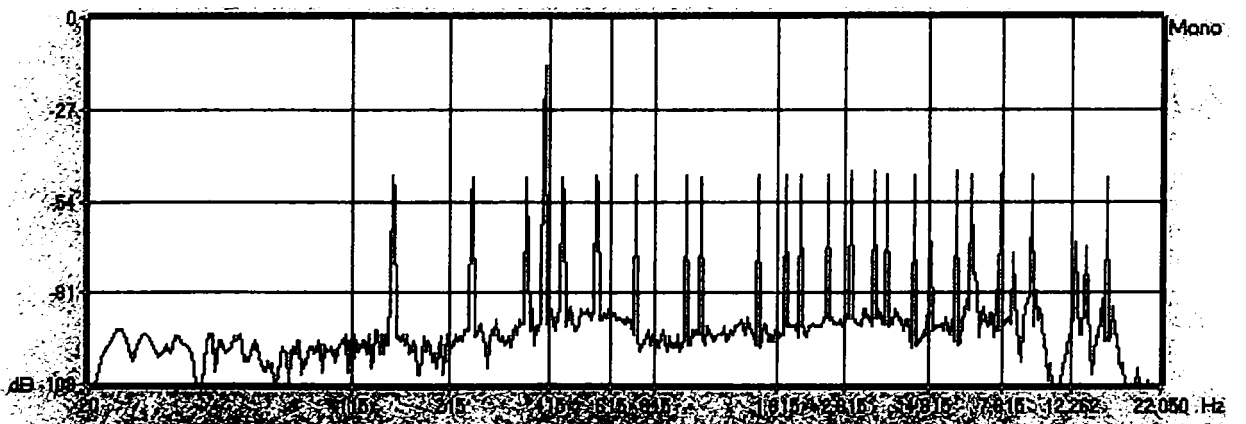


b)

Figura 7.8 Semnalul 2, 16 bit, 44,1 kHz, original PCM (a) și codat ATRAC (b)



a)



b)

Figura 7.9 Semnalul 3, 16 bit, 44,1 kHz, original PCM (a) și codat ATRAC (b)

Componentele sunt grupate în sub-benzi. Semnalul 1 (fig. 7.7) conține componente sub-benzile de frecvențe joase (0 - 5,5kHz), semnalul 2 (fig. 7.8) conține aceleași frecvențe dar cu o amplitudine mai mare și semnalul 3 (fig. 7.9) introduce frecvențe din sub-benzile medie și înaltă (peste 5,5kHz).

### Rezultate

Componentele spectrale conținând frecvențe din subbanda de jos sunt recreate fără o creștere aparentă a zgomotului de cuantizare, la o amplitudine de  $-60$  dB sub nivelul frecvenței de mascare.

Creșterea cu 20 dB a amplitudinii componentelor mascate duce la o creștere a zgomotului de cuantizare împrăștiat pe întreaga subbandă. Aceasta demonstrează că algoritmul de alocare a biților a distribuit biții în funcție de amplitudinea componentelor spectrale din aceeași subbandă.

Introducerea frecvențelor din subbenzile superioare produce o creștere semnificativă a zgomotului de cuantizare, până la 16 dB. În plus, componenta cu frecvența cea mai înaltă nu a fost deloc redată, ilustrând încă odată importanța frecvențelor joase în procesul de alocare a biților.

#### 7.1.4 Evaluarea apariției preecoului

ATRAC folosește lungimea blocului de transformare adaptivă, a cărei principală funcție este să prevină apariția preecoului. Preecoul poate apărea în timpul unor pasaje cu semnale tranzitorii puternice, datorită distribuției zgomotului de cuantizare pe întregul bloc de transformare.

Manifestarea audibilă a preecoului este zgomotul de cuantizare la atacul semnalului tranzitoriu. Acest efect nedorit poate fi vizualizat grafic în domeniul timp.

Următorul experiment folosește analiza în domeniul timp pentru a investiga prezența preecoului în cazul unui semnal puternic tranzitoriu (clopot) codat folosind algoritmul ATRAC. În plus același test s-a aplicat unui codor MPEG1-nivel3, la diferite debite binare, pentru măsurători comparative.

##### 1. Test 4

**Obiectiv** - evidențierea apariției preecoului ca rezultat al codării unui semnal puternic tranzitoriu folosind ATRAC și codorul Blade MP3 la următoarele debite: 96 kbit/s, 128 kbit/s și 192 kbit/s.

**Semnal** – clopot, original de pe CD (fig. 7.10)

### Rezultate

ATRAC cu un debit de 292 kbit/s nu prezintă un preecou măsurabil (fig. 7.11.a). Semnalele codate MP3 la 192 kbit/s, 128 kbit/s (fig. 7.11.b) și 96 kbit/s au un preecou de aprox. 150 ms, crescând în amplitudine cu scăderea debitului. Superioritatea algoritmului ATRAC din acest test reflectă eficiența dimensiunii blocului de transformare adaptivă.

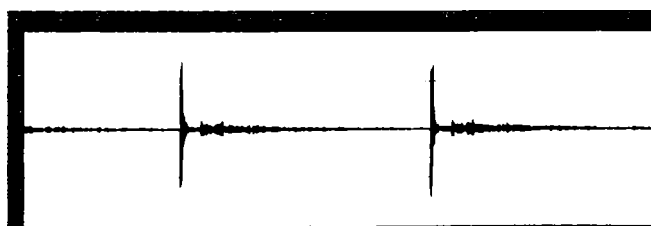


Figura 7.10 Clopot, 16 bit, 44,1 kHz, semnal PCM

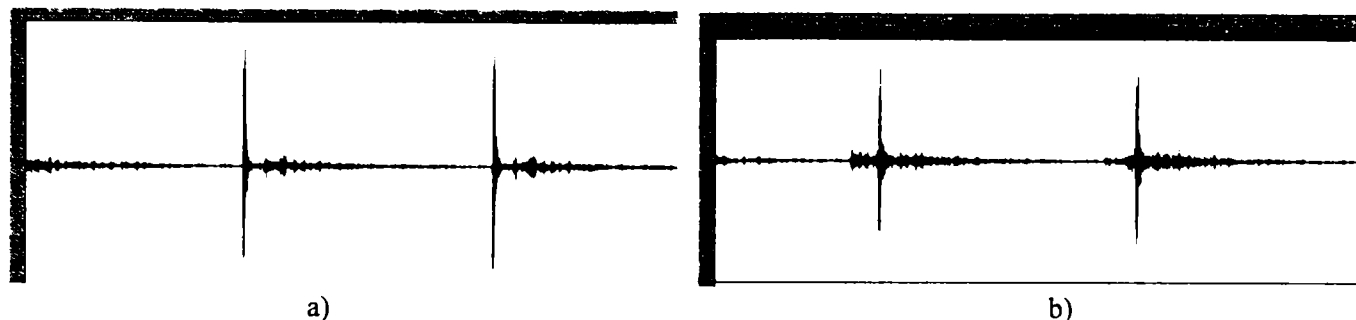


Figura 7.11 Clopot, semnal codat ATRAC (292 kbit/s) (a) și codat .mp3 la 128 kbit/s (b)

### 7.1.5 Codări ATRAC succesive

Unul dintre atuurile sistemului de înregistrare cu MD este capacitatea de a genera copii multiple pe/și de pe alt MD (echivalent cu copierea casetă - casetă).

Pentru a testa generații succesive de copii, s-a înregistrat pe MD un semnal generat de calculator și transferat înapoi în calculator. Aceasta a fost prima generație de înregistrări, care va deveni sursă pentru următoarea înregistrare pe MD, operația repetându-se de 10 ori. Semnalele rezultate în cele 10 generații au fost analizate atât în domeniul frecvență cât și în domeniul timp.

Crearea acestor fișiere de sunet a fost posibilă datorită faptului că interfața audio digitală a calculatorului folosită pentru transferul repetat al semnalelor de test, ignoră protecția la copiere din structura semnalului S/PDIF.

#### Test 5

**Obiectiv** - evidențierea degradării progresive a semnalului atât în domeniul frecvență cât și în domeniul timp datorită codărilor ATRAC succesive.

**Semnalul** – 16 kHz sinusoidal la 0 dB (ton de mascare) + 19,5 kHz sinusoidal (ton mascat) la -70 dB relativ față de mascator (fig. 7.12).

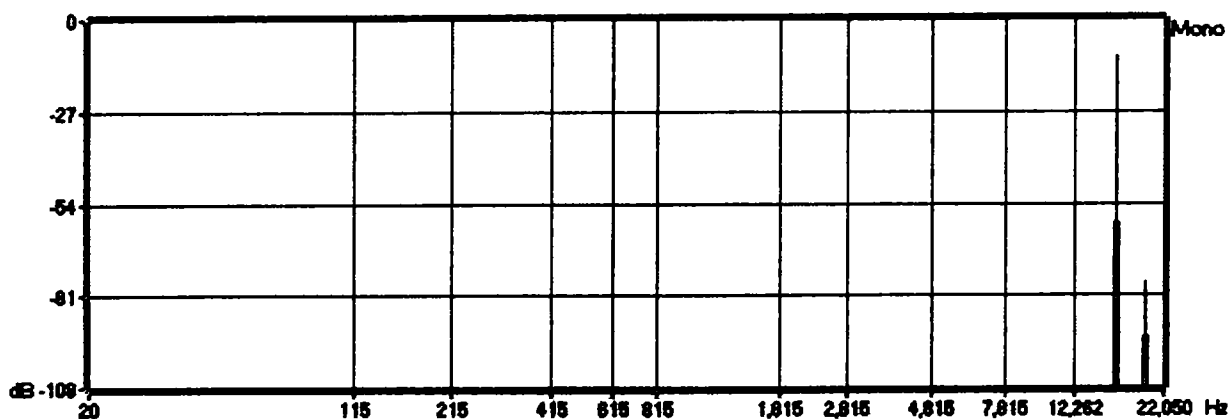


Figura 7.12 16 kHz (0dB) + 19,5 kHz (-70 dB) 16 bit, 44,1 kHz, semnal PCM

**Rezultate** - efectul înregistrărilor succesive folosind codarea ATRAC conduce la degradarea progresivă a componentelor spectrale mascate mai înalte, până la dispariția în zgomotul de cuantizare crescut (în acest caz la a 7-a generație). Creșterea zgomotului spectral de cuantizare la a 10-a generație este de aprox.12 dB față de zgomotul original. În figurile 7.13 – 7.15 este prezentată degradarea treptată a semnalului de test, atât în domeniul frecvență cât și în domeniul timp.

În plus față de modificările în conținutul spectral, analiza în domeniul timp a generațiilor succesive dezvăluie artefacte semnificative, de ex. vârfuri de semnal cu amplitudini excesive (în comparație cu amplitudinea medie), care nu sunt în concordanță ca timp și frecvență cu conținutul semnalului original.

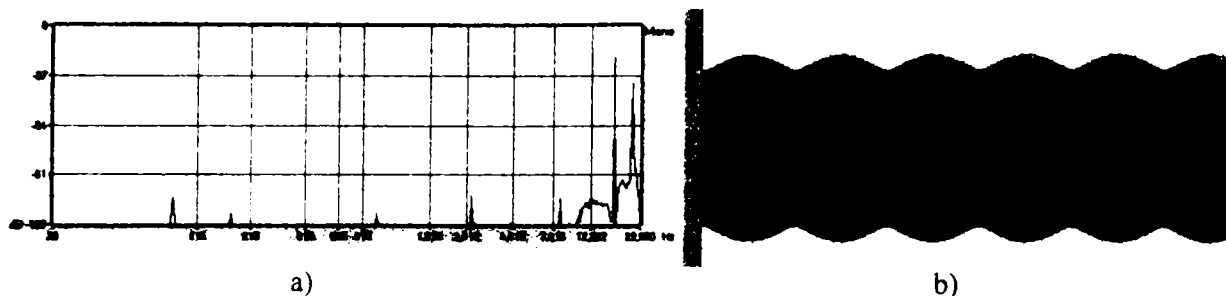


Figura 7.13 16 kHz (0 dB) + 19,5 kHz (-70 dB) 16 bit, 44,1 kHz, semnal codat ATRAC, a 2-a generație, transfer digital, frecvență (a) și timp (b)

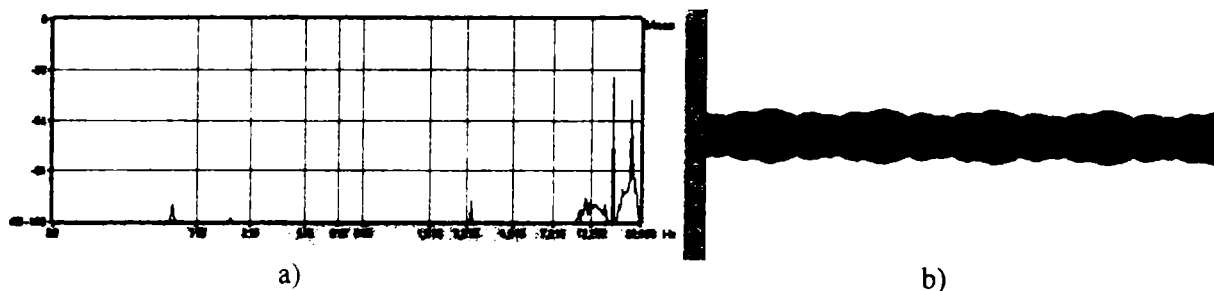


Figura 7.14 16 kHz (0 dB) + 19,5 kHz (-70 dB) 16bit, 44,1 kHz, semnal codat ATRAC, a 6-a generație, transfer digital, frecvență (a) și timp (b)

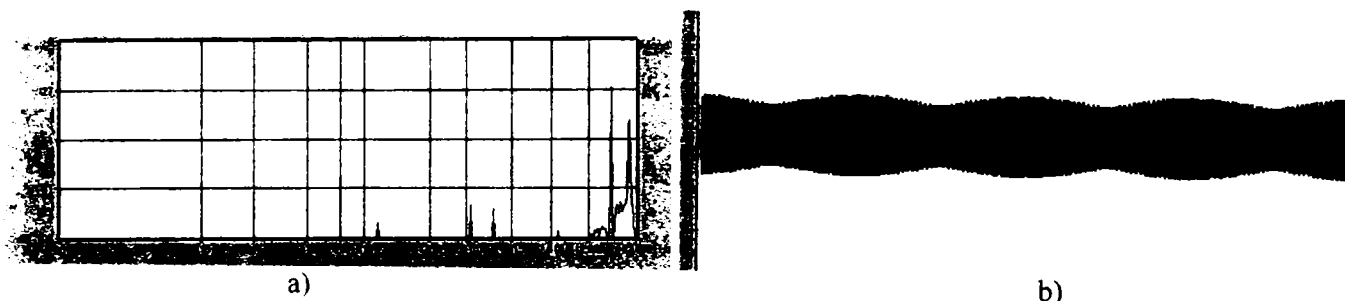


Figura 7.15 16 kHz (0 dB) + 19,5 kHz (-70 dB) 16 bit, 44,1 kHz, semnal codat ATRAC, a 2-a generație, transfer analogic, frecvență (a) și timp (b)

Testele efectuate conduc la concluzia că minidiscul, respectiv algoritmul ATRAC, este recomandabil pentru o gamă foarte largă de înregistrări, mai ales în domeniul semiprofesional. Având în vedere deteriorarea calității semnalului în cazul copiilor succesive, care devine sesizabilă în general de la a 6 generație (transfer digital), se recomandă ca minidiscul să se folosească pentru înregistrări primare (reportaje, interviuri, înregistrări muzicale simple), editări simple și un număr redus de copii succesive.

## 7.2 Evaluarea obiectivă a efectului frecvenței de eșantionare asupra calității audio percepute

Pentru această probă a fost aleasă o secvență de muzică funky cu durata de 16 s. Testele s-au efectuat cu metoda descrisă în capitolul 6, având ca referință fișierul cu frecvența de eșantionare de 44,1 kHz. Fișierele testate au fost eșantionate la următoarele frecvențe: 32 kHz, 22,05 kHz, 16 kHz, 11 kHz, 8 kHz și 6 kHz.

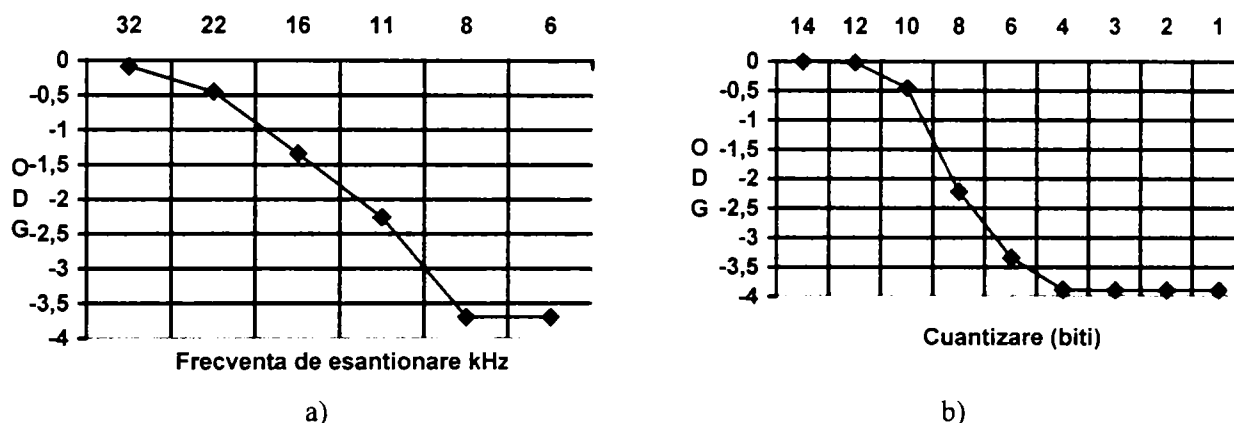


Figura 7.16 Efectul frecvenței de eșantionare (a) și al cuantizării (b) asupra calității audio

După cum se poate observa din figura 7.16.a, pentru frecvența de eșantionare de 32 kHz, diferența de calitate audio obiectivă față de referința de 44,1 kHz este insesizabilă (ODG = -0,09). De aceea frecvența de eșantionare de 32 kHz se folosește la radioul digital DAB (*Digital Audio Broadcasting*). Următoarea frecvență, 22,05 kHz, se încadrează în categoria sesizabil, dar nederanjant (ODG = -0,43), cu alte cuvinte “nu sună rău”. Începând cu frecvența de 16 kHz în jos, intrăm în domeniul vocal, deși valoarea ODG = -1,35 pentru un fragment muzical se încadrează la categoria “ușor deranjant”. Pentru un fișier muzical frecvența de eșantionare de 11 kHz este cam mică, calitatea audio evaluată, ODG = -2,27, încadrându-se la categoria deranjant. În sfârșit, ultimele două frecvențe, 8 kHz și 6 kHz intră la categoria bandă vocală limitată (telefonică), dând pentru fișierul muzical testat o valoare care se încadrează la categoria “foarte deranjant” (ODG = -3,68).

## 7.3 Evaluarea obiectivă a efectului cuantizării asupra calității audio percepute

Pentru această probă a fost aleasă aceeași secvență muzicală ca la proba anterioară. Testele s-au efectuat cu metoda descrisă în capitolul 6, având ca referință fișierul cu cuantizare pe 16 biți. Fișierele testate au fost cuantizate pe 14 biți, 12 biți, 10 biți, 8 biți, 6 biți, 4 biți, 3 biți, 2 biți și 1 bit.

Programul nostru de evaluare nu a sesizat diferențe audibile pentru primele două trepte de cuantizare, 14 și respectiv 12 biți (fig. 7.16.b). Chiar și pentru cuantizarea pe 10 biți valoarea ODG = -0,46 obținută se încadrează la categoria “sesizabil, dar nederanjant”. Cuantizarea pe 8 biți s-a folosit și se mai folosește acolo unde calitatea audio nu este foarte importantă, evaluarea calității încadrându-se la categoria “deranjant” (ODG = -2,23). Sub 6 biți, deteriorarea calității audio devine atât de deranjantă, încât iese din gama de măsură a metodei (ODG = -3,9).

## 7.4 Evaluarea obiectivă a calității audio pentru sistemele de codare folosite la transmisia prin Internet

### 7.4.1 Obiectiv

Scopul experimentelor executate a fost evaluarea obiectivă și compararea calității audio percepute a sistemelor de codare utilizate pentru transmisia audio prin Internet. Testarea tuturor codecurilor s-a făcut cu aceleași materiale, urmărindu-se efectul limbii, pentru o anumită schemă de codare, asupra calității audio și inteligibilității. De asemenea, s-a urmărit dacă există diferențe de calitate între codarea unei voci feminine sau masculine sau între calitatea unui material vocal sau muzical sau codarea unei voci singure în comparație cu o voce cu fundal muzical, care reprezintă modul cel mai uzual de transmitere a informației astăzi [STO97].

Au fost folosite următoarele codecuri:

1. MPEG-2 AAC (implementarea FhG-IIS)
2. MPEG-1 nivel-2
3. MPEG-1 nivel-3
4. RealNetworks Helix Producer 9
5. Microsoft Windows Media Technologies 9
6. Ogg Vorbis

Fiecare dintre aceste codecuri a fost testat la 5 debite binar diferite, corespunzătoare modemurilor folosite pe liniile de comunicații actuale:

- |                                  |                    |
|----------------------------------|--------------------|
| 1. Modem analogic 28,8 kbit/s    | - debit 16 kbit/s  |
| 2. Modem analogic 56 kbit/s      | - debit 32 kbit/s  |
| 3. ISDN (flux de bază) 64 kbit/s | - debit 48 kbit/s  |
| 4. ISDN dual 128 kbit/s          | - debit 96 kbit/s  |
| 5. DSL                           | - debit 128 kbit/s |

### 7.4.2 Semnale de test

Materialul folosit pentru teste a fost o selecție de fragmente de program obișnuite, neprocesate, care conțin texte citite de voci feminine și masculine în diferite limbi europene, fără și cu fundal muzical, și diferite genuri de muzică. Am folosit 8 materiale audio diferite, codate cu 6 sisteme de codare, durata secvențelor fiind cuprinsă între 11 s și 17 s (conform recomandărilor de la §4.3.1.B). Nu am folosit compresoare sau limitatoare, înregistrările digitale originale (referința) fiind liniare PCM, eșantionate la o frecvență de 44,1 kHz, 16 biți, mono sau stereo (.wav) [PFA03]. Pentru a putea aplica metoda de evaluare obiectivă a calității, materialele audio codate în diferite formate au fost convertite din nou în format .wav și reeșantionate cu frecvența de 44,1 kHz.

Materialele audio folosite și caracteristicile lor sunt prezentate în tabelul 7.1.

Tabelul 7.1 Materiale audio folosite la evaluarea calității la transmiterea prin Internet

Nr. crt.	Secvența audio	Conținut	Nume fișier	Dimensiune
1.	Muzică clasică	Brahms-Simfonia nr.1	brah_m/brah_s	1,21/2,43 MB
2.	Program radio	Voce masc. – l. daneză	denm_m/denm_s	1,17/2,35 MB
3.	Program radio	Voce fem.(daneză) + muz.	ebu_m/ebu_s	1,0/1,99 MB
4.	Transmisie directă	Comentariu sportiv	ice_m/ice_s	1,3/2,7 MB
5.	Program radio	Voce masc.(germ.) + muz.	irt_m/irt_s	1,14/2,29 MB
6.	Muzică de jazz	Lee Ritenour	jazz_m/jazz_s	1,23/2,46 MB
7.	Muzica pop	Fiesta	pop1_m/pop1_s	1,5/3,1 MB
8.	Muzica pop	Perfect World	pop2_m/pop2_s	1,32/2,64 MB

În figurile 7.17 – 7.24 sunt prezentate sonogramele celor 8 fragmente audio folosite la evaluarea calității audio în cazul transmiterii prin Internet. Aceste sonograme ne permit să avem o imagine asupra conținutului armonic și a complexității fragmentelor audio, de care depinde comportamentul și eficiența fiecărui sistem de codare audio.

