

ISO 11172-3:1993

ANNEXES C & D

3-ANNEX C (informative)

THE ENCODING PROCESS

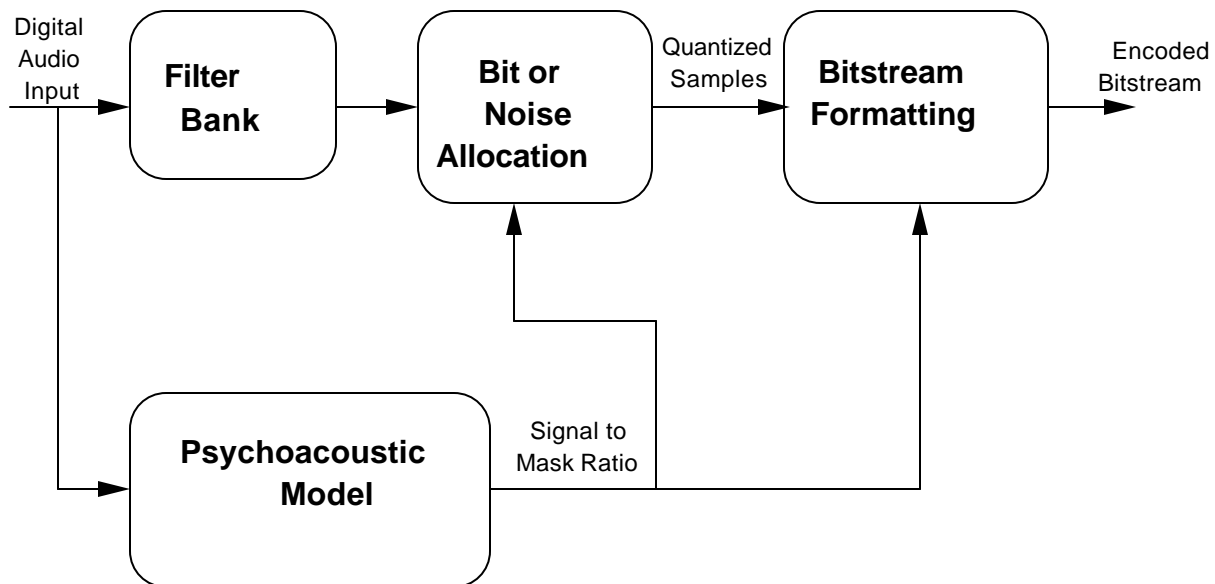
3-C.1 Encoder

3-C.1.1 Overview

For each of the Layers, an example of one suitable encoder with the corresponding flow-diagram is given in this annex. In subsequent clauses the analysis subband filter and the layer-specific encoding techniques are described. In Annex D two examples of psychoacoustic models, which are common to all layers, are described. A short introduction describes the overall philosophy.

INTRODUCTION

The MPEG-Audio algorithm is a psychoacoustic algorithm. The figure below shows the primary parts of a psychoacoustic algorithm.



The four primary parts of the psychoacoustic encoder are:

1) The Filterbank:

The filterbank does a time to frequency mapping. There are two filterbanks used in the MPEG-Audio algorithm, each providing a specific mapping in time and frequency. These filterbanks are critically sampled (i.e. there are as many samples in the analyzed domain as there are in the time domain). These filterbanks provide the primary frequency separation for the encoder, and the reconstruction filters for the decoder. The output samples of the filterbank are quantized.

2) The Psychoacoustic Model:

The psychoacoustic model calculates a just noticeable noise-level for each band in the filterbank. This noise level is used in the bit or noise allocation to determine the actual quantizers and quantizer levels. There are two psychoacoustic models presented in 3-Annex D. While they can both be applied to any layer of the MPEG-Audio algorithm, in practice Model 1 has been used for Layers I and II, and Model 2 for Layer III. In both psychoacoustic models, the final output of the model is a signal-to-mask ratio (SMR) for each band (Layers I and II) or group of bands (Layer III).

3) Bit or Noise Allocation:

The allocator looks at both the output samples from the and the SMR's from the psychoacoustic model, and adjusts the bit allocation (Layers I and II) or noise allocation (Layer III) in order to simultaneously meet both the bitrate requirements and the masking requirements. At low bitrates, these methods attempt to spend bits in a fashion that is psychoacoustically inoffensive when they cannot meet the psychoacoustic demand at the required bitrate.

4) The bitstream formatter:

The bitstream formatter takes the quantized filterbank outputs, the bit allocation (Layers I and II) or noise allocation (Layer III) and other required side information, and encodes and formats that information in an efficient fashion. In the case of Layer III, the Huffman codes are also inserted at this point.

The Filterbank

In Layers I and II, a filterbank with 32 subbands is used. In each subband, 12 or 36 samples are grouped for processing. In Layer III, the filterbank has a signal-dependant resolution, where there are either 6x32 or 18x32 frequency bands. In the case where there are 6x32 frequency samples, the 3 sets of each frequency are quantized separately.

Bit or Noise Allocation Method

There are two different bitrate control methods explained in this Annex. In Layers I and II this method is a bit allocation process, i.e. a number of bits is assigned to each sample (or group of samples) in each subband. The method for Layer III is a noise-allocation loop, where the quantizers are varied in an organized fashion, and the variable to be controlled is the actually injected noise. In either case, the result is a set of quantization parameters and quantized output samples that are given to the bitstream formatter.

Bitstream Formatting

The bitstream formatter varies from layer to layer. In Layers I and II, a fixed PCM code is used for each subband sample, with the exception that in Layer II quantized samples may be grouped. In Layer III, Huffman codes are used to represent the quantized frequency samples. These Huffman codes are variable-length codes that allow for more efficient bitstream representation of the quantized samples at the cost of additional complexity.

3-C.1.2 Input High-Pass Filter

The encoding algorithms provide a frequency response down to DC. However, in applications where this is not a requirement, it is recommended that a high-pass filter be included at the input of the encoder. The cut-off frequency should be in the range of 2 to 10Hz.

The application of such a high-pass filter avoids an unnecessarily high bitrate requirement for the lowest subband and increases the overall audio quality.

3-C.1.3 Analysis Subband Filter (norme ISO page 67)

An analysis subband filterbank is used to split the broadband signal with sampling frequency f_s into 32 equally spaced subbands with sampling frequencies $f_s/32$. The flow chart of this process with the appropriate formulas is given in Figure 3-C.1 "ANALYSIS SUBBAND FILTER FLOW CHART". The analysis subband filtering includes the following steps:

- Input 32 audio samples.
- Build an input sample vector, X , of 512 elements. The 32 audio samples are shifted in at positions 0 to 31, the most recent on at position 0, and the 32 oldest elements are shifted out.
- Window vector X by vector C . The coefficients are to be found in Table 3-C.1 "COEFFICIENTS C_i FOR THE ANALYSIS WINDOW".
- Calculate the 64 values Y_i according to the formula given in the flow chart.
- Calculate the 32 subband samples S_i by matrixing. The coefficients for the matrix can be calculated by the following formula:

$$M_{ik} = \cos [(2i + 1)(k - 16)\pi/64], \quad \text{for } i = 0 \text{ to } 31, \text{ and } k = 0 \text{ to } 63.$$

