LiveVideoStackCon

语音编解码技术演进和应用选型

掌门集团
ZENMEN

关于我

刘华平： 研究方向数字信号处理、全景音频重现理论(VR音频)。目前任掌门集团(WIFI万能钥匙)音视频技术研发总监，资源研究员。曾就职于行者悟空声学技术有限公司首席技术官(联合创始人)，阿里巴巴前高级技术专家(P8)， 阿里音乐音视频部门总监、Visualon 音频部门经理、盛大创新院研究员、Freescale 上海研发中心多媒体部门。早期Google Android SDK多媒体架构的贡献者，开源AMR_WB 编码器工程开发者；5项技术发明专利、二十余篇专业论文和多项软件著作权，参与过浙江省杭州重大专项项目，浙江省金华科委项目，上海市科委项目(球谐域全景音频关键技术研究)。

早期专业Blog： http://www.cnblogs.com/huaping-audio/ （音频刘品）

Tel： 13601842613 Email： liupin.2008@gmail.com

提纲

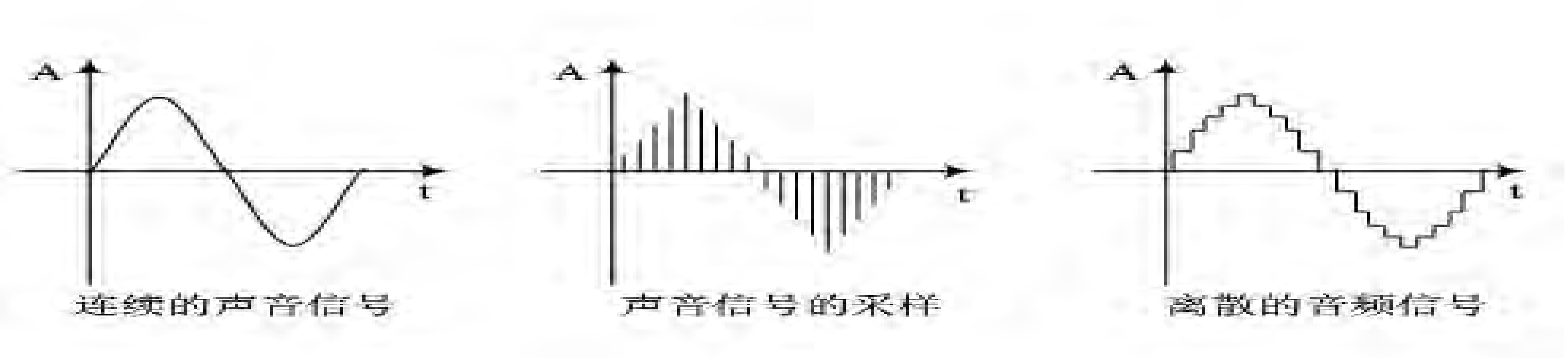
- 语音 / 音频编码总表
- 数字语音基本要素
- 为什么要压缩
- 编码器考虑的因素
- 语音经典编码模型
- ISO 编码模型
- USAC 编码
- 使用选型考虑的因素

语音 / 音频编码总表

Audio compression	**ISO/IEC**	MPEG-1 Layer III (MP3) · MPEG-1 Layer II (Multichannel) · MPEG-1 Layer I · AAC (HE-AAC · AAC-LD) · MPEG Surround · MPEG-4 ALS · MPEG-4 SLS · MPEG-4 DST · MPEG-4 HVXC · MPEG-4 CELP · MPEG-D USAC · MPEG-H 3D Audio
	ITU-T	G.711 (A-law, μ-law) · G.718 · G.719 · G.722 · G.722.1 · G.722.2 · G.723 · G.723.1 · G.726 · G.728 · G.729 · G.729.1
	IETF	Opus · iLBC
	3GPP	AMR · AMR-WB · AMR-WB+ · EVRC · EVRC-B · EVS · GSM-HR · GSM-FR · GSM-EFR
	Others	ACELP · AC-3 · AC-4 · ALAC · Asao · ATRAC · CELT · Codec2 · DRA · DTS · FLAC · iSAC · Monkey's Audio · TTA (True Audio) · MT9 · Musepack · OptimFROG · OSQ · QCELP · RCELP · RealAudio · RTAudio · SD2 · SHN · SILK · **Siren** · SMV · Speex · SVOPC · TwinVQ · VMR-WB · Vorbis · VSELP · WavPack · WMA · MQA · aptX

数字语音基本要素

数字声音三要素： 采样率，通道数和量化位数



- 采样:在时间轴上对信号数字化
- 量化:在幅度轴上对信号数字化
- 编码:按一定格式记录采样和量化后的数字数据

为什么要压缩

长度为4分钟，采样频率为44100Hz, 采样深度为16bits, 双声音 Wav文件大小： 44100Hz * 16 bits * 4 minutes * 2 = (44100 / 1 second) * 16 bits * (4 minutes * (60 seconds / 1 minutes) * 2 = 705600 bits / second * 240 seconds = 169344000 bits = 169344000 / (8 bits / 1 byte) * 2 = 42336000 bytes = 42336000 / (1048576 / 1M) bytes ≒ 40.37MB

MP3 128kbps压缩后文件大小：
128kbps * 4 minutes = (128k bits / 1 second) * (4 minutes * (60 seconds / 1 minutes)) = (128k bits / 1 second) * 240 seconds = 30720k bits = 30720k bits / (8 bits / 1 byte) = 3840k bytes = 3840k / (1024k / 1M) bytes = 3.75M bytes = 3.75MB