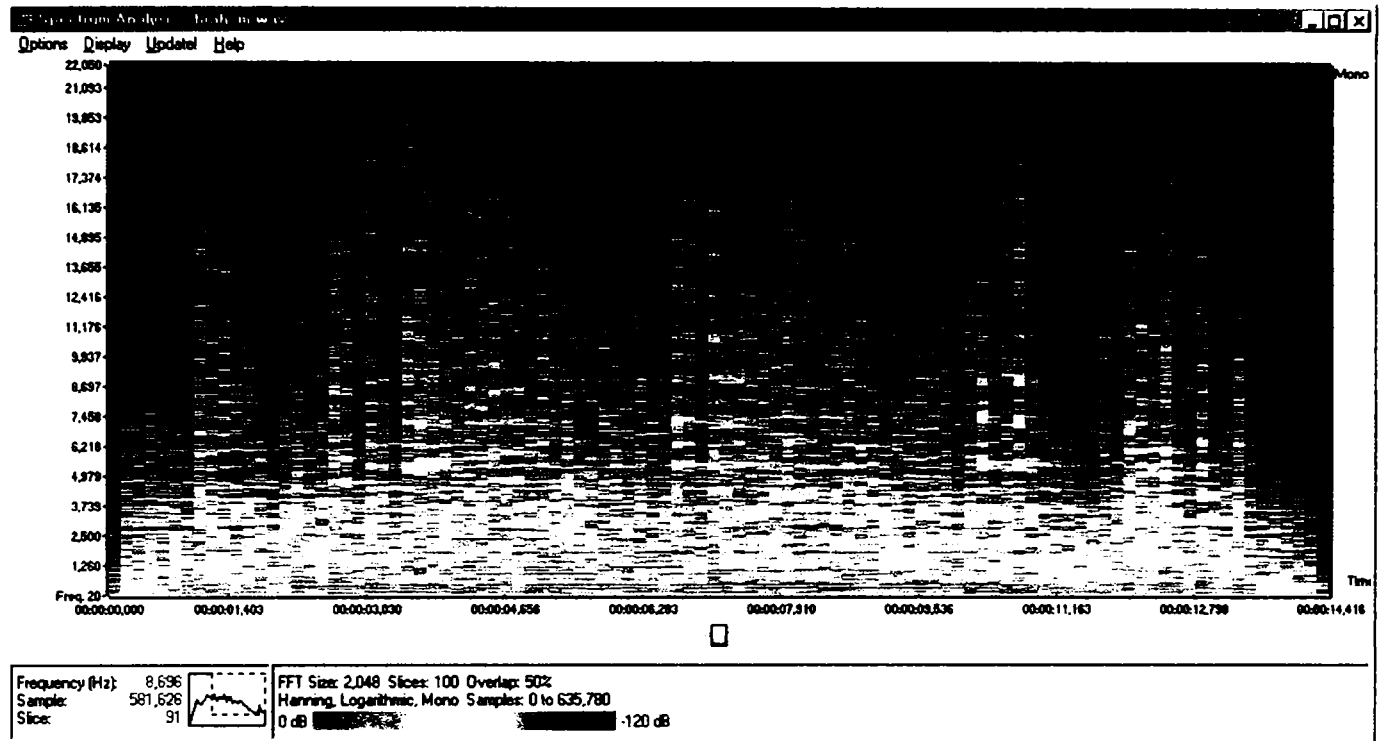


Figura 7.17 Sonograma fragmentului de muzică clasică *brah* - Johannes Brahms – Simfonia nr. 1

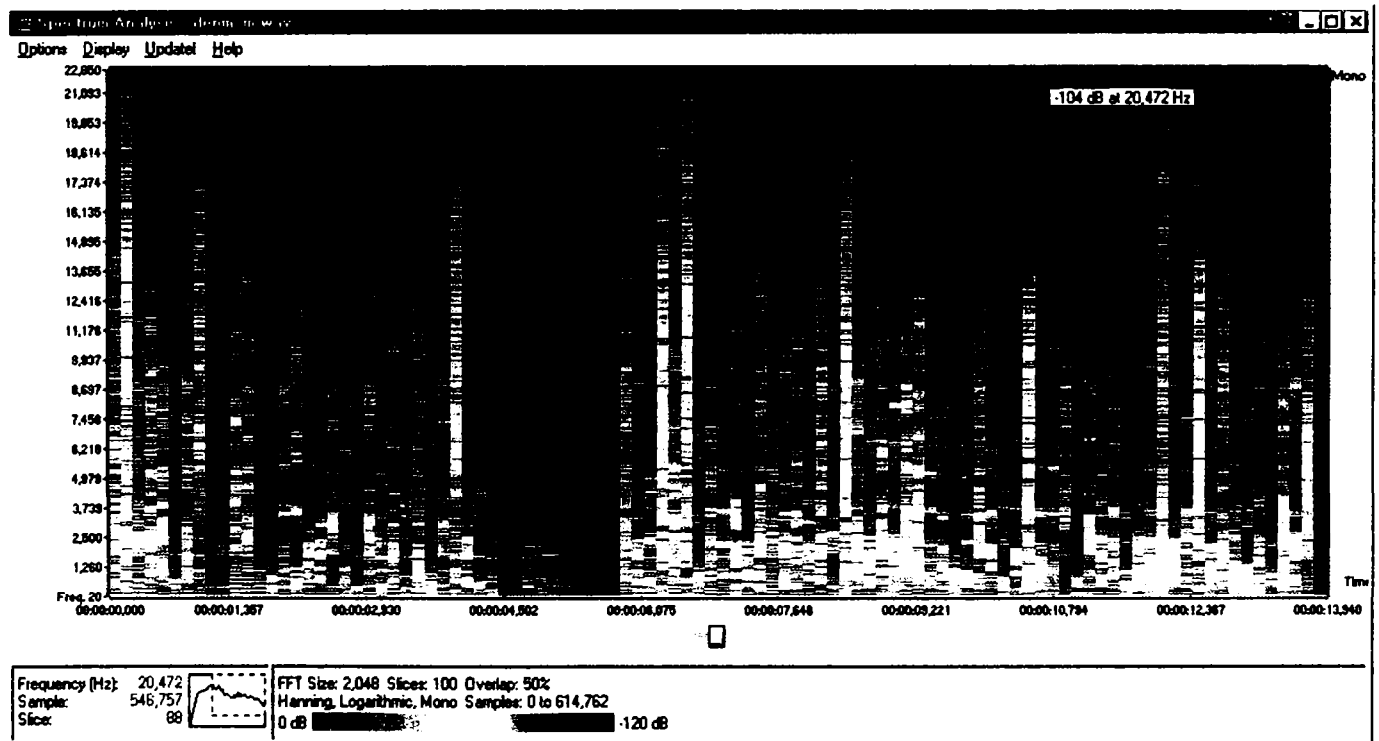


Figura 7.18 Sonograma fragmentului de program *denm* – voce masculină în limba daneză



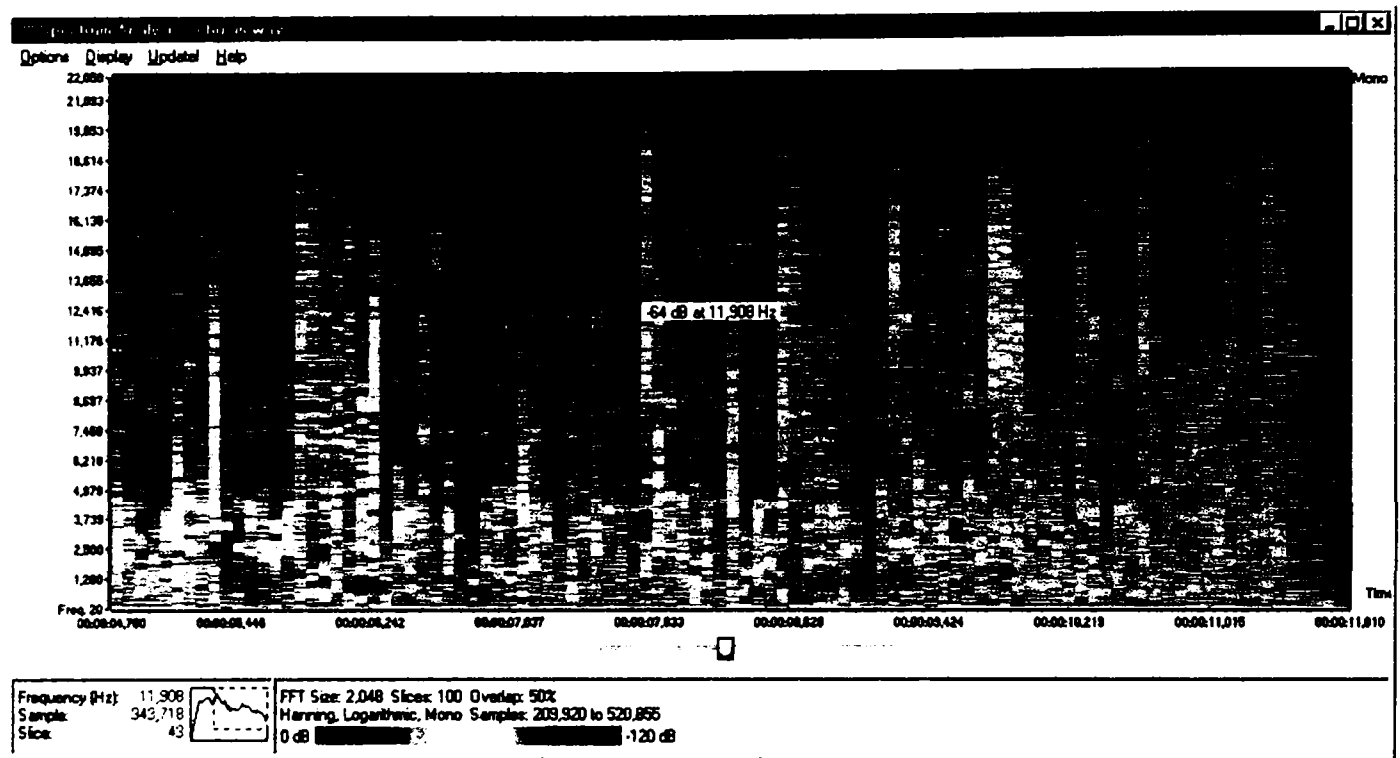


Figura 7.19 Sonograma fragmentului de program *ebu* – voce feminină în limba daneză, mixată cu muzică ritmată

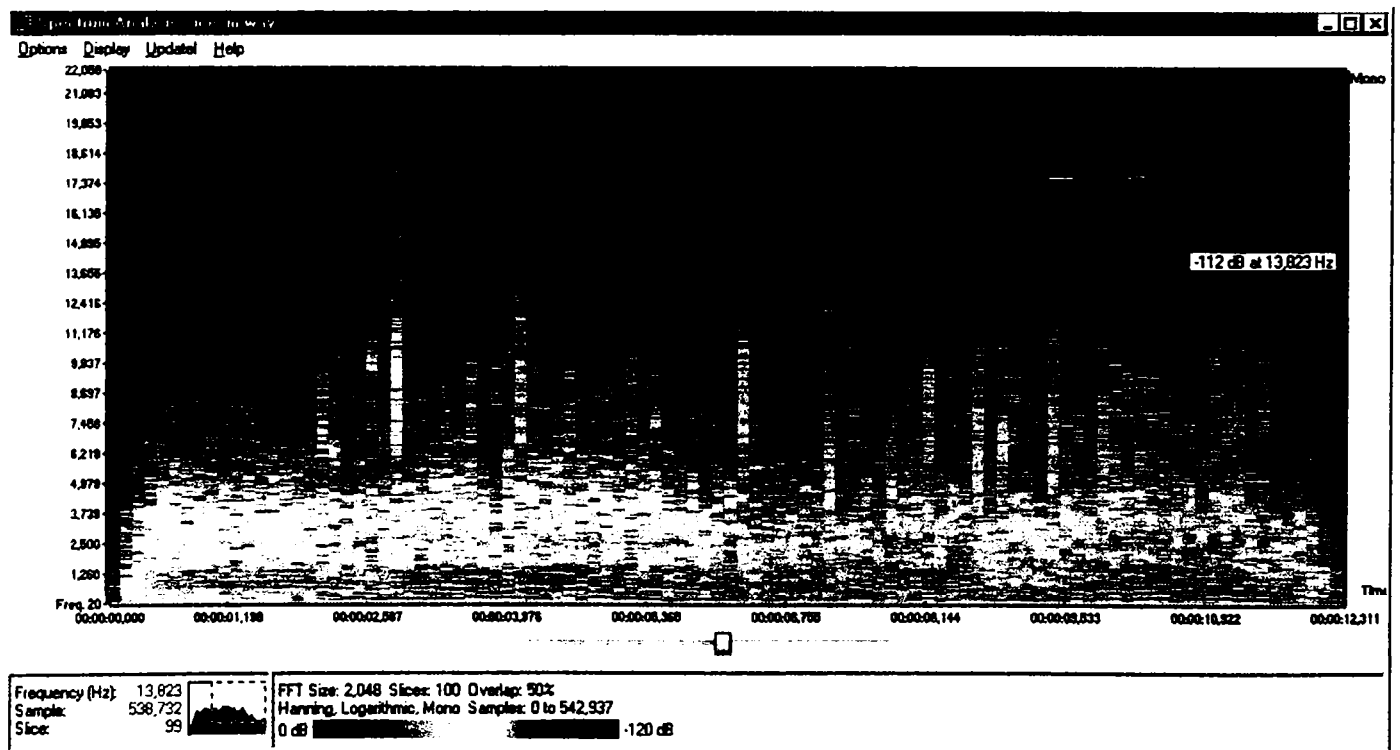


Figura 7.20 Sonograma fragmentului de transmisie sportivă în direct *ice* – comentariu voce feminină în limba engleză cu fundal de galerie suporteri

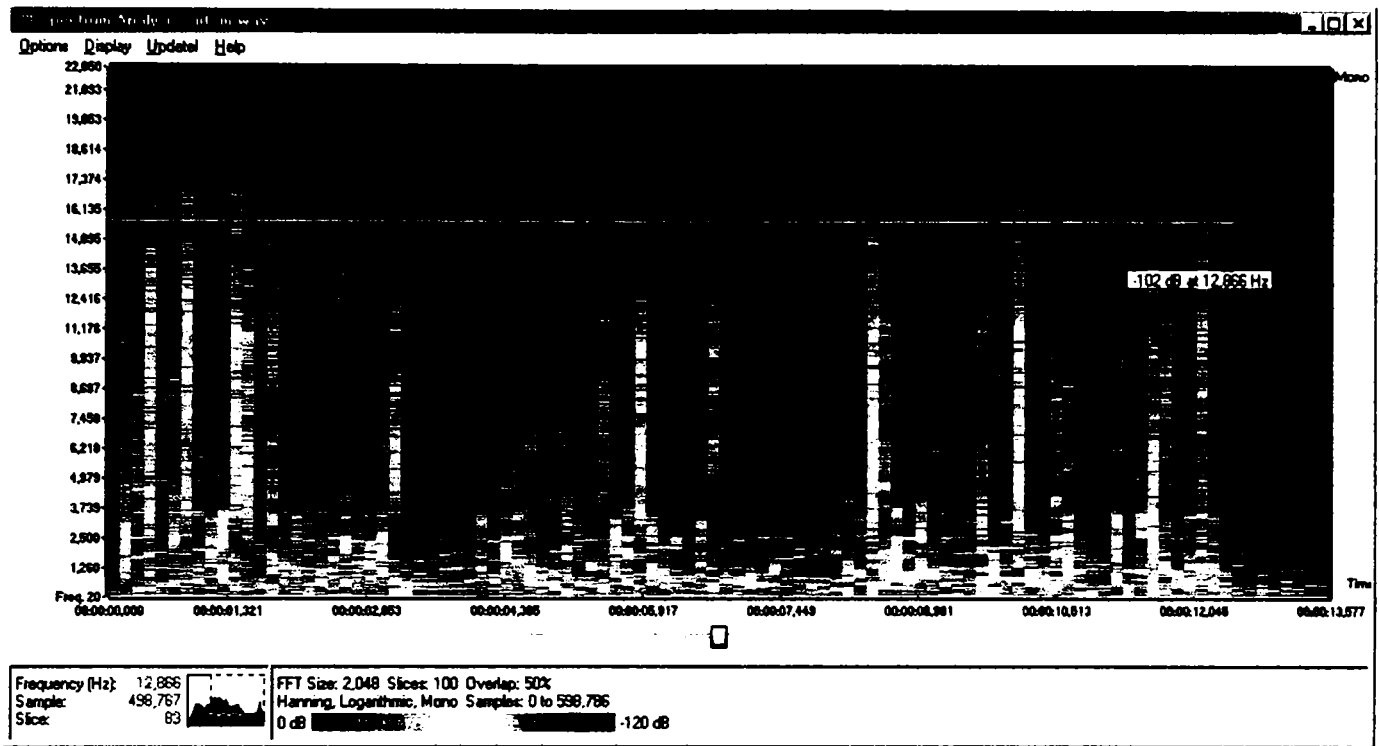


Figura 7.21 Sonograma fragmentului de program *irt* – voce masculină în limba germană cu fundal de muzică de flaut

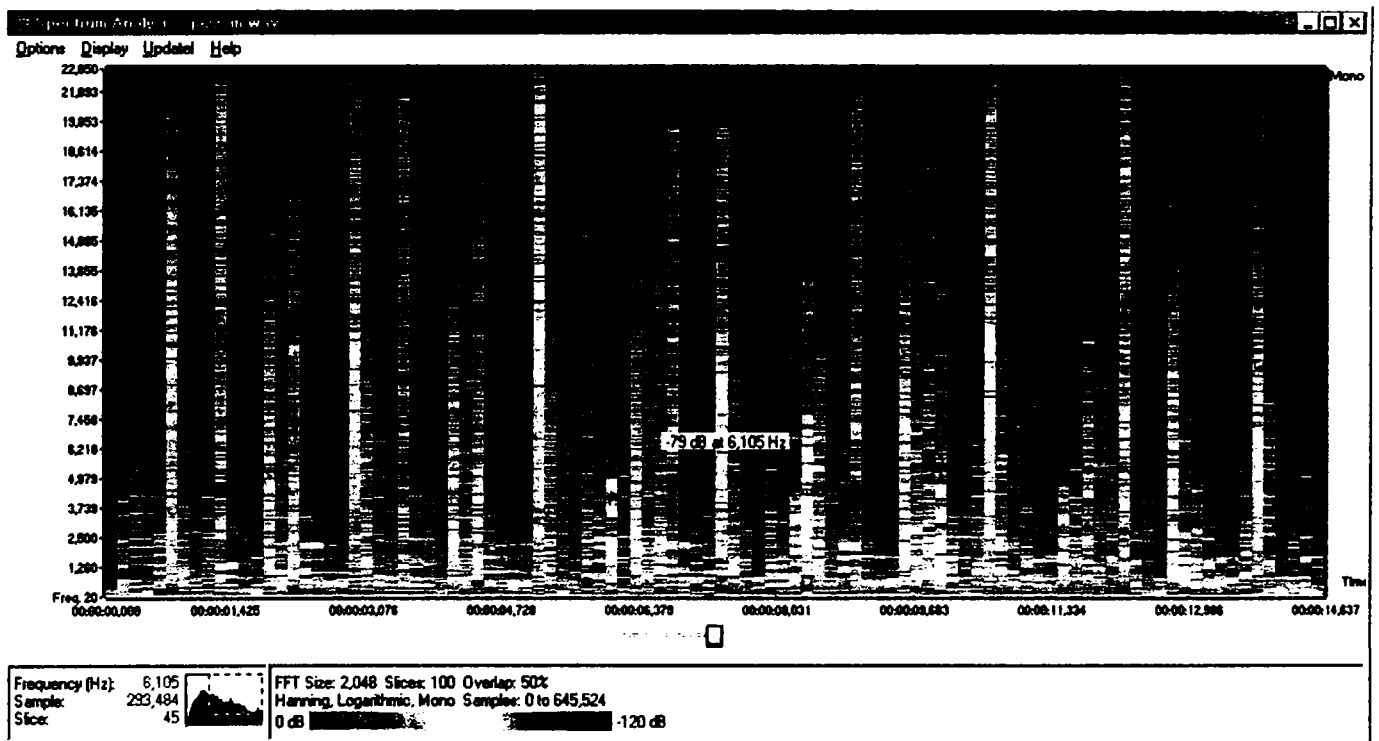


Figura 7.22 Sonograma fragmentului muzical *jazz* – Lee Ritenour

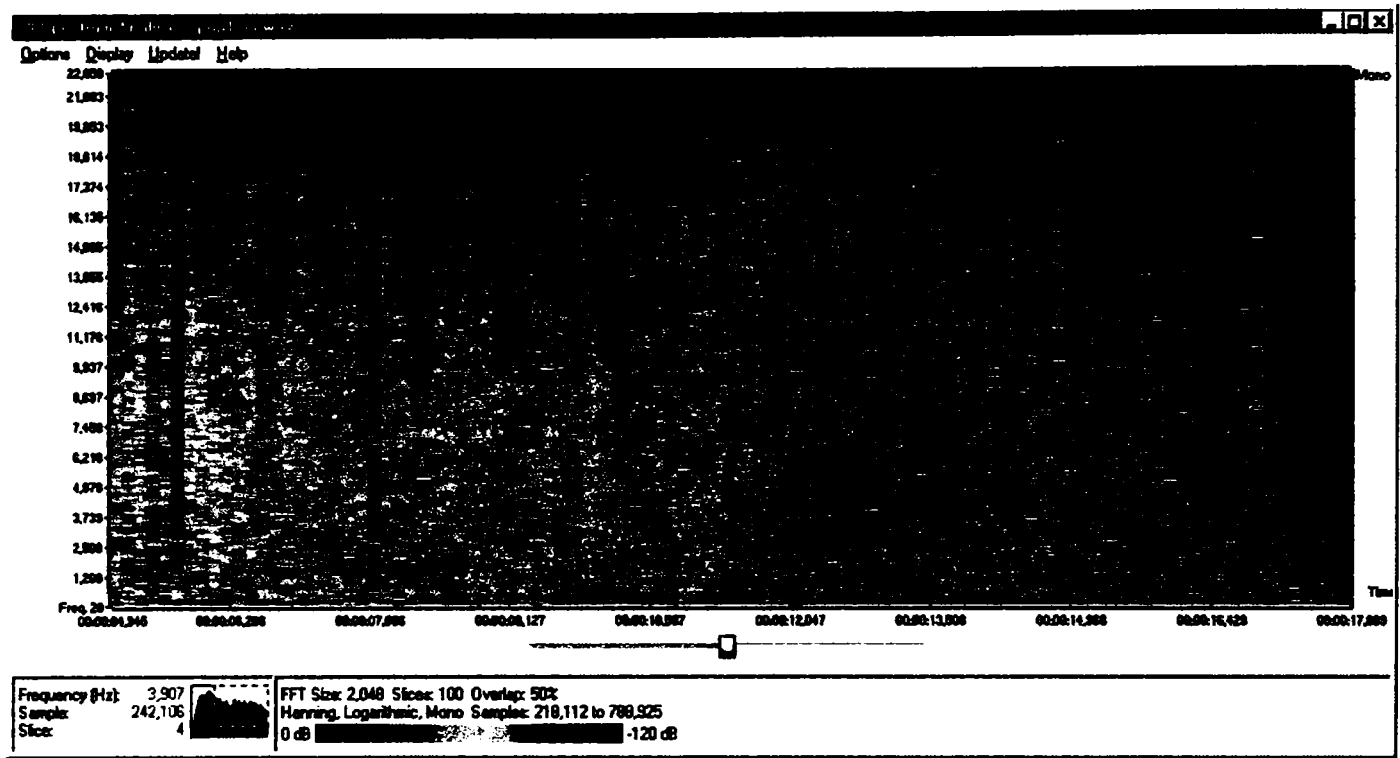


Figura 7.23 Sonograma fragmentului muzical *pop1* – Fiesta

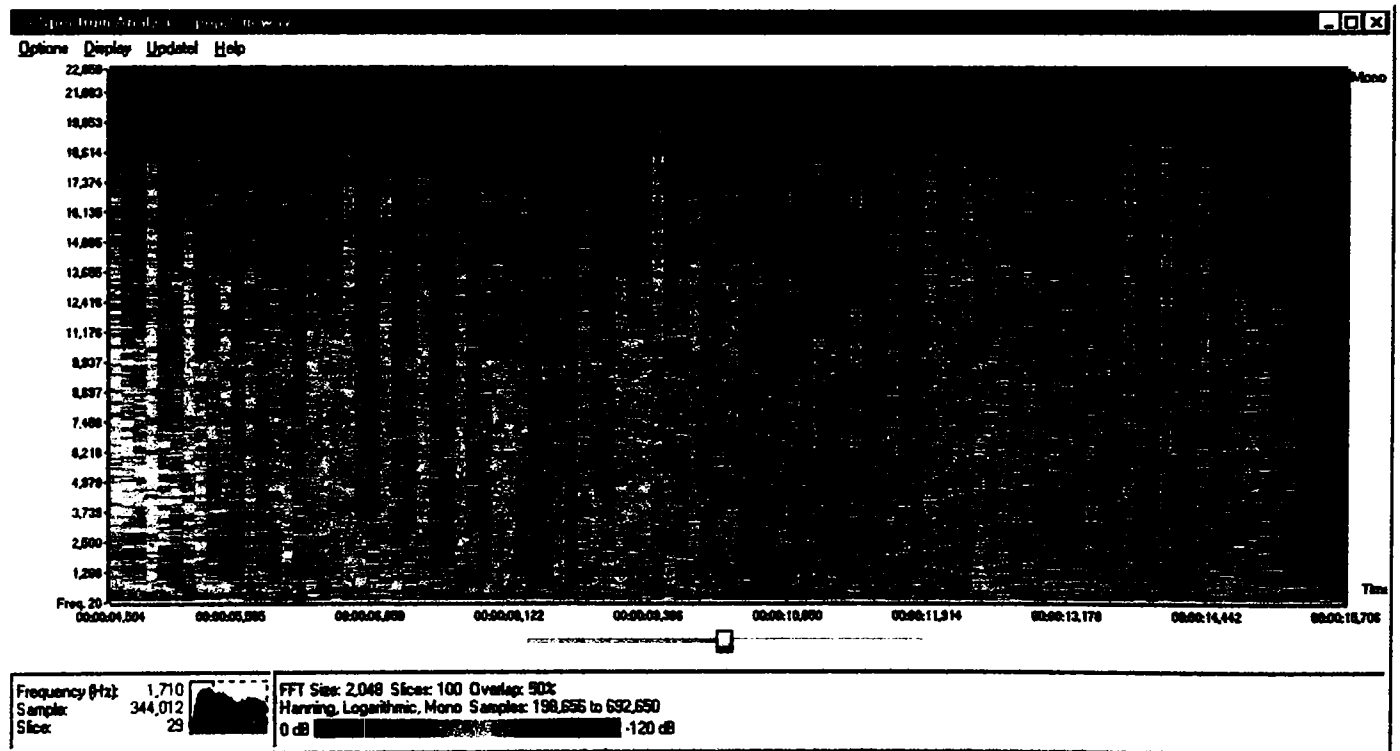


Figura 7.24 Sonograma fragmentului muzical *pop2* – Perfect World

### 7.4.3 Rezultatele testelor efectuate

#### ***A. Viteză foarte mică - Modem analogic, 28,8 kbit/s cu un debit de 16 kbit/s, mono***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 16 kbit/s a variantei mono a secvențelor de test sunt date în tabelul 7.2. După cum era de așteptat, calitatea audio este foarte scăzută, foarte deranjantă, cu un  $ODG_{\text{mediu}} = -3,73$ , codecul cu media cea mai bună a fost MP3 (-3,65), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,82). Valorile cele mai ridicate au fost obținute pentru secvența de muzică clasică, cu  $ODG_{\text{mediu}} = -3,41$  și un maxim pentru MS9 (-2,97). Diferențele dintre codecuri au fost de aprox. 1.0, adică -3,9 pentru cel mai rău și -2,9 pentru cel mai bun. În concluzie, la un debit binar de 16 kbit/s, codorul MPEG1 nivel-3 a fost evaluat cel mai bine la 4 din cele 8 secvențe de test mono, cu un conținut atât vocal cât și muzical. Totuși, având în vedere concentrarea rezultatelor în partea de jos a scalei de evaluare, am considerat că nici unul din algoritmi de codare testați nu este potrivit pentru acest debit redus.

#### ***B. Viteză foarte mică - Modem analogic, 28,8 kbit/s cu un debit de 16 kbit/s, stereo***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 16 kbit/s a variantei stereo a secvențelor de test sunt date în tabelul 7.3. După cum era de așteptat, calitatea audio este foarte scăzută, foarte deranjantă, cu un  $ODG_{\text{mediu}} = -3,78$ , mai scăzută decât varianta mono, codecul cu media cea mai bună a fost MS9 (-3,73), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,85). Valorile cele mai ridicate au fost obținute pentru secvența de muzică clasică, cu  $ODG_{\text{mediu}} = -3,57$  și un maxim pentru MS9 (-3,13). Diferențele dintre codecuri au fost de aprox. 0,8, adică -3,9 pentru cel mai rău și -3,1 pentru cel mai bun. În concluzie, la un debit binar de 16 kbit/s, codorul MPEG1 nivel-3 a fost evaluat cel mai bine la 4 din cele 8 secvențe de test stereo, cu un conținut atât vocal cât și muzical. Totuși, având în vedere concentrarea rezultatelor în partea de jos a scalei de evaluare, am considerat că nici unul din algoritmi de codare testați nu este potrivit pentru acest debit redus.

#### ***C. Viteză mică - Modem analogic, 56 kbit/s cu un debit de 32 kbit/s, mono***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 32 kbit/s a variantei mono a secvențelor de test sunt date în tabelul 7.4. După cum era de așteptat, calitatea audio este foarte scăzută, deranjantă, cu un  $ODG_{\text{mediu}} = -3,37$ , codecul cu media cea mai bună a fost MP3 (-2,84), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,65). Valorile cele mai ridicate au fost obținute pentru secvența de muzică clasică, cu două valori sub -3, MP3 (-2,72) și un maxim pentru MS9 (-2,68). Diferențele dintre codecuri au fost mai mari, de aprox. 1.2, adică -3,9 pentru cel mai rău și -2,7 pentru cel mai bun. În concluzie, la un debit binar de 32 kbit/s, codorul AAC a fost evaluat cel mai bine la 6 din cele 8 secvențe de test mono, cu un conținut atât vocal cât și muzical.

#### ***D. Viteză mică - Modem analogic, 56 kbit/s cu un debit de 32 kbit/s, stereo***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 32 kbit/s a variantei stereo a secvențelor de test sunt date în tabelul 7.5. După cum era de așteptat, calitatea audio este foarte scăzută, deranjantă, cu un  $ODG_{\text{mediu}} = -3,56$ , codecul cu media cea mai bună a fost MS9 (-3,25), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,89). Valorile cele mai ridicate au fost obținute pentru secvența de muzică clasică, cu  $ODG_{\text{mediu}} = -3,32$  și un maxim pentru MS9 (-2,28). Diferențele dintre codecuri au fost mai mari, de aprox. 1.6, adică -3,9 pentru cel mai rău și -2,3 pentru cel mai bun. Codorul MS9 a obținut cele mai bune evaluări la 4 din cele 8 secvențe de test, cu conținut atât vocal, cât și muzical, urmat de codorul AAC, care a fost cel mai bine cotate la cele 3 secvențe cu conținut predominant vocal.