Table 3-C.1 Coefficients C_i of the Analysis Window

$C[0] = 0.000000000$	$C[1] = -0.000000477$	$C[2] = -0.000000477$	$C[3] = -$
0.000000477			
$C[4] = -0.000000477$	$C[5] = -0.000000477$	$C[6] = -0.000000477$	$C[7] = -$
0.000000954			
$C[8] = -0.000000954$	$C[9] = -0.000000954$	$C[10] = -0.000000954$	$C[11] = -$
0.000001431			
$C[12] = -0.000001431$	$C[13] = -0.000001907$	$C[14] = -0.000001907$	$C[15] = -$
0.000002384			
$C[16] = -0.000002384$	$C[17] = -0.000002861$	$C[18] = -0.000003338$	$C[19] = -$
0.000003338			
$C[20] = -0.000003815$	$C[21] = -0.000004292$	$C[22] = -0.000004768$	$C[23] = -$
0.000005245			
$C[24] = -0.000006199$	$C[25] = -0.000006676$	$C[26] = -0.000007629$	$C[27] = -$
0.000008106			
$C[28] = -0.000009060$	$C[29] = -0.000010014$	$C[30] = -0.000011444$	$C[31] = -$
0.000012398			
$C[32] = -0.000013828$	$C[33] = -0.000014782$	$C[34] = -0.000016689$	$C[35] = -$
0.000018120			
$C[36] = -0.000019550$	$C[37] = -0.000021458$	$C[38] = -0.000023365$	$C[39] = -$
0.000025272			
$C[40] = -0.000027657$	$C[41] = -0.000030041$	$C[42] = -0.000032425$	$C[43] = -$
0.000034809			
$C[44] = -0.000037670$	$C[45] = -0.000040531$	$C[46] = -0.000043392$	$C[47] = -$
0.000046253			
$C[48] = -0.000049591$	$C[49] = -0.000052929$	$C[50] = -0.000055790$	$C[51] = -$
0.000059605			
$C[52] = -0.000062943$	$C[53] = -0.000066280$	$C[54] = -0.000070095$	$C[55] = -$
0.000073433			
$C[56] = -0.000076771$	$C[57] = -0.000080585$	$C[58] = -0.000083923$	$C[59] = -$
0.000087261			
$C[60] = -0.000090599$	$C[61] = -0.000093460$	$C[62] = -0.000096321$	$C[63] = -$
0.000099182			
$C[64] = 0.000101566$	$C[65] = 0.000103951$	$C[66] = 0.000105858$	$C[67] =$
0.000107288			
$C[68] = 0.000108242$	$C[69] = 0.000108719$	$C[70] = 0.000108719$	$C[71] =$
0.000108242			
$C[72] = 0.000106812$	$C[73] = 0.000105381$	$C[74] = 0.000102520$	$C[75] =$
0.000099182			

C[76]= 0.000095367	C[77]= 0.000090122	C[78]= 0.000084400	C[79]=
0.000077724			
C[80]= 0.000069618	C[81]= 0.000060558	C[82]= 0.000050545	C[83]=
0.000039577			
C[84]= 0.000027180	C[85]= 0.000013828	C[86]= -0.000000954	C[87]= -
0.000017166			
C[88]= -0.000034332	C[89]= -0.000052929	C[90]= -0.000072956	C[91]= -
0.000093937			
C[92]= -0.000116348	C[93]= -0.000140190	C[94]= -0.000165462	C[95]= -
0.000191212			
C[96]= -0.000218868	C[97]= -0.000247478	C[98]= -0.000277042	C[99]= -
0.000307560			
C[100]= -0.000339031	C[101]= -0.000371456	C[102]= -0.000404358	C[103]= -
0.000438213			
C[104]= -0.000472546	C[105]= -0.000507355	C[106]= -0.000542164	C[107]= -
0.000576973			
C[108]= -0.000611782	C[109]= -0.000646591	C[110]= -0.000680923	C[111]= -
0.000714302			
C[112]= -0.000747204	C[113]= -0.000779152	C[114]= -0.000809669	C[115]= -
0.000838757			
C[116]= -0.000866413	C[117]= -0.000891685	C[118]= -0.000915051	C[119]= -
0.000935555			
C[120]= -0.000954151	C[121]= -0.000968933	C[122]= -0.000980854	C[123]= -
0.000989437			
C[124]= -0.000994205	C[125]= -0.000995159	C[126]= -0.000991821	C[127]= -
0.000983715			
C[128]= 0.000971317	C[129]= 0.000953674	C[130]= 0.000930786	C[131]=
0.000902653			
C[132]= 0.000868797	C[133]= 0.000829220	C[134]= 0.000783920	C[135]=
0.000731945			
C[136]= 0.000674248	C[137]= 0.000610352	C[138]= 0.000539303	C[139]=
0.000462532			
C[140]= 0.000378609	C[141]= 0.000288486	C[142]= 0.000191689	C[143]=
0.000088215			
C[144]= -0.000021458	C[145]= -0.000137329	C[146]= -0.000259876	C[147]= -
0.000388145			
C[148]= -0.000522137	C[149]= -0.000661850	C[150]= -0.000806808	C[151]= -
0.000956535			
C[152]= -0.001111031	C[153]= -0.001269817	C[154]= -0.001432419	C[155]= -
0.001597881			
C[156]= -0.001766682	C[157]= -0.001937389	C[158]= -0.002110004	C[159]= -
0.002283096			
C[160]= -0.002457142	C[161]= -0.002630711	C[162]= -0.002803326	C[163]= -
0.002974033			
C[164]= -0.003141880	C[165]= -0.003306866	C[166]= -0.003467083	C[167]= -
0.003622532			
C[168]= -0.003771782	C[169]= -0.003914356	C[170]= -0.004048824	C[171]= -
0.004174709			
C[172]= -0.004290581	C[173]= -0.004395962	C[174]= -0.004489899	C[175]= -
0.004570484			
C[176]= -0.004638195	C[177]= -0.004691124	C[178]= -0.004728317	C[179]= -
0.004748821			
C[180]= -0.004752159	C[181]= -0.004737377	C[182]= -0.004703045	C[183]= -
0.004649162			
C[184]= -0.004573822	C[185]= -0.004477024	C[186]= -0.004357815	C[187]= -
0.004215240			
C[188]= -0.004049301	C[189]= -0.003858566	C[190]= -0.003643036	C[191]= -
0.003401756			
C[192]= 0.003134727	C[193]= 0.002841473	C[194]= 0.002521515	C[195]=
0.002174854			
C[196]= 0.001800537	C[197]= 0.001399517	C[198]= 0.000971317	C[199]=
0.000515938			