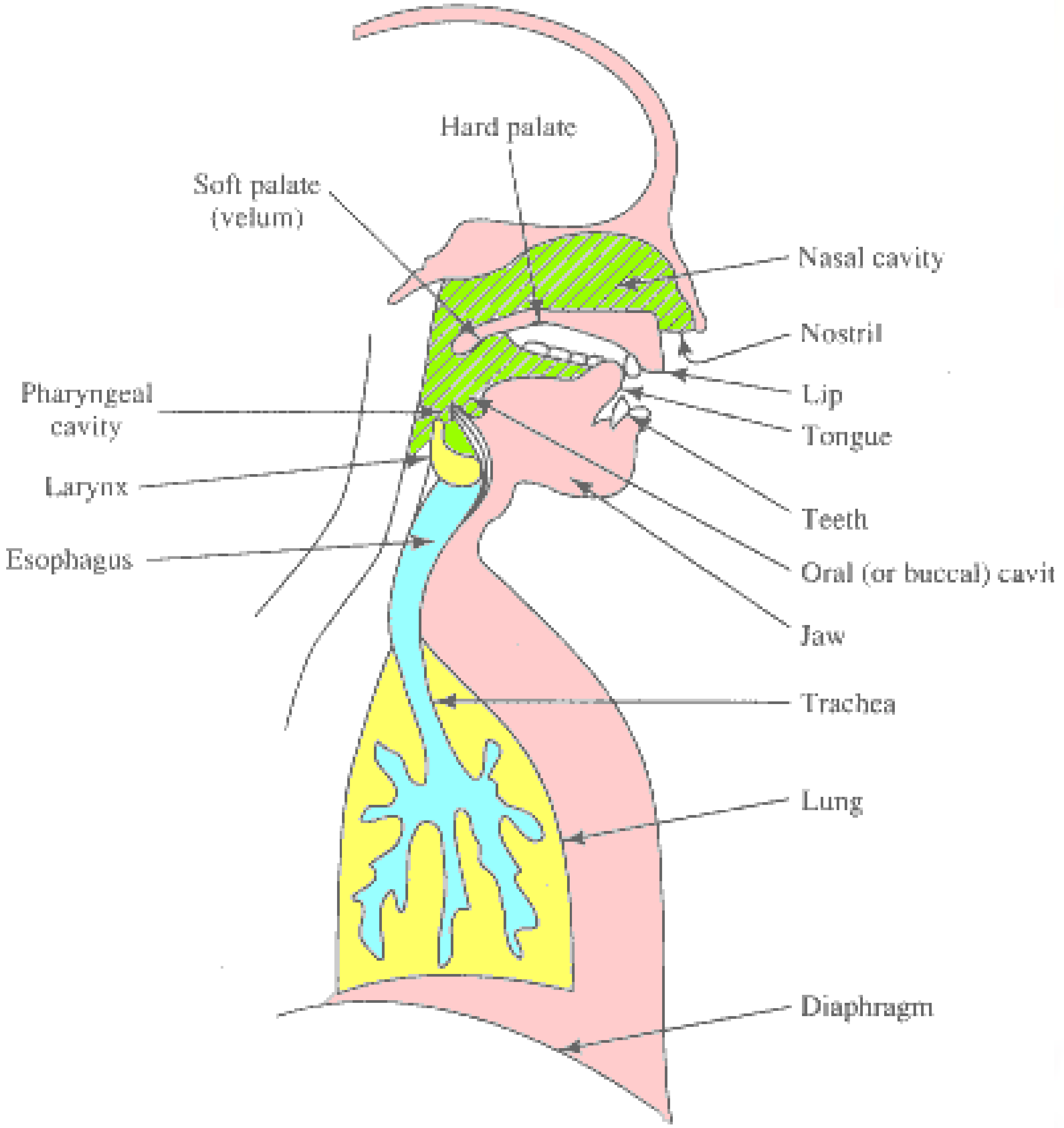
1、存储和带宽二大因素决定了语音压缩的必要性；

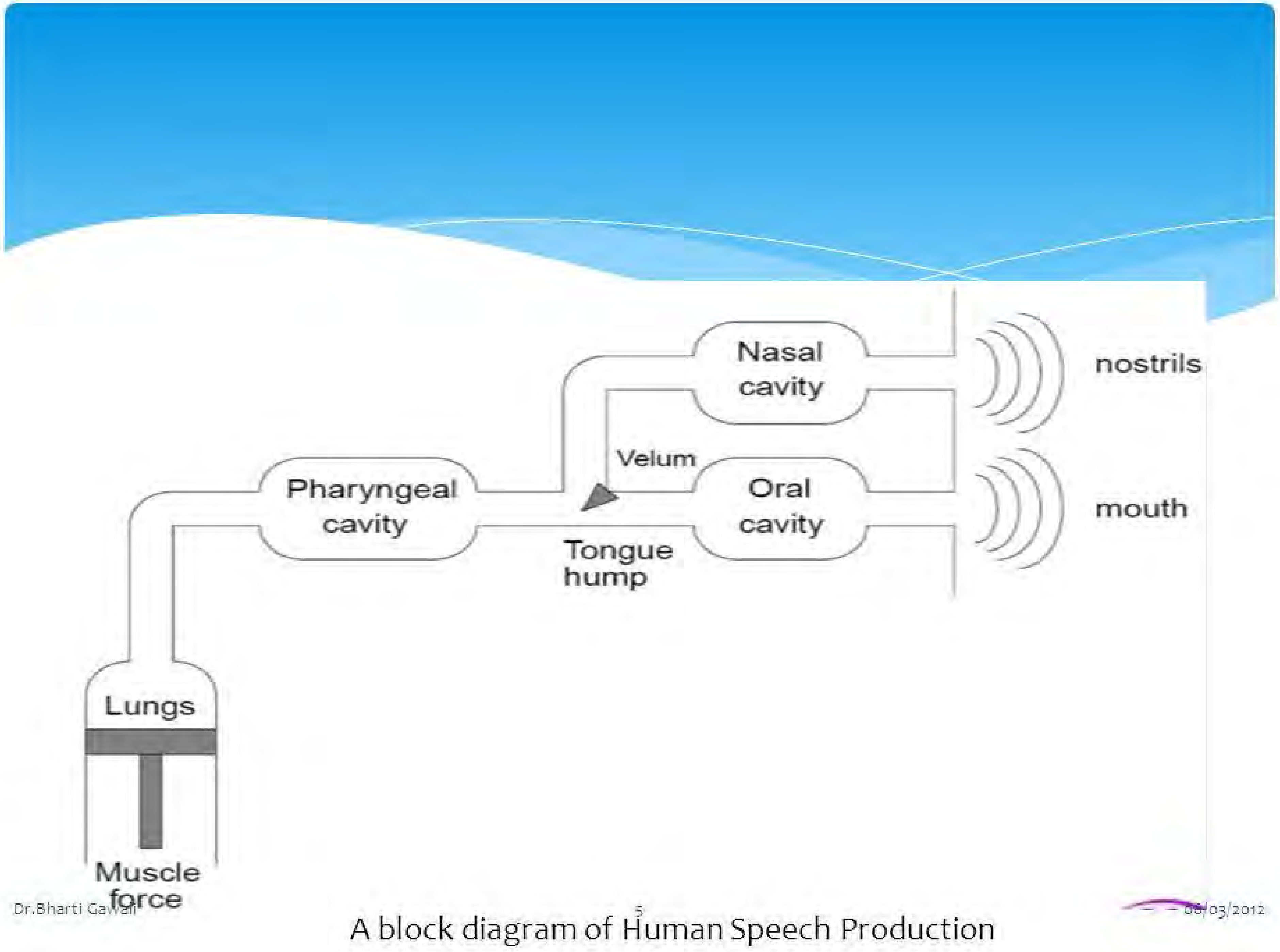
2、压缩理论上的支持是语音压缩的充分条件；

编码器考虑的因素

- 最佳压缩比
- 算法的复杂度
- 算法延时
- 针对特殊场景下的特定设计
- 兼容性

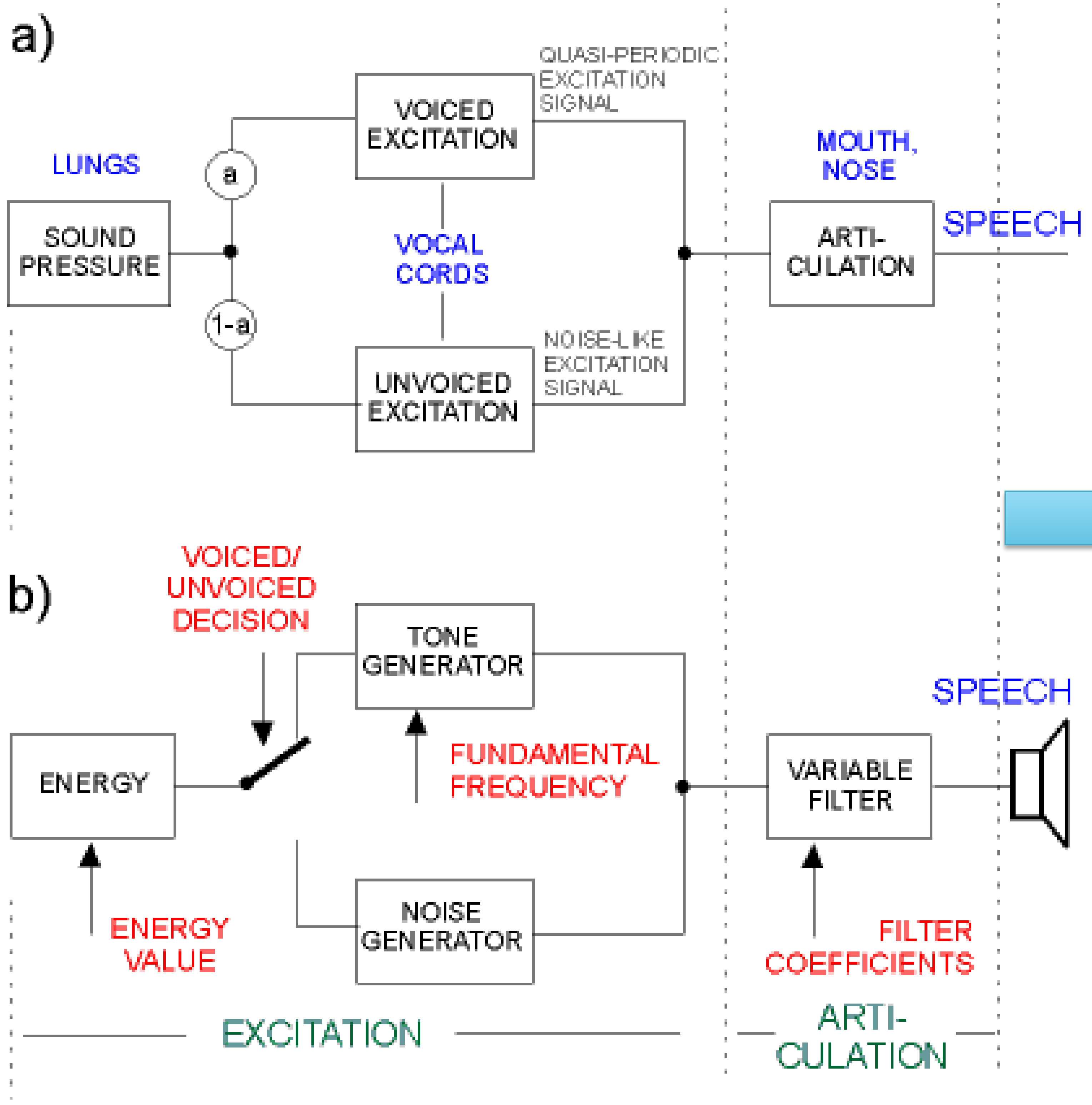
语音经典编码模型-----发音模型*

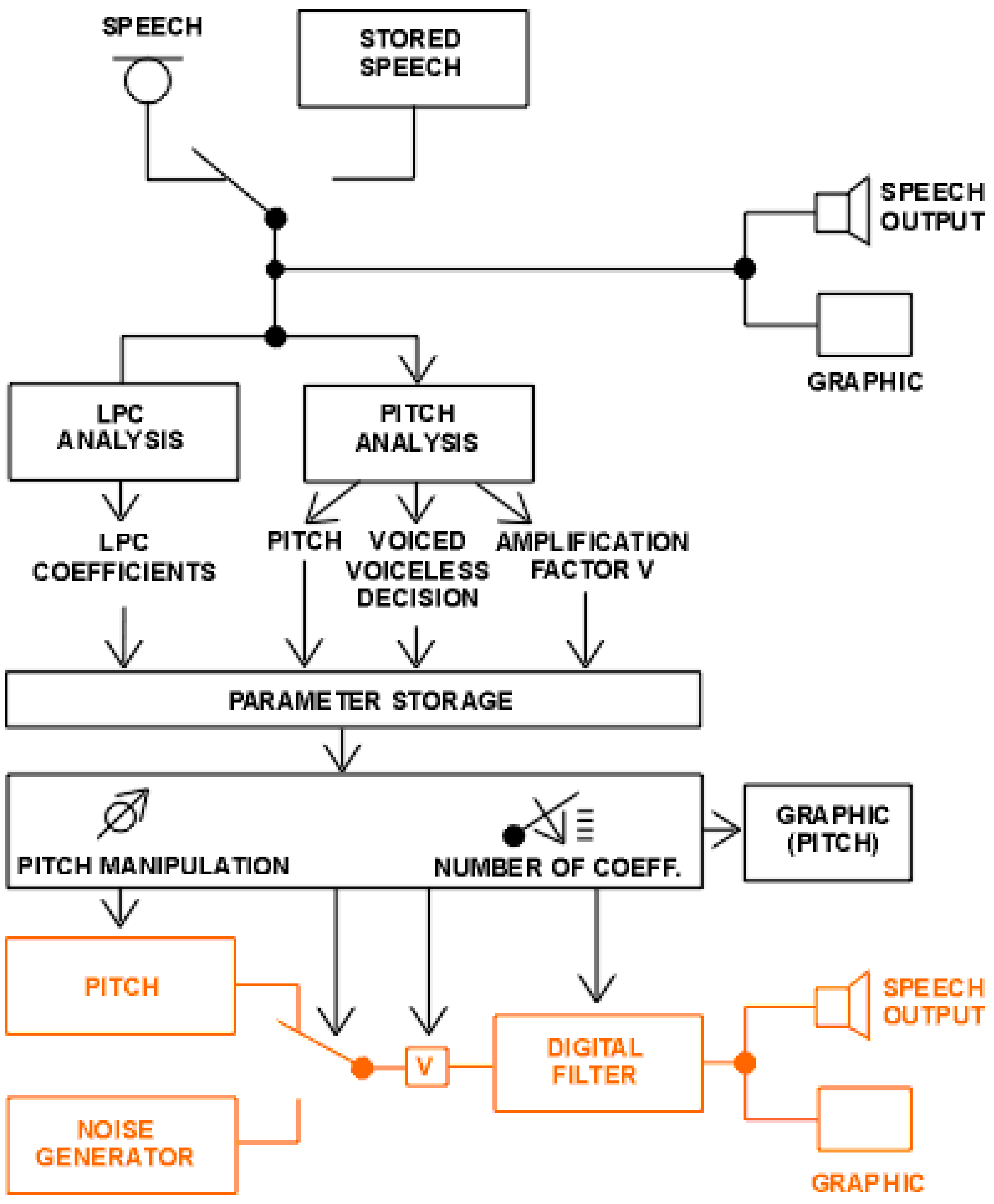




* http://mirlab.org/jang/books/audioSignalProcessing/humanVoiceProduction.asp?title=3-3%20Human%20Voice%20Production%20(%A4H%C1n%

经典语音编码模型：LPC*





* https://www2.spsc.tugraz.at/add_material/courses/scl/vocoder/tech.html

LPC 数学表达：

思想x[n]之前的N个样本的线性组合来预测x[n],

$$y[n]=\sum_{i=1}^{N}a_i x[n-i]$$

其中y[n]就是x[n]的线性预测。预测误差为:

$$e[n]=x[n]-y[n]=x[n]-\sum_{i=1}^{N}a_i x[n-i]$$

线性预测的目的就是寻找使得误差平方和函数最小的最优预测系数：

$$E=\sum_{n=0}^{L-1}[e[n]]^2=\sum_{n=0}^{L-1}\left[x[n]-\sum_{i=1}^{N}a_i x[n-i]\right]^2$$

对于一个N阶滤波器，滤波器系数ai通过求解一个N*N线性系统**Ra=r**求得，其中：
如右上角公式：

$$\mathbf{R}=\begin{bmatrix} R(0) & R(1) & \cdots & R(N-1) \\ R(1) & R(0) & \cdots & R(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(N-1) & R(N-2) & \cdots & R(0) \end{bmatrix}$$

$$\mathbf{r}=\begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(N) \end{bmatrix}$$