***E. ISDN (flux de bază) – 64 kbit/s cu un debit de 48 kbit/s, mono***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 48 kbit/s a variantei mono a secvențelor de test sunt date în tabelul 7.6. Calitatea audio este scăzută, deranjantă, cu un  $ODG_{\text{mediu}} = -2,845$ , codecul cu media cea mai bună a fost AAC (-2,04), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,59). Valorile cele mai ridicate au fost obținute pentru secvența de comentariu sportiv, un  $ODG_{\text{mediu}} = -2,26$  și un maxim pentru AAC (-1,9). Diferențele dintre codecuri au fost mai mari, de aprox. 2, adică -3,9 pentru cel mai rău și -1,9 pentru cel mai bun. În concluzie, la un debit binar de 48 kbit/s, codorul AAC a fost evaluat cel mai bine la 5 din cele 8 secvențe de test mono, cu un conținut atât vocal cât și muzical.

***F. ISDN (flux de bază) – 64 kbit/s cu un debit de 48 kbit/s, stereo***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 48 kbit/s a variantei stereo a secvențelor de test sunt date în tabelul 7.7. Calitatea audio este scăzută, deranjantă, cu un  $ODG_{\text{mediu}} = -2,86$ , codecul cu media cea mai bună a fost AAC (-0,638), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,88). Pentru prima dată un codec obține o evaluare din categoria “sesizabil, dar nederanjant”. Valorile cele mai ridicate au fost obținute pentru secvența de program cu voce masculină în limba daneză, cu  $ODG_{\text{mediu}} = -2,41$ . Maximul acestei secvențe de evaluări a fost obținut de AAC (-0,5) și în mod foarte interesant, pentru fragmentul de jazz, unul dintre cele mai critice materiale audio. În concluzie, la un debit binar de 48 kbit/s, codorul AAC a fost evaluat cel mai bine la toate cele 8 secvențe de test stereo.

***G. ISDN dual 128 kbit/s cu un debit de 96 kbit/s, stereo***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 96 kbit/s a variantei stereo a secvențelor de test sunt date în tabelul 7.8. Se observă o creștere semnificativă a calității, cu un  $ODG_{\text{mediu}} = -2,23$ , codecurile cu media cea mai bună fiind AAC și MS9 (-1,33), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,84). Valorile cele mai ridicate au fost obținute pentru secvența de program cu voce masculină în limba daneză, cu  $ODG_{\text{mediu}} = -1,49$ , cu un maxim obținut de AAC (-0,28). În concluzie, la un debit binar de 96 kbit/s, codorul AAC a fost evaluat cel mai bine la 5 din cele 8 secvențe de test, cu un conținut atât vocal cât și muzical.

***H. Viteză mare - DSL cu un debit de 128 kbit/s, stereo***

Rezultatele evaluării obiective a calității audio pentru codare cu un debit de 128 kbit/s a variantei stereo a secvențelor de test sunt date în tabelul 7.9. Se observă o creștere semnificativă a calității, cu un  $ODG_{\text{mediu}} = -1,734$ , codecul cu media cea mai bună fiind MP3 (-0,498), iar media cea mai slabă îi aparține lui Real Audio G2 (-3,83). Valorile cele mai ridicate au fost obținute pentru secvența de program cu voce masculină în limba daneză, cu  $ODG_{\text{mediu}} = -1,16$ , cu un maxim obținut de MP3 (-0,035). În concluzie, la un debit binar de 128 kbit/s, codorul MPEG1 nivel-3 a fost evaluat cel mai bine la 6 din cele 8 secvențe de test, cu un conținut atât vocal cât și muzical. Aceste rezultate sunt în concordanță cu practica, sistemul de codare .mp3 fiind cel mai utilizat la ora actuală la acest debit.

Tabelul 7.2 Evaluarea calității audio pentru un semnal mono la un debit de 16 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier [kB]	Debit [kbit/s]	Raport de compresie	ODG
<i>brah_m</i>	AAC	14 s	24 kHz	28.4	16	43:1	-3,322
	MPEG1, L2		16 kHz	28.1	16	44:1	-3,736
	MPEG1, L3		11 kHz	28.4	16	43:1	-3,247
	Real Audio G2		22,05 kHz	32.2	19	38:1	-3,780
	MS9		16 kHz	33.8	19	37:1	-2,975
<i>denm_m</i>	AAC	13 s	24 kHz	27.6	17	43:1	-3,789
	MPEG1, L2		16 kHz	27.2	17	44:1	-3,666
	MPEG1, L3		11 kHz	27.5	17	43:1	-3,882
	Real Audio G2		22,05 kHz	32.2	19	37:1	-3,765
	MS9		16 kHz	33.1	19	36:1	-3,851
<i>ebu_m</i>	AAC	11 s	24 kHz	23.4	17	43:1	-3,768
	MPEG1, L2		16 kHz	23.0	17	44:1	-3,835
	MPEG1, L3		11 kHz	23.3	17	44:1	-3,892
	Real Audio G2		22,05 kHz	27.7	20	37:1	-3,908
	MS9		16 kHz	28.5	21	36:1	-3,852
<i>ice_m</i>	AAC	12 s	24 kHz	24.4	16	54:1	-
	MPEG1, L2		16 kHz	24.0	16	55:1	-3,701
	MPEG1, L3		11 kHz	24.3	16	54:1	-3,369
	Real Audio G2		22,05 kHz	27.7	19	48:1	-3,595
	MS9		16 kHz	29.8	20	44:1	-3,438
<i>irt_m</i>	AAC	13 s	24 kHz	26.9	17	43:1	-3,849
	MPEG1, L2		16 kHz	26.5	17	44:1	-3,734
	MPEG1, L3		11 kHz	26.8	17	43:1	-3,760
	Real Audio G2		22,05 kHz	32.2	19	36:1	-3,801
	MS9		16 kHz	31.1	18	37:1	-3,870
<i>jazz_m</i>	AAC	14 s	24 kHz	29.0	17	43:1	-3,891
	MPEG1, L2		16 kHz	28.5	17	44:1	-3,802
	MPEG1, L3		11 kHz	28.8	17	44:1	-3,767
	Real Audio G2		22,05 kHz	32.2	19	39:1	-3,908
	MS9		16 kHz	33.8	19	37:1	-3,807
<i>pop1_m</i>	AAC	17 s	24 kHz	35.4	17	43:1	-3,742
	MPEG1, L2		16 kHz	35.0	17	44:1	-3,841
	MPEG1, L3		11 kHz	35.1	17	44:1	-3,698
	Real Audio G2		22,05 kHz	41.2	18	37:1	-3,896
	MS9		16 kHz	41.0	18	37:1	-3,766
<i>pop2_m</i>	AAC	15 s	24 kHz	31.0	17	44:1	-3,871
	MPEG1, L2		16 kHz	30.6	17	44:1	-3,801
	MPEG1, L3		11 kHz	30.9	17	44:1	-3,629
	Real Audio G2		22,05 kHz	36.7	19	37:1	-3,906
	MS9		16 kHz	36.4	19	37:1	-3,875

Codorul cel mai bun: 1. MPEG1 nivel-3 – *ice, jazz, pop1, pop2*  
2. AAC – *ebu*; MPEG1 nivel-2 – *denm*; MS9 - *brah*

Tabelul 7.3 Evaluarea calității audio pentru un semnal stereo la un debit de 16 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier [kB]	Debit [kbit/s]	Raport de compresie	ODG
<i>brah_s</i>	AAC	14 s	8 kHz	28.8	17	86:1	-3.667
	MPEG1, L3		11 kHz	31.9	18	78:1	-3.617
	RA 8		22,05 kHz	34.3	20	72.5:1	-3.879
	MS9		16 kHz	33.8	19	74:1	<b>-3.131</b>
<i>denm_s</i>	AAC	13 s	8 kHz	28.1	17	86:1	-3.802
	MPEG1, L3		11 kHz	31.0	19	78:1	<b>-3.771</b>
	RA 8		22,05 kHz	34.3	22	70:1	-3.887
	MS9		16 kHz	32.5	20	74:1	-3.854
<i>ebu_s</i>	AAC	11 s	8 kHz	23.8	18	86:1	-3.878
	MPEG1, L3		11 kHz	26.5	20	77:1	-3.826
	RA 8		22,05 kHz	29.5	22	69:1	-3.910
	MS9		16 kHz	28.5	21	71.5:1	<b>-3.487</b>
<i>ice_s</i>	AAC	12 s	8 kHz	24.8	17	86:1	<b>-3.642</b>
	MPEG1, L3		11 kHz	27.4	19	78:1	-3.699
	RA 8		22,05 kHz	29.5	20	72:1	-3.905
	MS9		16 kHz	29.8	20	71.5:1	-3.870
<i>irt_s</i>	AAC	13 s	8 kHz	27.2	17	86:1	-3.871
	MPEG1, L3		11 kHz	30.1	19	78:1	-3.854
	RA 8		22,05 kHz	29.5	19	79.5:1	<b>-3.844</b>
	MS9		16 kHz	31.1	20	75.5:1	-3.877
<i>jazz_s</i>	AAC	14 s	8 kHz	29.2	17	87:1	-3.857
	MPEG1, L3		11 kHz	32.4	19	78:1	<b>-3.793</b>
	RA 8		22,05 kHz	34.3	20	74:1	-3.909
	MS9		16 kHz	33.8	20	75:1	-3.870
<i>pop1_s</i>	AAC	17 s	8 kHz	35.6	17	87:1	-3.884
	MPEG1, L3		11 kHz	39.7	19	78:1	<b>-3.816</b>
	RA 8		22,05 kHz	39.2	19	79:1	-3.907
	MS9		16 kHz	41.0	20	75:1	-3.868
<i>pop2_s</i>	AAC	15 s	8 kHz	31.3	17	87:1	-3.811
	MPEG1, L3		11 kHz	34.7	19	78:1	<b>-3.776</b>
	RA 8		22,05 kHz	36.7	20	74:1	-3.910
	MS9		16 kHz	36.4	20	74.5:1	-3.887

Codorul cel mai bun: 1. MPEG1 nivel-3 – *denm, jazz, pop1, pop2*  
2. MS9 – *brah, ebu*  
3. AAC – *ice*; RA8 - *irt*

Tabelul 7.4 Evaluarea calității audio pentru un semnal mono la un debit de 32 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier kB	Debit kbit/s	Raport de compresie	ODG
<i>brah_m</i>	AAC	14 s	32 kHz	56.9	33	22:1	-3.173
	MPEG1, L2		44,1 kHz	56.4	33	22:1	-3.481
	MPEG1, L3		22,05 kHz	56.5	33	22:1	-2.725
	Real Audio G2		44,1 kHz	64.2	37.5	19,5:1	-3.798
	MS9		44,1 kHz	64.6	38	19:1	<b>-2.680</b>
<i>denm_m</i>	AAC	13 s	32 kHz	54.9	34.5	22:1	-3.443
	MPEG1, L2		44,1 kHz	54.5	34	22:1	-3.348
	MPEG1, L3		22,05 kHz	54.6	34	22:1	<b>-2.711</b>
	Real Audio G2		44,1 kHz	64.2	40	19:1	-3.362
	MS9		44,1 kHz	61.6	39	19,5:1	-3.657
<i>ebu_m</i>	AAC	11 s	32 kHz	46.5	35	22:1	-3.491
	MPEG1, L2		44,1 kHz	46.2	34	22:1	-3.848
	MPEG1, L3		22,05 kHz	46.4	35	22:1	<b>-2.978</b>
	Real Audio G2		44,1 kHz	55.1	41	18,5:1	-3.906
	MS9		44,1 kHz	52.7	39	19:1	-3.741
<i>ice_m</i>	AAC	12 s	32 kHz	48.6	33	22:1	-3.238
	MPEG1, L2		44,1 kHz	48.2	33	22:1	-3.238
	MPEG1, L3		22,05 kHz	48.3	33	22:1	-2.804
	Real Audio G2		44,1 kHz	55.1	38	19:1	<b>-2.795</b>
	MS9		44,1 kHz	57.2	39	19:1	-3.170
<i>irt_m</i>	AAC	13 s	32 kHz	53.3	33.5	22:1	-3.413
	MPEG1, L2		44,1 kHz	53.1	33	22:1	-3.410
	MPEG1, L3		22,05 kHz	53.3	33.5	22:1	<b>-2.781</b>
	Real Audio G2		44,1 kHz	55.1	35	21:1	-3.786
	MS9		44,1 kHz	57.2	36	20,5:1	-3.598
<i>jazz_m</i>	AAC	14 s	32 kHz	57.6	34	22:1	-3.324
	MPEG1, L2		44,1 kHz	57.3	33.5	22:1	-3.804
	MPEG1, L3		22,05 kHz	57.3	33.5	22:1	<b>-2.689</b>
	Real Audio G2		44,1 kHz	64.2	37.5	20:1	-3.904
	MS9		44,1 kHz	64.6	38	19,5:1	-3.218
<i>pop1_m</i>	AAC	17 s	32 kHz	70.3	34	22:1	-3.218
	MPEG1, L2		44,1 kHz	70.0	34	22:1	-3.749
	MPEG1, L3		22,05 kHz	70.0	34	22:1	<b>-3.044</b>
	Real9		44,1 kHz	73.2	35	21:1	-3.893
	MS9		44,1 kHz	77.9	37.5	20:1	-3.729
<i>pop2_m</i>	AAC	15 s	32 kHz	61.7	34	22:1	-3.313
	MPEG1, L2		44,1 kHz	61.5	34	22:1	-3.789
	MPEG1, L3		22,05 kHz	61.5	34	22:1	<b>-3.010</b>
	Real Audio G2		44,1 kHz	64.2	35	21:1	-3.789
	MS9		44,1 kHz	69.0	38	20:1	-3.773

Codorul cel mai bun: 1. MPEG1 nivel-3 – *denm, ebu, irt, jazz, pop1, pop2*  
2. RA G2 – *ice*; MS9 - *brah*



Tabelul 7.5 Evaluarea calității audio pentru un semnal stereo la un debit de 32 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier kB	Debit kbit/s	Raport de compresie	ODG
<i>brah_s</i>	AAC	14 s	24 kHz	57.4	33	43:1	-3.343
	MPEG1, L2		22,05 kHz	56.3	33	44:1	-3.823
	MPEG1, L3		11 kHz	56.8	33	44:1	-3.278
	RA8		22,05 kHz	60.4	35	41:1	-3.886
	MS9		32 kHz	68.1	39	37:1	<b>-2.285</b>
<i>denm_s</i>	AAC	13 s	24 kHz	55.3	35	43,5:1	<b>-3.031</b>
	MPEG1, L2		22,05 kHz	54.4	34	44:1	-3.742
	MPEG1, L3		11 kHz	55	34	44:1	-3.451
	RA8		22,05 kHz	60.4	38	40:1	-3.875
	MS9		32 kHz	65.11	41	37:1	-3.398
<i>ebu_s</i>	AAC	11 s	24 kHz	47.0	34	43,5:1	<b>-3.227</b>
	MPEG1, L2		22,05 kHz	46.1	34	44:1	-3.900
	MPEG1, L3		11 kHz	46.7	34	44:1	-3.597
	RA8		22,05 kHz	52.9	39	39:1	-3.909
	MS9		32 kHz	55.9	41	36,5:1	-3.438
<i>ice_s</i>	AAC	12 s	24 kHz	48.8	33	44:1	-2.970
	MPEG1, L2		22,05 kHz	48.1	33	44:1	-3.796
	MPEG1, L3		11 kHz	48.7	33	44:1	-3.31
	RA8		22,05 kHz	52.9	35	41:1	-3.886
	MS9		32 kHz	58.9	40	36:1	<b>-2.919</b>
<i>irt_s</i>	AAC	13 s	24 kHz	53.5	34	44:1	<b>-3.540</b>
	MPEG1, L2		22,05 kHz	53.0	34	44:1	-3.797
	MPEG1, L3		11 kHz	58.7	34	44:1	-3.718
	RA8		22,05 kHz	60.4	38	39:1	-3.841
	MS9		32 kHz	60.5	38	39:1	-3.645
<i>jazz_s</i>	AAC	14 s	24 kHz	57.8	33	44:1	-3.464
	MPEG1, L2		22,05 kHz	57.3	34	44:1	-3.869
	MPEG1, L3		11 kHz	57.6	33	44:1	-3.616
	RA8		22,05 kHz	60.4	35	42:1	-3.907
	MS9		32 kHz	66.6	39	38:1	<b>-3.278</b>
<i>pop1_s</i>	AAC	17 s	24 kHz	70.7	33	44:1	-3.744
	MPEG1, L2		22,05 kHz	70.0	34	44:1	-3.853
	MPEG1, L3		11 kHz	70.2	33	44:1	-3.608
	RA8		22,05 kHz	75.3	36	41:1	-3.903
	MS9		32 kHz	80.4	39	38,5:1	<b>-3.429</b>
<i>pop2_s</i>	AAC	15 s	24 kHz	48	33	-	-3.577
	MPEG1, L2		22,05 kHz	61.4	34	44:1	-3.872
	MPEG1, L3		11 kHz	61.7	33	44:1	<b>-3.339</b>
	RA8		22,05 kHz	67.8	37	40:1	-3.908
	MS9		32 kHz	71.2	39	3833:1	-3.620

Codorul cel mai bun: 1. MS9 – *brah, ice, jazz, pop1*  
2. AAC – *denm, ebu, irt*  
3. MPEG1 nivel-3 – *pop2*

Tabelul 7.6 Evaluarea calității audio pentru un semnal mono la un debit de 48 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier kB	Debit kbit/s	Raport de compresie	ODG
<i>brah_m</i>	AAC	14 s	32 kHz	85.1	49	14:1	-1.951
	MPEG1, L2		44,1 kHz	84.6	50	15:1	-3.243
	MPEG1, L3		44,1 kHz	84.5	48	15:1	-2.420
	Real Audio G2		44,1 kHz	82.2	48	15:1	-3.754
	MS9		44,1 kHz	97.2	57	13:1	<b>-1.354</b>
<i>denm_m</i>	AAC	13 s	32 kHz	82.3	51	14:1	-2.175
	MPEG1, L2		44,1 kHz	81.8	52	15:1	-2.995
	MPEG1, L3		44,1 kHz	81.8	50	15:1	<b>-2.120</b>
	Real Audio G2		44,1 kHz	90.6	57	13:1	-3.274
	MS9		44,1 kHz	84	53	14:1	-2.759
<i>ebu_m</i>	AAC	11 s	32 kHz	69.8	51	15:1	<b>-2.203</b>
	MPEG1, L2		44,1 kHz	69.3	52	15:1	-3.84
	MPEG1, L3		44,1 kHz	69.4	51	15:1	-2.36
	Real Audio G2		44,1 kHz	72.0	54	14:1	-3.904
	MS9		44,1 kHz	77.4	57	13:1	-3.096
<i>ice_m</i>	AAC	12 s	32 kHz	72.7	50	15:1	<b>-1.899</b>
	MPEG1, L2		44,1 kHz	72.3	49	15:1	-2.875
	MPEG1, L3		44,1 kHz	72.3	49	15:1	-2.015
	Real Audio G2		44,1 kHz	72.0	49	15:1	-2.402
	MS9		44,1 kHz	84	57	13:1	-2.112
<i>irt_m</i>	AAC	13 s	32 kHz	80.0	49	15:1	-2.501
	MPEG1, L2		44,1 kHz	79.7	50	15:1	-3.171
	MPEG1, L3		44,1 kHz	79.6	49	15:1	<b>-2.339</b>
	Real Audio G2		44,1 kHz	82.2	52	14:1	-3.745
	MS9		44,1 kHz	84	53	14:1	-2.736
<i>jazz_m</i>	AAC	14 s	32 kHz	86.4	49	15:1	<b>-1.733</b>
	MPEG1, L2		44,1 kHz	86.0	50	15:1	-3.805
	MPEG1, L3		44,1 kHz	85.9	49	15:1	-2.095
	Real Audio G2		44,1 kHz	82.2	48	15:1	-3.903
	MS9		44,1 kHz	95	55	13:1	-2.042
<i>pop1_m</i>	AAC	17 s	32 kHz	105	49	15:1	<b>-1.853</b>
	MPEG1, L2		44,1 kHz	105	51	15:1	-3.702
	MPEG1, L3		44,1 kHz	104	49	15:1	-2.211
	Real Audio G2		44,1 kHz	102	49	15:1	-3.883
	MS9		44,1 kHz	117	56	13:1	-3.043
<i>pop2_m</i>	AAC	15 s	32 kHz	92.7	49	15:1	<b>-2.055</b>
	MPEG1, L2		44,1 kHz	92.2	50	15:1	-3.744
	MPEG1, L3		44,1 kHz	92.1	49	15:1	-2.146
	Real Audio G2		44,1 kHz	92.5	48	15:1	-3.900
	MS9		44,1 kHz	101	55	13:1	-3.176

- Codorul cel mai bun:
1. AAC – *ebu, ice, jazz, pop1, pop2*
  2. MPEG1 nivel-3 – *denm, irt*
  3. MS9 - *brah*

Tabelul 7.7 Evaluarea calității audio pentru un semnal stereo la un debit de 48 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier kB	Debit real kbit/s	Raport de compresie	ODG
<i>brah_s</i>	AAC	14 s	32 kHz	84.8	49	29:1	-1.162
	MPEG1, L2			84.6	50	29:1	-3.857
	MPEG1, L3		22,05 kHz	84.8	49	29:1	-3.017
	Real Audio G2		22,05 kHz	101	59	25:1	-3.796
	MS9		44,1 kHz	97.2	57	26:1	-2.953
<i>denm_s</i>	AAC	13 s	32 kHz	82.2	51	29:1	-0.046
	MPEG1, L2			81.8	52	29:1	-3.815
	MPEG1, L3		22,05 kHz	81.9	50	29:1	-2.424
	Real Audio G2		22,05 kHz	82.2	52	29:1	-3.886
	MS9		44,1 kHz	95	60	25:1	-1.908
<i>ebu_s</i>	AAC	11 s	32 kHz	69.6	49	29:1	-0.216
	MPEG1, L2			69.3	52	29:1	-3.904
	MPEG1, L3		22,05 kHz	69.6	52	29:1	-2.723
	Real Audio G2		22,05 kHz	72	54	28:1	-3.908
	MS9		44,1 kHz	81.8	52	25:1	-3.227
<i>ice_s</i>	AAC	12 s	32 kHz	72.7	49	29:1	-0.693
	MPEG1, L2			72.3	49	29:1	-3.813
	MPEG1, L3		22,05 kHz	72.5	48	29:1	-2.764
	Real Audio G2		22,05 kHz	72	49	29:1	-3.898
	MS9		44,1 kHz	90.6	62	23:1	-2.800
<i>irt_s</i>	AAC	13 s	32 kHz	79.7	49	29:1	-0.665
	MPEG1, L2			79.7	50	29:1	-3.854
	MPEG1, L3		22,05 kHz	79.9	49	29:1	-2.841
	Real Audio G2		22,05 kHz	82.2	52	29:1	-3.865
	MS9		44,1 kHz	90.6	57	26:1	-3.486
<i>jazz_s</i>	AAC	14 s	32 kHz	86.0	49	29:1	-0.509
	MPEG1, L2			86.0	50	29:1	-3.890
	MPEG1, L3		22,05 kHz	86.0	49	29:1	-2.805
	Real Audio G2		22,05 kHz	82.2	48	31:1	-3.901
	MS9		44,1 kHz	99.5	58	25:1	-3.068
<i>pop1_s</i>	AAC	17 s	32 kHz	105	49	29:1	-0.957
	MPEG1, L2			105	51	29:1	-3.851
	MPEG1, L3		22,05 kHz	105	49	29:1	-3.329
	Real Audio G2		22,05 kHz	102	50	30:1	-3.887
	MS9		44,1 kHz	119	57	26:1	-3.698
<i>pop2_s</i>	AAC	15 s	32 kHz	92.4	49	29:1	-0.858
	MPEG1, L2			92.2	50	29:1	-3.898
	MPEG1, L3		22,05 kHz	92.3	49	29:1	-2.996
	Real Audio G2		22,05 kHz	92.4	51	29:1	-3.904
	MS9		44,1 kHz	10649	58	26:1	-3.693

Codorul cel mai bun: AAC – toate secvențele

Tabelul 7.8 Evaluarea calității audio pentru un semnal mono la un debit de 96 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier kB	Debit real kbit/s	Raport de compresie	ODG
<i>brah_s</i>	AAC	14 s	32 kHz	171	99	15:1	-2.416
	MPEG1, L2			170	99	15:1	-3.729
	MPEG1, L3		44,1 kHz	170	99	15:1	-2.278
	RealAudio G2		44,1 kHz	175	102	14:1	-3.699
	MS9		44,1 kHz	191	112	13:1	<b>-1.381</b>
	Ogg		44,1 kHz	113	66	22:1	-3.552
<i>denm_s</i>	AAC	13 s	32 kHz	165	104	15:1	<b>-0.284</b>
	MPEG1, L2			164	104	15:1	-3.040
	MPEG1, L3		44,1 kHz	165	104	15:1	-0.36
	RealAudio G2		44,1 kHz	175	110	14:1	-3.857
	MS9		44,1 kHz	178	112	13,5:1	-0.308
	Ogg		44,1 kHz	142	87-107	17:1	-1.095
<i>ebu_s</i>	AAC	11 s	32 kHz	140	104	15:1	<b>-0.648</b>
	MPEG1, L2			139	104	15:1	-3.880
	MPEG1, L3		44,1 kHz	140	104	15:1	-0.669
	RealAudio G2		44,1 kHz	150	112	14:1	-3.904
	MS9		44,1 kHz	152	112	13:1	-0.944
	Ogg		44,1 kHz	127	94-112	16:1	-1.306
<i>ice_s</i>	AAC	12 s	32 kHz	146	100	15:1	<b>-1.232</b>
	MPEG1, L2			145	100	15:1	-2.995
	MPEG1, L3		44,1 kHz	146	100	15:1	-1.695
	RealAudio G2		44,1 kHz	150	102	14:1	-3.891
	MS9		44,1 kHz	165	113	13:1	-1.360
	Ogg		44,1 kHz	144	93-101	15:1	-1.531
<i>irt_s</i>	AAC	13 s	32 kHz	161	101	15:1	-1.254
	MPEG1, L2			160	101	15:1	-3.236
	MPEG1, L3		44,1 kHz	160	101	15:1	<b>-1.249</b>
	RealAudio G2		44,1 kHz	175	110	13:1	-3.814
	MS9		44,1 kHz	163	103	14:1	-1.432
	Ogg		44,1 kHz	147	94-109	16:1	-1.899
<i>jazz_s</i>	AAC	14 s	32 kHz	173	101	15:1	<b>-1.211</b>
	MPEG1, L2			173	101	15:1	-3.810
	MPEG1, L3		44,1 kHz	173	101	15:1	-1.635
	RealAudio G2		44,1 kHz	200	117	13:1	-3.892
	MS9		44,1 kHz	187	109	13,5:1	-1.371
	Ogg		44,1 kHz	164	99-108	15:1	-1.505
<i>pop1_s</i>	AAC	17 s	32 kHz	211	102	15:1	-2.158
	MPEG1, L2			210	102	15:1	-3.768
	MPEG1, L3		44,1 kHz	211	102	15:1	-2.736
	RealAudio G2		44,1 kHz	224	108	14:1	-3.837
	MS9		44,1 kHz	226	109	14:1	-2.050
	Ogg		44,1 kHz	214	100-119	14:1	<b>-1.858</b>
<i>pop2_s</i>	AAC	15 s	32kHz	186	101	15:1	<b>-1.512</b>
	MPEG1, L2			185	101	15:1	-3.854
	MPEG1, L3		44,1 kHz	185	101	15:1	-1.472
	RealAudio G2		44,1 kHz	200	109	14:1	-3.893
	MS9		44,1 kHz	200	109	14:1	-1.839
	Ogg		44,1 kHz	172	97-105	16:1	-1.674