C[200]= 0.000033379	C[201]= -0.000475883	C[202]= -0.001011848	C[203]= -
0.001573563			
C[204]= -0.002161503	C[205]= -0.002774239	C[206]= -0.003411293	C[207]= -
0.004072189			
C[208]= -0.004756451	C[209]= -0.005462170	C[210]= -0.006189346	C[211]= -
0.006937027			
C[212]= -0.007703304	C[213]= -0.008487225	C[214]= -0.009287834	C[215]= -
0.010103703			
C[216]= -0.010933399	C[217]= -0.011775017	C[218]= -0.012627602	C[219]= -
0.013489246			
C[220]= -0.014358521	C[221]= -0.015233517	C[222]= -0.016112804	C[223]= -
0.016994476			
C[224]= -0.017876148	C[225]= -0.018756866	C[226]= -0.019634247	C[227]= -
0.020506859			
C[228]= -0.021372318	C[229]= -0.022228718	C[230]= -0.023074150	C[231]= -
0.023907185			
C[232]= -0.024725437	C[233]= -0.025527000	C[234]= -0.026310921	C[235]= -
0.027073860			
C[236]= -0.027815342	C[237]= -0.028532982	C[238]= -0.029224873	C[239]= -
0.029890060			
C[240]= -0.030526638	C[241]= -0.031132698	C[242]= -0.031706810	C[243]= -
0.032248020			
C[244]= -0.032754898	C[245]= -0.033225536	C[246]= -0.033659935	C[247]= -
0.034055710			
C[248]= -0.034412861	C[249]= -0.034730434	C[250]= -0.035007000	C[251]= -
0.035242081			
C[252]= -0.035435200	C[253]= -0.035586357	C[254]= -0.035694122	C[255]= -
0.035758972			
C[256]= 0.035780907	C[257]= 0.035758972	C[258]= 0.035694122	C[259]=
0.035586357			
C[260]= 0.035435200	C[261]= 0.035242081	C[262]= 0.035007000	C[263]=
0.034730434			
C[264]= 0.034412861	C[265]= 0.034055710	C[266]= 0.033659935	C[267]=
0.033225536			
C[268]= 0.032754898	C[269]= 0.032248020	C[270]= 0.031706810	C[271]=
0.031132698			
C[272]= 0.030526638	C[273]= 0.029890060	C[274]= 0.029224873	C[275]=
0.028532982			
C[276]= 0.027815342	C[277]= 0.027073860	C[278]= 0.026310921	C[279]=
0.025527000			
C[280]= 0.024725437	C[281]= 0.023907185	C[282]= 0.023074150	C[283]=
0.022228718			
C[284]= 0.021372318	C[285]= 0.020506859	C[286]= 0.019634247	C[287]=
0.018756866			
C[288]= 0.017876148	C[289]= 0.016994476	C[290]= 0.016112804	C[291]=
0.015233517			
C[292]= 0.014358521	C[293]= 0.013489246	C[294]= 0.012627602	C[295]=
0.011775017			
C[296]= 0.010933399	C[297]= 0.010103703	C[298]= 0.009287834	C[299]=
0.008487225			
C[300]= 0.007703304	C[301]= 0.006937027	C[302]= 0.006189346	C[303]=
0.005462170			
C[304]= 0.004756451	C[305]= 0.004072189	C[306]= 0.003411293	C[307]=
0.002774239			
C[308]= 0.002161503	C[309]= 0.001573563	C[310]= 0.001011848	C[311]=
0.000475883			
C[312]= -0.000033379	C[313]= -0.000515938	C[314]= -0.000971317	C[315]= -
0.001399517			
C[316]= -0.001800537	C[317]= -0.002174854	C[318]= -0.002521515	C[319]= -
0.002841473			
C[320]= 0.003134727	C[321]= 0.003401756	C[322]= 0.003643036	C[323]=
0.003858566			

C[324]= 0.004049301 0.004477024	C[325]= 0.004215240	C[326]= 0.004357815	C[327]=
C[328]= 0.004573822 0.004737377	C[329]= 0.004649162	C[330]= 0.004703045	C[331]=
C[332]= 0.004752159 0.004691124	C[333]= 0.004748821	C[334]= 0.004728317	C[335]=
C[336]= 0.004638195 0.004395962	C[337]= 0.004570484	C[338]= 0.004489899	C[339]=
C[340]= 0.004290581 0.003914356	C[341]= 0.004174709	C[342]= 0.004048824	C[343]=
C[344]= 0.003771782 0.003306866	C[345]= 0.003622532	C[346]= 0.003467083	C[347]=
C[348]= 0.003141880 0.002630711	C[349]= 0.002974033	C[350]= 0.002803326	C[351]=
C[352]= 0.002457142 0.001937389	C[353]= 0.002283096	C[354]= 0.002110004	C[355]=
C[356]= 0.001766682 0.001269817	C[357]= 0.001597881	C[358]= 0.001432419	C[359]=
C[360]= 0.001111031 0.000661850	C[361]= 0.000956535	C[362]= 0.000806808	C[363]=
C[364]= 0.000522137 0.000137329	C[365]= 0.000388145	C[366]= 0.000259876	C[367]=
C[368]= 0.000021458 0.000288486	C[369]= -0.000088215	C[370]= -0.000191689	C[371]= -
C[372]= -0.000378609 0.000610352	C[373]= -0.000462532	C[374]= -0.000539303	C[375]= -
C[376]= -0.000674248 0.000829220	C[377]= -0.000731945	C[378]= -0.000783920	C[379]= -
C[380]= -0.000868797 0.000953674	C[381]= -0.000902653	C[382]= -0.000930786	C[383]= -
C[384]= 0.000971317 0.000995159	C[385]= 0.000983715	C[386]= 0.000991821	C[387]=
C[388]= 0.000994205 0.000968933	C[389]= 0.000989437	C[390]= 0.000980854	C[391]=
C[392]= 0.000954151 0.000891685	C[393]= 0.000935555	C[394]= 0.000915051	C[395]=
C[396]= 0.000866413 0.000779152	C[397]= 0.000838757	C[398]= 0.000809669	C[399]=
C[400]= 0.000747204 0.000646591	C[401]= 0.000714302	C[402]= 0.000680923	C[403]=
C[404]= 0.000611782 0.000507355	C[405]= 0.000576973	C[406]= 0.000542164	C[407]=
C[408]= 0.000472546 0.000371456	C[409]= 0.000438213	C[410]= 0.000404358	C[411]=
C[412]= 0.000339031 0.000247478	C[413]= 0.000307560	C[414]= 0.000277042	C[415]=
C[416]= 0.000218868 0.000140190	C[417]= 0.000191212	C[418]= 0.000165462	C[419]=
C[420]= 0.000116348 0.000052929	C[421]= 0.000093937	C[422]= 0.000072956	C[423]=
C[424]= 0.000034332 0.000013828	C[425]= 0.000017166	C[426]= 0.000000954	C[427]= -
C[428]= -0.000027180 0.000060558	C[429]= -0.000039577	C[430]= -0.000050545	C[431]= -
C[432]= -0.000069618 0.000090122	C[433]= -0.000077724	C[434]= -0.000084400	C[435]= -
C[436]= -0.000095367 0.000105381	C[437]= -0.000099182	C[438]= -0.000102520	C[439]= -
C[440]= -0.000106812 0.000108719	C[441]= -0.000108242	C[442]= -0.000108719	C[443]= -
C[444]= -0.000108242 0.000103951	C[445]= -0.000107288	C[446]= -0.000105858	C[447]= -