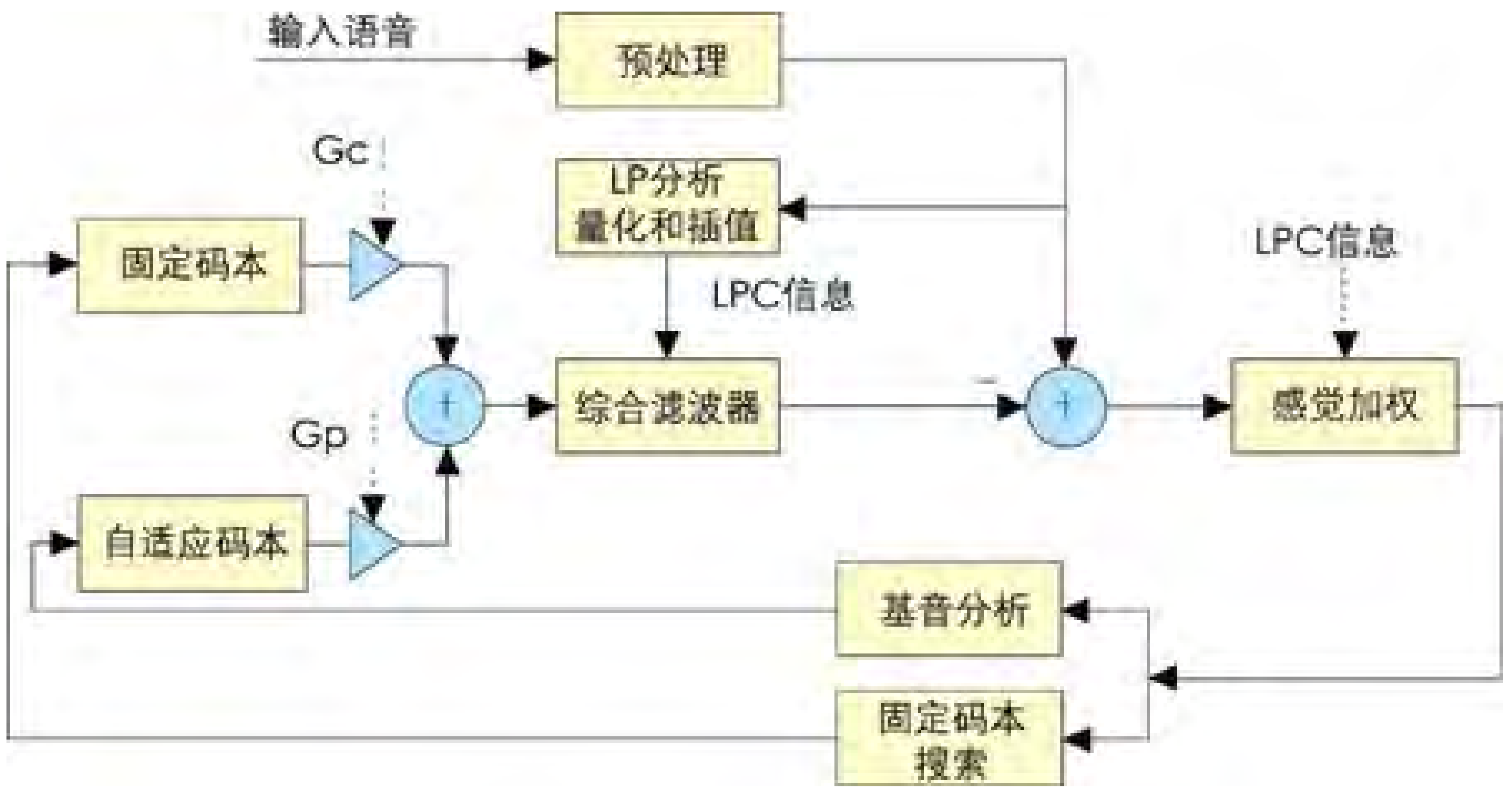
R（m）为信号x[n]的自相关：

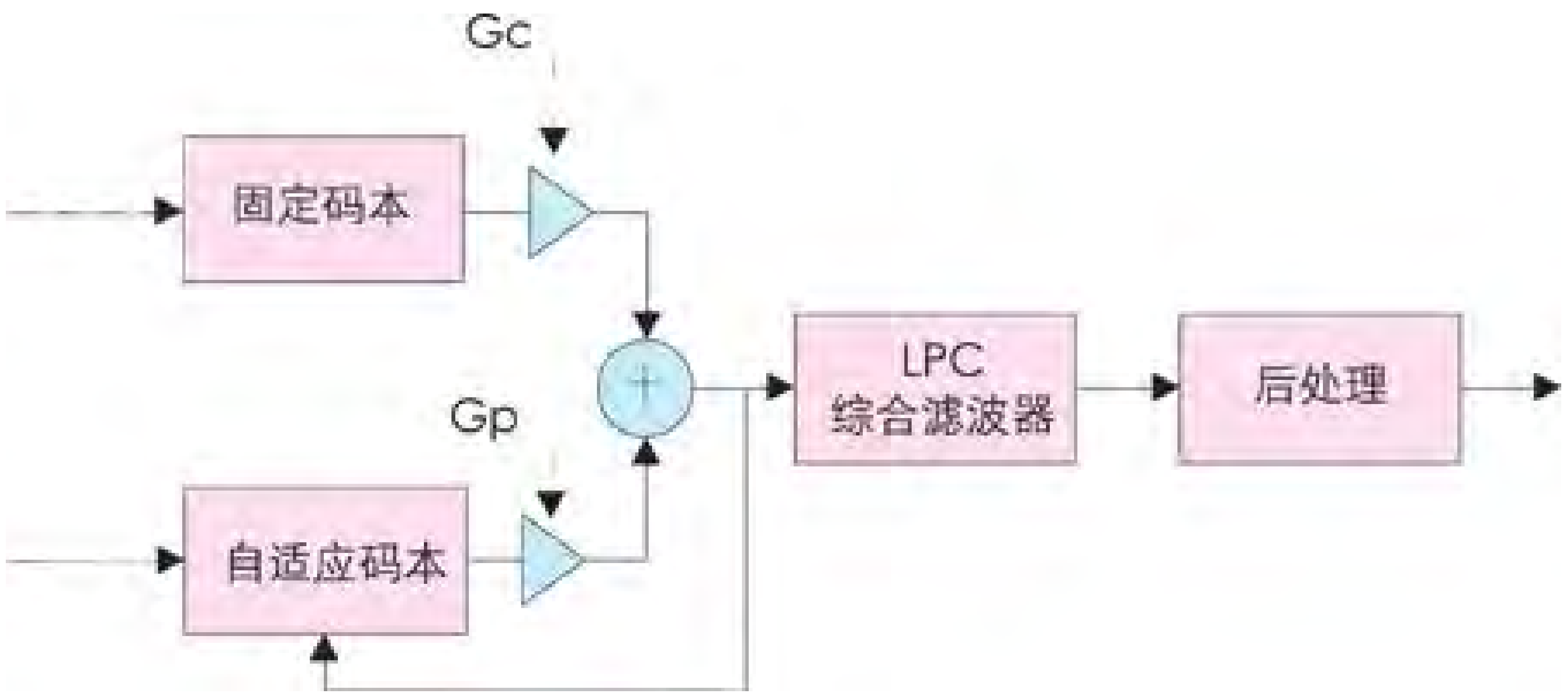
$$R(m)=\sum_{i=0}^{N-1}x[i]x[i-m]$$

因为**R**是Hermitian Toeplitz矩阵，
所以可以利用Levinson-Durbin算法，使得复杂度由O(N^3)降低为O(N^2)

LPC以及改进算法是很多Codec基石，例如：SILK，MPEG-4 ALS，FLAC，G7xx

经典语音编码模型: G.729(CELP)

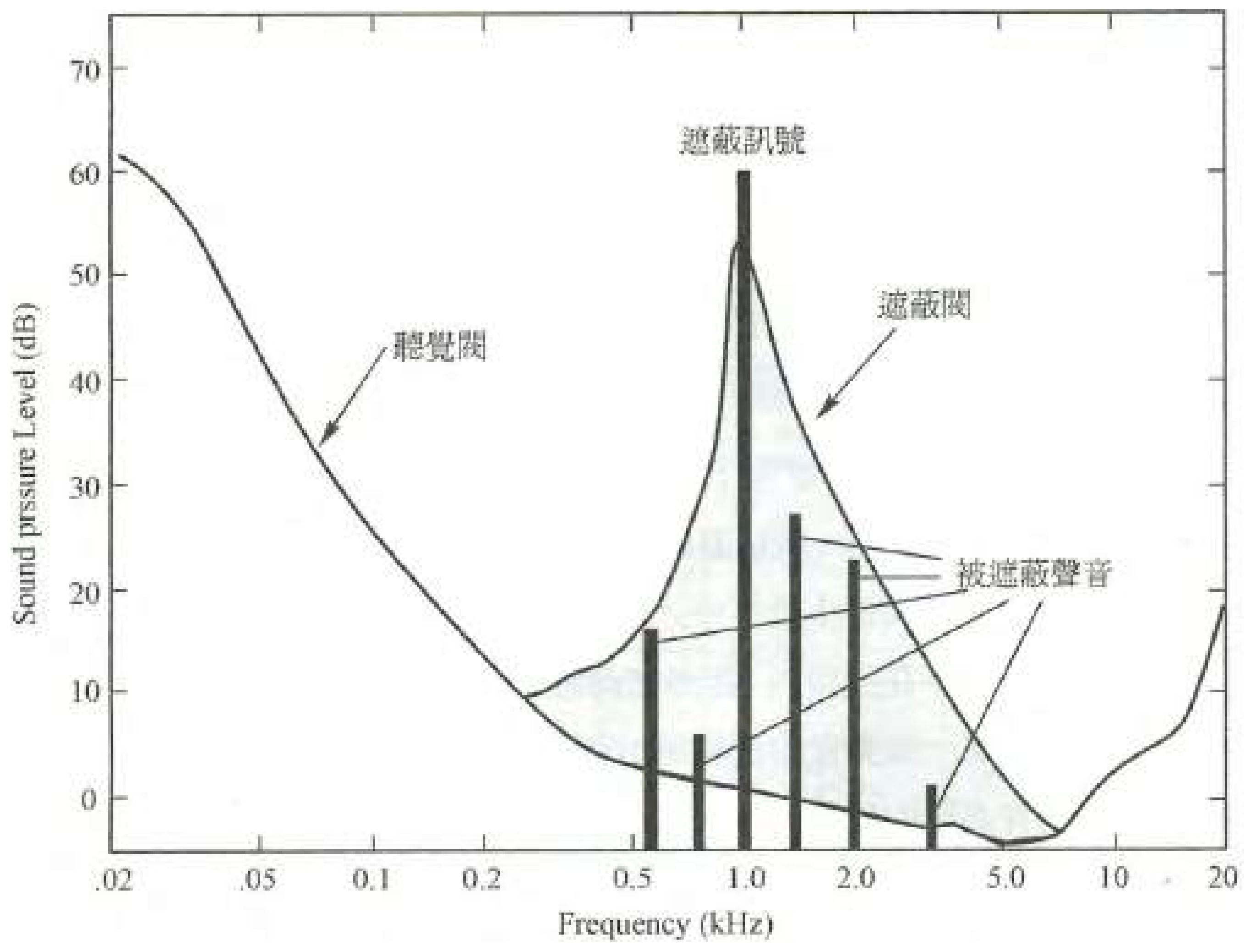


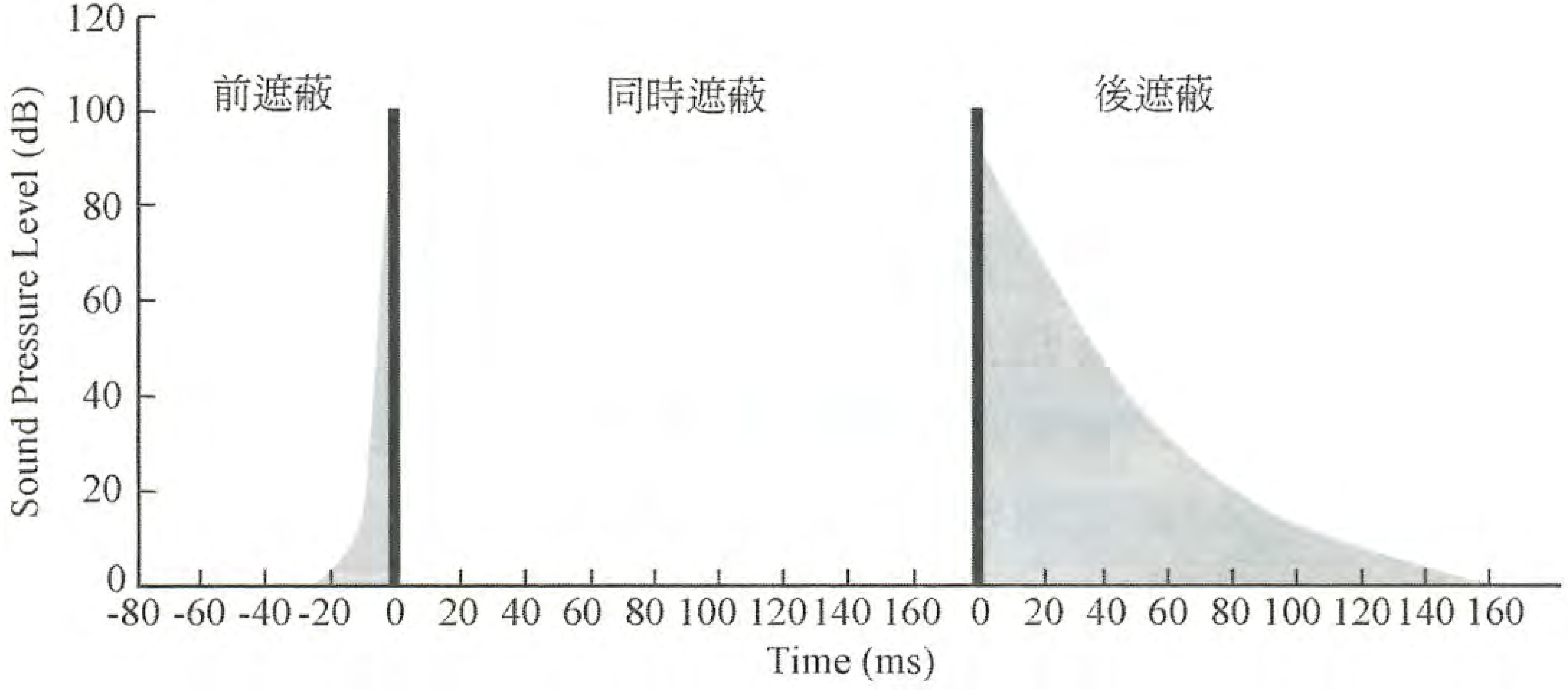


G.729 建议了共轭结构的算术码本激励线性预测(CS-ACELP)编码方案。G.729算法的帧长为 10ms, 编码器含5ms 前瞻，算法时延15ms， 语音质量MOS分可达4.0.