Codorul cel mai bun: 1. AAC – *denm, ebu, ice, jazz, pop2*  
2. MPEG1 nivel-3 – *irt*; MS9 – *brah*; OGG – *pop1*

Tabelul 7.9 Evaluarea calității audio pentru un semnal stereo la un debit de 128 kbit/s

Nume fișier	Codec	Durata	Frecvența de eșantionare	Dimensiune fișier kB	Debit real kbit/s	Raport de compresie	ODG
<i>brah_s</i>	AAC	14 s	48 kHz	227	133	11:1	-1.162
	MPEG1, L2			226	133	11:1	-3.775
	MPEG1, L3		44,1 kHz	227	133	11:1	-1.053
	RA8-ATRAC3		44,1 kHz	239.4	140	10:1	-3.725
	MS9		44,1 kHz	248	145	10:1	<b>-0.501</b>
	Ogg		44,1 kHz	113	89-96	22:1	-2.289
<i>denm_s</i>	AAC	13 s	32 kHz	220	138	11:1	-0.047
	MPEG1, L2			219	138	11:1	-1.813
	MPEG1, L3		44,1 kHz	219	138	11:1	<b>-0.035</b>
	RA8-ATRAC3		44,1 kHz	240	151	10:1	-3.842
	MS9		44,1 kHz	236	148	10:1	-0.138
	Ogg		44,1 kHz	142	96-108	17:1	-1.095
<i>ebu_s</i>	AAC	11 s	32 kHz	186	138	11:1	-0.214
	MPEG1, L2			185	138	11:1	-3.877
	MPEG1, L3		44,1 kHz	186	138	11:1	<b>-0.188</b>
	RA8-ATRAC3		44,1 kHz	206	153	10:1	-3.902
	MS9		44,1 kHz	207	153	10:1	-0.445
	Ogg		44,1 kHz	127	99-108	16:1	-1.306
<i>ice_s</i>	AAC	12 s	32 kHz	194	132	11:1	-0.693
	MPEG1, L2			194	132	11:1	-2.351
	MPEG1, L3		44,1 kHz	194	132	11:1	<b>-0.610</b>
	RA8-ATRAC3		44,1 kHz	206	140	10:1	-3.883
	MS9		44,1 kHz	219	149	10:1	-0.634
	Ogg		44,1 kHz	144	96-101	15:1	-1.531
<i>irt_s</i>	AAC	13 s	32 kHz	214	135	11:1	-0.665
	MPEG1, L2			213	135	11:1	-2.980
	MPEG1, L3		44,1 kHz	213	135	11:1	<b>-0.347</b>
	RA8-ATRAC3		44,1 kHz	240	151	10:1	-3.744
	MS9		44,1 kHz	225	141	10:1	-0.542
	Ogg		44,1 kHz	147	94-109	16:1	-1.899
<i>jazz_s</i>	AAC	14 s	32 kHz	230	135	11:1	-0.509
	MPEG1, L2			230	135	11:1	-3.830
	MPEG1, L3		44,1 kHz	230	135	11:1	<b>-0.441</b>
	RA8-ATRAC3		44,1 kHz	274	160	9:1	-3.875
	MS9		44,1 kHz	248	145	10:1	-0.457
	Ogg		44,1 kHz	164	95-107	15:1	-1.505
<i>pop1_s</i>	AAC	17 s	32 kHz	281	135	11:1	-0.957
	MPEG1, L2			280	135	11:1	-3.670
	MPEG1, L3		44,1 kHz	281	135	11:1	-0.774
	RA8-ATRAC3		44,1 kHz	308	148	10:1	-3.846
	MS9		44,1 kHz	306	148	10:1	<b>-0.755</b>
	Ogg		44,1 kHz	214	105-120	14:1	-1.858
<i>pop2_s</i>	AAC	15 s	32 kHz	247	135	11:1	-0.858
	MPEG1, L2			247	135	11:1	-3.843
	MPEG1, L3		44,1 kHz	247	135	11:1	<b>-0.540</b>
	RA8-ATRAC3		44,1 kHz	274	150	10:1	-3.876
	MS9		44,1 kHz	265	144	10:1	-0.730
	Ogg		44,1 kHz	172	97-104	16:1	-1.674

Codorul cel mai bun: 1. MPEG1 nivel-3 – *denm, ebu, ice, irt, jazz, pop2*  
2. MS9 – *brah, pop1*

În figura 7.25 sunt prezentate valorile medii ale evaluării calității audio de bază (ODG) pentru toate fragmentele audio testate în funcție de codecul utilizat. De aici se pot desprinde câteva concluzii asupra eficienței fiecărui sistem de codare, în funcție de debitul binar utilizat:

1. La un debit binar de 128 kbit/s cele mai eficiente sisteme de codare au fost MPEG1 nivel-3 și Microsoft Windows Media 9, cu o evaluare medie ODG = -0,5, urmate la mică distanță de AAC cu o evaluare medie ODG = -0,63;

2. La un debit binar de 96 kbit/s cele mai eficiente sisteme de codare au fost AAC și Microsoft Windows Media 9, cu o evaluare medie ODG = -1,33, urmate la mică distanță de MPEG1 nivel-3 cu o evaluare medie ODG = -1,51;

3. La un debit binar de 48 kbit/s cel mai eficient sistem de codare au fost AAC cu o evaluare medie ODG = -0,64, mai bună decât în cazul debitului de 96 kbit/s;

4. La un debit binar de 32 kbit/s toate evaluările medii au fost sub -3,1, intrând la categoria “foarte deranjante”, evaluările cele mai bune fiind obținute de sistemele de codare Microsoft Windows Media 9 (ODG = -3,25) și AAC (ODG = -3,36);

5. La debitul binar de 16 kbit/s toate evaluările medii au fost sub -3,7, cu diferențe între codoare sub 5%, neputându-se face o clasificare semnificativă.

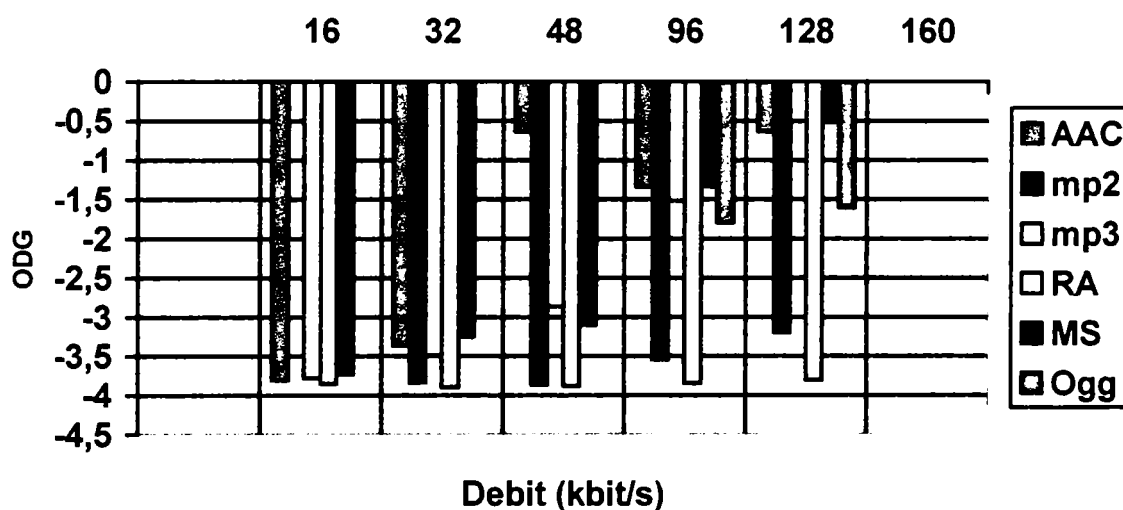


Figura 7.25 Valorile medii ale evaluării calității audio în funcție de codecul utilizat

În figura 7.26 sunt prezentate valorile medii ale evaluării calității audio obținute cu sistemele de codare testate pentru fiecare fragment audio. De aici rezultă ca fragmentul *denm* (voce masculină în limba daneză) a obținut evaluarea medie cea mai bună pentru 3 debite (48, 96 și 128 kbit/s), iar fragmentul *pop1* (Fiesta) a fost cel mai dificil de codat.

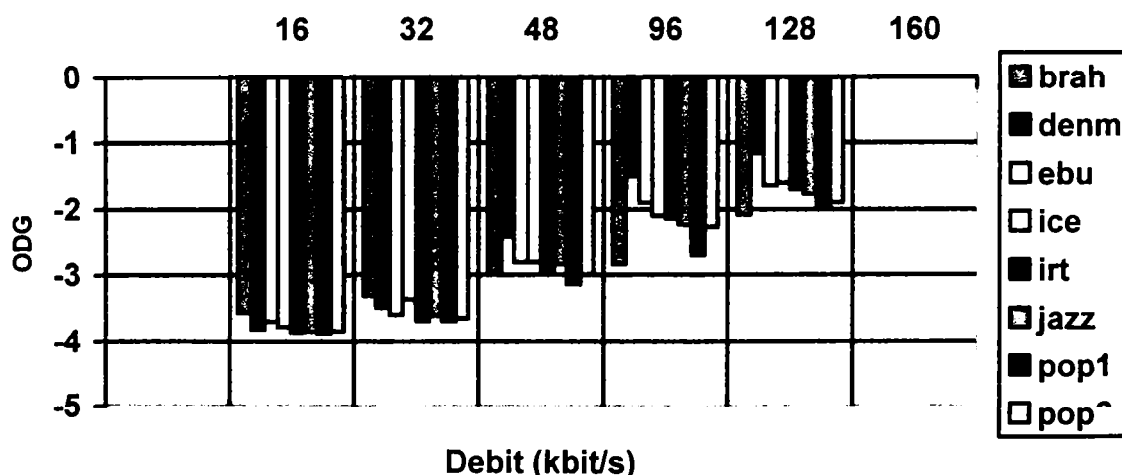


Figura 7.26 Valorile medii ale evaluării calității audio în funcție de secvența audio

## 7.5 Studiu comparativ asupra caracteristicilor unor codoare .mp3 uzuale

### 7.5.1 Obiectiv

Scopul experimentelor executate a fost evaluarea obiectivă și compararea calității audio percepute pentru câteva implementări uzuale ale algoritmului audio de codare MPEG-1 nivel-3. Au fost luate în considerare cele două tehnologii de bază: codarea cu debit binar constant (CBR) și codarea cu debit binar variabil (VBR). În cazul CBR s-a folosit gama maximă de debite binare disponibile, de la 320 kbit/s și până la 20 kbit/s.

Testarea tuturor codecurilor s-a făcut cu aceleași material sonor, urmărindu-se calitatea audio de bază a fiecărei a secvenței audio codate cu diferiți algoritmi și parametri, precum și evoluția dimensiunii fișierului.

Au fost folosite următoarele codecuri:

1. WaveLab 4.0, Steinberg, Lame MP3 encoder;
2. Sound Forge 5.0, Sonic Foundry, Fraunhofer IIS MPEG1 nivel-3;
3. Audio Catalyst 2.0, Xing Technology;
4. AudioActive, Telos Systems, Fraunhofer IIS MPEG1 nivel-3.

### 7.5.2 Semnale de test

Materialul folosit pentru teste a fost un fragment de muzică folclorică românească, "Hora de la Grănicești", cu durata de 10 s și dimensiunea de 1.716 kB. Nu s-au folosit compresoare sau limitatoare, înregistrarea originală (referința) fiind codată liniar PCM, cu o frecvență de eșantionare de 44,1 kHz, 16 biți, mono sau stereo (.wav). Pentru a putea aplica metoda de evaluare obiectivă a calității, materialele audio codate în diferite formate au fost convertite din nou în format .wav și reeșantionate cu frecvența de 44,1 kHz.

### 7.5.3 Rezultate experimentale

Rezultatele evaluării obiective a calității audio pentru codare cu un debit binar constant a sunt date în tabelul 7.10. Unele codoare (WaveLab, Audio Active) dispun de diferite trepte de calitate pentru același debit binar. După cum se poate observa, la debite binare mari ODG are și câteva valori pozitive (vezi §6.6.1).

Tabel 7.10 Codare cu debit binar constant - CBR

Debit kbit/s	WaveLab								Sound Forge 5.0		Audio Catalyst 2.0		AudioActive			
	WMA		SQ		HQ		Fast		kB	ODG	kB	ODG	HQ		Faster	
	kB	ODG	kB	ODG	kB	ODG	kB	ODG					kB	ODG	kB	ODG
320	-	-	391	-1.43	391	-1.43	391	-1.44	391	<b>0.04</b>	390	-0.05	-	-	-	-
256	-	-	314	-0.02	314	-0.03	313	-1.48	313	<b>0.07</b>	-	-	313	<b>0.07</b>	313	<b>0.07</b>
224	-	-	275	-0.08	275	-0.05	274	-1.52	274	<b>0.02</b>	273	-0.07	-	-	-	-
192	-	-	236	-0.18	236	<b>-0.13</b>	235	-1.61	235	-0.21	234	-0.19	235	-0.21	235	-0.23
160	198	-1.28	197	-0.50	197	<b>-0.43</b>	196	-1.83	196	-0.46	195	-0.46	-	-	-	-
128	159	-1.47	157	-1.15	157	-0.96	157	-0.90	157	<b>-0.73</b>	156	-0.98	156	<b>-0.73</b>	156	-0.80
112	-	-	138	-2.15	138	-1.93	137	<b>-1.72</b>	-	-	-	-	-	-	-	-
96	120	<b>-2.15</b>	118	-2.32	118	-2.22	118	-2.41	118	-2.57	117	-2.50	118	-2.28	118	-2.51
80	100	-2.79	99	-2.77	99	-2.76	99	<b>-2.48</b>	-	-	-	-	-	-	-	-
64	81	-3.28	79	-3.12	79	-3.12	79	-3.12	79	<b>-2.93</b>	78	-3.31	78	<b>-2.92</b>	78	-2.98
56	-	-	69	<b>-3.12</b>	69	<b>-3.12</b>	69	-3.13	69	-3.19	69	-3.31	69	-3.19	69	-3.17
48	61	-3.58	60	-3.37	60	-3.37	60	-3.36	59	<b>-3.15</b>	59	-3.32	59	-3.16	59	<b>-3.15</b>
40	-	-	50	-3.45	50	-3.45	50	-3.46	49	<b>-3.39</b>	-	-	49	<b>-3.39</b>	49	<b>-3.39</b>
32	41	-3.67	40	<b>-3.42</b>	40	<b>-3.42</b>	40	-3.43	40	-3.46	39	-3.48	40	-3.46	40	-3.47
24	29	-3.79	30	-3.57	30	-3.57	30	-3.57	30	<b>-3.55</b>	-	-	30	<b>-3.55</b>	30	-3.57
20	26	-3.84	25	-3.64	25	-3.64	25	-3.65	25	<b>-3.62</b>	-	-	25	<b>-3.62</b>	25	<b>-3.62</b>

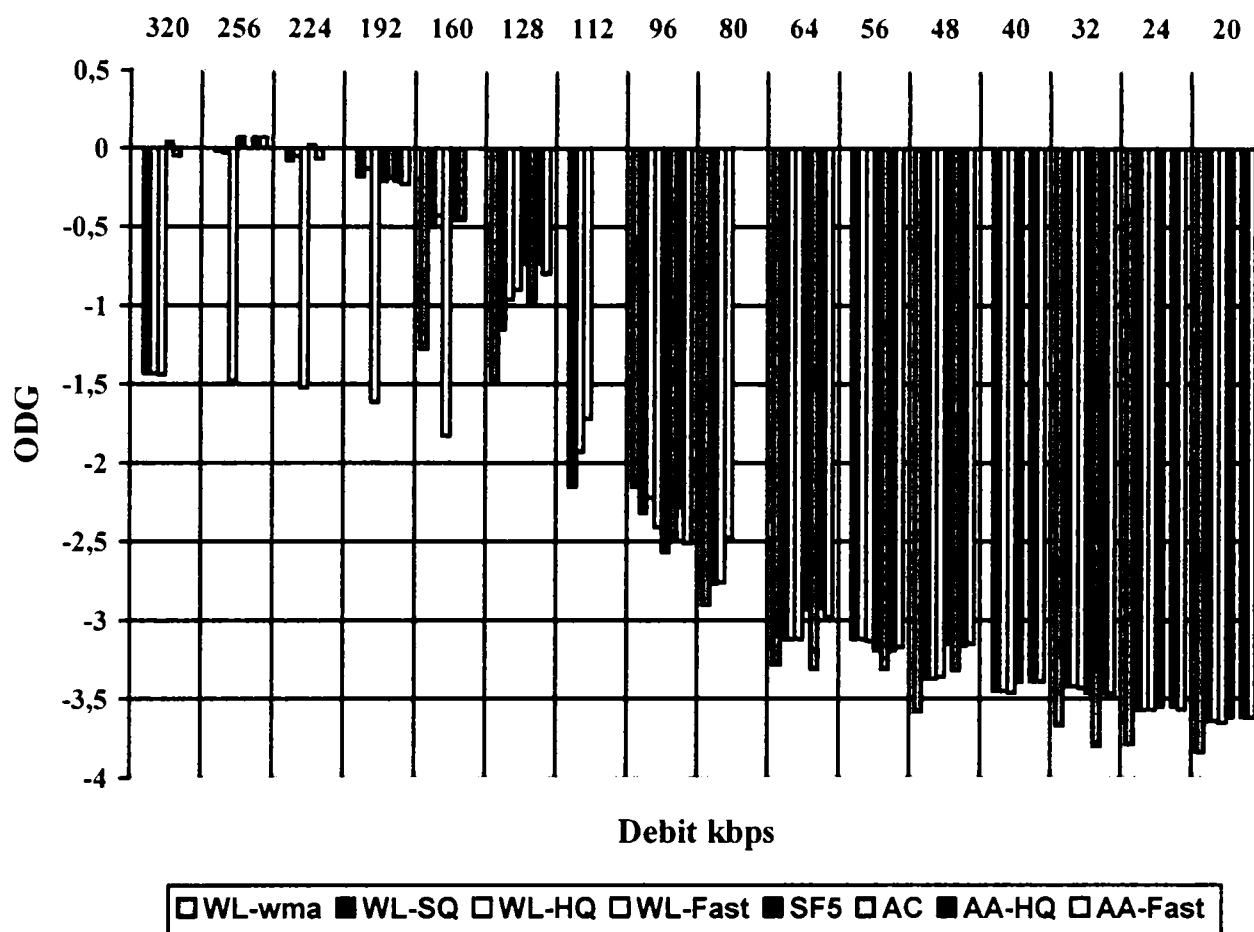


Figura 7.27 Evoluția calității audio obiective pentru diferite codoare .mp3 în funcție de debitul binar (codare cu debit binar constant)

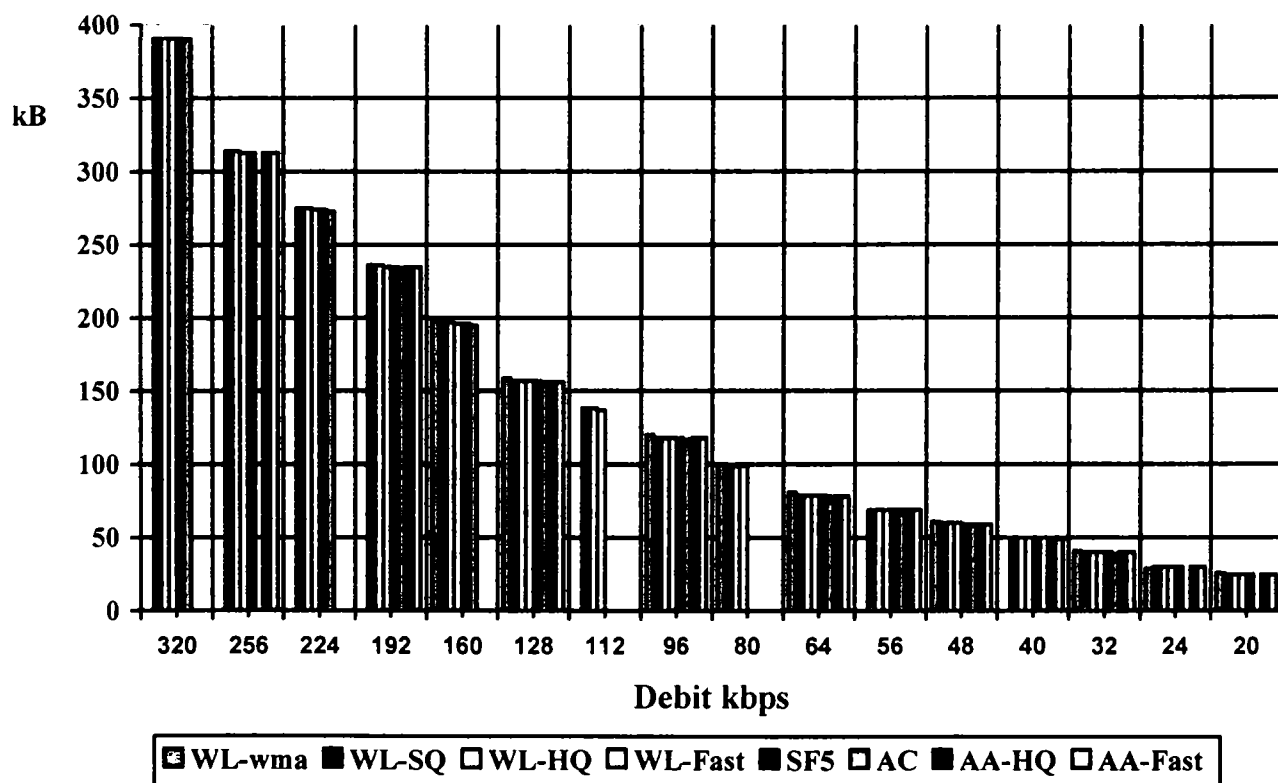


Figura 7.28 Evoluția dimensiunii fișierului audio pentru diferite codoare .mp3 în funcție de calitatea codării (codare cu debit binar constant)



În general, se poate spune că variantele de codor .mp3 implementate în programele Sound Forge și Audio Active (Fraunhofer IIS) s-au comportat cel mai bine în majoritatea cazurilor, urmate de codorul Lame implementat în programul WaveLab. În cazul codării cu debit binar constant, codorul Xing, deși cel mai răspândit, a avut evaluările cele mai slabe în comparație cu contracandidații săi.

Din figura 7.28 se poate observa că diferențele dintre dimensiunile fișierelor la același debit binar sunt nesemnificative (1 – 2 kB). Pentru gama de debite binare studiată, raportul de compresie variază de la 1:4 (320 kbit/s) până la 1:68 (20 kbit/s).

În concluzie, în funcție de debitele binare utilizate am realizat o clasificare a calității audio obținute în urma codării mp3 cu debit binar constant și câteva recomandări privind domeniile de utilizare:

- 320 – 224 kbit/s – ODG<-0,1, MPEG1, calitate arhivă, redare pe instalații HiFi High End
- 192 - 160 kbit/s - -0,5<ODG<-0,1, MPEG1, calitate arhivă, redare pe instalații HiFi
- 128 kbit/s – MPEG1, calitate CD, debitul ideal pentru majoritatea utilizatorilor
- 112 kbit/s – MPEG1, calitate aproape CD, alegere ideală pentru playere MP3 portabile
- 96 kbit/s – MPEG1, calitate aproape CD, alegere bună pentru playere MP3 portabile
- 80 - 64 kbit/s – MPEG2, calitate MF, dimensiune mică
- 56 kbit/s – MPEG2, calitate MF, dimensiune minimă
- 48 kbit/s – MPEG2, calitate MA, dimensiune mică
- 32 kbit/s – MPEG2, calitate MA, dimensiune minimă
- 24 kbit/s – MPEG2, calitate telefonică, dimensiune mică
- 16 kbit/s – MPEG2, calitate telefonică, dimensiune minimă

Rezultatele evaluării obiective a calității audio pentru codare cu un debit binar variabil sunt prezentate în tabelul 7.11. Ca și în cazul codării cu CBR, pentru fiecare nivel de compresie, codorul Lame din programul WaveLab dispune de trei niveluri de calitate.

Tabel 7.11 Codare cu debit binar variabil - VBR

Calitate	WaveLab						Audio Catalyst 2.0	
	SQ		HQ		Fast		kB	ODG
Compresie	kB	ODG	kB	ODG	kB	ODG		
Low	272	-0.19	272	-0.19	281	-0.16	262	-0.08
N/L	225	-0.35	225	-0.35	232	-0.31	219	-0.26
Normal	173	-0.76	173	-0.76	177	-0.71	178	-0.58
N/H	142	-1.39	142	-1.39	144	-1.35	149	-1.07
High	117	-2.44	117	-2.44	119	-2.42	125	-2.58

Rezultatele obținute mi-au permis să fac o recomandare a categoriilor de utilizatori în funcție de cele 5 niveluri de compresie disponibile:

- Low – calitate maximă, arhivarea fișierelor muzicale originale;
- Normal/Low – arhivarea fișierelor muzicale, redare pe instalații audio Hi-Fi;
- Normal – calitate audio ideală pentru majoritatea utilizatorilor;
- Normal/High – calitate aproape CD, ideal pentru playere MP3 portabile;
- High – calitate FM, poate fi o alegere acceptabilă pentru dispozitive portabile de redare, unde primează cantitatea de informație muzicală stocată (Memory Stick, telefoane mobile etc.).

Din figura 7.29.a se poate observa că, în cazul codării cu debit binar variabil, codorul Xing din cadrul programului Audio Catalyst, a avut evaluarea cea mai bună în primele 4 niveluri de compresie.

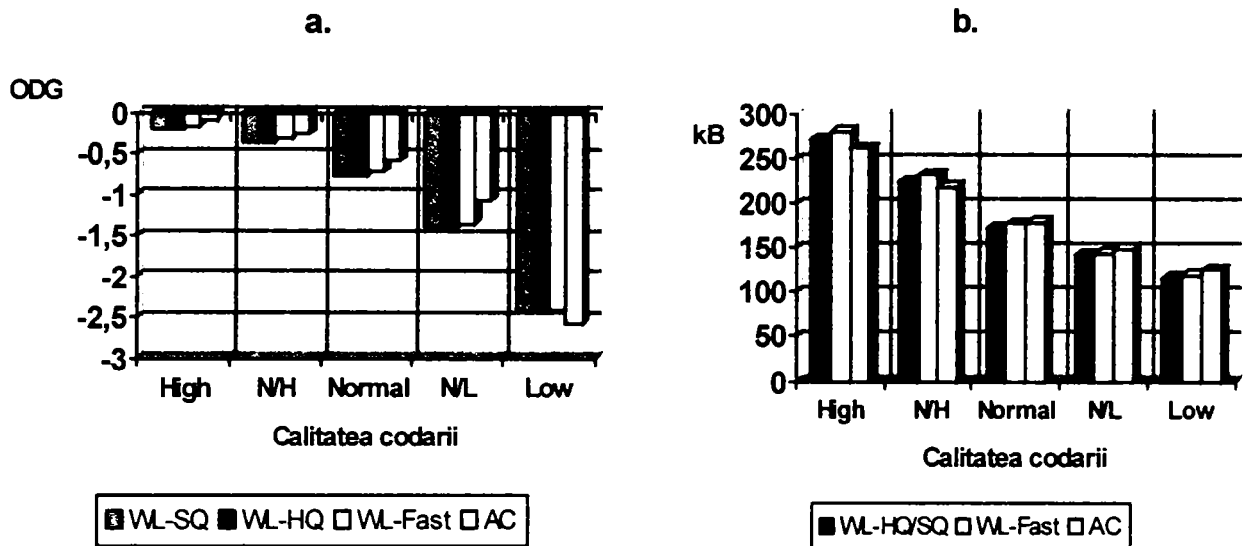


Figura 7.29 Evoluția calității audio obiective (a) și a dimensiunii fișierului audio (b) pentru diferite codoare .mp3 în funcție de calitatea codării (codare cu debit binar variabil)

Din tabelul 7.11 și figura 7.29.b se poate observa evoluția dimensiunii fișierului, în funcție de gradul de compresie impus, raportul de compresie variind între 1:6 până la 1:14. La același grad de compresie, diferențele dintre dimensiunile fișierelor în funcție de codor și de viteza codării sunt cuprinse între 2 și 20 kB.