C[448]= 0.000101566	C[449]= 0.000099182	C[450]= 0.000096321	C[451]=
0.000093460			
C[452]= 0.000090599	C[453]= 0.000087261	C[454]= 0.000083923	C[455]=
0.000080585			
C[456]= 0.000076771	C[457]= 0.000073433	C[458]= 0.000070095	C[459]=
0.000066280			
C[460]= 0.000062943	C[461]= 0.000059605	C[462]= 0.000055790	C[463]=
0.000052929			
C[464]= 0.000049591	C[465]= 0.000046253	C[466]= 0.000043392	C[467]=
0.000040531			
C[468]= 0.000037670	C[469]= 0.000034809	C[470]= 0.000032425	C[471]=
0.000030041			
C[472]= 0.000027657	C[473]= 0.000025272	C[474]= 0.000023365	C[475]=
0.000021458			
C[476]= 0.000019550	C[477]= 0.000018120	C[478]= 0.000016689	C[479]=
0.000014782			
C[480]= 0.000013828	C[481]= 0.000012398	C[482]= 0.000011444	C[483]=
0.000010014			
C[484]= 0.000009060	C[485]= 0.000008106	C[486]= 0.000007629	C[487]=
0.000006676			
C[488]= 0.000006199	C[489]= 0.000005245	C[490]= 0.000004768	C[491]=
0.000004292			
C[492]= 0.000003815	C[493]= 0.000003338	C[494]= 0.000003338	C[495]=
0.000002861			
C[496]= 0.000002384	C[497]= 0.000002384	C[498]= 0.000001907	C[499]=
0.000001907			
C[500]= 0.000001431	C[501]= 0.000001431	C[502]= 0.000000954	C[503]=
0.000000954			
C[504]= 0.000000954	C[505]= 0.000000954	C[506]= 0.000000477	C[507]=
0.000000477			
C[508]= 0.000000477	C[509]= 0.000000477	C[510]= 0.000000477	C[511]=
0.000000477			

3-C.1.4 Psychoacoustic Models

Two examples of psychoacoustic models are presented in Annex D, "PSYCHOACOUSTIC MODELS".

3-C.1.5 Encoding

3-C.1.5.1 Layer I Encoding

1. Introduction

This clause describes a possible Layer I encoding method. The description is made according to Figure 3-C.2, "LAYER I, II ENCODER FLOW CHART".

2. Psychoacoustic Model

The calculation of the psychoacoustic parameters can be done either with Psychoacoustic Model I described in Annex D, clause 3-D.1. or with Psychoacoustic Model II as described in Annex D, clause 3-D.2. The FFT shiftlength equals 384 samples. Either model provides the signal-to-mask ratio for every subband.

3. Analysis Subband Filtering

The subband analysis is described in the clause 3-C.1.3, "ANALYSIS SUBBAND FILTER".

4. Scalefactor Calculation (norme ISO page 70)

The calculation of the scalefactor for each subband is performed every 12 subband samples. The maximum of the absolute value of these 12 samples is determined. The next largest value in 3-Annex B, Table 3-B.1., "LAYER I, II SCALEFACTORS" is used as the scalefactor.

5. Coding of Scalefactors

The index in the 3-Annex B, Table 3-B.1., "LAYER I, II SCALEFACTORS" is represented by 6 bits, MSB first. The scalefactor is transmitted only if a non-zero number of bits has been allocated to the subband.

6. Bit Allocation

Before adjustment to a fixed bitrate, the number of bits that are available for coding the samples and the scalefactors must be determined. This number can be obtained by subtracting from the total number of bits available "cb", the numbers of bits needed for bit allocation "bbal", and the number of bits required for ancillary data "banc":

$$adb = cb - (bbal + banc)$$

The resulting number of bits can be used to code the subband samples and the scalefactors. The principle used in the allocation procedure is minimization of the total noise-to-mask ratio over the frame with the constraint that the number of bits used does not exceed the number of bits available for that frame. The possible number of bits allocated to one sample can be found in the table in clause 2.4.2.5 of the main part of the audio standard (Audio data, LayerI); the range is 0...15 bits, excluding an allocation of 1 bit.

The allocation procedure is an iterative procedure, where in each iteration step the number of levels of the subband samples of greatest benefit is increased.

First the mask-to-noise ratio "MNR" for each subband is calculated by subtracting from the signal-to-noise-ratio "SNR" the signal-to-mask-ratio "SMR":

$$MNR = SNR - SMR$$

The signal-to-noise-ratio can be found in the 3-Annex C, Table 3-C.2., "LAYER I SIGNAL-TO-NOISE-RATIOS". The signal-to-mask-ratio is the output of the psychoacoustic model.

Then zero bits are allocated to the samples and the scalefactors. The number of bits for the samples "bspl" and the number of bits for the scalefactors "bscf" are set to zero. Next an iterative procedure is started. Each iteration loop contains the following steps :

- Determination of the minimal MNR of all subbands.
- The accuracy of the quantization of the subband with the minimal MNR is increased by using the next higher number of bits.
- The new MNR of this subband is calculated.
- bspl is updated according to the additional number of bits required. If a non-zero number of bits is assigned to a subband for the first time, bscf has to be incremented by 6 bits. Then adb is calculated again using the formula: $adb = cb - (bbal + bscf + bspl + banc)$

The iterative procedure is repeated as long as adb is not less than any possible increase of bspl and bscf within one loop.

7. Quantization and Encoding of Subband Samples

A linear quantizer with a symmetric zero representation is used to quantize the subband samples. This representation prevents small value changes around zero from quantizing to different levels. Each of the subband samples is normalized by dividing its value by the scalefactor to obtain X, and quantized using the following formula :

- Calculate $AX+B$
- Take the N most significant bits.
- Invert the MSB.

A and B can be found in 3-Annex C, Table 3-C.3, "LAYER I QUANTIZATION COEFFICIENTS". N represents the necessary number of bits to encode the number of steps. The inversion of the most significant bit (MSB) is done in order to avoid the all '1' representation of the code, because the all '1' code is used for the synchronization word.

8. Coding of Bit Allocation

The 4-bit code for the allocation is given in clause 2.4.2.5, "Audio data LayerI", of the main part of the audio standard.

9. Formatting

The encoded subband information is transferred in frames (See also clauses 2.4.1.2, 2.4.1.3, 2.4.1.5 and 2.4.1.8 of the clause 2.4.1 "Specification of the Coded Audio Bitstream Syntax " of the main part of the audio standard. The number of slots in a frame varies with the sample frequency (Fs) and bitrate. Each frame contains information on 384 samples of the original input signal, so the frame rate is $F_s/384$.

Fs (kHz)	Frame size (ms)
48	8
44.1	8.7074...
32	12

A frame may carry audio information from one or two channels.

The length of a slot in LayerI is 32 bits. The number of slots in a frame can be computed by this formula :

Number of slots/frame (N) = * 12

If this does not give an integer number the result is truncated and 'padding' is required. This means that the number of slots may vary between N and N + 1.