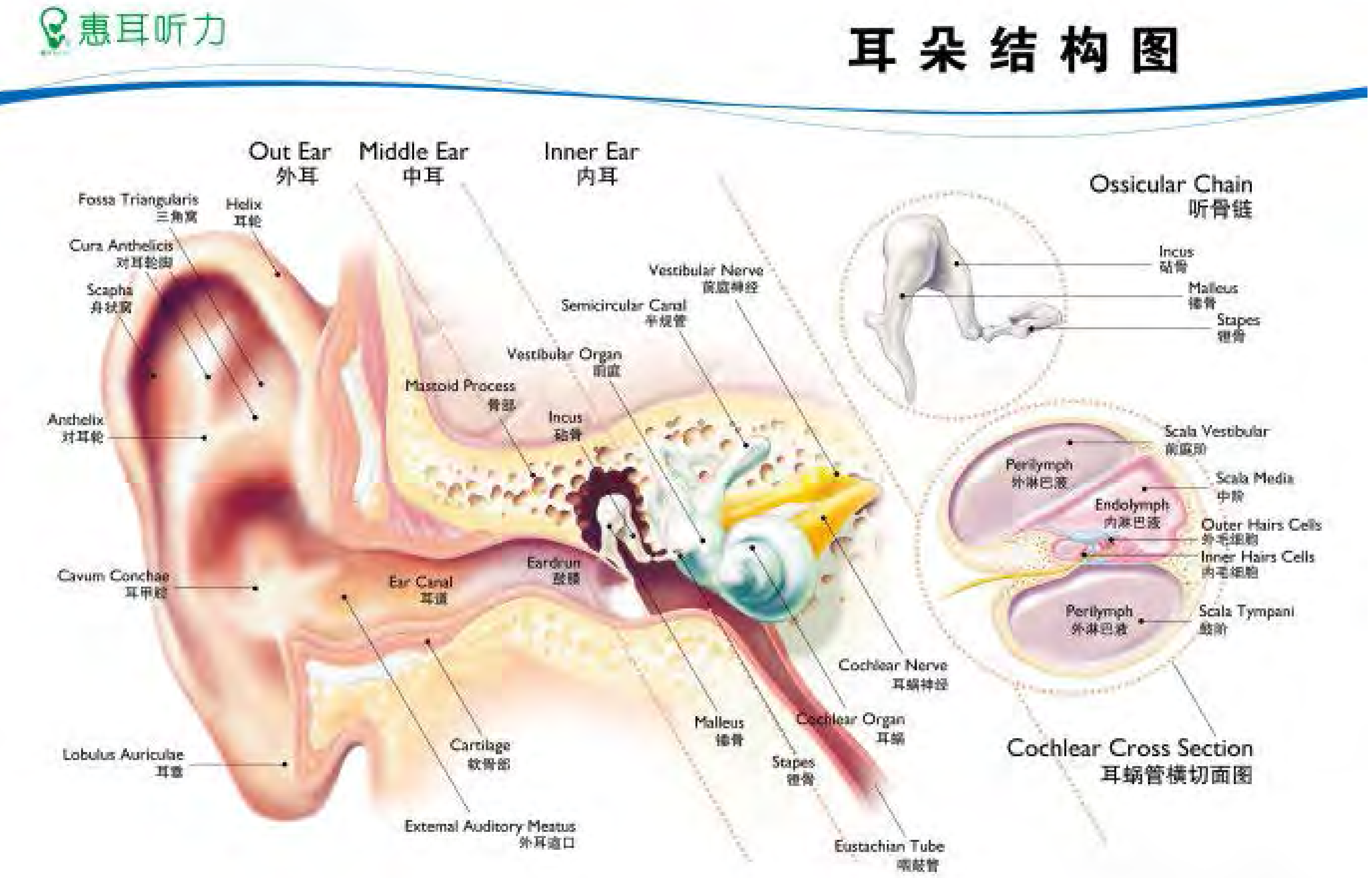
ISO编码模型——--心理声学模型

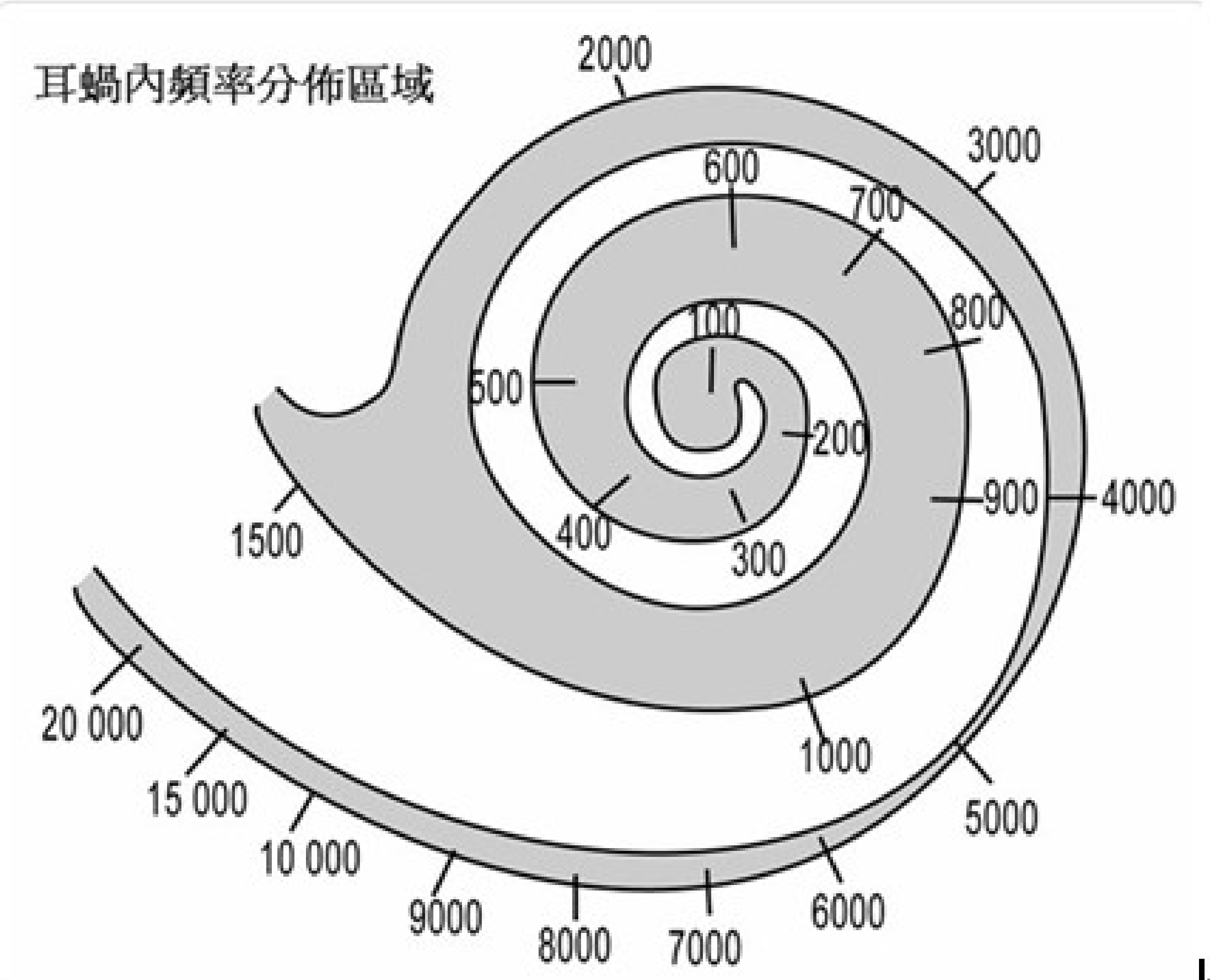
一个声音信号能否被人耳听到取决于它的频率，强度以及其他音的干扰，心理声学模型就是找出音频信号中存在的冗余信息，以达到对其压缩去不影响听觉效果的目的，覆盖的声学原理有：临界频带，绝对听觉阈值，频域掩蔽和时域掩蔽。心理声学理论的成熟为感知编码系统奠定了理论基础。





听觉模型


惠耳听力
耳朵结构图
Out Ear
外耳
Middle Ear
中耳
Inner Ear
内耳
Fossa Triangularis
三角窝
Helix
耳轮
Cura Anthelicis
对耳轮脚
Scapha
舟状窝
Anthelix
对耳轮
Cavum Conchae
耳甲腔
Lobulus Auriculae
耳垂
Cartilage
软骨部
External Auditory Meatus
外耳道口
Ear Canal
耳道
Eardrum
鼓膜
Mastoid Process
骨部
Incus
砧骨
Vestibular Organ
前庭
Semicircular Canal
半规管
Vestibular Nerve
前庭神经
Cochlear Nerve
耳蜗神经
Cochlear Organ
耳蜗
Malleus
锤骨
Stapes
镫骨
Eustachian Tube
咽鼓管
Ossicular Chain
听骨链
Incus
砧骨
Malleus
锤骨
Stapes
镫骨
Scala Vestibular
前庭阶
Perilymph
外淋巴液
Scala Media
中阶
Endolymph
内淋巴液
Outer Hairs Cells
外毛细胞
Inner Hairs Cells
内毛细胞
Perilymph
外淋巴液
Scala Tympani
鼓阶
Cochlear Cross Section
耳蜗管横切面图


耳蝸內頻率分佈區域
2000
3000
600
700
800
100
500
200
900
4000
400
300
1500
1000
20 000
15 000
10 000
9000
8000
7000
6000
5000

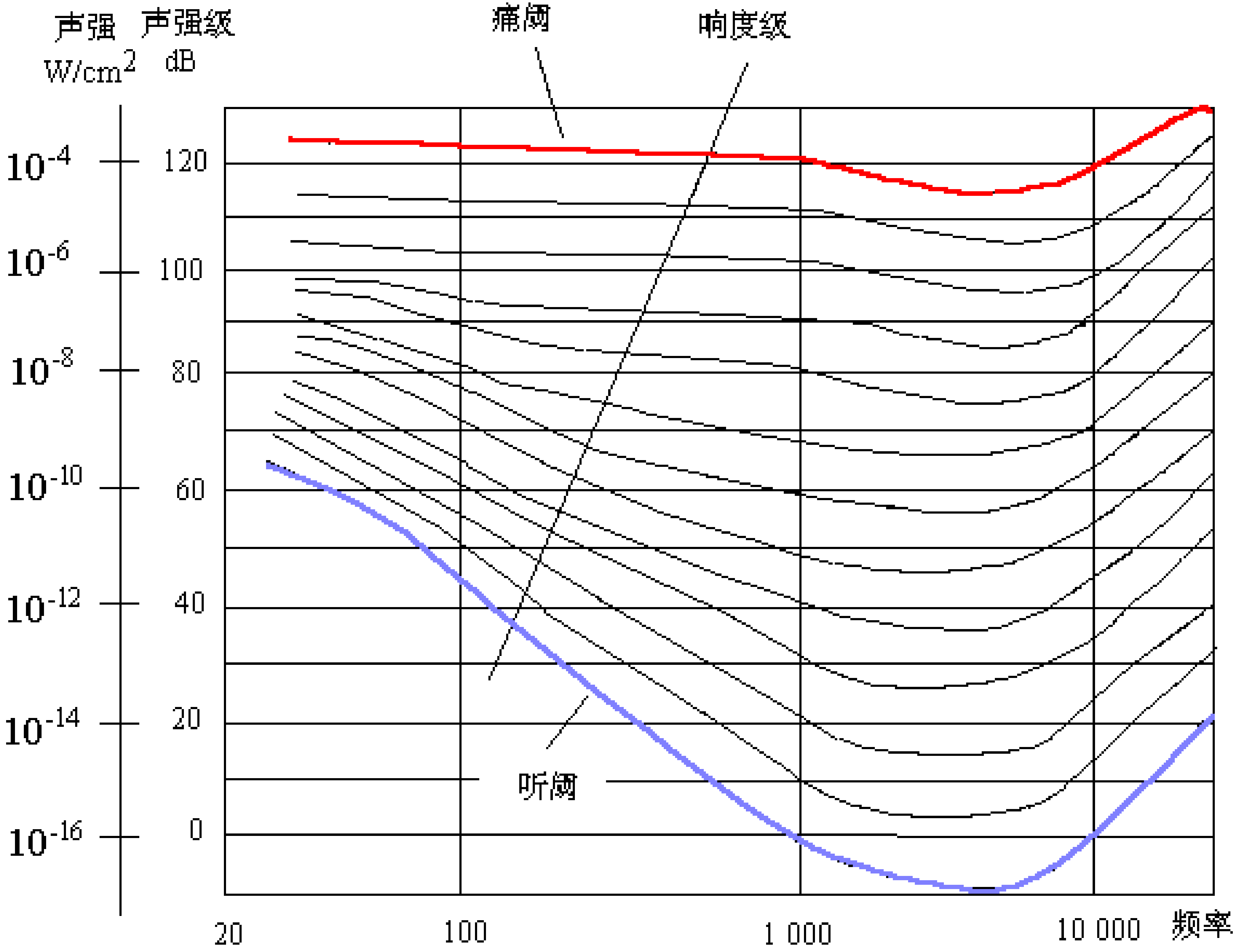
临界频带

由于声音频率与掩蔽曲线不是线性关系，为从感知上来统一度量声音频率，引入了“临界频带”的概念。通常认为，在20Hz到16kHz范围内有24个监界频带。临界频带的单位叫Bark(巴克)

$$B(f) = 13\tan^{-1}(0.76\frac{f}{1000}) + 3.5\tan^{-1}(\frac{f}{7500})^2$$

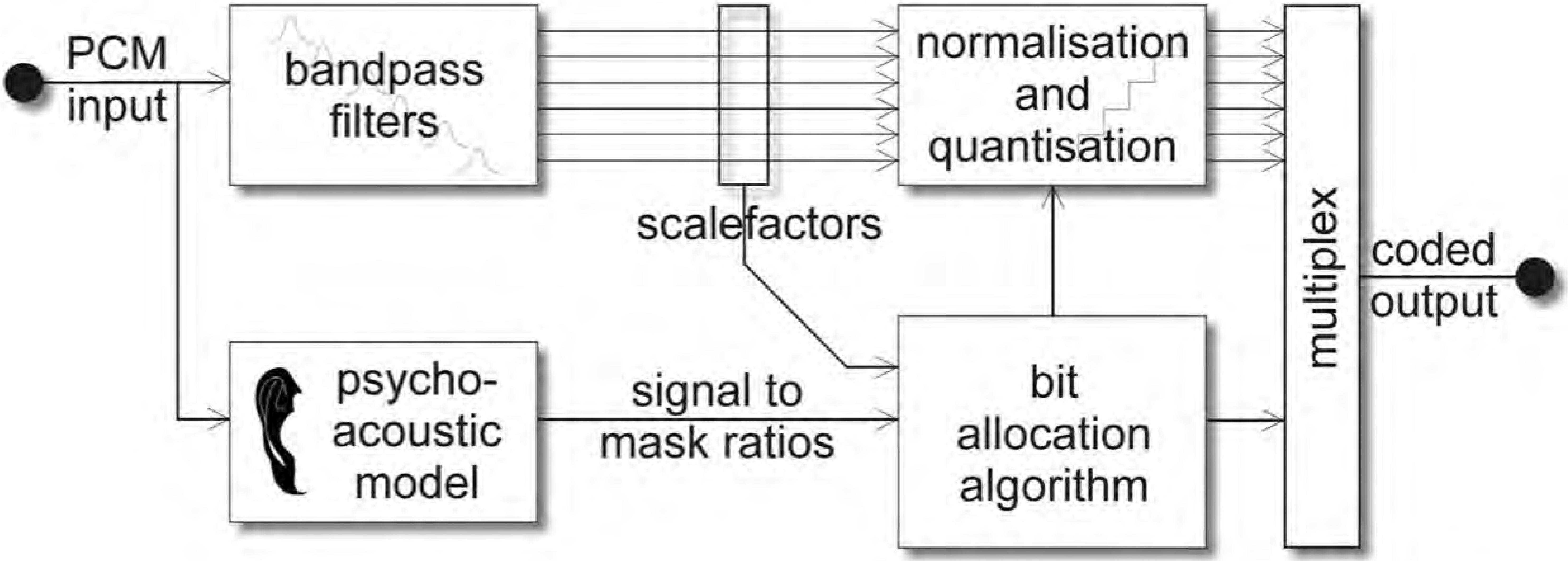
临界	频率 (Hz)			临界	频率 (Hz)		
频带	低端	高端	宽度	频带	低端	高端	宽度
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