## 7.6 Evaluarea efectului codărilor succesive asupra calității audio percepute

### 7.6.1 Obiectiv

Scopul testelor executate a fost evaluarea obiectivă a degradării progresive a calității audio percepute în cazul unui tandem omogen la același debit binar (clasa A<sub>1</sub>, conform §5.1). Pentru acest experiment am folosit unul dintre cele mai răspândite codoare MPEG1 nivel-3, Audio Catalyst 2.0 de la Xing Technologies. Am efectuat 10 cicluri succesive complete codare/decodare din formatul inițial .wav în format .mp3 și invers, denumite generația 1 la generația 10, la același debit binar. Operațiunile succesive de codare au fost realizate la 6 debite binare diferite – 320, 256, 224, 192, 160 și 128 kbit/s.

### 7.6.2 Semnal de test

Materialul folosit pentru teste a fost un fragment de muzical din extremul orient, interpretat la instrumente de coarde și de percuție, care conține o gamă foarte largă de frecvențe (până la 19kHz), cu durata de 14 s. În figura 7.30 este prezentată sonograma acestui fragment sonor. Nu s-au folosit compresoare sau limitatoare, înregistrarea originală (referința) fiind codată liniar PCM, cu o frecvență de eșantionare de 44,1 kHz, 16 biți, stereo (.wav). Pentru a putea aplica metoda de evaluare obiectivă a calității audio, materialul audio codat .mp3, la diferite debite binare, a fost convertit din nou în format .wav și reeșantionat cu frecvența de 44,1 kHz.

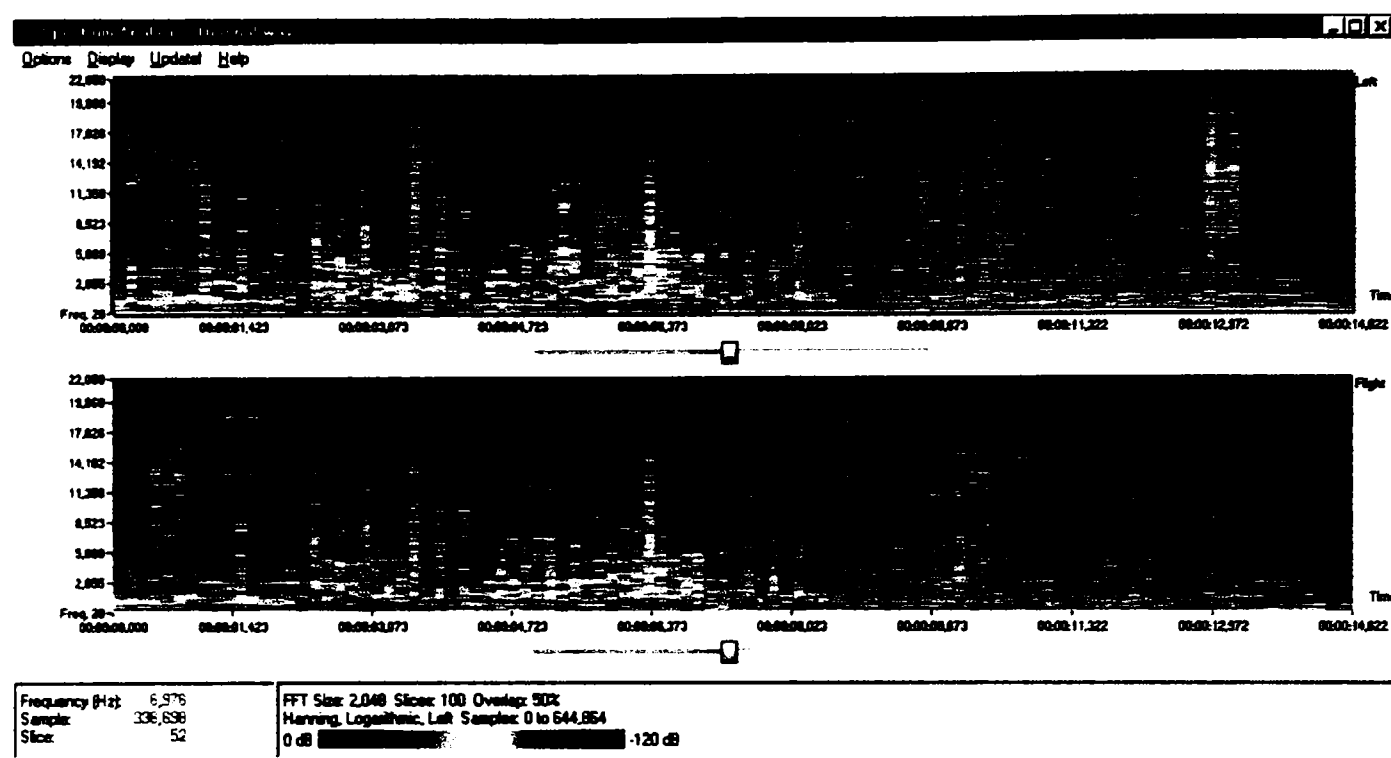


Figura 7.30 Sonograma fragmentului folosit la evaluarea efectului codărilor succesive

### 7.6.3 Rezultate experimentale

Rezultatele evaluării obiective a efectului celor 10 generații de codări succesive asupra calității audio percepute sunt prezentate în figura 7.31. După cum se poate observa, scăderea calității audio este aproape liniară, odată cu creșterea generației de codare succesivă. De asemenea, se observă o creștere a pantei degradării în funcție de scăderea debitului binar.

În cazul debitelor binare peste 200 kbit/s (224, 256 și 320 kbit/s), degradarea calității audio datorată celor 10 generații de codare succesivă este relativ redusă, încadrându-se în prima parte a scalei de diferențe subiective percepute (ODG = -1, tabel 4.1), respectiv “modificări perceptibile, dar nederanjante”.

Pentru debitul binar de 192 kbit/s, până la generația a 5-a, modificările se încadrează în prima categorie (ODG = -1), iar până la generația a 10-a (ODG = -1,889) se încadrează în a doua categorie, respectiv “modificări ușor deranjante”.

În cazul debitului binar de 160 kbit/s, degradarea calității audio este mai accentuată, ultimele 5 generații având o evaluare cuprinsă în cea de-a treia categorie, adică “modificări deranjante”.

În sfârșit, pentru debitul binar de 128 kbit/s, în general cel mai utilizat, degradările sunt mult mai evidente, prima generație având deja o evaluare ODG = -0,9. De aici rezultă foarte clar că fișierele audio codate la acest debit binar nu trebuie să mai sufere alte cicluri de codare/decodare pentru că degradarea calității audio devine semnificativă.

Acest experiment conduce la concluzia că, în cazul codării MPEG1 nivel-3 a unor fișiere audio susceptibile de a suferi un mare număr de prelucrări ulterioare succesive ( $n \leq 10$ ) se recomandă utilizarea unor debite binare mari (256 sau 320 kbit/s), pentru a evita degradări sesizabile ale calității audio. Totuși, debitul binar de 192 kbit/s este soluția care permite codări de calitate, ce pot fi folosite la arhivarea materialelor audio, care pot suferi chiar un număr mediu ( $n \leq 5$ ) de prelucrări ulterioare succesive.

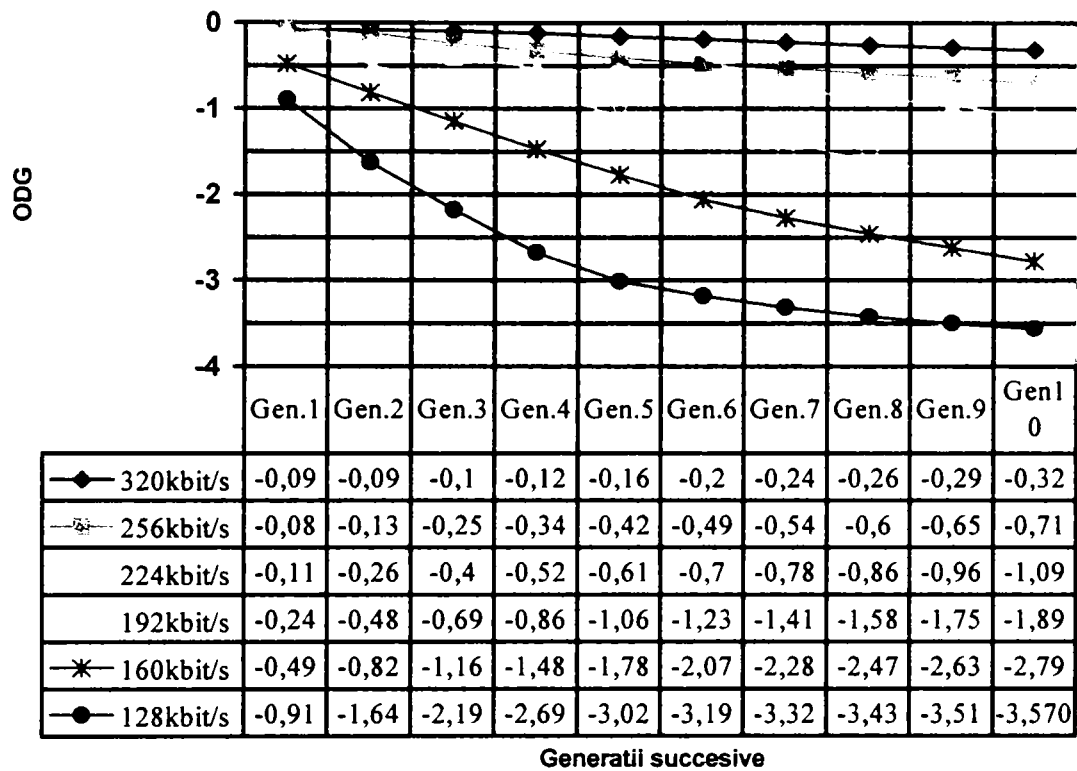


Figura 7.31 Efectul codărilor succesive asupra calității audio percepute



# CAPITOLUL 8

## CONTRIBUȚII PERSONALE ȘI CONCLUZII

Lucrarea analizează și compară diferiți algoritmi de compresie a sunetului, efectele ce se produc asupra calității audio percepute și metodele de evaluare subiectivă și obiectivă a calității audio. Originalitatea lucrării constă în abordarea cu prioritate a rezultatelor obținute la prelucrarea unei game foarte largi de fișiere audio prin intermediul a diferite sisteme și algoritmi de compresie utilizate în domeniul radiodifuziunii, din punct de vedere al calității sunetului perceput de ascultătorul final, beneficiar al muncii redactorilor și tehnicienilor. Literatura internă studiată a abordat domeniul compresiei sunetului mai mult din punct de vedere teoretic sau parțial, insistându-se numai asupra domeniului vocal, utilizat cu precădere în comunicații, care apelează la principii, metode și algoritmi diferiți de compresie.

Efectele compresiei sunetului asupra calității audio perceptuale au fost analizate în funcție de debitul binar utilizat, de conținutul (complexitatea) armonic al fragmentului analizat, de algoritmul de compresie folosit și de numărul de codări succesive pe care un fișier audio le poate suferi. Rezultatele obținute pot fi considerate ca recomandări utile tuturor celor care lucrează în domeniul radiodifuziunii sau multimedia, pentru a realiza un raport optim între raportul de compresie (mărimea fișierului audio) și calitatea audio perceptuală, în funcție de scopul utilizării și publicul țintă căruia i se adresează.

În lucrarea de față se pot evidenția următoarele contribuții personale, cu caracter de noutate în domeniul analizei calității sunetului rezultat în urma compresiei audio digitale și a ciclurilor succesive de codare/decodare la care poate fi supus un fragment sonor, fiind însoțite de referințe bibliografice la articolele publicate de doctorand.

### □ Evidențierea și analiza distorsiunilor specifice sistemelor de codare perceptuale

Pentru a putea analiza calitatea audio subiectivă și obiectivă a unui fragment muzical digitalizat trebuie să știi ce să asculți. Artefactele produse de codarea audio digitală diferă mult prin natura și audibilitatea lor de binecunoscutele distorsiuni audio analogice, fiind mai greu de identificat într-un sistem audio format dintr-un lanț de diferite echipamente de prelucrare și codare/decodare a sunetului.

În cadrul prezentei lucrări au fost evidențiate și analizate următoarele tipuri de artefacte de codare: preecoul, artefacte produse în codarea vocală, diferența de nivel datorată mascării binaurale, pierderea imaginii stereofonice, limitarea în frecvență (sus), erori de subeșantionare, artefacte datorate codării succesive (în tandem) și cele datorate codării muzicii cu codoare vocale [PFA02]. Secvențele audio originale, precum și cele prezentând artefactele analizate au fost înregistrate pe un CD.

### □ Metode subiective și obiective de măsurare a calității audio

Reprezentările digitale ale formelor de undă analogice au ca efect apariția unor distorsiuni, numite *erori de codare* sau *zgomot de codare*. Aceste distorsiuni pot fi evaluate prin:

- criterii subiective;
- criterii obiective simple precum raportul semnal/zgomot, ca o măsură a *similarității formei de undă* dintre sursă și semnalul reconstruit;
- criterii obiective complexe, ca măsură a similarității perceptuale, care țin cont de caracteristicile percepției auditive umane.

Măsurările obiective încearcă să modeleze percepția umană a semnalelor acustice. Metoda standard propusă ITU-R BS.1387-1, numită **PEAQ** (*Perceptual Evaluation of Audio Quality* - evaluarea perceptuală a calității audio), se bazează pe principii psihoacustice general acceptate. Un semnal prelucrat printr-o anumită metodă este comparat cu semnalul original, aliniat în timp. Cadrele corespunzătoare ale semnalului original și prelucrat sunt transformate într-o reprezentare a membranei basulare și diferențele sunt apoi analizate ca o funcție de frecvență și timp pe baza unui model cognitiv. Sunt extrase caracteristicile perceptuale relevante, care se folosesc apoi pentru a calcula o măsură a calității, în schemă fiind disponibile mai multe modele de variabile de ieșire - **MOV** (*model output variables*). Fiabilitatea metodelor obiective de măsurare a calității audio poate fi estimată numai prin compararea rezultatelor obținute prin teste de ascultare subiective aplicate aceluiași semnale audio. Recomandarea ITU-R BS.116 descrie în amănunt metodologia de testare bazată pe judecăți subiective pentru a obține așa numita "calitate audio de bază" a unui dispozitiv. În cazul PEAQ se dorește evaluarea aceleiași calități audio de bază, dar prin metode de măsurare obiective.

În lucrare s-a efectuat o implementare cu ajutorul programului MATCAD a metodei obiective de evaluare a calității sunetului, pornind de standardul ITU-R BS.1387-1. Având în vedere că multe aspecte ale standardului sunt insuficient specificate, au fost propuse de autor diferite interpretări alternative și soluții mai eficiente pentru implementarea segmentelor cu calcule intensive ale algoritmului. Înțelegerea și explicarea tehnicilor folosite în standard pentru evaluarea calității audio au permis o mai bună înțelegere a mecanismului dinamic de alocare a biților într-un codor audio și a audibilității efectelor reducerii debitului binar.

Rezultatele evaluărilor obiective din cadrul testelor efectuate în lucrare, au fost incluse într-o lucrare [PFA04], care va fi prezentată la o manifestare internațională de specialitate.

### □ Studiu comparativ a modului de răspuns a diferitelor codoare mp3

Analiza comparativă s-a făcut pe un fragment de muzică populară românească, cu o mare complexitate armonică și dinamică, folosind patru codoare diferite – WaveLab, SoundForge 5.0, Audio Catalyst 2.0 și AudioActive, la gama maximă de debite binare disponibile 20 – 320 kbit/s și diferite calități de codare. Studiul calității audio obiective ODG s-a făcut cu implementarea proprie a metodei obiective de evaluare a calității audio și a relevat faptul că debite binare cuprinse între 320 – 160 kbit/s permit codări de calitate CD, cu diferențe puțin perceptibile și recomandabile pentru transmisii de calitate și stocări de arhivă, 128 kbit/s este debitul binar ideal pentru amatori, permițând codări de calitate apropiată CD-ului, cu diferențe audio nederanjabile, perceptibile numai de către ascultători experimentați. Urmează o categorie de debite binare intermediare (112, 96, 80 kbit/s), mai puțin folosite, care produc codări cu diferențe perceptibile și o calitate audio obiectivă ODG cuprinsă între -2.0 și -3.0 și în sfârșit grupul debitelor binare reduse 64 – 20 kbit/s, care produc codări cu diferențe supărătoare și calitate audio obiectivă ODG sub -3.0, dar cu rapoarte de compresie mari.

## □ **Recomandări privind transmiterea programelor de radiodifuziune prin Internet**

Dezvoltarea spectaculoasă a Internetului, datorată creșterii vitezei de acces în special pentru utilizatorii casnici, a dus la răspândirea utilizării lui ca mediu de difuzare a programelor posturilor de radio. Unul din factorii principali care concură la răspândirea și acceptarea radioului prin Internet este calitatea audio a informației recepționate. Instrumentul de evaluare a calității audio a schemelor de compresie folosite pentru Internet Radio a fost implementarea proprie a metodei obiective de evaluare a calității audio. Testele s-au efectuat pe un grup de opt fișiere audio, mono și stereo, cu un variat conținut armonic și dinamic, care să reprezinte cât mai bine complexitatea programelor radiofonice transmise, în special de un post public. Aceste fișiere audio au fost codate cu ajutorul a 6 algoritmi de compresie perceptuală (AAC, MPEG1-L2, MPEG1-L3, RealAudio 9, MS Media 9 și Ogg Vorbis) a sunetului, la diferite debite binare care să simuleze diferite condiții de recepție (de la 16 kbit/s la 128 kbit/s). Analiza rezultatelor obținute, calitatea audio obiectivă, raportul de compresie, debitul binar real, au dus la concluzia că nu există, încă, un algoritm de compresie audio ideal, care să ofere rezultate optime pe toată gama de debite testate. Având în vedere tendința de specializare a conținutului programelor transmise – știri, transmisii sportive, muzică - clasică, pop, rock, jazz etc. – am făcut recomandări asupra codoarelor cele mai potrivite genului de informație transmis și, în funcție de publicul țintă ales, pretențiile sale de calitate a recepției și debitul binar disponibil la locul recepției, programul poate fi codat și transmis cu unul sau mai multe debite simultan [PFA03].

## □ **Analiza efectelor codărilor succesive**

Datorită răspândirii echipamentelor ce folosesc tehnici de codare audio cu debit binar redus (dbr) în activitatea studiourilor de radiodifuziune, de la contribuție, prin studioul de producție până la distribuție și emisie, a apărut problema codărilor multiple, succesive sau în “tandem” cum mai sunt denumite. Într-un lanț digital de transmitere a materialelor audio pot exista combinații diferite de scheme perceptuale de codare succesive și s-a observat că în funcție de algoritmi și debitele binare folosite calitatea finală a sunetului poate fi serios afectată.

Deși sistemele care utilizează un algoritm de codare cu dbr funcționează bine când sunt testate singure, calitatea sunetului scade semnificativ în cazul unor codări succesive.

În lucrare au fost analizate efectele înregistrărilor succesive în cazul unor tandemuri omogene, pentru algoritmi .mp3 și ATRAC.

În cazul tandemului omogen cu codor mp3, s-a ales o secvență audio cu muzică de jazz, pentru care codarea s-a făcut în zece generații succesive cu ajutorul unui codor XING din cadrul unui program larg răspândit, Audio Catalyst 2.0, cu șase debite diferite din gama 320 – 128 kbit/s. S-a analizat calitatea audio obiectivă a fiecărei generații, rezultând faptul că debitele între 320 – 192 kbit/s permit generații multiple cu modificări acustice nederanjante ( $ODG = 0, -2$ ), putând fi folosite la stocarea materialelor susceptibile de a suferi diferite prelucrări ulterioare, debitul de 128 kbit/s fiind recomandabil numai pentru o stocare finală, arhivare, difuzare, fără prelucrări ulterioare.

În cazul algoritmului ATRAC (minidiscul) s-a urmărit degradarea progresivă a semnalului atât în domeniul frecvență cât și în domeniul timp, ce poate apărea ca rezultat al codărilor succesive. Pentru a testa generații succesive de copii, s-a înregistrat pe minidisc un semnal generat de calculator și transferat înapoi în calculator. Aceasta a fost prima generație de înregistrări, care va deveni sursă pentru următoarea înregistrare pe minidisc, operația repetându-se de 10 ori. Semnalele rezultate în cele 10 generații au fost analizate atât în domeniul frecvență cât și în domeniul timp. Efectul înregistrărilor succesive folosind codarea ATRAC conduce la degradarea progresivă a componentelor spectrale mascate mai înalte, până la dispariția în zgomotul de cuantizare crescut (în acest caz la a 7-a generație). Creșterea zgomotului spectral de cuantizare la a 10-a generație este de aprox.12 dB față de zgomotul original. În plus față de



modificările în conținutul spectral, analiza în domeniul timp a generațiilor succesive dezvăluie artefacte semnificative, de ex. vârfuri de semnal cu amplitudini excesive (în comparație cu amplitudinea medie), care nu sunt în concordanță ca timp și frecvență cu conținutul semnalului original [PFA01].

#### □ **Soluții de optimizare a activităților de producție în studiourile de radiodifuziune**

Soluțiile de optimizare propuse de autor au rezultat din contactul direct cu activitatea din mai multe studiouri regionale și naționale de radiodifuziune și se referă la introducerea tehnicilor digitale de înregistrare, stocare, prelucrare, redare și transmitere a sunetului.

Testele și analizele efectuate în cadrul lucrării, au permis autorului să propună un mod de organizare a circulației și prelucrării fișierelor audio, care să mențină la un nivel acceptabil calitatea audio a unui material pe tot parcursul său de la contribuție și până la emisie.

#### □ **Proiectarea și realizarea unui CD cu scop educațional**

Proiectarea unui CD cu scop educațional, o premieră absolută în România, combină informațiile teoretice asupra codării audio perceptuale, cu rezultatele diferitelor teste efectuate de-a lungul lucrării și cu exemple audio specifice, care pot fi ascultate pe orice CD player audio de calitate.

Utilizatorul poate asculta și învăța care sunt efectele diferiților parametri ce afectează o înregistrare audio digitală, poate compara (asculta critic) rezultatele unui anumit tip de codare, cu anumiți parametri, asupra unor secvențe audio cu un variat conținut sonor.

În final sunt prezentate mai multe adrese de internet care furnizează informații de referință asupra codării audio perceptuale și a evaluării calității audio subiective și obiective.