An overview of the Layer I format is given below:

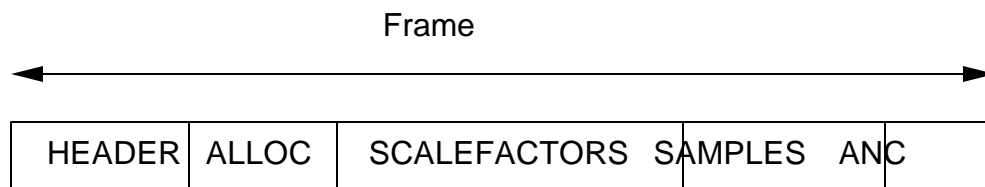


TABLE 3-C.2 LAYER I SIGNAL-TO-NOISE RATIOS

No. of steps	SNR (dB)
--------------	----------

0	0.00
3	7.00
7	16.00
15	25.28
31	31.59
63	37.75
127	43.84
255	49.89
511	55.93
1023	61.96
2047	67.98
4095	74.01
8191	80.03
16383	86.05
32767	92.01

TABLE 3-C.3 LAYER I QUANTIZATION COEFFICIENTS

No. of steps	A	B
3	0.750000000	-0.250000000
7	0.875000000	-0.125000000
15	0.937500000	-0.062500000
31	0.968750000	-0.031250000
63	0.984375000	-0.015625000
127	0.992187500	-0.007812500
255	0.996093750	-0.003906250
511	0.998046875	-0.001953125
1023	0.999023438	-0.000976563
2047	0.999511719	-0.000488281
4095	0.999755859	-0.000244141
8191	0.999877930	-0.000122070
16383	0.999938965	-0.000061035
32767	0.999969482	-0.000030518

FIGURE 3-C.1 Analysis subband filter flow chart

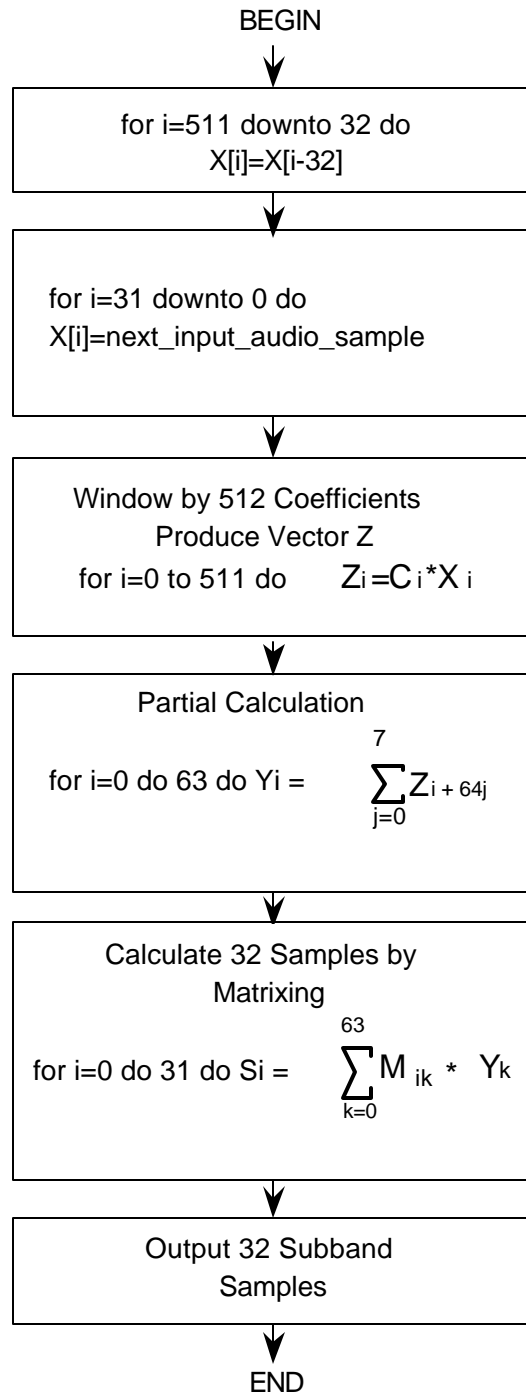
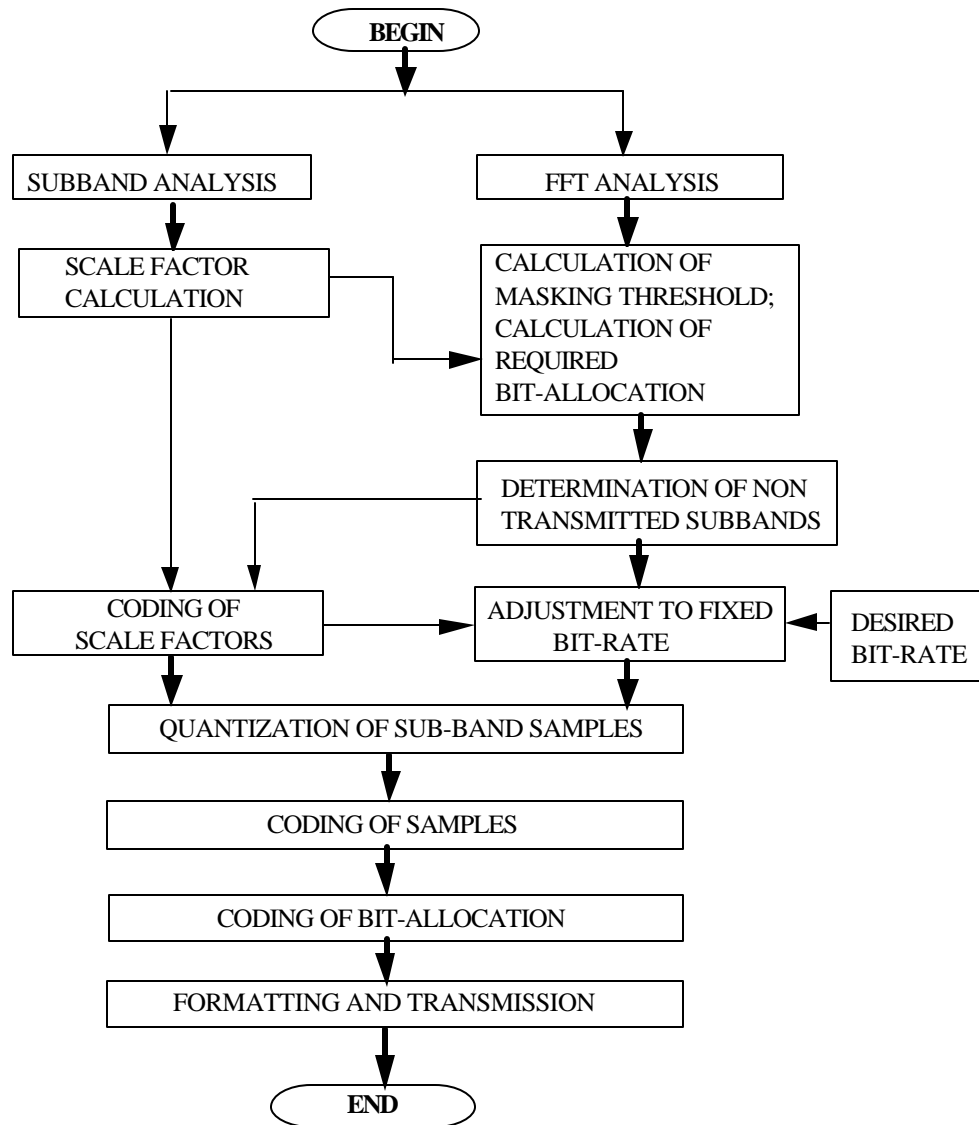


FIGURE 3-C.2 Layer I, II encoder flow chart



3-ANNEX D (informative)

PSYCHOACOUSTIC MODELS

3-D.1. Psychoacoustic Model I

Layer I: A new bit allocation is calculated for each block of 12 subband or 384 input PCM samples.

The bit allocation of the 32 subbands is calculated on the basis of the signal-to-mask ratios of all the subbands. Therefore it is necessary to determine, for each subband the maximum signal level and the minimum masking

threshold. The minimum masking threshold is derived from an FFT of the input PCM signal, followed by a psychoacoustic model calculation.