绝对听觉阈值

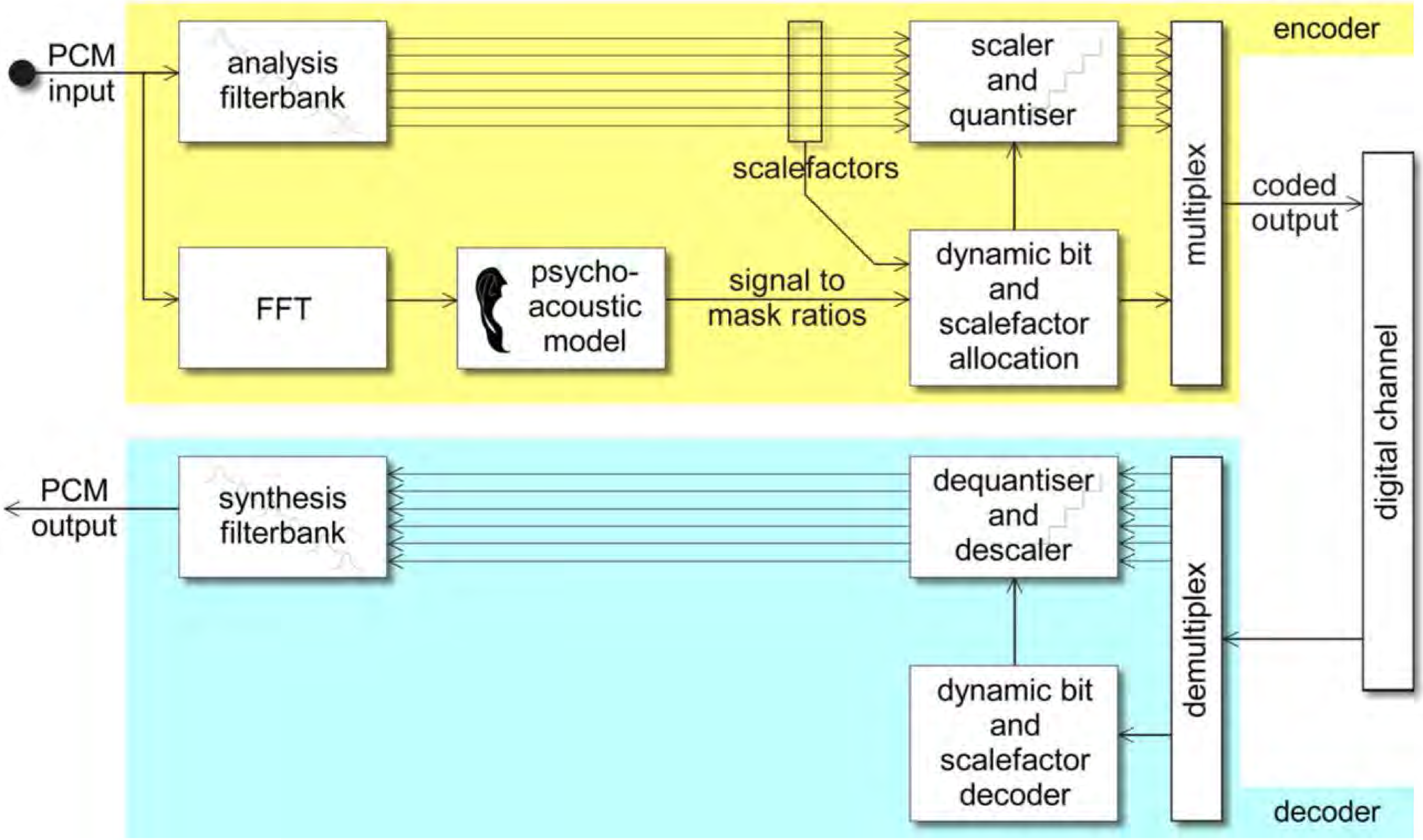


声强
W/cm2
声强级
dB
痛阈
响度级
听阈
10-4
10-6
10-8
10-10
10-12
10-14
10-16
120
100
80
60
40
20
0
20
100
1 000
10 000
频率

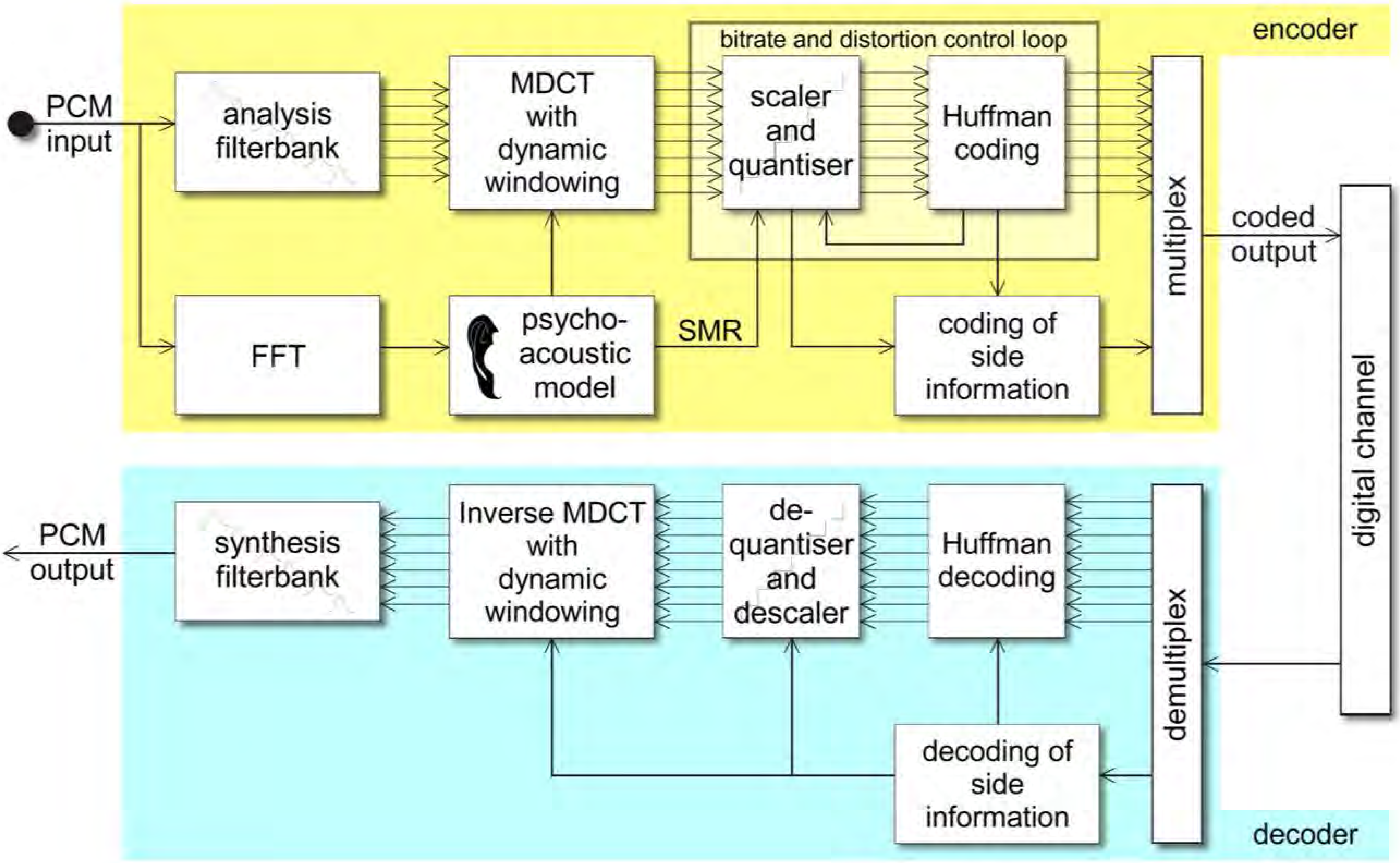
经典音频编码：ISO



PCM
input
bandpass
filters
normalisation
and
quantisation
scalefactors
multiplex
coded
output
psycho-
acoustic
model
signal to
mask ratios
bit
allocation
algorithm