# Metodă de evaluare a calității audio percepute Implementare și simulare în MATCAD

K. Pfaff Rev.1.0

## 1. Modelul urechii periferice

### 1.1 Generalități

Nivelul maxim de presiune sonoră	$L_p := 92$
Numărul de cadre	$nn := 4$
Index pentru cadrul curent din semnal	$n := 0..nn$
Factorul de normalizare	$Norm := 10$
Dimensiunea cadrului (eșantioane)	$N := 2048$
	$esant := 0..191999$
Index pentru cadru de semnal	$kt := 0..2047$
Nr. de benzi de frecvență	$k := 0..108$

### 1.2. Prelucrarea în domeniul timp

Uref := READWAV("pop1\_m\_o.wav")

Utest := READWAV("pop1\_m\_28k\_modem\_mp3.wav")

$$tnref := \begin{cases} tn \leftarrow 0 \\ \text{for } kt \in 0..2047 \\ \quad \text{for } n \in 0..nn \\ \quad \quad tn_{kt,n} \leftarrow Uref_{2048 \cdot n + kt} \\ tn \end{cases}$$

$$tntest := \begin{cases} tn \leftarrow 0 \\ \text{for } kt \in 0..2047 \\ \quad \text{for } n \in 0..nn \\ \quad \quad tn_{kt,n} \leftarrow Utest_{2048 \cdot n + kt} \\ tn \end{cases}$$

$$tw_{kt} := \frac{\sqrt{8}}{2\sqrt{3}} \cdot \left( 1 - \cos\left(2 \cdot \pi \cdot \frac{kt}{N-1}\right) \right)$$

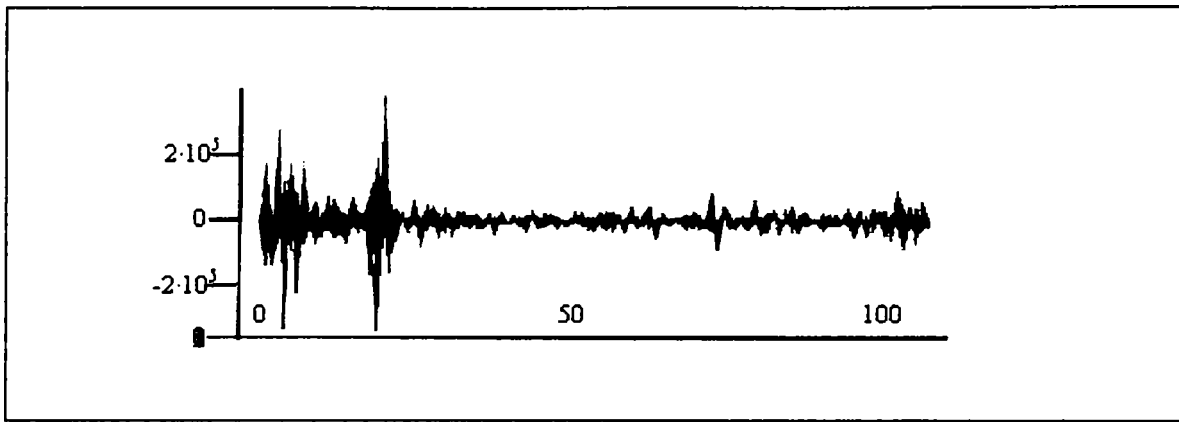
### 1.3. Transformarea FFT

Factorul de scalare pentru fereastra Hann  $fac := \frac{L_p}{10^{\frac{20}{Norm}}}$   $fac = 3.981 \times 10^3$

Transformarea Fourier scalată

$$FFTref_{kf,n} := \frac{fac}{N} \cdot \sum_{kt=0}^{N-1} \left( tw_{kt} \cdot tnref_{kt,n} \cdot e^{-i \cdot 2 \cdot \frac{\pi}{N} \cdot kf \cdot kt} \right) \quad kf := 1..108$$

$$FFTtest_{kf,n} := \frac{fac}{N} \cdot \sum_{kt=0}^{N-1} \left( tw_{kt} \cdot tntest_{kt,n} \cdot e^{-i \cdot 2 \cdot \frac{\pi}{N} \cdot kf \cdot kt} \right)$$



FFTref – FFTtest

### 1.4. Urechea externă și mijlocie

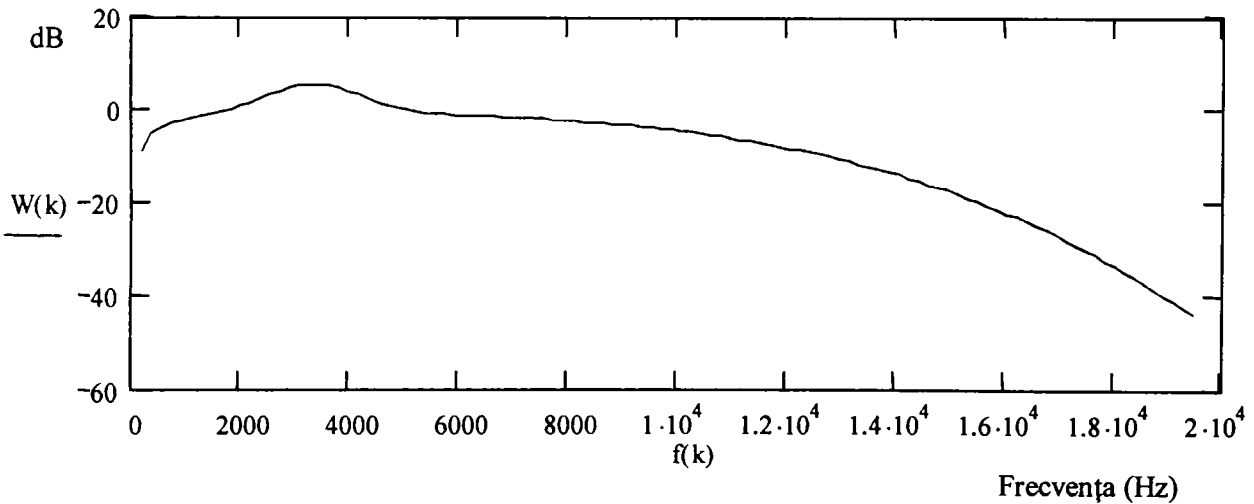
Răspunsul în frecvență al urechii externe și mijlocii

Funcția de ponderare

$$W(k) := -2.184 \left( \frac{f(k)}{1000} \right)^{-0.8} + 6.5 \cdot e^{-0.6 \cdot \left( \frac{f(k)}{1000} - 3.3 \right)^2} - 10^{-3} \cdot \left( \frac{f(k)}{1000} \right)^{3.6} \quad f(k) := k \cdot 180$$

Ieșirea ponderată a FFT pentru urechea externă

$$F_{\text{test}}_{kf,n} := \left| \text{FFT}_{\text{test}}_{kf,n} \right| \cdot 10^{\frac{W(\text{FF}_{kf,2})}{20}} \quad F_{\text{ref}}_{kf,n} := \left| \text{FFT}_{\text{ref}}_{kf,n} \right| \cdot 10^{\frac{W(\text{FF}_{kf,2})}{20}}$$



### 1.5. Gruparea în benzi critice

Distanța dintre benzile filtrelor

$$dZ := 0.2^z \text{ Bark}$$

Banda de frecvență inferioară, corespunzătoare  $f_l = 80 \text{ Hz}$

$$Z_l := 7 \cdot \text{asinh} \left( \frac{80}{650} \right)$$

Banda de frecvență superioară, corespunzătoare  $f_s = 18.000$  Hz

$$Z_u := 7 \cdot \operatorname{asinh} \left( \frac{18000}{650} \right)$$

Număr total de benzi de frecvență

$$x := \frac{(Z_u - Z_l)}{dZ}$$

$$x = 108.972$$

Calculul mărimilor caracteristice benzilor de frecvență în Hz

ZZ0 – frecvența inferioară      ZZ1 – frecvența superioară

ZZ2 – frecvența centrală      ZZ3 – lățimea benzii

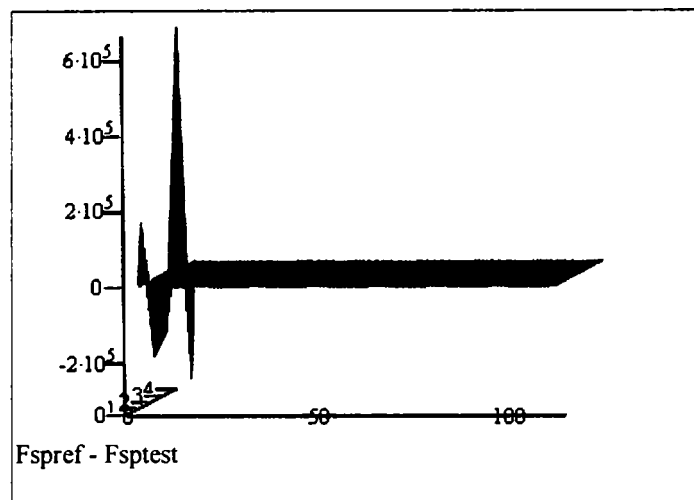
```

FF := | ZZ ← 0
      | i ← 0
      | while i ≤ (Zu - Zl) / dZ
      |   | ZZi,0 ← 650 · sinh( (Zl + i · dZ) / 7 )
      |   | ZZi,1 ← 650 · sinh( [Zl + (i + 1) · dZ] / 7 )
      |   | ZZi,2 ← (ZZi,0 + ZZi,1) / 2
      |   | ZZi,3 ← ZZi,1 - ZZi,0
      |   | i ← i + 1
      | ZZ
  
```

Reprezentarea energiei ieșirii ponderate a transformatei FFT pentru urechea externă

$$F_{sptest}_{kf,n} := \left( |F_{etest}_{kf,n}| \right)^2$$

$$F_{spref}_{kf,n} := \left( |F_{eref}_{kf,n}| \right)^2$$



Gruparea vectorului energie în benzile critice

Indexul grupurilor de frecvențe       $i := 0..108$

Constantă pentru rezoluția frecvenței       $F_{res} := \frac{48000}{2048}$

$$f_{i,1} := FF_{i,0} \quad f_{c,i} := FF_{i,2} \quad f_{s,i} := FF_{i,1}$$

## Calculul energiilor grupurilor de frecvențe Pe

```

Peref := Pe ← 0
for n ∈ 0..nn
  for i ∈ 0..108
    Pei,n ← 0
    for k ∈ 0..108
      Pei,n ← Pei,n + Fsprefk,n if [[(k - .5)·Fres ≥ fi,1] ∧ [(k + .5)·Fres ≤ fs,i]]
      Pei,n ← Pei,n + Fsprefk,n ·  $\frac{(f_{s,i} - f_{i,1})}{Fres}$  if [(k - .5)·Fres < fi,1 ∧ (k + .5)·Fres > fs,i]
      Pei,n ← Pei,n + Fsprefk,n ·  $\frac{[(k + .5)·Fres - f_{i,1}]}{Fres}$  if [(k - .5)·Fres < fi,1 ∧ (k + .5)·Fres > fi,1]
      Pei,n ← Pei,n + Fsprefk,n ·  $\frac{[f_{s,i} - (k - .5)·Fres]}{Fres}$  if [(k - .5)·Fres < fs,i ∧ (k + .5)·Fres > fs,i]
      Pei,n ← Pei,n otherwise
    Pei,n ← max(Pei,n, .0000000000000000)
  Pe

```

```

Petest := Pe ← 0
for n ∈ 0..nn
  for i ∈ 0..108
    Pei,n ← 0
    for k ∈ 0..108
      Pei,n ← Pei,n + Fsptestk,n if [[(k - .5)·Fres ≥ fi,1] ∧ [(k + .5)·Fres ≤ fs,i]]
      Pei,n ← Pei,n + Fsptestk,n ·  $\frac{(f_{s,i} - f_{i,1})}{Fres}$  if [(k - .5)·Fres < fi,1 ∧ (k + .5)·Fres > fs,i]
      Pei,n ← Pei,n + Fsptestk,n ·  $\frac{[(k + .5)·Fres - f_{i,1}]}{Fres}$  if [(k - .5)·Fres < fi,1 ∧ (k + .5)·Fres > fi,1]
      Pei,n ← Pei,n + Fsptestk,n ·  $\frac{[f_{s,i} - (k - .5)·Fres]}{Fres}$  if [(k - .5)·Fres < fs,i ∧ (k + .5)·Fres > fs,i]
      Pei,n ← Pei,n otherwise
    Pei,n ← max(Pei,n, .0000000000000000)
  Pe

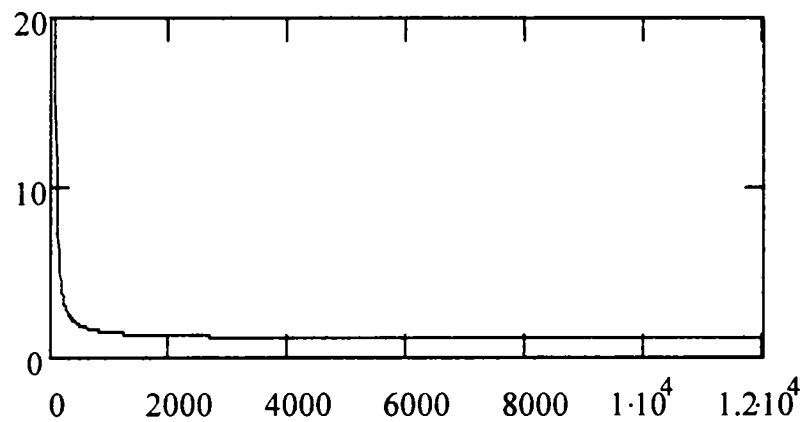
```

## 1.6. Adăugarea zgomotului intern

kk:= 0.. 108

Offset dependent de frecvență pentru compensarea zgomotului intern generat de ureche

$$Ptresh_{kk} := 10 \cdot 0.4 \cdot 0.364 \cdot \left( \frac{fc_{kk}}{1000} \right)^{-8}$$



Înălțimile (*Pitch patterns*)

$$Ppref_{kk,n} := Peref_{kk,n} + Ptresh_{kk}$$

$$Pptest_{kk,n} := Petest_{kk,n} + Ptresh_{kk}$$

## 1.7. Împrăștierea în frecvență

$$Lref_{kk,n} := 10 \cdot \log(Ppref_{kk,n})$$

$$Ltest_{kk,n} := 10 \cdot \log(Pptest_{kk,n})$$

Panta descendentă

$$Suref_{kk,n} := -24 - \frac{230}{fc_{kk}} + 0.2 \cdot Lref_{kk,n}$$

$$Sutest_{kk,n} := -24 - \frac{230}{fc_{kk}} + 0.2 \cdot Ltest_{kk,n}$$

Panta ascendentă

$$Sl_{kk,n} := 27$$

Z := 10<sup>9</sup>

res := 0.2<sup>ε</sup>

j := 0.. 10<sup>8</sup>

$$EEref_{j,kk} := \left[ \begin{array}{l} \frac{-res \cdot (j-kk) \cdot Sl_{j,0}}{10^{10}} \\ \frac{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,0}}{10^{10}} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot Suref_{j,0}}{10^{10}}}{10^{10}} \end{array} \right] \text{ if } kk < j$$

$$\left[ \begin{array}{l} \frac{res \cdot (kk-j) \cdot Suref_{j,0}}{10^{10}} \\ \frac{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,0}}{10^{10}} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot Suref_{j,0}}{10^{10}}}{10^{10}} \end{array} \right] \text{ otherwise}$$

$$EE_{test,j,kk} := \left[ \begin{array}{l} \frac{\frac{-res \cdot (j-kk) \cdot Sl_{j,0}}{10}}{10} \\ \frac{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,0}}{10} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot S_{utest,j,0}}{10}}{10} \end{array} \right] \text{ if } kk < j$$

$$\left[ \begin{array}{l} \frac{res \cdot (kk-j) \cdot S_{utest,j,0}}{10} \\ \sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,0}}{10} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot S_{utest,j,0}}{10} \end{array} \right] \text{ otherwise}$$

$$Normspref_{kk} := \left[ \sum_{j=0}^{Z-1} (EE_{ref,j,kk})^4 \right]^{\frac{1}{.4}}$$

$$Normspstest_{kk} := \left[ \sum_{j=0}^{Z-1} (EE_{test,j,kk})^4 \right]^{\frac{1}{.4}}$$

```

Elineref :=
  Eline ← .0
  for j ∈ 0..108
    for kk ∈ 0..108
      for n ∈ 0..nn
        Elinej, kk+n·108 ←  $\left[ \begin{array}{l} \frac{\frac{Lref_{j,n}}{10} \cdot \frac{-res \cdot (j-kk) \cdot Sl_{j,n}}{10}}{10} \\ \frac{\sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,n}}{10} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot S_{uref,j,n}}{10}}{10} \end{array} \right]$  if kk < j
        Elinej, kk+n·108 ←  $\left[ \begin{array}{l} \frac{Lref_{j,n}}{10} \cdot \frac{res \cdot (kk-j) \cdot S_{uref,j,n}}{10}}{10} \\ \sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,n}}{10} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot S_{uref,j,n}}{10} \end{array} \right]$  otherwise
      Eline
    
```

```

Elinetest := | Eline ← .0
              | for j ∈ 0..108
              |   for kk ∈ 0..108
              |     for n ∈ 0..nn
              |       Elinej, kk+n·108 ← 
$$\left[ \frac{\frac{Ltest_{j,n}}{10} \frac{-res \cdot (j-kk) \cdot Sl_{j,n}}{10}}{10} + \sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,n}}{10} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot Sute_{j,n}}{10} \right] \text{ if } kk < j$$

              |       Elinej, kk+n·108 ← 
$$\left[ \frac{\frac{Ltest_{j,n}}{10} \frac{res \cdot (kk-j) \cdot Sute_{j,n}}{10}}{10} + \sum_{\mu=0}^{j-1} \frac{-res \cdot (j-\mu) \cdot Sl_{j,n}}{10} + \sum_{\mu=j}^Z \frac{res \cdot (\mu-j) \cdot Sute_{j,n}}{10} \right] \text{ otherwise}$$

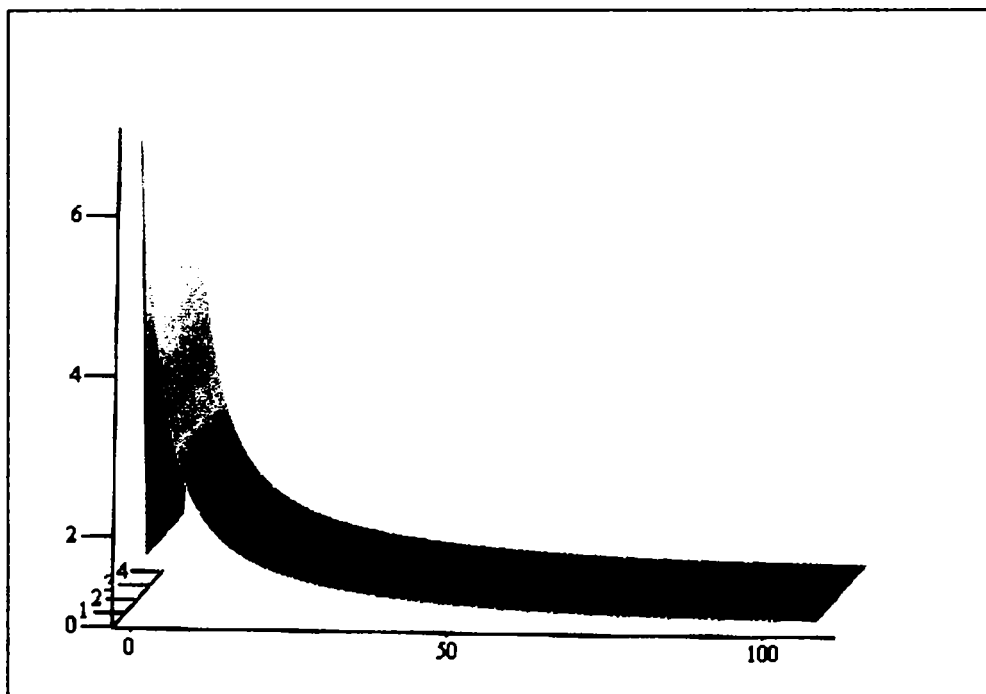
              |     Eline

```

Mărimile de excitație

$$E2ref_{kk,n} := \frac{\left[ \sum_{j=0}^{Z-1} (Elineref_{j, kk+n \cdot 108})^4 \right]^{2.5}}{Normspref_{kk}}$$

$$E2test_{kk,n} := \frac{\left[ \sum_{j=0}^{Z-1} (Elinetest_{j, kk+n \cdot 108})^4 \right]^{2.5}}{Normsptest_{kk}}$$



E2ref



## 1.8. Împrăștierea în domeniul timp

$$\tau_{100} := 0.02$$

$$\tau_{\min} := 0.001$$

$$\tau_{kk} := \tau_{\min} + \frac{100(\tau_{100} - \tau_{\min})}{f_{c_{kk}}}$$

Constanta de timp

$$a_{kk} := e^{\frac{-4}{187.5 \cdot \tau_{kk}}}$$

Energia în domeniul Bark

```

Eref := | Ef ← 0
        | for kk ∈ 0..108
        |   for n ∈ 0..nn
        |     Efkk,n ← 0
        |   for kk ∈ 0..108
        |     for n ∈ 1..nn
        |       Efkk,n ← akk · Efkk,n-1 + (1 - akk) · E2refkk,n
        | Ef

```

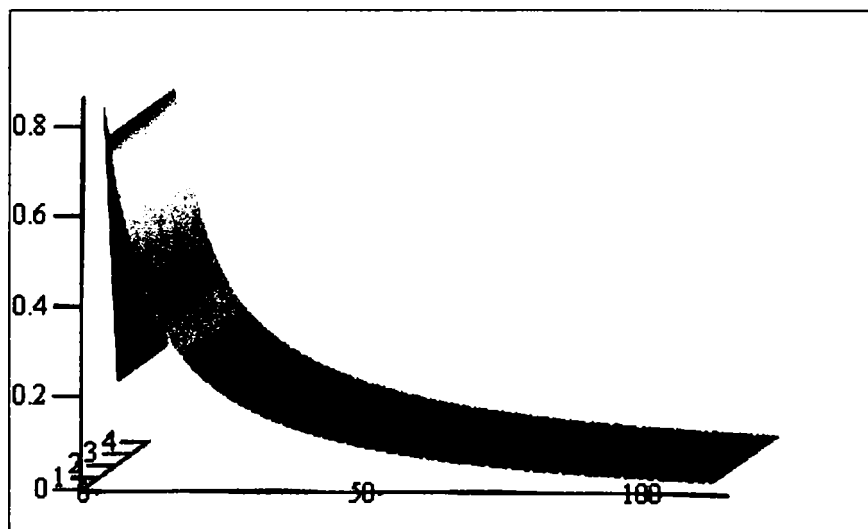
```

Etest := | Ef ← 0
         | for kk ∈ 0..108
         |   for n ∈ 0..nn
         |     Efkk,n ← 0
         |   for kk ∈ 0..108
         |     for n ∈ 1..nn
         |       Efkk,n ← akk · Efkk,n-1 + (1 - akk) · E2testkk,n
         | Ef

```

$$Elogref_{kk,n} := \log(\max(Efref_{kk,n}, E2ref_{kk,n}))$$

$$Elogtest_{kk,n} := \log(\max(Eftest_{kk,n}, E2test_{kk,n}))$$



Elogref

## Mărimile de excitație

$$E_{ref_{kk,n}} := \max(E_{ref_{kk,n}}, E_{2ref_{kk,n}})$$

$$E_{test_{kk,n}} := \max(E_{test_{kk,n}}, E_{2test_{kk,n}})$$

## 1.9. Pragul de mascare

Mărimi de excitație

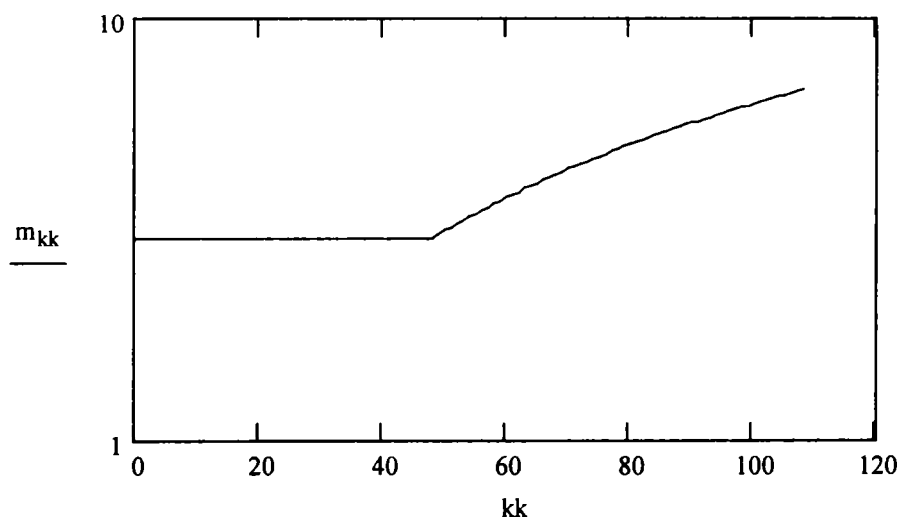
$$E := E_{test}$$

$$E_{log} := E_{logtest}$$

Parametru de ponderare (dB)

$$m_{kk} := \begin{cases} 3 & \text{if } kk \cdot 25 \leq 12 \\ 0.25 \cdot kk \cdot 25 & \text{otherwise} \end{cases}$$

Offsetul de frecvență funcție de banda critică



Pragul de mascare

$$M_{kk,n} := \frac{E_{kk,n}}{10^{\frac{m_{kk}}{10}}}$$

$$M_{log_{kk,n}} := \frac{E_{log_{kk,n}}}{10^{\frac{m_{kk}}{10}}}$$

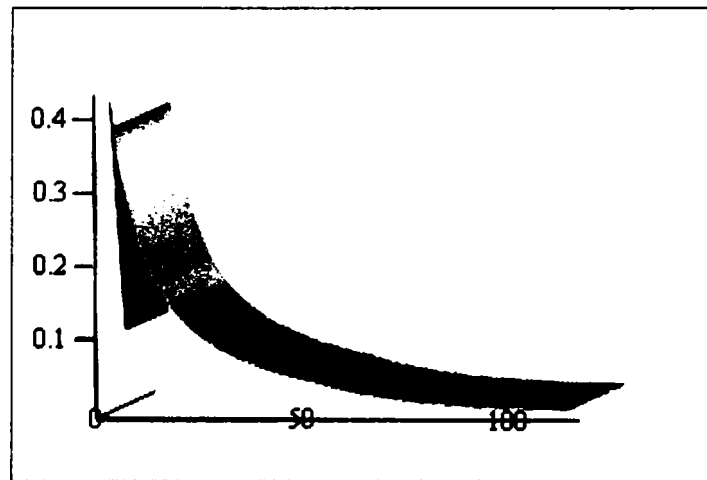
Mărimile de mascare

$$M_{ref_{kk,n}} := \frac{E_{ref_{kk,n}}}{10^{\frac{m_{kk}}{10}}}$$

$$M_{logref_{kk,n}} := \frac{E_{logref_{kk,n}}}{10^{\frac{m_{kk}}{10}}}$$

$$M_{test_{kk,n}} := \frac{E_{test_{kk,n}}}{10^{\frac{m_{kk}}{10}}}$$

$$M_{logtest_{kk,n}} := \frac{E_{logtest_{kk,n}}}{10^{\frac{m_{kk}}{10}}}$$



Mlog

## 2. Preprocesarea mărimilor de excitație

### 2.1 Adaptarea nivelului și mărimilor

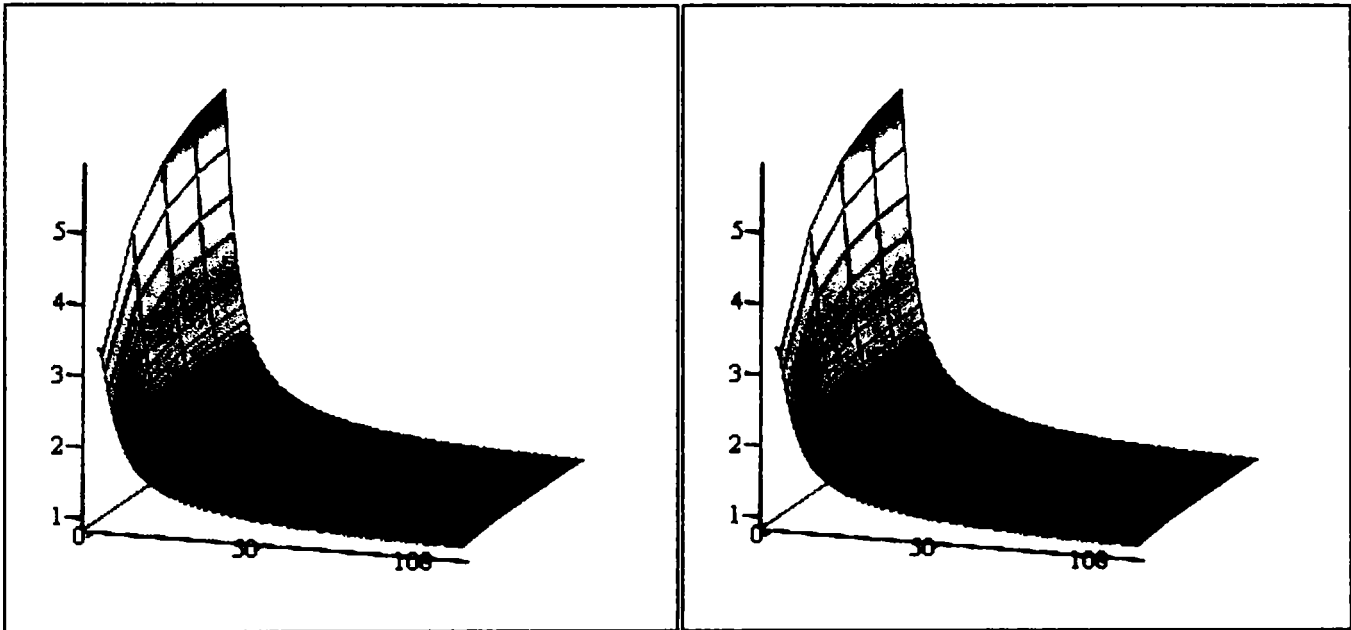
Calculul mărimilor de intrare filtrate pe benzi de frecvență

```
Prefl := Pref ← 0
for kk ∈ 0..108
  for n ∈ 0..nn + 1
    Prefkk,n ← 0
for kk ∈ 0..108
  for n ∈ 1..nn + 1
    Prefkk,n ← akk · Prefkk,n-1 + (1 - akk) · Erefkk,n-1
Pref
```

```
Pref := for kk ∈ 0..108
for n ∈ 0..nn
  Prefkk,n ← Preflkk,n+1
Pref
```