The FFT in parallel with the subband filter compensates for the lack of spectral selectivity obtained at low frequencies by the subband filterbank. This technique provides both a sufficient time resolution for the coded audio signal (Polyphase filter with optimized window for minimal pre-echoes) and a sufficient spectral resolution for the calculation of the masking thresholds.

The frequencies and levels of aliasing distortions can be calculated. This is necessary for calculating a minimum bit rate for those subbands which need some bits to cancel the aliasing components in the decoder. The additional complexity to calculate the better frequency resolution is necessary only in the encoder, and introduces no additional delay or complexity in the decoder.

The calculation of the signal-to-mask-ratio is based on the following steps:

Step 1

- Calculation of the FFT for time to frequency conversion.

Step 2

- Determination of the sound pressure level in each subband.

Step 3

- Determination of the threshold in quiet (absolute threshold).

Step 4

- Finding of the tonal (more sinusoid-like) and non-tonal (more noise-like) components of the audio signal.

Step 5

- Decimation of the maskers, to obtain only the relevant maskers.

Step 6

- Calculation of the individual masking thresholds.

Step 7

- Determination of the global masking threshold.

Step 8

- Determination of the minimum masking threshold in each subband.

Step 9

- Calculation of the signal-to-mask ratio in each subband.

Step 1: FFT Analysis (norme ISO page 110)

The masking threshold is derived from an estimate of the power density spectrum that is calculated by a 512-point FFT for Layer I. The FFT is calculated directly from the input PCM signal, windowed by a Hann window.

For a coincidence in time between the bit-allocation and the corresponding subband samples, the PCM-samples entering the FFT have to be delayed:

1. The delay of the analysis subband filter is 256 samples, corresponding to **5.8ms at 44.1kHz** sampling rate. This corresponds to a window shift of 256 samples.

2. The Hann window must coincide with the subband samples of the frame. For Layer I this amounts to an additional window shift of 64 samples.

Technical data of the FFT:

	Layer I
- transform length	512 samples
Window size if $f_s = 44.1$ kHz	11.6 ms
- Frequency resolution	$f_s/512$
- Hann window, $h(i)$:	
$h(i) = 0.5 * \{1 - \cos[2 * \pi * (i)/(N-1)]\}$	$0 \leq i \leq N-1$
- power density spectrum $X(k)$:	
$X(k) = 10 * \log 1/Nh(l) * s(l) * e^{-j * k * l * 2 * \pi / N} ^2$ dB	$k = 0 \dots N/2$

A normalization to the reference level of 96 dB SPL (Sound Pressure Level) has to be done in such a way that the maximum value corresponds to 96dB.

Step 2: Determination of the sound pressure level (norme ISO page 110)

The sound pressure level L_{sb} in subband n is computed by:

$$L_{sb}(n) = \text{MAX}[X(k), 20 * \log(\text{scfmax}(n) * 32768) - 10] \text{ dB}$$

$X(k)$ in subband n

where $X(k)$ is the sound pressure level of the spectral line with index k of the FFT with the maximum amplitude in the frequency range corresponding to subband n . The expression $\text{scfmax}(n)$ is in Layer I the scalefactor, and in Layer II the maximum of the three scalefactors of subband n within a frame. The "-10 dB" term corrects for the difference between peak and RMS level. The sound pressure level $L_{sb}(n)$ is computed for every subband n .

Step 3: Considering the threshold in quiet

The threshold in quiet $LT_q(k)$, also called absolute threshold, is available in the tables "Frequencies, Critical Band Rates and Absolute Threshold" (Tables 3-D.1a, 3-D.1b, 3-D.1c for Layer I). These tables depend on the sampling rate of the input PCM signal. Values are available for each sample in the frequency domain where the masking threshold is calculated.

An offset depending on the overall bit rate is used for the absolute threshold. This offset is -12 dB for bit rates ≥ 96 kbit/s and 0 dB for bit rates < 96 kbit/s per channel.

Step 4: Finding of tonal and non-tonal components (norme ISO page 111-113)

The tonality of a masking component has an influence on the masking threshold. For this reason, it is worthwhile to discriminate between tonal and non-tonal components. For calculating the global masking threshold it is necessary to derive the tonal and the non-tonal components from the FFT spectrum.

This step starts with the determination of local maxima, then extracts tonal components (sinusoids) and calculates the intensity of the non-tonal components within a bandwidth of a critical band. The boundaries of the critical bands are given in the tables "CRITICAL BAND BOUNDARIES" (Tables 3-D.2a, 3-D.2b, 3-D.2c for LayerI;).

The bandwidth of the critical bands varies with the center frequency with a bandwidth of about only 0.1 kHz at low frequencies and with a bandwidth of about 4 kHz at high frequencies. It is known from psychoacoustic experiments that the ear has a better frequency resolution in the lower than in the higher frequency region. To determine if a local maximum may be a tonal component a frequency range df around the local maximum is examined. The frequency range df is given by:

Sampling rate: 44.1kHz

Layer I:	$df = 172.266 \text{ Hz}$	0 kHz	$< f \leq$	5.512kHz
	$df = 281.25 \text{ Hz}$	5.512 kHz	$< f \leq$	11.024 kHz
	$df = 562.50 \text{ Hz}$	11.024 kHz	$< f \leq$	19.982kHz

To make lists of the spectral lines $X(k)$ that are tonal or non-tonal, the following three operations are performed:

(i) Labelling of local maxima

A spectral line $X(k)$ is labelled as a local maximum if

$$X(k) > X(k-1) \text{ and } X(k) > X(k+1)$$

(ii) Listing of tonal components and calculation of the sound pressure level

A local maximum is put in the list of tonal components if

$$X(k) - X(k+j) \geq 7 \text{ dB,}$$

where j is chosen according to

Layer I:

$j = -2, +2$	for $2 < k < 63$
$j = -3, -2, +2, +3$	for $63 \leq k < 127$
$j = -6, \dots, -2, +2, \dots, +6$	for $127 \leq k \leq 250$

If $X(k)$ is found to be a tonal component, then the following parameters are listed:

- Index number k of the spectral line.
- Sound pressure level $X_{tm}(k) = X(k-1) + X(k) + X(k+1)$, in dB
- Tonal flag.

Next, all spectral lines within the examined frequency range are set to -8 dB.

(iii) Listing of non-tonal components and calculation of the power

The non-tonal (noise) components are calculated from the remaining spectral lines. To calculate the non-tonal components from these spectral lines $X(k)$, the critical bands $z(k)$ are determined using the tables, "Critical Band Boundaries" (Tables 3-D.2a, 3-D.2b, 3-D.2c for LayerI). In LayerI, 24 critical bands are used for the sampling rate of 44.1kHz. Within each critical band, the power of the spectral lines are summed to form the sound pressure level of the new non-tonal component corresponding to that critical band.

The following parameters are listed:

- Index number k of the spectral line nearest to the geometric mean of the critical band.
- Sound pressure level $X_{nm}(k)$ in dB.
- Non-tonal flag.

Step 5: Decimation of tonal and non-tonal masking components

Decimation is a procedure that is used to reduce the number of maskers which are considered for the calculation of the global masking threshold.