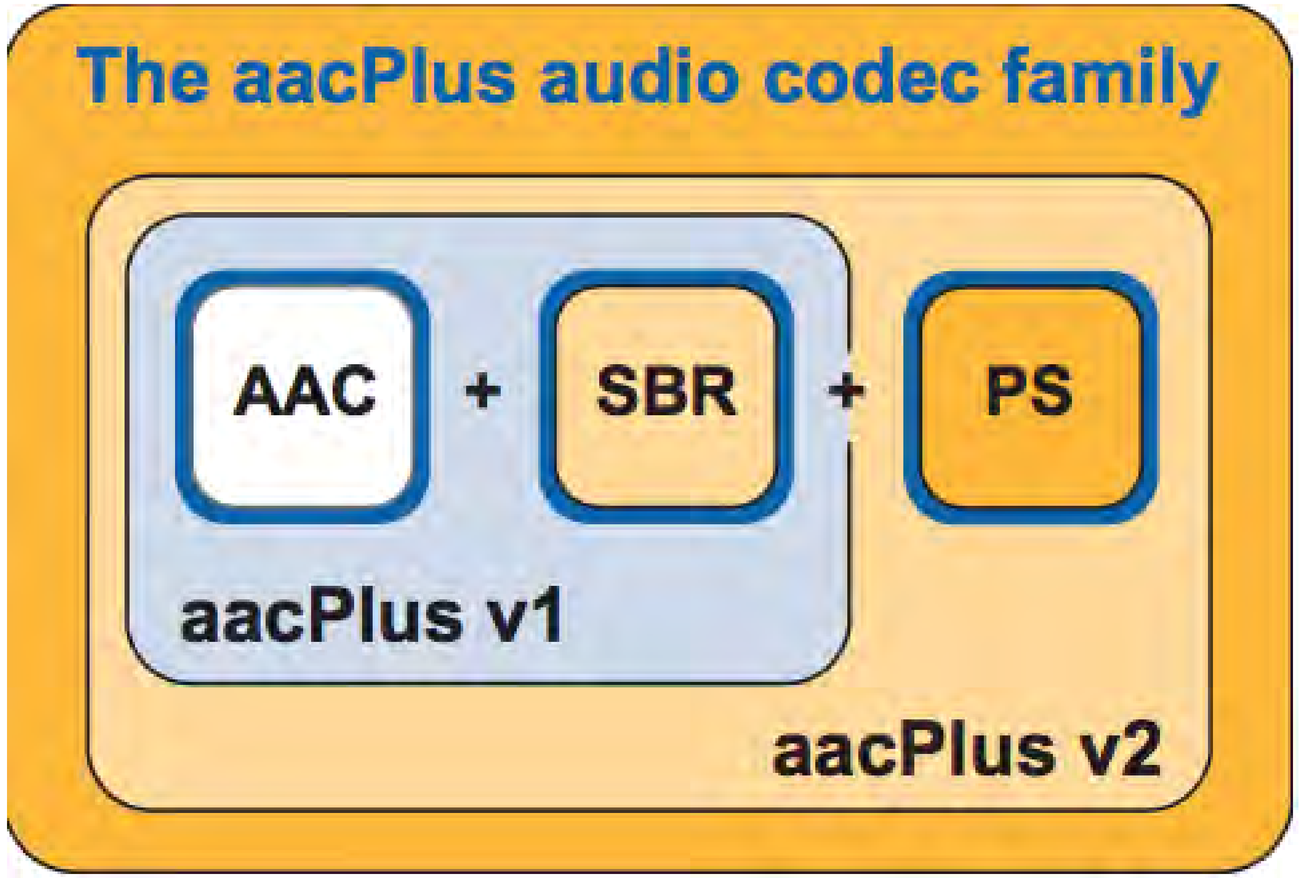


MPEG1 LayerI
Codec

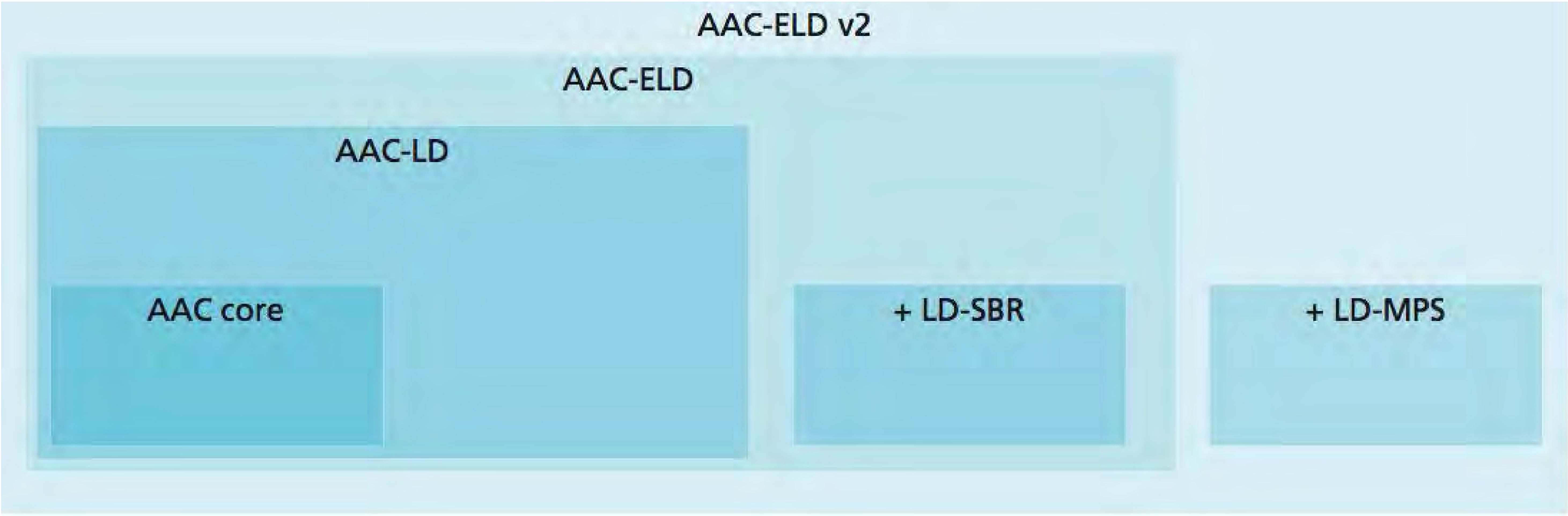


MPEG1 LayerIII Codec

AAC 家族

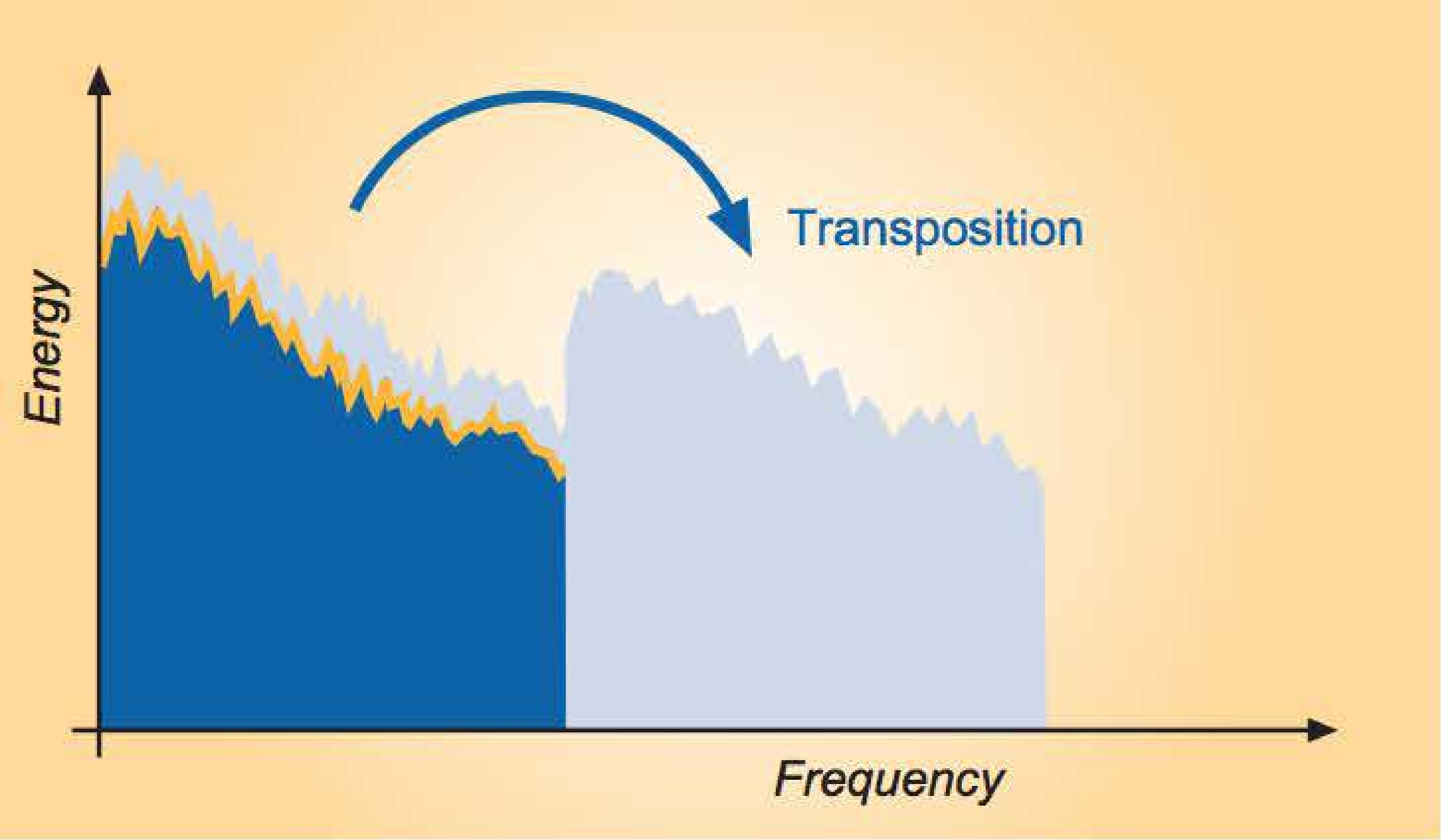


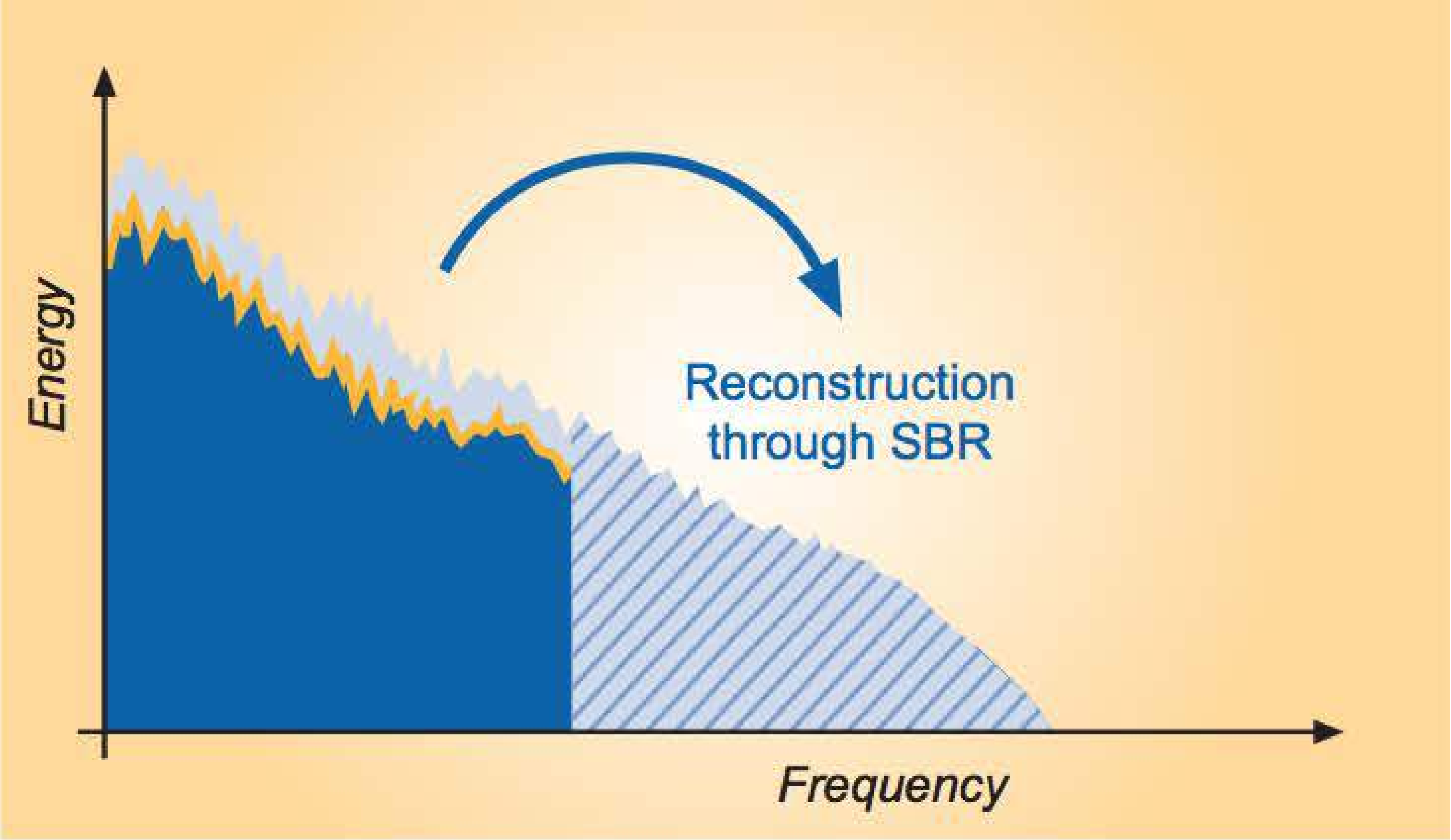
AACplus 协议族



AAC LD 协议族

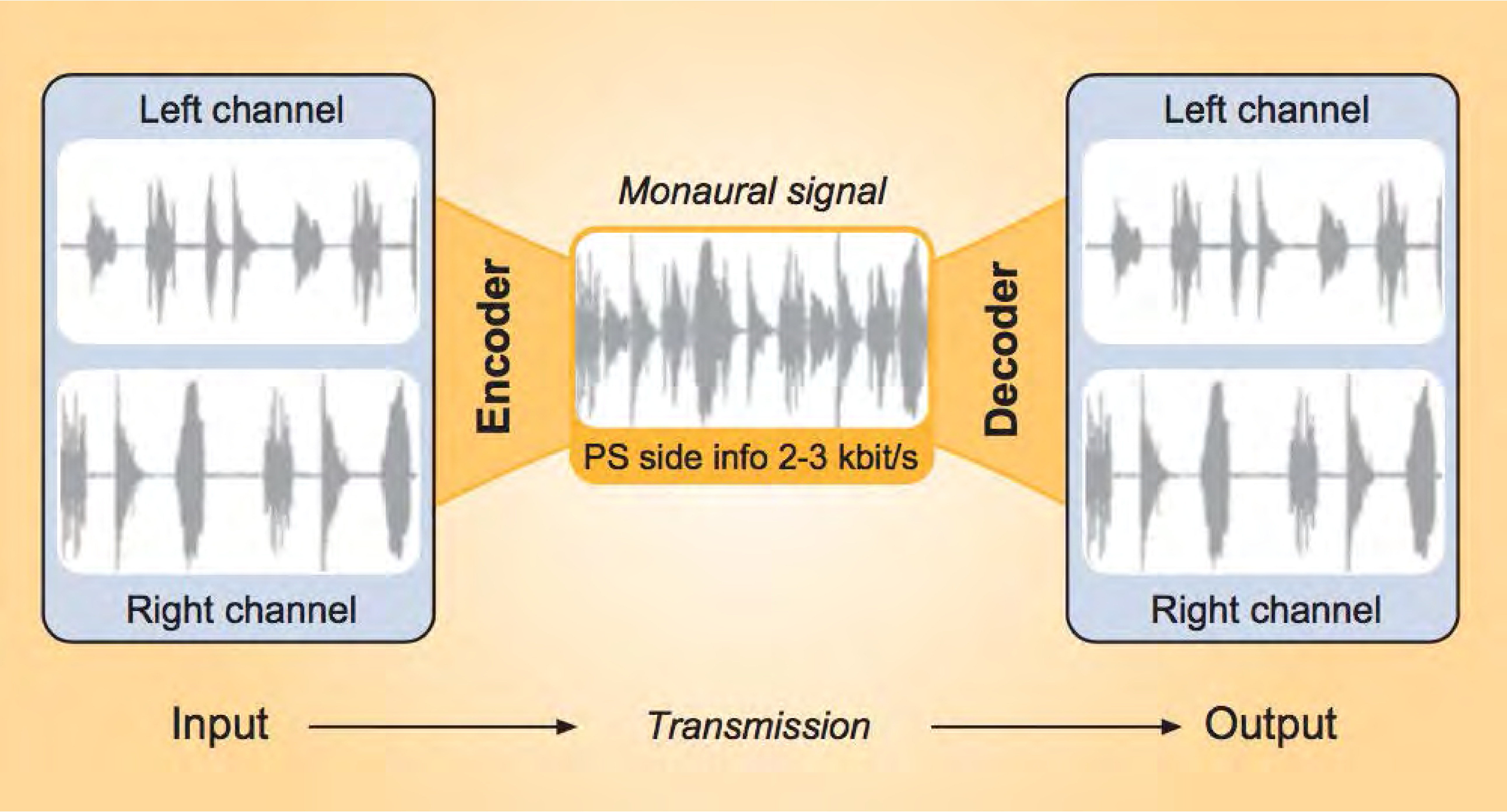
AACPlus 核心模块之一：SBR(Spectral Band Replication)