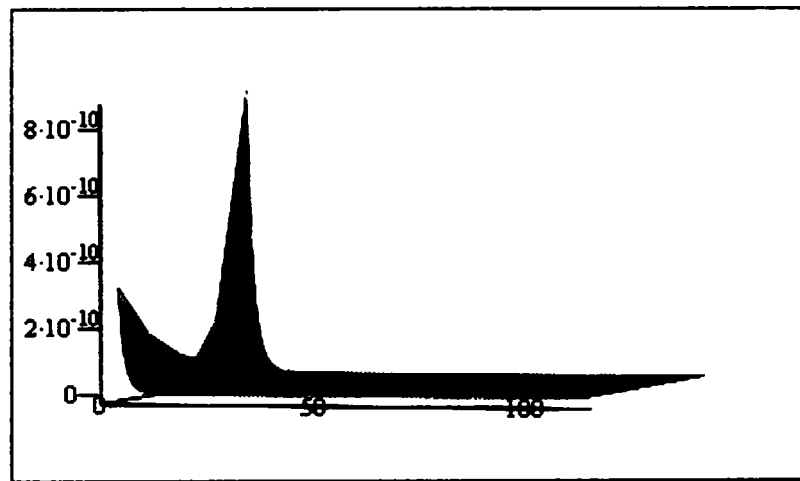
```
Ptestl := Pref ← 0
for kk ∈ 0..108
  for n ∈ 0..nn + 1
    Prefkk,n ← 0
for kk ∈ 0..108
  for n ∈ 1..nn + 1
    Prefkk,n ← akk · Prefkk,n-1 + (1 - akk) · Etestkk,n-1
Pref
```

```
Ptest := for kk ∈ 0..108
for n ∈ 0..nn
  Prefkk,n ← Ptestlkk,n+1
Pref
```



Pref

Ptest



Erroare

$$\tau_{100} := 0.0\% \quad \tau_{\min} := 0.00\% \quad \tau_{kk} := \tau_{\min} + \frac{100(\tau_{100} - \tau_{\min})}{f_{c,kk}}$$

### 2.1.1 Adaptarea nivelului

Factor de corecție momentan

$$\text{LevCorr}_n := \left( \frac{\sum_{kk=0}^{Z-1} \sqrt{\text{Pref}_{kk,n} \cdot \text{Ptest}_{kk,n}}}{\sum_{kk=0}^{Z-1} \text{Ptest}_{kk,n}} \right)^2$$

## Corelarea nivelului

$$ELRef_{kk,n} := \begin{cases} \frac{Eref_{kk,n}}{LevCorr_n} & \text{if } [(LevCorr_n > 1) \wedge (n \neq 0)] \\ Eref_{kk,n} & \text{otherwise} \end{cases}$$

$$ELTest_{kk,n} := \begin{cases} E_{kk,n} \cdot LevCorr_n & \text{if } LevCorr_n \leq 1 \\ E_{kk,n} & \text{otherwise} \end{cases}$$

## 2.1.2 Adaptarea modelului

Factorii de corecție pentru fiecare canal se calculează prin compararea anvelopei temporale a ieșirilor

$$R_{kk,n} := \frac{\sum_{i=0}^n (a_{kk})^i \cdot ELTest_{kk,n-i} \cdot ELRef_{kk,n-i}}{\sum_{i=0}^n (a_{kk})^i \cdot ELRef_{kk,n-i} \cdot ELRef_{kk,n-i}}$$

$$RTest_{kk,n} := \begin{cases} \frac{1}{R_{kk,n}} & \text{if } R_{kk,n} \geq 1 \wedge n \neq 0 \\ 0 & \text{if } \left[ \sum_{i=0}^n (a_{kk})^i \cdot ELRef_{kk,n-i} \cdot ELRef_{kk,n-i} \right] = 0 \wedge \left[ \sum_{i=0}^n (a_{kk})^i \cdot ELTest_{kk,n-i} \cdot ELRef_{kk,n-i} \right] > 0 \\ RTest_{kk-1,n} & \text{if } \left[ \sum_{i=0}^n (a_{kk})^i \cdot ELRef_{kk,n-i} \cdot ELRef_{kk,n-i} \right] = 0 \wedge \sum_{i=0}^n (a_{kk})^i \cdot ELTest_{kk,n-i} \cdot ELRef_{kk,n-i} = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$RRef_{kk,n} := \begin{cases} R_{kk,n} & \text{if } R_{kk,n} < 1 \\ 1 & \text{if } \left[ \sum_{i=0}^n (a_{kk})^i \cdot ELRef_{kk,n-i} \cdot ELRef_{kk,n-i} \right] = 0 \wedge \left[ \sum_{i=0}^n (a_{kk})^i \cdot ELTest_{kk,n-i} \cdot ELRef_{kk,n-i} \right] > 0 \\ RRef_{kk-1,n} & \text{if } \left[ \sum_{i=0}^n (a_{kk})^i \cdot ELRef_{kk,n-i} \cdot ELRef_{kk,n-i} \right] = 0 \wedge \sum_{i=0}^n (a_{kk})^i \cdot ELTest_{kk,n-i} \cdot ELRef_{kk,n-i} = 0 \\ 1 & \text{otherwise} \end{cases}$$

Lățimea ferestrei de frecvență

MM := 8

$$M1 := \begin{cases} \frac{(MM - 1)}{2} & \text{if } \text{trunc}\left(\frac{MM}{2}\right) - \frac{MM}{2} \neq 0 \\ \frac{MM}{2} - 1 & \text{otherwise} \end{cases} \quad M1 = 3$$

$$M2 := \begin{cases} \frac{(MM - 1)}{2} & \text{if } \text{trunc}\left(\frac{MM}{2}\right) - \frac{MM}{2} \neq 0 \\ \frac{MM}{2} & \text{otherwise} \end{cases} \quad M2 = 4$$

Factorii de corecție mediați pe M benzi și curățați în timp

```
PattCorrTest1 := PCR ← 0
                for kk ∈ 0..108
                  for n ∈ 0..nn
                    PCRkk,n ← 0
                  for kk ∈ 0..108
                    M11 ← min(M1, kk)
                    M22 ← min(M2, Z - kk - 1)
                    for n ∈ 1..nn + 1
                      PCRkk,n ← akk · PCRkk,n-1 + (1 - akk) ·  $\frac{\sum_{i=-M11}^{M22} RTest_{kk+i,n-1}}{MM}$ 
                    PCR
```

```
PattCorrTest := for kk ∈ 0..108
                 for n ∈ 0..nn
                   Prefkk,n ← PattCorrTest1kk,n+1
                 Pref
```

```
PattCorrRef1 := PCR ← 0
                for kk ∈ 0..108
                  for n ∈ 0..nn
                    PCRkk,n ← 0
                  for kk ∈ 0..108
                    M11 ← min(M1, kk)
                    M22 ← min(M2, Z - kk - 1)
                    for n ∈ 1..nn + 1
                      PCRkk,n ← akk · PCRkk,n-1 + (1 - akk) ·  $\frac{\sum_{i=-M11}^{M22} RRef_{kk+i,n-1}}{MM}$ 
                    PCR
```

```

PattCorrRef := | for kk ∈ 0..108
                |   for n ∈ 0..nn
                |     Ptestkk,n ← PattCorrRefkk,n+1
                | Ptest

```

Mărimile spectrale adaptate

$$E\text{PRef}_{kk,n} := E\text{LRef}_{kk,n} \cdot \text{PattCorrRef}_{kk,n}$$

$$E\text{PTest}_{kk,n} := E\text{LTest}_{kk,n} \cdot \text{PattCorrTest}_{kk,n}$$

## 2.2 Calculul modulației

Diferențele medii

```

EderTest := | EderTest ← 0
              | for kk ∈ 0..108
              |   for n ∈ 0..nn
              |     EderTestkk,n ← 0
              |   for kk ∈ 0..108
              |     for n ∈ 1..nn
              |       EderTestkk,n ← akk · EderTestkk,n-1 + (1 - akk) ·  $\frac{48000 \left| (E2\text{test}_{kk,n})^{0.3} - (E2\text{test}_{kk,n-1})^{0.3} \right|}{1024}$ 
              | EderTest

```

```

EderRef := | EderRef ← 0
            | for kk ∈ 0..108
            |   for n ∈ 0..nn
            |     EderRefkk,n ← 0
            |   for kk ∈ 0..108
            |     for n ∈ 1..nn
            |       EderRefkk,n ← akk · EderRefkk,n-1 + (1 - akk) ·  $\frac{48000 \left| (E2\text{ref}_{kk,n})^{0.3} - (E2\text{ref}_{kk,n-1})^{0.3} \right|}{1024}$ 
            | EderRef

```

## Tăria medie

```

EbaratTest := | EbaratTest ← 0
                | for kk ∈ 0..108
                |   for n ∈ 0..nn
                |     EbaratTestkk,n ← 0
                |   for kk ∈ 0..108
                |     for n ∈ 1..nn
                |       EbaratTestkk,n ← akk·EbaratTestkk,n-1 + (1 - akk)·(E2testkk,n)0.3
                | EbaratTest

```

```

EbaratRef := | EbaratRef ← 0
              | for kk ∈ 0..108
              |   for n ∈ 0..nn
              |     EbaratRefkk,n ← 0
              |   for kk ∈ 0..108
              |     for n ∈ 1..nn
              |       EbaratRefkk,n ← akk·EbaratRefkk,n-1 + (1 - akk)·(E2refkk,n)0.3
              | EbaratRef

```

## Modulația înfășurătoarei pentru ieșirea fiecărui filtru

$$\text{Modtest}_{kk,n} := \frac{\text{EderTest}_{kk,n}}{1 + \frac{\text{EbaratTest}_{kk,n}}{0.3}} \quad \text{Modref}_{kk,n} := \frac{\text{EderRef}_{kk,n}}{1 + \frac{\text{EbaratRef}_{kk,n}}{0.3}}$$

## 2.3 Calculul tăriei

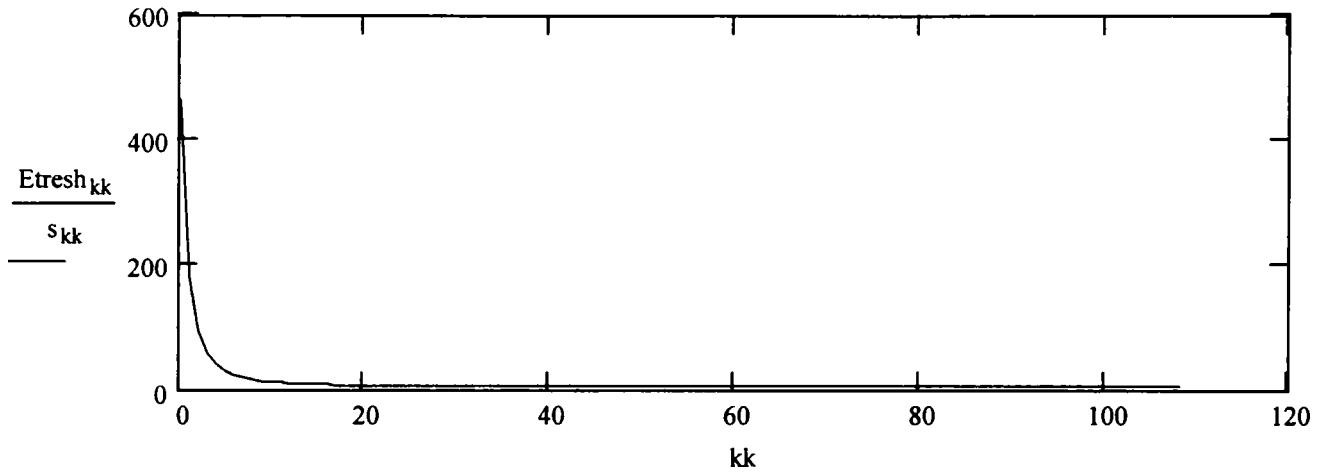
## Indexul pragului

$$s_{kk} := 10 \frac{\left( -2 - 2.05 \cdot \text{atan}\left(\frac{fc_{kk}}{4000}\right) - 0.75 \cdot \text{atan}\left(\frac{fc_{kk}}{1600}\right)^2 \right)}{10}$$

$$\text{Excitația la prag} \quad \text{Etrash}_{kk} := 10 \cdot 0.364 \cdot \left(\frac{fc_{kk}}{1000}\right)^{-0.8}$$

$$\text{const} := 1.0766$$





Tăria specifică

$$N_{kk,n} := \text{const} \cdot \left( \frac{\text{Etresh}_{kk}}{10000 s_{kk}} \right)^{.23} \cdot \left[ \left( 1 - s_{kk} + \frac{s_{kk} \cdot E_{kk,n}}{\text{Etresh}_{kk}} \right)^{0.23} - 1 \right]$$

Tăria totală a semnalului

$$N_{\text{total}_n} := \frac{24 \sum_{k=0}^{108} \max(N_{k,n}, 0)}{109}$$

## 2.4 Calculul semnalului de eroare

kk := 0..108

n := 0..4

$$F_{\text{noise}_{kk,n}} := \left| |F_{\text{eref}_{kk,n}}| - |F_{\text{etest}_{kk,n}}| \right|$$

Pnoise := Pnoise ← 0

for n ∈ 0..nn

for i ∈ 0..108

Pnoise<sub>i,n</sub> ← 0

for k ∈ 0..108

Pnoise<sub>i,n</sub> ← Pnoise<sub>i,n</sub> + Fnoise<sub>k,n</sub> if  $\left[ \left[ (k - .5) \cdot \text{Fres} \geq \text{fi}_i \right] \wedge \left[ (k + .5) \cdot \text{Fres} \leq \text{fs}_i \right] \right]$

Pnoise<sub>i,n</sub> ← Pnoise<sub>i,n</sub> + Fnoise<sub>k,n</sub> ·  $\frac{(\text{fs}_i - \text{fi}_i)}{\text{Fres}}$  if  $\left[ (k - .5) \cdot \text{Fres} < \text{fi}_i \wedge (k + .5) \cdot \text{Fres} > \text{fs}_i \right]$

Pnoise<sub>i,n</sub> ← Pnoise<sub>i,n</sub> + Fnoise<sub>k,n</sub> ·  $\frac{\left[ (k + .5) \cdot \text{Fres} - \text{fi}_i \right]}{\text{Fres}}$  if  $\left[ (k - .5) \cdot \text{Fres} < \text{fi}_i \wedge (k + .5) \cdot \text{Fres} > \text{fs}_i \right]$

Pnoise<sub>i,n</sub> ← Pnoise<sub>i,n</sub> + Fnoise<sub>k,n</sub> ·  $\frac{\left[ \text{fs}_i - (k - .5) \cdot \text{Fres} \right]}{\text{Fres}}$  if  $\left[ (k - .5) \cdot \text{Fres} < \text{fs}_i \wedge (k + .5) \cdot \text{Fres} > \text{fs}_i \right]$

Pnoise<sub>i,n</sub> ← Pnoise<sub>i,n</sub> otherwise

Pnoise<sub>i,n</sub> ← max(Pnoise<sub>i,n</sub>, .00000000000000000000)

Pnoise

### 3. Calculul variabilelor de ieșire ale modelului

#### 3.2 Diferența de modulație

Parametrii

$$\text{offset} := \begin{pmatrix} 1 \\ 0.01 \end{pmatrix} \quad \text{levWt} := 100 \quad \text{negWt} := \begin{pmatrix} 1 \\ 0.1 \end{pmatrix} \quad q := 0..1$$

q - parametru pentru selectarea varabilelor de modulație diferențială

q=0 Diif1

q=1 Diff2

$$w\text{Diff1}_{kk,n} := \begin{cases} 1 & \text{if } \text{Modtest}_{kk,n} > \text{Modref}_{kk,n} \\ 1 & \text{otherwise} \end{cases} \quad w\text{Diff2}_{kk,n} := \begin{cases} 1 & \text{if } \text{Modtest}_{kk,n} > \text{Modref}_{kk,n} \\ 0.1 & \text{otherwise} \end{cases}$$

Diferența instantanee de modulație

$$\text{ModDiff1}_{kk,n} := w\text{Diff1}_{kk,n} \cdot \frac{|\text{Modtest}_{kk,n} - \text{Modref}_{kk,n}|}{1 + \text{Modref}_{kk,n}}$$

$$\text{ModDiff2}_{kk,n} := w\text{Diff2}_{kk,n} \cdot \frac{|\text{Modtest}_{kk,n} - \text{Modref}_{kk,n}|}{0.01 + \text{Modref}_{kk,n}}$$

Diferența momentană de modulație

$$\text{ModDiff1Total}_n := \frac{100 \sum_{k=0}^{108} \text{ModDiff1}_{k,n}}{109}$$

$$\text{ModDiff2Total}_n := \frac{100 \sum_{k=0}^{108} \text{ModDiff2}_{k,n}}{109}$$

Factor de ponderare

$$\text{TempWt}_n := \sum_{k=0}^{108} \frac{\text{EbaratRef}_{k,n}}{\text{EbaratRef}_{k,n} + \text{levWt} \cdot (\text{Etrash}_k)^{0.3}}$$

#### 3.3 Tăria zgomotului

Parametrii

$$\text{ThresFac0} := 0.15 \quad qq := 0..1 \quad \text{NLmin} := 0 \quad s0 := \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \quad \alpha := 1.5$$

s0 - parametru pentru selectarea varabilelor de ieșire pentru tăria totală a zgomotului

s0=0 Missing Components

s0=1 Noise Loud

$$\text{stest}_{kk,n} := \text{ThresFac0} \cdot \text{Modtest}_{kk,n} + s0$$

$$\text{sref}_{kk,n} := \text{ThresFac0} \cdot \text{Modref}_{kk,n} + s0$$

Coefficient pentru determinarea cantității de mascare

$$\beta_{kk,n} := e^{-\alpha \cdot \frac{EPTest_{kk,n} - EPref_{kk,n}}{EPref_{kk,n}}}$$

Tăria parțială a zgomotului

$$NL_{kk,n} := \left( \frac{E_{resh_{kk}}}{stest_{kk,n} \cdot E_0} \right)^{.23} \cdot \left[ 1 + \frac{\max(stest_{kk,n} \cdot EPTest_{kk,n} - sref_{kk,n} \cdot EPref_{kk,n}, 0)}{E_{resh_{kk}} + sref_{kk,n} \cdot (EPref_{kk,n}) \cdot \beta_{kk,n}} \right]^{.23} - 1$$

$E_0 := 1$

$L := 4$  lungimea ferestrei de eșantionare în timp

$$WinModDiff1 := \sqrt{\frac{\sum_{n=L-1}^{nn-1} \left( \frac{\sum_{i=0}^{L-1} \sqrt{ModDiff1Total_{n-i}}}{L} \right)^4}{nn - L + 1}}$$

$$WinModDiff2 := \sqrt{\frac{\sum_{n=L-1}^{nn-1} \left( \frac{\sum_{i=0}^{L-1} \sqrt{ModDiff2Total_{n-i}}}{L} \right)^4}{nn - L + 1}}$$

$$WinNoiseLoud := \sqrt{\frac{\sum_{n=L-1}^{nn-1} \left( \frac{\sum_{i=0}^{L-1} \sum_{kk=0}^{108} \sqrt{Pnoise_{kk,n-i}}}{L} \right)^4}{nn - L + 1}}$$

$$WinMissingComponents := \sqrt{\frac{\sum_{n=L-1}^{nn-1} \left( \frac{\sum_{i=0}^{L-1} \sqrt{ModDiff2Total_{n-i}}}{L} \right)^4}{nn - L + 1}}$$

$$AvgModDiff1 := \frac{\sum_{n=0}^{nn} ModDiff1Total_n}{nn + 1}$$

$$AvgModDiff2 := \frac{\sum_{n=0}^{nn} ModDiff2Total_n}{nn + 1}$$

$$\text{AvgNoiseLoud} := \frac{\sum_{kk=0}^{108} \sum_{n=0}^{nn} (\text{Pnoise}_{kk,n})}{(nn+1) \cdot 108}$$

$$\text{AvgMissingComponents} := \frac{\sum_{n=0}^{nn} \text{ModDiff2Total}_n}{nn+1}$$

$$\text{RmsModDiff1} := \sqrt{\frac{\sum_{n=0}^{nn} (\text{ModDiff1Total}_n)^2}{nn+1}}$$

$$\text{RmsModDiff2} := \sqrt{\frac{\sum_{n=0}^{nn} (\text{ModDiff2Total}_n)^2}{nn+1}}$$

$$\text{RmsNoiseLoud} := \sqrt{\frac{\sum_{kk=0}^{108} \sum_{n=0}^{nn} (\text{Pnoise}_{kk,n})^2}{(nn+1) \cdot 108}}$$

$$\text{RmsMissingComponents} := \sqrt{\frac{\sum_{n=0}^{nn} (\text{ModDiff2Total}_n)^2}{nn+1}}$$

$$\text{RmsWModDiff1} := \sqrt{nn+1} \cdot \sqrt{\frac{\sum_{n=0}^{nn} (\text{TempWt}_n \cdot \text{ModDiff1Total}_n)^2}{\sum_{n=0}^{nn} (\text{TempWt}_n)^2}}$$

$$\text{RmsWModDiff2} := \sqrt{nn+1} \cdot \sqrt{\frac{\sum_{n=0}^{nn} (\text{TempWt}_n \cdot \text{ModDiff2Total}_n)^2}{\sum_{n=0}^{nn} (\text{TempWt}_n)^2}}$$

$$\text{RmsWNoiseLoud} := \sqrt{nn+1} \cdot \sqrt{\frac{\sum_{n=0}^{nn} \left( \text{TempWt}_n \cdot \sum_{kk=0}^{108} \text{Pnoise}_{kk,n} \right)^2}{\sum_{n=0}^{nn} (\text{TempWt}_n)^2}}$$

$$\text{RmsWMissingComponents} := \sqrt{nn+1} \cdot \sqrt{\frac{\sum_{n=0}^{nn} (\text{TempWt}_n \cdot \text{ModDiff2Total}_n)^2}{\sum_{n=0}^{nn} (\text{TempWt}_n)^2}}$$

$$\begin{aligned} \text{WinModDiff1} &= 1.377 \times 10^{-10} \\ \text{WinModDiff2} &= 1.068 \times 10^{-9} \\ \text{AvgModDiff1} &= 3.643 \times 10^{-10} \\ \text{AvgModDiff2} &= 3.314 \times 10^{-9} \\ \text{RmsModDiff1} &= 5.131 \times 10^{-10} \\ \text{RmsModDiff2} &= 4.907 \times 10^{-9} \\ \text{RmsWModDiff1} &= 1.31 \times 10^{-9} \\ \text{RmsWModDiff2} &= 1.256 \times 10^{-8} \end{aligned}$$

$$\begin{aligned} \text{WinNoiseLoud} &= 1.097 \times 10^{-6} \\ \text{AvgNoiseLoud} &= 2.15 \times 10^{-8} \\ \text{RmsNoiseLoud} &= 3.127 \times 10^{-7} \\ \text{RmsWNoiseLoud} &= 7.876 \times 10^{-6} \\ \text{WinMissingComponents} &= 1.068 \times 10^{-9} \\ \text{AvgMissingComponents} &= 3.314 \times 10^{-9} \\ \text{RmsMissingComponents} &= 4.907 \times 10^{-9} \\ \text{RmsWMissingComponents} &= 1.256 \times 10^{-8} \end{aligned}$$

### 3.4 Lățimea benzii

$$kx_{kk} := \text{round}\left(\frac{fc_{kk} \cdot 2048}{48000}\right)$$

$$\text{FLevelTest}_{kf, n} := 10 \cdot \log(\text{FFT}_{kf, n})$$

$$kx1 := \text{round}\left(8109 \frac{2048}{48000}\right) \quad kx1 = 346$$

$$kx2 := \text{round}\left(21586 \frac{2048}{48000}\right) \quad kx2 = 921$$

$$\text{ZeroThreshold} := \text{FLevelTest}(921)$$

$$\text{BwRef} := 0 \quad \text{BwTest} := 0$$

$$\text{ZeroThreshold}_n := \begin{cases} \text{for } k \in 921..1023 \\ \text{ZeroThreshold} \leftarrow \max(\text{ZeroThreshold}, \text{FLevelTest}_{kf, n}) \\ \text{ZeroThreshold} \end{cases}$$

### 3.5 Raportul zgomot mascare

Valoarea locală NMR a cadrului curent

$$\text{NMRlocal}_n := 10 \cdot \log\left(\frac{\sum_{k=0}^{Z-1} \frac{\text{Pnoise}_{k, n}}{M_{k, n}}}{Z}\right)$$

TotalNMR

$$\text{NMRtotal} := 10 \cdot \log \left( \frac{\sum_{n=0}^{nn} \frac{\sum_{k=0}^{Z-1} \frac{\text{Pnoise}_{k,n}}{M_{k,n}}}{Z}}{nn + 1} \right)$$

$$\text{NMRlocal} = \begin{pmatrix} -81.824 \\ -83.155 \\ -83.302 \\ -77.07 \\ -71.219 \end{pmatrix}$$

$$\text{NMRtotal} = -76.527$$

### 3.6 Relative Disturbed Frames

$$\text{RelDistFramesB} := \begin{array}{l} i \leftarrow 0 \\ \text{for } n \in 0..nn \\ \quad \text{for } kk \in 0..108 \\ \quad \quad \text{MK}_{kk,n} \leftarrow 10 \cdot \log \left( \frac{\text{Pnoise}_{kk,n}}{M_{kk,n}} \right) \\ \quad \quad i \leftarrow i + 1 \text{ if } \max(\text{MK}) \geq 1.5 \\ i \end{array}$$

### 3.7 Probabilitatea detecției

Mărimi logaritmice de excitație

$$\text{Edb\_test}_{kk,n} := 10 \cdot \log(\text{Etest}_{kk,n}) \quad \text{Edb\_ref}_{kk,n} := 10 \cdot \log(\text{Eref}_{kk,n})$$

Excitația medie asimetrică  $S := 0$ 

$$L_{kk,n} := 0.3 \cdot \max(\text{Edb\_ref}_{kk,n}, \text{Edb\_test}_{kk,n}) + 0.7 \cdot \text{Edb\_test}_{kk,n}$$

$$S_{kk,n} := \begin{cases} S_{kk,n} \leftarrow \left[ 5.95072 \left( \frac{6.39468}{L_{kk,n}} \right)^{1.71332} + 9.0103310^{-11} \cdot (L_{kk,n})^4 + 5.0562210^{-6} \cdot (L_{kk,n})^3 - 0.00102438(L_{kk,n})^2 + 0.0550197L_{kk,n} - 0.198719 \right] & \text{if } L_{kk,n} > 0 \\ S_{kk,n} \leftarrow 10^{30} & \text{otherwise} \end{cases}$$

Eroarea cu semn  $e_{kk,n} := Edb\_ref_{kk,n} - Edb\_test_{kk,n}$

$$b_{kk,n} := \begin{cases} b_{kk,n} \leftarrow 4 & \text{if } Edb\_ref_{kk,n} > Edb\_test_{kk,n} \\ b_{kk,n} \leftarrow 6 & \text{otherwise} \end{cases}$$

Factorul de scală  $a_{kk,n} := \frac{10^{\frac{\log(\log(2))}{b_{kk,n}}}}{S_{kk,n}}$

Probabilitatea detecției

$$pc1_{kk,n} := 1 - 10^{-a_{kk,n} \cdot (e_{kk,n})^{b_{kk,n}}} \quad pc_{kk,n} := 2 - .5 \left( \frac{e_{kk,n}}{S_{kk,n}} \right)^{b_{kk,n}}$$

Numărul total de pași peste prag

$$qc_{kk,n} := \frac{\lceil \text{ceil}(e_{kk,n}) \rceil}{S_{kk,n}}$$

Probabilitatea binaurală de detecție

$$pbin_{kk,n} := \max(pcst_{kk,n}, pcdr_{kk,n}) \quad qbin_{kk,n} := \max(qst_{kk,n}, qdr_{kk,n})$$

Probabilitatea de detecție totală a canalului  $c$  al cadrului  $n$

$$Pc_n := 1 - \prod_{kk=0}^{108} (1 - pc_{kk,n}) \quad Qc_n := \sum_{kk=0}^{108} qc_{kk,n} \quad Qc = \begin{pmatrix} 1.993 \\ 0.442 \\ 0.024 \\ 1.787 \\ 1.811 \end{pmatrix}$$