- (i) Tonal $X_{tm}(k)$ or non-tonal components $X_{nm}(k)$ are considered for the calculation of the masking threshold only if:

$$X_{tm}(k) \geq LT_q(k) \quad \text{or} \quad X_{nm}(k) \geq LT_q(k)$$

In this expression, $LT_q(k)$ is the absolute threshold (or threshold in quiet) at the frequency of index k . These values are given in the Tables 3-D.1a, 3-D.1b, 3-D.1c for LayerI;

- (ii) Decimation of two or more tonal components within a distance of less than 0.5 Bark: Keep the component with the highest power, and remove the smaller component(s) from the list of tonal components. For this operation, a sliding window in the critical band domain is used with a width of 0.5 Bark.

In the following, the index j is used to indicate the relevant tonal or non-tonal masking components from the combined decimated list.

Step 6: Calculation of individual masking thresholds

Of the original $N/2$ frequency domain samples, indexed by k , only a subset of the samples, indexed by i , are considered for the global masking threshold calculation. The samples used are shown in Tables 3-D.1a, 3-D.1b, 3-D.1c for LayerI;

Layer I:

For the frequency lines corresponding to the frequency region which is covered by the first six subbands no subsampling is used. For the frequency region corresponding to the next six subbands every second spectral line is considered. Finally, in the case of 44.1 and 48 kHz sampling rates, in the frequency region corresponding to the remaining subbands, every fourth spectral line is considered up to 20 kHz. In the case of 32 kHz sampling rate, in the frequency region corresponding to the remaining subbands, every fourth spectral line is considered up to 15 kHz (See also Tables 3-D.1a, 3-D.1b, 3-D.1c "Frequencies, Critical Band Rates and Absolute Threshold" for LayerI.)

The number of samples, i , in the subsampled frequency domain is different depending on the sampling rates and layers.

44.1 kHz sampling rate: $i = 106$ for Layer I

To every tonal and non-tonal component the index i in the subsampled frequency domain is assigned, which is closest in frequency to the original spectral line $X(k)$. This index i is given in Tables 3-D.1a, 3-D.1b, 3-D.1c for LayerI; "Frequencies, Critical Band Rates and Absolute Threshold".

The individual masking thresholds of both tonal and non-tonal components are given by the following expression:

$$\begin{aligned} LT_{tm}[z(j), z(i)] &= X_{tm}[z(j)] + av_{tm}[z(j)] + vf[z(j), z(i)] \quad \text{dB} \\ LT_{nm}[z(j), z(i)] &= X_{nm}[z(j)] + av_{nm}[z(j)] + vf[z(j), z(i)] \quad \text{dB} \end{aligned}$$

In this formula LT_{tm} and LT_{nm} are the individual masking thresholds at critical band rate z in Bark of the masking component at the critical band rate z_m in Bark. The values in dB can be either positive or negative. The term $X_{tm}[z(j)]$ is the sound pressure level of the masking component with the index number j at the corresponding critical band rate $z(j)$. The term a_v is called the masking index and v_f the masking function of the masking component $X_{tm}[z(j)]$. The masking index a_v is different for tonal and non-tonal masker ($a_{v_{tm}}$ and $a_{v_{nm}}$).

For tonal maskers it is given by

$$a_{v_{tm}} = -1.525 - 0.275 * z(j) - 4.5 \text{ dB},$$

and for non-tonal maskers

$$a_{v_{nm}} = -1.525 - 0.175 * z(j) - 0.5 \text{ dB}.$$

The masking function v_f of a masker is characterized by different lower and upper slopes, which depend on the distance in Bark $dz = z(i) - z(j)$ to the masker. In this expression i is the index of the spectral line at which the masking function is calculated and j that of the masker. The critical band rates $z(j)$ and $z(i)$ can be found in Tables 3-D.1a, 3-D.1b, 3-D.1c for Layer I; "Frequencies, Critical Band Rates and Absolute Threshold". The masking function, which is the same for tonal and non-tonal maskers, is given by:

$$v_f = 17 * (dz + 1) - (0.4 * X[z(j)] + 6) \text{ dB} \quad \text{for } -3 \leq dz < -1 \text{ Bark}$$

$$v_f = (0.4 * X[z(j)] + 6) * dz \text{ dB} \quad \text{for } -1 \leq dz < 0 \text{ Bark}$$

$$v_f = -17 * dz \text{ dB} \quad \text{for } 0 \leq dz < 1 \text{ Bark}$$

$$v_f = -(dz - 1) * (17 - 0.15 * X[z(j)]) - 17 \text{ dB} \quad \text{for } 1 \leq dz < 8 \text{ Bark}$$

In these expressions $X[z(j)]$ is the sound pressure level of the j 'th masking component in dB.

If $dz < -3$ Bark, or $dz \geq 8$ Bark, the masking is no longer considered (LT_{tm} and LT_{nm} are set to -8dB outside this range).

Step 7: Calculation of the global masking threshold LT_g

The global masking threshold $LT_g(i)$ at the i 'th frequency sample is derived from the upper and lower slopes of the individual masking threshold of each of the j tonal and non-tonal maskers, and in addition from the threshold in quiet $LT_q(i)$. This is also given in Tables 3-D.1a, 3-D.1b, 3-D.1c for Layer I; "Frequencies, Critical Band Rates and Absolute Threshold". The global masking threshold is found by summing the powers corresponding to the individual masking thresholds and the threshold in quiet.

$$LT_g(i) = 10 \log (10^{LT_q(i)/10} + \dots)$$

The total number of tonal maskers is given by m , and the total number of non-tonal maskers is given by n . For a given i , the range of j can be reduced to just encompass those masking components that are within -8 to +3 Bark from i . Outside of this range LT_{tm} and LT_{nm} are -8 dB.

Step 8: Determination of the minimum masking threshold

The minimum masking level $LT_{min}(n)$ in subband n is determined by the following expression:

$$LT_{min}(n) = \text{MIN}[LT_g(i)] \text{ dB} \\ \text{f(i) in subband n}$$

where $f(i)$ is the frequency of the i 'th frequency sample. The $f(i)$ are tabulated in the Tables 3-D.1a, 3-D.1b, 3-D.1c for LayerI; Tables 3-D.1d, 3-D.1e, 3-D.1f for LayerII of "Frequencies, Critical Band Rates and Absolute Threshold". A minimum masking level $LT_{min}(n)$ is computed for every subband.

Step 9: Calculation of the signal-to-mask-ratio

The signal-to-mask ratio

$$SMR_{sb}(n) = L_{sb}(n) - LT_{min}(n) \text{ dB}$$

is computed for every subband n .

Table 3-D.1b.: Frequencies, Critical Band Rates and Absolute Threshold

Table is valid for Layer I at a sampling rate of 44.1 kHz.