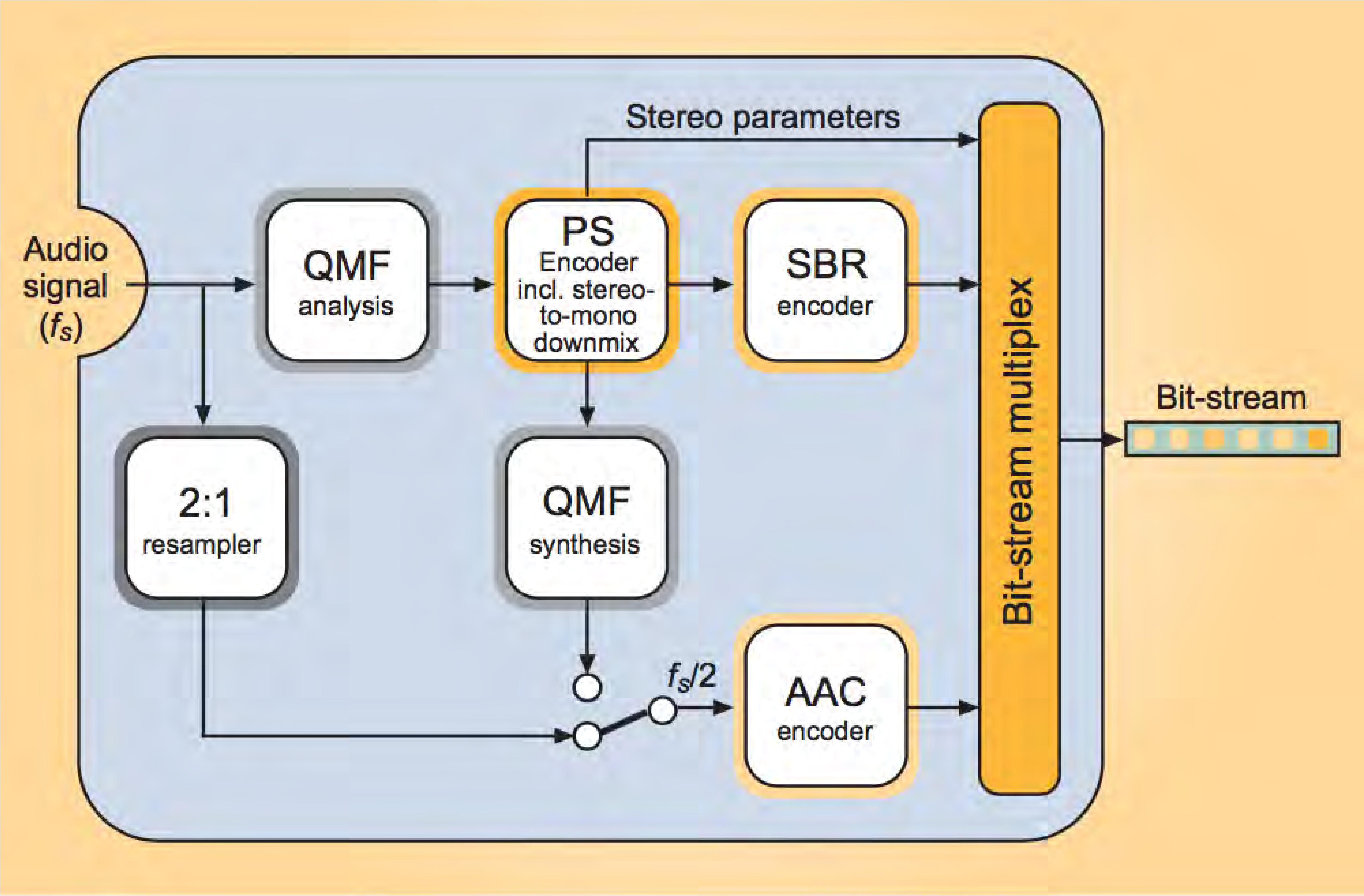
Stereo bitrate (bit/s)	AAC frequency range (Hz)	SBR frequency range (Hz)
20 000	0 - 4 500	4 500 - 15 400
32 000	0 - 6 800	6 800 - 16 900
48 000	0 - 8 300	8 300 - 16 900

AACPlus 核心模块之二：PS(Parametric Stereo)

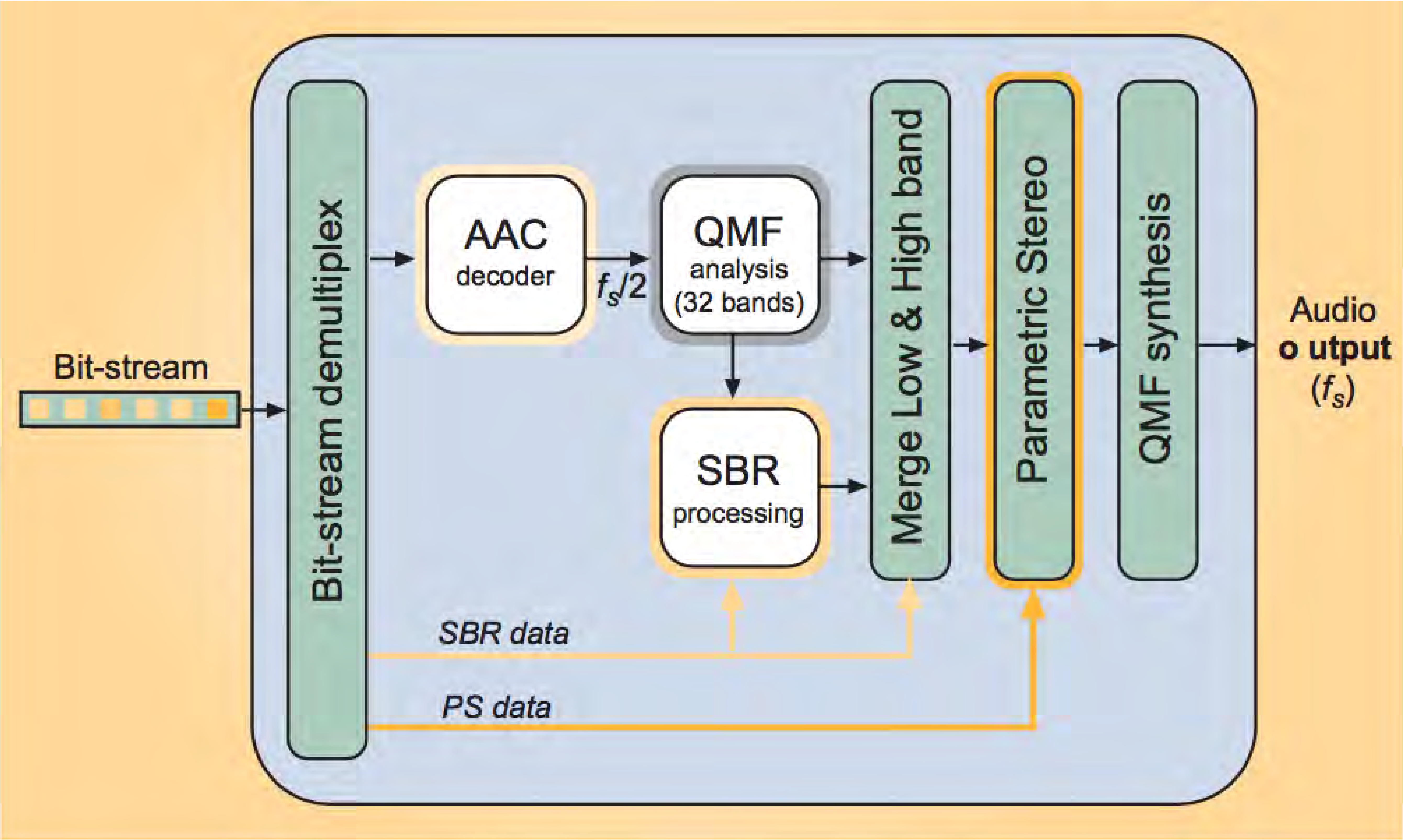


PS 描述参数：IID(Inter-channel Intensity Difference), ICC(Inter-channel Cross-Correlation)
IPD(Inter-channel Phase Difference)