### 3.7.1. Probabilitatea maximă de detecție filtrată

$$\text{StepSize} := 1024 \quad c0 := 0.9^{\frac{\text{StepSize}}{1024}} \quad c0 = 0.9$$

$$Pbaratc := \begin{cases} \text{for } n \in 0..nn + 1 \\ \quad Pbaratc_n \leftarrow 0 \\ \text{for } n \in 1..nn + 1 \\ \quad Pbaratc_{n-1} \leftarrow (1 - c0) \cdot Pc_{n-1} + c0 \cdot Pbaratc_{n-1} \\ Pbaratc \end{cases}$$

$$c1 := 0.99 \frac{\text{StepSize}}{1024} \quad c1 = 0.99$$

$$PMc := \begin{cases} \text{for } n \in 0..nn + 1 \\ \quad PMc_n \leftarrow 0 \\ \text{for } n \in 1..nn + 1 \\ \quad PMc_n \leftarrow \max(c1 \cdot PMc_{n-1}, Pbaratc_{n-1}) \\ PMc \end{cases}$$

$$MFPD := PMc_{nn}$$

$$MFPD = 0.1$$

### 3.7.2. Blocul mediu distorsionat

$$Q := \sum_{n=0}^{nn} Qc_n \quad Q = 6.057$$

$$ndist := \begin{cases} ndist \leftarrow 0 \\ \text{for } n \in 0..nn \\ \quad ndist \leftarrow ndist + 1 \text{ if } Pc_n > 0.5 \\ ndist \end{cases} \quad ndist = 5$$

$$ADP := \begin{cases} ADP \leftarrow 0 \text{ if } ndist = 0 \\ ADP \leftarrow \log\left(\frac{Q}{ndist}\right) \text{ if } ndist > 0 \wedge Q > 0 \\ ADP \leftarrow -0.5 \text{ if } ndist > 0 \wedge Q = 0 \\ ADP \end{cases} \quad ADP = 0.083$$

### 3.8. Structura armonică a erorii

$$B1 := \begin{cases} \text{for } i \in 0..806327 \quad \text{Corelatia} := \text{corr}(Uref, B1) \\ \quad B1_i \leftarrow Utest_i \\ B1 \end{cases} \quad \text{Corelatia} = 0.946$$



## 4. Estimarea calității audio de bază

### 4.1. Rețeaua neuronală

I := 11    numărul de intrări

J := 3    numărul de noduri în stratul ascuns

EHS := 0            BandwidthTest := 0            BandwidthRef := 0

A := READWAV("pop1\_m\_o.wav" )

Factori de scală de intrare            amin := min(A)            amax := max(A)

Factori de scală de ieșire            bmin := -3.98            bmax := 0.22

$$\begin{array}{l}
 \left( \begin{array}{c}
 \text{BandwidthRef} \\
 \text{BandwidthTest} \\
 \text{NMRtotal} \\
 \text{WinModDiff1} \\
 \text{ADB} \\
 \text{EHS} \\
 \text{AvgModDiff1} \\
 \text{AvgModDiff2} \\
 \text{RmsNoiseLoud} \\
 \text{MFPD} \\
 \text{RelDistFrames}
 \end{array} \right) \\
 X :=
 \end{array}
 \begin{array}{l}
 \left( \begin{array}{c}
 400 \\
 100 \\
 -11 \\
 5 \\
 1.4 \\
 6 \\
 31 \\
 1 \\
 7 \\
 0.5 \\
 0.4
 \end{array} \right) \\
 X1 :=
 \end{array}
 \begin{array}{l}
 \left( \begin{array}{c}
 393.91665 \\
 361.96533 \\
 -24.04511 \\
 1.11066 \\
 -0.20662 \\
 0.07431 \\
 1.11368 \\
 0.95034 \\
 0.02998 \\
 0.00010 \\
 0
 \end{array} \right) \\
 \text{amirr} =
 \end{array}
 \begin{array}{l}
 \left( \begin{array}{c}
 921 \\
 881.13122 \\
 16.21203 \\
 107.13777 \\
 2.88601 \\
 13.93335 \\
 63.25787 \\
 1145.01855 \\
 14.81974 \\
 1 \\
 1
 \end{array} \right) \\
 \text{amax} =
 \end{array}$$

esant := length(A)

$$s := \frac{1023}{\text{amax} - \text{amin}}$$

Set de ponderi de intrare

$$\begin{array}{l}
 \left( \begin{array}{ccc}
 -0.502657 & 0.436333 & 1.219602 \\
 4.307481 & 3.246017 & 1.123743 \\
 4.984241 & -2.211189 & -0.192096 \\
 0.051056 & -1.762424 & 4.331315 \\
 2.321580 & 1.789971 & -0.754560 \\
 -5.303901 & -3.452257 & -10.814982 \\
 2.730991 & -6.111805 & 1.519223 \\
 0.624950 & -1.331523 & -5.955151 \\
 3.102889 & 0.871260 & -5.922878 \\
 -1.051468 & -0.939882 & -0.142913 \\
 -1.804679 & -0.503610 & -0.620456 \\
 -2.518254 & 0.654841 & -2.207228
 \end{array} \right) \\
 Wx :=
 \end{array}$$

$$W_y := \begin{pmatrix} -3.817048 \\ 4.107138 \\ 4.629582 \\ -0.307594 \end{pmatrix}$$

Set de ponderi de ieșire

Indexul de distorsiuni

$$ID := W_{y_j} + \sum_{j=0}^{J-1} \left[ (W_{y_j}) \cdot \frac{1}{1 + e^{-\left[ (W_{x_{i-1},j}) + \sum_{i=0}^{I-1} (W_{x_{i,j}}) \cdot \frac{X_i - \text{amin}_i}{\text{amax}_i - \text{amin}_i} \right]}} \right]$$

Gradul de diferență obiectivă

$$ODG := b_{\min} + (b_{\max} - b_{\min}) \cdot \frac{1}{1 + e^{ID}}$$

ODG = -3.945 - pentru fișierele alese la simulare



## BIBLIOGRAFIE

- [AES00] - AES information document for digital audio - *Personal computer audio quality measurements*, în J. Audio Eng. Soc., Vol.48, No.5, 2000 May, pp.450-482
- [ALE99] - Alexa, F. - *Introducere în tehnica sunetului*, Editura de Vest, Timișoara 1999
- [ALL00] - Allard, L., Rossi, M., Zufferey, M.-A. - *Assessment of the Pertinence of Perceived Attributes for Listening Tests*, presented at the AES 108th Convention 2000 February 19-22 Paris, France, preprint nr. 5089 (C-6)
- [BAU00] - Baumgarte F. - *Ein psychophysiologisches Gehörmodell zur Nachbildung von Wahrnehmungsschwellen für Audiocodierung*, Dissertation, Universität Hannover, 2000
- [BAU95] - Baumgarte F., Charalampos F. - *A Nonlinear Psychoacoustic Model Applied to the ISO MPEG Layer 3 Coder*, presented at 99<sup>th</sup> AES Convention, New York, October 1995, Preprint 4087
- [BAU97] - Baumgarte F. - *A Physiological Ear Model for Auditory Masking Applicable to Perceptual Coding*, presented at 103<sup>rd</sup> AES Convention, New York, September 1997, Preprint 4511
- [BAU98] - Baumgarte F. - *Evaluation of a Physiological Ear Model Considering Masking Effects Relevant to Audio Coding*, presented at 105<sup>th</sup> AES Convention, San Francisco, September 1998, Preprint 4789
- [BAU99] - Baumgarte F. - *Application of a Physiological Ear Model to Irrelevance Reduction in Audio Coding*, AES 17<sup>th</sup> International Conference on High Quality Audio Coding, Florence, Italy, September 1999, p.171-181
- [BEA94] - Beaton, R.J. - *Network Implications of Low Bitrate Audio Codecs*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.553-566
- [BEE92] - Beerends, J.G., Stemerdink, J.A. - *A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation*, în J. Audio Eng. Soc., Vol.40, No.12, 1992 December, pp.963-978
- [BEE94] - Beerends, J.G., Stemerdink, J.A. - *Modelling A Cognitive Aspect in the Measurement of the Quality of Music Codecs*, presented at the AES 96th Convention 1994 February 26 - March 01, Amsterdam, preprint nr. 3800 (P4.1)
- [BEE96] - Beerends, J.G., van den Brink, W.A.C., Rodger, B. - *The role of informational masking and perceptual streaming in the measurement of music codec quality*, presented at the AES 100th Convention 1996 May, Copenhagen, preprint nr. 4176
- [BER71] - Berindei, M., Rebreanu, N., Mișcă, A. - *Tehnica sunetului - Captarea*, Ed. Tehnică, București - 1971
- [BOL98] - Boland S., Deriche M. - *Hybrid LPC and discrete wavelet transform audio coding with a novel bit allocation algorithm*, în Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP-98), May 1998, pp. 3657-3660
- [BOS92] - Bosi, M., Davidson, G. - *High Quality, Low-Rate Audio Transform Coding for Transmission and Multimedia Applications*, presented at the AES 93th Convention, 1992 October 1-4
- [BRA92] - Brandenburg, K., Stoll, G. - *The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio*, presented at 92nd AES Convention, Vienna, Austria, 24-27 March, 1992
- [BRA94] - Brandenburg, K., Buchta, R., Eberlein, E. etc. - *Second Generation ISO/MPEG Audio Layer III Coding*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.501-509
- [BRA94] - Brandenburg, K., Stoll, G. - *The ISO-MPEG-1 audio: A Generic Standard for Coding of High Quality Digital Audio*, J. Audio Eng. Soc., Oct. 1994, pp.780-792
- [BRA98] - Braun, D., Boland, M. - *On the Perception of Data Reduction with MiniDisc*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.710-723

- [BRA99] - Brandenburg, K. - *MP3 and AAC Explained*, presented at AES 17<sup>th</sup> International Conference on High Quality Audio Coding, Florence, Italy, 1999, September 2 -5
- [CEL94] - Cellier, C., Chenes, P., Rossi, M. - *Lossless coding*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.464-483
- [COL94] - Colomes, C., Lever, M., Dehery, Y-F., Faucon G. - *A Perceptual Objective Measurement system (POM) for the Quality Assessment of Perceptual Codecs*, presented at the AES 96th Convention 1994 February 26 - March 01, Amsterdam, preprint nr. 3801 (P4.2)
- [CRA94] - Craig, C., Davidson, G., Davis, M.F., Fielder, L.D., Link, B.D., Vernon, S. - *AC3: Flexible Perceptual Coding for Audio Transmission and Storage*, presented at the AES 96th Convention, 1994 February 26 - March 1, Paris, Preprint nr.3796
- [DAV93] - Davis, M.F. - *The AC-3 Multichannel Coder*, presented at the AES 95th Convention, 1993 October 7-10, Paris
- [DAV94] - Davidson, G.A., Fielder, L.D., Link, B.D. - *Parametric Bit Allocation in a Perceptual Audio Coder*, presented at the AES 97<sup>th</sup> Convention, 10-13 November 1994
- [DIE94] - Dietz, M., Eberlein, E. - *Transcoding Digital Audio Signals from ISO/MPEG Layer III to Layer II*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.577-586
- [DIE96] - Dietz, M., Popp, H., Brandenburg, K., Friedrich R. - *Audio Compression for Network Transmission*, în AES Journal of The Audio Engineering Society, Vol. 44, No.1/2, 1996 January/February
- [DIE00] – Dietz, M., Meltzer, S. – *CT-aacPlus – a state-of-the-art*, EBU Technical Review, July 2002
- [DUM99] - Dumitru, M. - *Asigurarea parametrilor de calitate ai semnalelor în condițiile capacității limitate de transmisie și stocare*, Societatea Inginerilor Audio, București 1999
- [ECO96] – Edler, B., Contin, L. – *MPEG-4 audio test results (MOS test)*, ISO/IEC JTC1/SC29/WG11 N1144, Jan. 1996
- [EDL00] - Edler, B., Purnhagen, H. - *Parametric audio coding*, 5th International Conference on Signal Processing (ICSP 2000), Beijing, August 2000
- [EDL96] – Edler, B., Purnhagen, H., Ferekidis, C. – *ASAC – Analysis/Synthesis audio codec for very low bit rates*, presented at the AES 100<sup>th</sup> Convention, May 1996, preprint nr.4179
- [EDL98] - Edler, B., Purnhagen, H. – *Concepts for hybrid audio coding schemes based on parametric techniques*, presented at the AES 105<sup>th</sup> Convention, May 1998, preprint nr.4808
- [EHR98] - Ehret, A., Dietz, M., Kunz, O. - *High Quality Audio Transmission Using MPEG-2 AAC*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.693-709
- [FEH93] - Feher, K. – *Comunicații digitale avansate*, Editura Tehnică, București 1993
- [GÂR80] - Gârbea, Șt. - *Otologie*, Ed. științifică și enciclopedică, București
- [GRE93] - Grewin, C. - *Methods for Quality Assessment of Low Bit-Rate Audio Codecs*, în AES 12th International Conference: Perception of Reproduced sound, Copenhagen, 1993, June 28 - 30, pp.97-108
- [GRE90] – Greenwood D. – *A cochlear frequency-position function for several species: 29 years later*, J. Acoust. Soc. Amer., vol.87, June 1990, pp.2592-2605
- [GRE94] - Grewin, C., Jonsson, L. - *Subjective Assessment of the PC-Card PCX5*, Swedish Broadcasting Corporation, Research & Development, 1994-09-03
- [GRI91] - Grigore, G. - *Stereofonia*, Ed. Tehnică, București – 1991

- [GUY96] - Guyton, A.C. - *Fiziologie*, Ed. Medicala Amaltea, București – 1996
- [HEI94] - Heimbecher, S. - *Practical knowledge with ISO/MPEG Layer II editing systems*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.343-362
- [HEL97] - Hellsten, B. - *Decreasing Distorsion over ATRAC Generations*, în Elektronikvaerlden (The Electronic World), 6/97, Sweden
- [HER94] – Herre, J., Brandenburg, K., Lederer, D. – *Intensity Stereo Coding*, 96<sup>th</sup> AES Convention, Amsterdam 1994, Preprint 3799
- [HOE03] – Hoeg, W., Lauterbach, T. – *Digital Audio Broadcasting, Principles and Applications of Digital Radio*, Second Edition, John Wiley, 2003
- [HOL94] - Hollier, M.P., Hawksford, M.O., Guard, D.R. - *Objective Perceptual Analysis: Comparing the Audible Performance of Data Reduction Schemes*, presented at the AES 96th Convention 1994 February 26 - March 01, Amsterdam, preprint nr. 3797 (P3.6)
- [HOR94] - Horbach, U., Zimmermann, H. - *Quality Criteria for Digital Audio Signal Processing*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.614-622
- [ISO90] - ISO/IEC JTC 1/SC2/WG11 MPEG/Audio test report, Document MPEG90/N0030, October 1990
- [ISO91] - ISO/IEC JTC 1/SC2/WG11 MPEG/Audio test report, Document MPEG91/N0010, June 1991
- [ISO94] - ISO/IEC JTC1/SC29/WG11, MPEG94/063 - *Coding of Moving Pictures and Associated Audio - Report on the MPEG/Audio Multichannel Formal Subjective Listening Tests* - March 1994
- [ISO98/1] - ISO/IEC JTC1/SC29/WG11 N2460 - *MPEG-7 Context and Objectives*, October 1998
- [ISO98/2] - ISO/IEC JTC1/SC29/WG11 N2461 - *MPEG-7 Requirements*, October 1998
- [ISO98/3] - ISO/IEC JTC1/SC29/WG11 N2462 - *MPEG-7 Applications*, October 1998
- [ITU93] - International Telecommunication Union, Radiocommunication Study Groups, Task Group 10/3 - *Methods For The Subjective Assessment Of Small Impairments In Audio Systems Including Multichannel Sound Systems*, Novemeber 1, 1993
- [ITU94] - International Telecommunication Union, Radiocommunication Study Groups - *New Results Using the POM System*
- [JAY97] - Jayant, N. - *Signal Compression*, în Int. Journal of High Speed Electronics and Systems, vol.8, No.1, 1997, pp.1-12
- [JES82] – Jesteadt W., Bacon S., Lehman J. – *Forward masking as a function of frequency, masker level and signal delay*, J. Acoust. Soc. Amer. vol.71, 1982, pp.950-962
- [KER94] - van de Kerkhof, L. - *MPEG Audio Layer II at Lower Sampling Frequencies: Low Complexity, High Quality Speech and Music Coding at Bit Rates of 64kbit/s and Lower*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.520-527
- [KEY94] - Keyhl, M., Herre, J., Schmidmer, C. - *NMR Measurements on Multiple Generations Audio Coding*, presented at the AES 96th Convention, 1994 February 26 - March 01, Amsterdam, preprint 3803 (P4.6)
- [KIR96] - Kirby, D., Watanabe, K. - *Overview of the Report on the Formal Subjective Listening Tests of MPEG-2 NBC multichannel audio coding*, ISO/IEC JTC1/SC29/WG11, N1420, November 1996
- [KOZ02] - Kozamernik, F. - *Media over the Internet – an overview of delivery technologies*, EBU Technical Review, October 2002

- [KUR99] - Kurth, F. - *An Audio Codec for Multiple Generations Compression without Loss of Perceptual Quality*, presented at AES 17th Int. Conference on High Quality Audio Coding, Florence, Italy, September 2-5, 1999
- [LAG99] - Lagadec, R. - *Digital Audio and the Challenge of the Internet*, în J. Audio Eng. Soc., Vol.47, No.11, 1999 November, pp.1013-1017
- [LIN94] - Link, M., Stoll, G. - *ISO/MPEG Layer II Compared to 16 bit PCM: Which Coding is the Best Choice?*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.587-597
- [MAS00] - Mason, A.J., McParland A.K., Yeadon, N.J. - *The Atlantic Audio Demonstration Equipment*, în J. Audio Eng. Soc., Vol.48, No.1/2, 2000 January/February, pp.106-118
- [MCL94] - McLeish, R. - *Radio Production: Manual for Broadcasters*, 3 Rev. ed, Focal Press, Butterworth-Heinemann Ltd, 1994
- [MEA98] - Meares, D., Watanabe, K., Scheirer, E. - *Report on the MPEG-2 AAC Stereo Verification Tests*, ISO/IEC JTC1/SC29/WG11, N2006, February 1998
- [MOO78] - *Psychophysical tuning curves measured in simultaneous and forward masking*, J. Acoust. Soc. Amer., vol.63, 1978, pp.524-532
- [MOU99] - Moulton, D. - *What Do We Mean By Audibility?*, Moulton Laboratories, Groton, MA, 1999
- [MTS94] - Talbot-Smith, M. - *Audio Engineer's Reference Book*, Focal Press, Butterworth-Heinemann Ltd, 1994
- [NAF95] - Naforniță, I., Câmpeanu, A., Isar, A. - *Semnale Circuite și Sisteme*, Partea I, U.T.Timișoara, Fac. de Electronică și Telecomunicații, 1996
- [NOL97] - Noll, P., Davis, P. - *ISO/MPEG Audio Coding*, în Int. Journal of High Speed Electronics and Systems, vol.8, No.1, 1997, pp.69-118
- [OLI99] - Olive, S., Castro, B., Toole, F. - *A New Laboratory and Methodology for the Subjective Evaluation of Workstation Audio Systems*, presented at the AES 106th Convention, 8-11 May 1999, preprint 4876 (B8)
- [OTE97] - Oteșteanu, M., Alexa, F., Ianăși, C. - *Sisteme de înregistrare audio și video*, Editura de Vest, Timișoara 1997
- [PAI00] - Painter T., Spanias A. - *Perceptual Coding of Digital Audio*, Proceedings of the IEEE, vol.88, No.4, April 2000, pp.451-513
- [PFA98] - Pfaff, K.V. - *Matrici de redare și transmisie stereofonică multidifuzor*, Universitatea "Politehnica" din Timișoara, Departamentul de Comunicații, Referat doctorat, 1998
- [PFA01] - Pfaff, K.V. - *Evaluating Objective Audio Quality for Minidisc and Influence of Multiple Generations Audio Coding for Perceptual Quality*, în Buletinul Științific al Universității "Politehnica" din Timișoara, Seria Electronică și Telecomunicații, Tom 46 (60), Fascicola 1, 2001, p.51-58
- [PFA02] - Pfaff, K.V., Toma, C. - *Specific Artifacts for Perceptual Audio Coders*, Proceedings of "Etc.2002", Timișoara, September 19-20, 2002, vol. II, p.126-131
- [PFA03] - Pfaff, K.V. - *Radioul în Internet*, în Anale Universitatea Tibiscus, seria Informatică, vol. I, fasc. I, 2003, p.184-198
- [PFA04] - Pfaff, K.V. - *Method for Objective Evaluation of Perceived Audio Quality for use in Broadcasting Networks*, trimis spre prezentare la 23. Tonmeistertagung, Leipzig, 5 - 8 Nov. 2004
- [PFT03] - Pfaff, K.V., Toma C. - *Audio Quality of Internet Radio Systems*, prezentat la SCS 2003, Iași 10-11 iulie 2003, vol. II, p.301-304

- [POH97] - Pohlmann, K.C., - *Format Fisticuffs Recording Rivalry MiniDisc vs. DolbyS*, in Stereo Review, March 1997
- [POP00] - Popescu, D. - *Viitorul radiodifuziunii în lumea multimediei*, prezentat la Soc. Română de Radiodifuziune, sala "Catinca Ralea", 11.01.2000
- [RAN02] - Ranada, D. *Facing the Codec Challenge*, Sound&Vision, July 2002, 98-100
- [RAU93] - Rault, J.B., Philippe P., Lever M. - *MUSICAM (ISO/MPEG audio) very low bit-rate coding at reduced sampling frequency*, presented at 95<sup>th</sup> AES Convention, San Francisco, October 1993, Preprint 3741
- [RIB03] - Ribas-Corbera, J. *Windows Media 9 Series*, EBU Technical Revue, January 2003
- [RIT94] - Ritscher, S., Felderhoff, U., Nielsen, S.H. - *Cascading of Perceptually Based Audio Coding Systems*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.567-576
- [ROB98] - Robjohns, H. - *Cutting it Fine*, în Sound on Sound, august 1998
- [RUM97] - Rumsey, F., McCormick, T. - *Sound and Recording: An Introduction*, Focal Press, Third Edition 1997
- [SCA94] - Schaffner, H. - *ISDN Sound Transmission: Advantages and Drawbacks of Bit-Rate Coded Audio Signals*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.540-552
- [SCR94] - Schröder, E.F., Spille, J. - *MPEG-2 Audio: Refined ISO-Standard Developments for Audio Data Reduction*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.484-500
- [SIN90] - Singhal, S. - *High quality audio coding using multipulse LPC*, în Proc. ICASSP - 90, May 1990, pp. 1101-1104
- [SMI99] - Smith, J.O., Abel, J.S. - *Bark and ERB Bilinear Transforms*, IEEE Transactions on Speech and Audio Processing, November, 1999
- [SPE00] - Spector, D. - *Evaluating Adaptive Transform Acoustic Coding for Mini Disc*, presented at the AES 108th Convention, 2000 February 19-22, Paris, Preprint nr. 5082 (B-4)
- [SPO95] - Sporer, Th., Gbur, U., Herre, J., Kapust, R. - *Evaluating a Measurement System*, J. Audio Eng. Soc., Vol.43, No.5, 1995 March, pp.107-116
- [STA99] - Stanomir, D., Tincu, L. - *Acustică Aplicată, Volumul I. Structuri și sisteme mecano - acustice*, Editura Tincu și Stanomir, 1999
- [STO97] - Stoll, G., Felderhoff, U., Spikofski, G. - *Internet Radio and Excellent Audio Quality: Dreamboat or Reality?*, IBC'97, Amsterdam, 12-16 September 1997
- [STO99] - Stoll, G. - *Streaming-Audio@Internet: Perspectives for the Broadcasters*, Proceedings of the AES 17-th International Conference on High Quality Audio Coding, Florence, September 2-5, 1999, pp.19-33
- [STO00] - Stoll, G., Kozamernik, F. - *EBU Listenings Tests On Internet Audio Codecs*, EBU Technical Revue - June 2000
- [THI98] - Thiede, T., Treurniet, W.C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J.G., Colomes, C., Keyhl, M., Stoll, G., Brandenburg, K., Feiten, B. - *PEAQ - The Future ITU Standard for Objective Measurement of Perceived Audio Quality*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.724-766
- [THI00] - Thiede, T., Treurniet, W.C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J.G., Colomes, C., Keyhl, M., Stoll, G., Brandenburg, K., Feiten, B. - *PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality*, în J. Audio Eng. Soc., Vol.48, No.1/2, 2000 January/February, pp.3-29
- [THT98] - Thimet, T. - *The Internet as a Broadcast Medium*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.1137-1160



- [TRE00] - Treurniet, W.C., Souloudre, G. - *Evaluation of the ITU-R Objective Audio Quality Measurement Method*, în J. Audio Eng. Soc., Vol.48, No.3, 2000 March, pp.164-173
- [TRO98] - Troge, T.A. - *The Sound makes the Music - also in the Age of Multimedia*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.22-31
- [TSU92] - Tsutsui, K., Suzuki, H., Shimoyoshi, O. - *ATRAC: Adaptive Transform Acoustic Coding for Minidisc*, presented at the AES 93<sup>rd</sup> Convention, San Francisco, 1992 October 1-4
- [TÜR98] - Türke, F. - *Digitalisation of Radio Broadcasting Production. A Report on Experiences of the Deutsche Welle*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.292-296
- [URM82] - Urmă, D. - *Acustică și muzică*, Ed. științifică și enciclopedică, București - 1982
- [VLA97] - Vlaicu, A. - *Prelucrarea numerică a imaginilor*, Universitatea Tehnică Cluj-Napoca, ianuarie 1997
- [VOE00] - Voelker, E.-J., Fischer, S. - *Acoustical Conditions in Computerized Working Places - Situation, Requirements and Proposals*, presented at the AES 108th Convention, 2000 February 19-22, Paris, Preprint nr. 5131 (K-4)
- [WAT95] - Watkinson, J. - *Compression in Video and Audio*, Focal Press 1995
- [WAT97] - Watkinson, J. - *Audio for Television*, Focal Press, Butterworth-Heinemann Ltd, 1997
- [WIE94] - Wiese, D., Hinderks, L.W. - *New Concepts for Digital Multi Purpose Realtime Systems Using ISO/MPEG Layer II (MUSICAM) with Scaling Audio Quality*, în Bericht 18. Tonmeistertagung, Karlsruhe 1994, pp.510-519
- [WIE02] - Wiese, D., Ludwig, W. - *Zu viele Codierverfahren: Lösung in Sicht für den Rundfunk?*, în Bericht 22. Tonmeistertagung, Hannover 2002
- [WIN98] - Winduratna, B. - *FM analysis/synthesis based audio coding*, presented at 104<sup>th</sup> AES Convention, May 1998, preprint 4746
- [WPG99] - EBU Webcasting Project Group, *Practical Webcasting*, BPN022, June 30<sup>th</sup> 1999
- [WÜS98] - Wüstenhagen, U., Feiten, B., Hoeg, W. - *International Listening Test of Multichannel Audio Systems MPEG-2 Layer II and Dolby AC-3*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.680-692
- [YOS94] - Yoshida, T. - *The Rewritable MiniDisc System*, în Proceedings of the IEEE, USA, vol.82, no.10, oct. 1994, pp.1492-1500
- [ZIE98] - Zieglmeier, W. - *Streaming Audio@Internet Versus Audio Quality: How Far Did We Come?*, în Bericht 20. Tonmeistertagung, Karlsruhe 1998, pp.1101-1109