Index Number	Frequency	Crit.Band Rate	Absolute Thresh.
i	[Hz]	[z]	[dB]
1	86.13	.850	25.87
2	172.27	1.694	14.85
3	258.40	2.525	10.72
4	344.53	3.337	8.50
5	430.66	4.124	7.10
6	516.80	4.882	6.11
7	602.93	5.608	5.37
8	689.06	6.301	4.79
9	775.20	6.959	4.32
10	861.33	7.581	3.92
11	947.46	8.169	3.57
12	1033.59	8.723	3.25
13	1119.73	9.244	2.95
14	1205.86	9.734	2.67
15	1291.99	10.195	2.39
16	1378.13	10.629	2.11
17	1464.26	11.037	1.83
18	1550.39	11.421	1.53
19	1636.52	11.783	1.23
20	1722.66	12.125	.90
21	1808.79	12.448	.56
22	1894.92	12.753	.21
23	1981.05	13.042	-.17
24	2067.19	13.317	-.56
25	2153.32	13.578	-.96
26	2239.45	13.826	-1.38
27	2325.59	14.062	-1.79
28	2411.72	14.288	-2.21
29	2497.85	14.504	-2.63
30	2583.98	14.711	-3.03
31	2670.12	14.909	-3.41
32	2756.25	15.100	-3.77
33	2842.38	15.284	-4.09
34	2928.52	15.460	-4.37

35	3014.65	15.631	-4.60
36	3100.78	15.796	-4.78
37	3186.91	15.955	-4.91
38	3273.05	16.110	-4.97
39	3359.18	16.260	-4.98
40	3445.31	16.406	-4.92
41	3531.45	16.547	-4.81
42	3617.58	16.685	-4.65
43	3703.71	16.820	-4.43
44	3789.84	16.951	-4.17
45	3875.98	17.079	-3.87
46	3962.11	17.205	-3.54
47	4048.24	17.327	-3.19
48	4134.38	17.447	-2.82
49	4306.64	17.680	-2.06
50	4478.91	17.905	-1.32
51	4651.17	18.121	-.64
52	4823.44	18.331	-.04
53	4995.70	18.534	.47
54	5167.97	18.731	.89
55	5340.23	18.922	1.23
56	5512.50	19.108	1.51
57	5684.77	19.289	1.74
58	5857.03	19.464	1.93
59	6029.30	19.635	2.11
60	6201.56	19.801	2.28
61	6373.83	19.963	2.46
62	6546.09	20.120	2.63
63	6718.36	20.273	2.82
64	6890.63	20.421	3.03
65	7062.89	20.565	3.25
66	7235.16	20.705	3.49
67	7407.42	20.840	3.74
68	7579.69	20.972	4.02
69	7751.95	21.099	4.32
70	7924.22	21.222	4.64
71	8096.48	21.342	4.98
72	8268.75	21.457	5.35
73	8613.28	21.677	6.15
74	8957.81	21.882	7.07
75	9302.34	22.074	8.10
76	9646.88	22.253	9.25
77	9991.41	22.420	10.54
78	10335.94	22.576	11.97
79	10680.47	22.721	13.56
80	11025.00	22.857	15.31
81	11369.53	22.984	17.23
82	11714.06	23.102	19.34
83	12058.59	23.213	21.64
84	12403.13	23.317	24.15
85	12747.66	23.415	26.88
86	13092.19	23.506	29.84
87	13436.72	23.592	33.05
88	13781.25	23.673	36.52

89	14125.78	23.749	40.25
90	14470.31	23.821	44.27
91	14814.84	23.888	48.59
92	15159.38	23.952	53.22
93	15503.91	24.013	58.18
94	15848.44	24.070	63.49
95	16192.97	24.125	68.00
96	16537.50	24.176	68.00
97	16882.03	24.225	68.00
98	17226.56	24.271	68.00
99	17571.09	24.316	68.00
100	17915.63	24.358	68.00
101	18260.16	24.398	68.00
102	18604.69	24.436	68.00
103	18949.22	24.473	68.00
104	19293.75	24.508	68.00
105	19638.28	24.542	68.00
106	19982.81	24.574	68.00

Table 3-D.2b. Critical Band Boundaries

This table is valid for Layer I at a sampling rate of 44.1 kHz.
The frequencies represent the top end of each critical band.

no	index of	frequency [Hz]	Bark [z]
	table F&CB		
0	1	86.133	.850
1	2	172.266	1.694
2	3	258.398	2.525
3	5	430.664	4.124
4	6	516.797	4.882
5	8	689.063	6.301
6	9	775.195	6.959
7	11	947.461	8.169
8	13	1119.727	9.244
9	15	1291.992	10.195
10	17	1464.258	11.037
11	20	1722.656	12.125
12	23	1981.055	13.042
13	27	2325.586	14.062
14	32	2756.250	15.100
15	37	3186.914	15.955
16	45	3875.977	17.079
17	50	4478.906	17.904
18	55	5340.234	18.922
19	61	6373.828	19.963
20	68	7579.688	20.971
21	75	9302.344	22.074
22	81	11369.531	22.984
23	93	15503.906	24.013
24	106	19982.813	24.573

Table 3-D.4b. Absolute Threshold Table

This table is valid at a sampling rate of 44.1 kHz.

A value of 0 dB represents a level in the absolute threshold calculation of 96 dB below the energy of a sine wave of amplitude ± 32760 .

index (line)		absthr
lower	higher	dB

1	1	45.05
2	2	25.87
3	3	18.70
4	4	14.85
5	5	12.41
6	6	10.72
7	7	9.47
8	8	8.50
9	9	7.73
10	10	7.10
11	11	6.56
12	12	6.11
13	13	5.72
14	14	5.37
15	15	5.07
16	16	4.79
17	17	4.55
18	18	4.32
19	19	4.11
20	20	3.92
21	21	3.74
22	22	3.57
23	23	3.40
24	24	3.25
25	25	3.10
26	26	2.95
27	27	2.81
28	28	2.67
29	29	2.53
30	30	2.39
31	31	2.25
32	32	2.11
33	33	1.97
34	34	1.83
35	35	1.68
36	36	1.53
37	37	1.38
38	38	1.23
39	39	1.07
40	40	.90

41	41	.74
42	42	.56
43	43	.39
44	44	.21
45	45	.02
46	46	-.17
47	47	-.36
48	48	-.56
49	50	-.96
51	52	-1.37
53	54	-1.79
55	56	-2.21
57	58	-2.63
59	60	-3.03
61	62	-3.41
63	64	-3.77
65	66	-4.09
67	68	-4.37
69	70	-4.60
71	72	-4.78
73	74	-4.91
75	76	-4.97
77	78	-4.98
79	80	-4.92
81	82	-4.81
83	84	-4.65
85	86	-4.43
87	88	-4.17
89	90	-3.87
91	92	-3.54
93	94	-3.19
95	96	-2.82
97	100	-2.06
101	104	-1.33
105	108	-.64
109	112	-.04
113	116	.47
117	120	.89
121	124	1.23
125	128	1.51
129	132	1.74
133	136	1.93
137	140	2.11
141	144	2.28
145	148	2.45
149	152	2.63
153	156	2.82
157	160	3.03
161	164	3.25
165	168	3.49
169	172	3.74
173	176	4.02
177	180	4.32
181	184	4.64

185	188	4.98
189	192	5.35
193	200	6.15
201	208	7.07
209	216	8.10
217	224	9.25
225	232	10.54
233	240	11.97
241	248	13.56
249	256	15.30
257	264	17.23
265	272	19.33
273	280	21.64
281	288	24.15
289	296	26.88
297	304	29.84
305	312	33.04
313	320	36.51
321	328	40.24
329	336	44.26
337	344	48.58
345	352	53.21
353	360	58.17
361	368	63.48
369	376	69.13
377	384	69.13
385	392	69.13
393	400	69.13
401	408	69.13
409	416	69.13
417	424	69.13
425	432	69.13
433	440	69.13
441	448	69.13
449	456	69.13
457	464	69.13