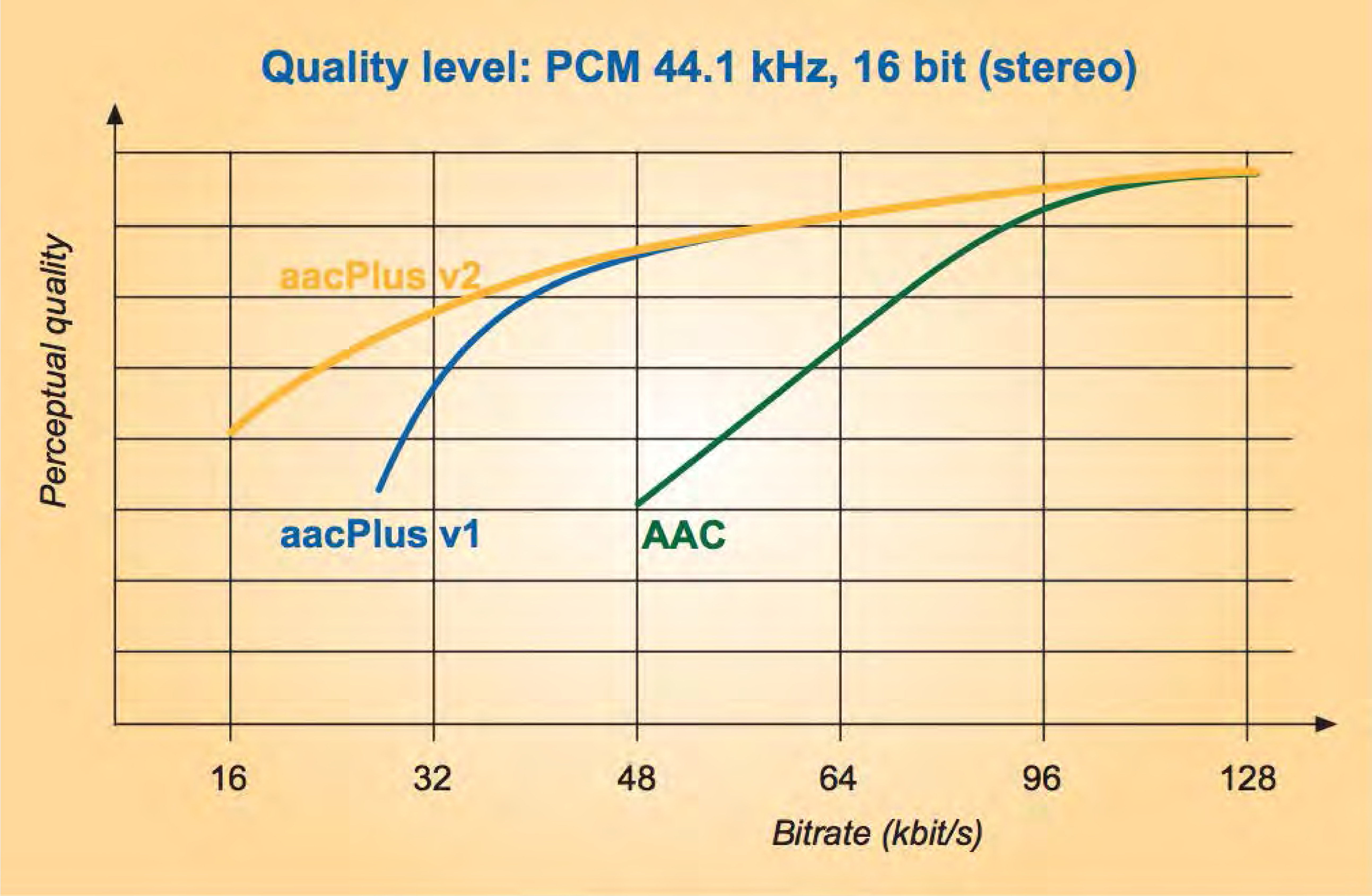
AACPlus v2编码框图



AACPlus v2解码框图



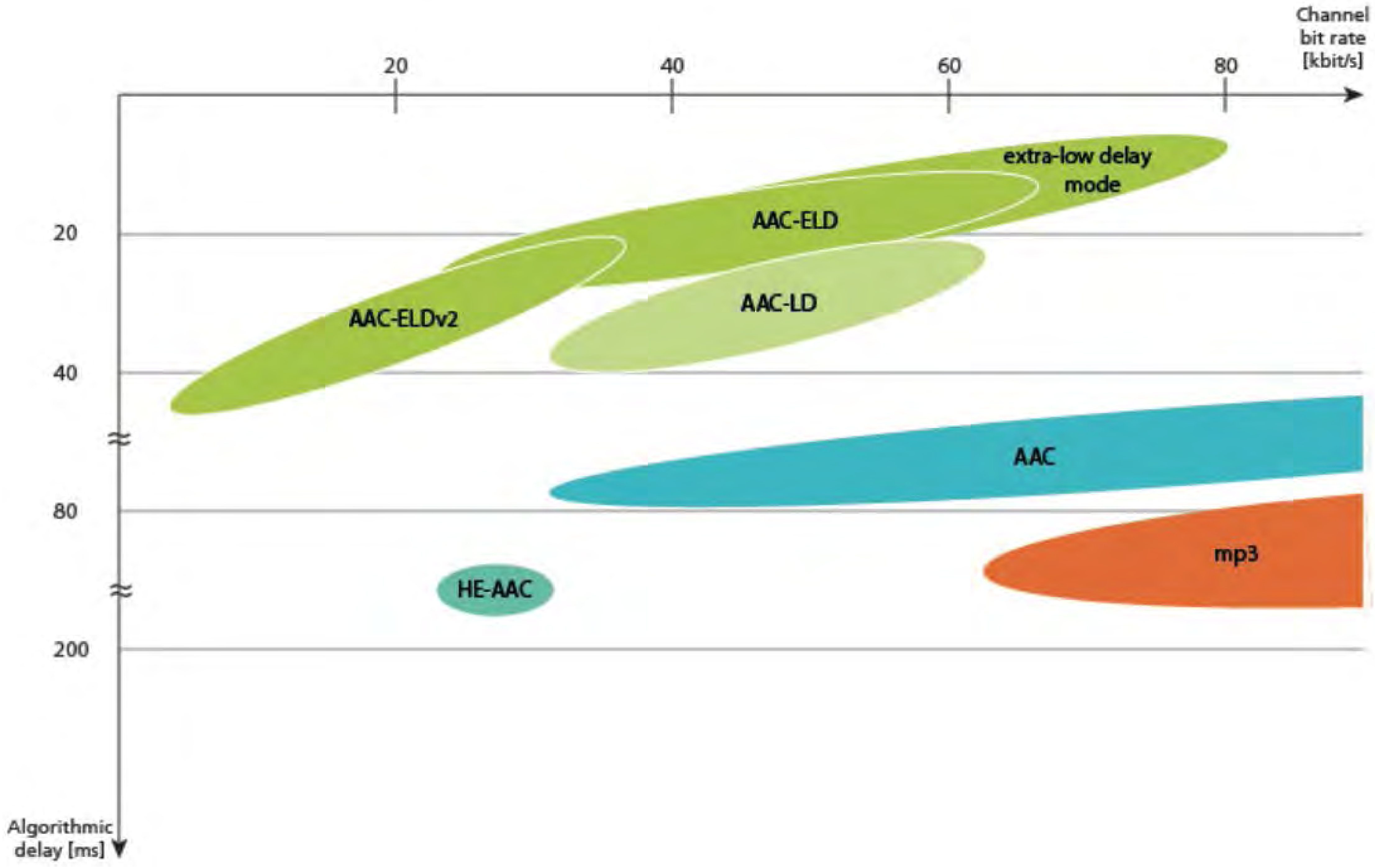
AAC 甜点码率



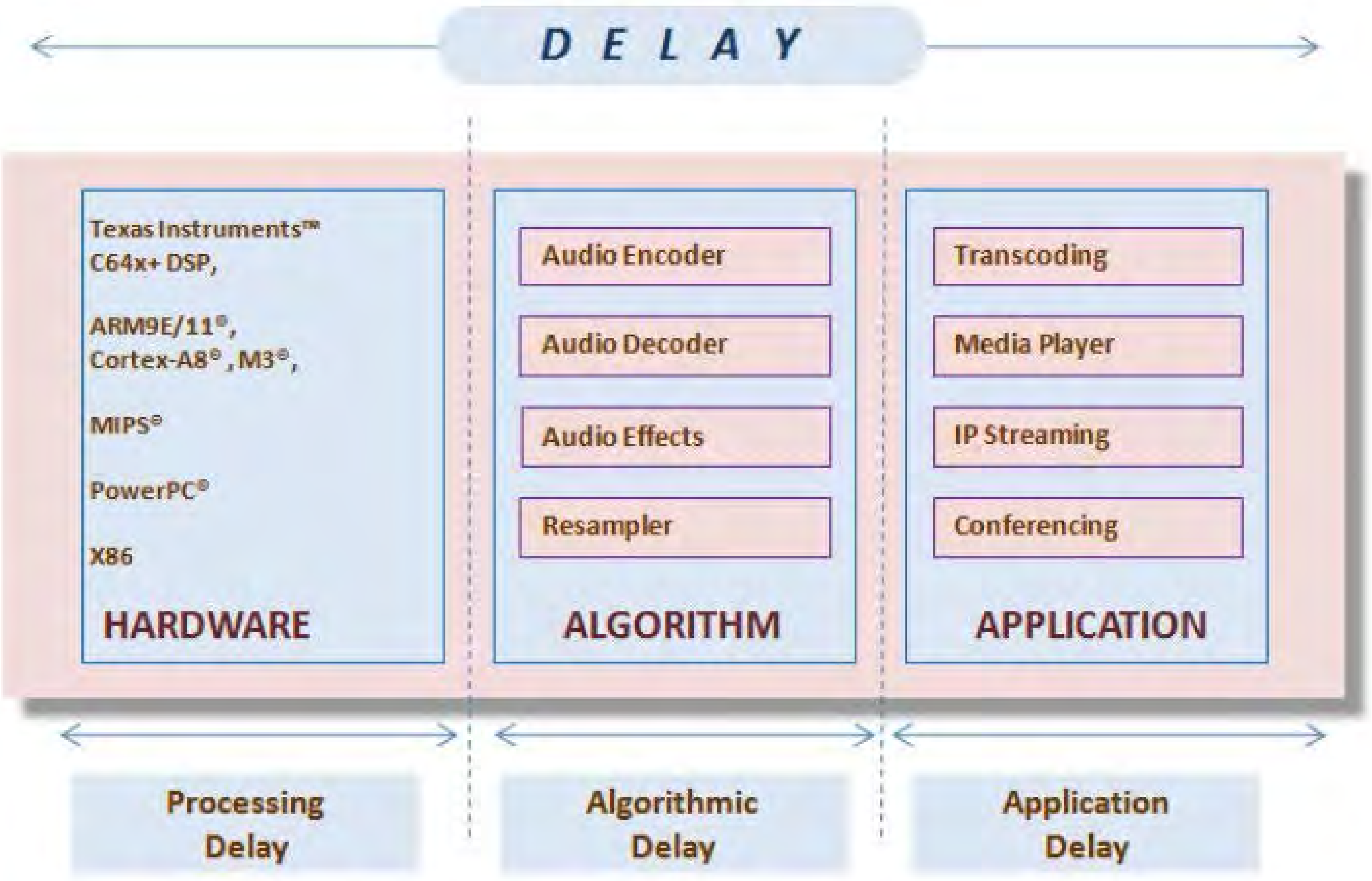
Quality level: PCM 44.1 kHz, 16 bit (stereo)
Perceptual quality
aacPlus v2
aacPlus v1
AAC
16
32
48
64
96
128
Bitrate (kbit/s)

AAC-ELD 家族

产生背景：aacplus v2 已经在压缩和音质方面做到了近似于极致，但由于算法实现上的长达100ms左右的延时极大的阻碍aacplus v2在实时通讯领域的应用。Fraunhofer IIS 为了解决这个问题，对AAC进行相关改进，形成了AAC-ELD协议族。



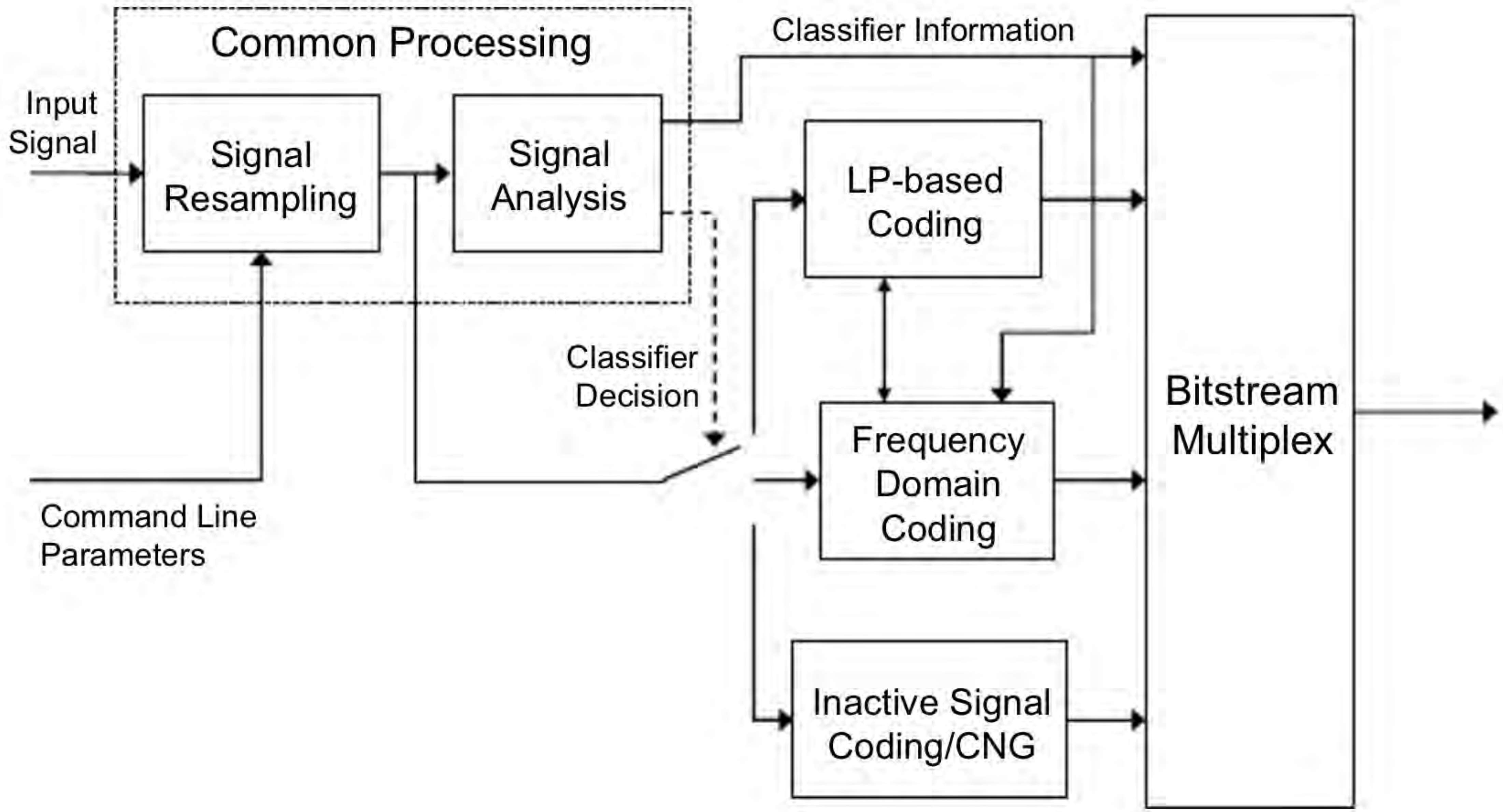
应用中端到端的延时



DELAY
Texas Instruments™
C64x+ DSP,
ARM9E/11®,
Cortex-A8®, M3®,
MIPS®
PowerPC®
X86
HARDWARE
Audio Encoder
Audio Decoder
Audio Effects
Resampler
ALGORITHM
Transcoding
Media Player
IP Streaming
Conferencing
APPLICATION
Processing Delay
Algorithmic Delay
Application Delay

USAC

- EVS (Enhanced Voice Services): 主要是VoiceAge, Dolby, Fraunhofer, 华为 联合开发的 USAC编码器，低速率音乐编码质量很好



IETF系列

	子系列	支持采样率	通道数	支持码率	甜点压缩比	备注
Opus	窄带单声语音 (SILK)	8kHz （128kbps）	1chs	变码率, 最佳码率: 11kbps	8.5%	Opus特点： 1. 支持码率：6～510kbps 2. 采样率 8kHz(Narrowband) to 48kHz(fullband) 3. 帧大小：2.5ms ~ 60ms 4. 支持CBR/VBR 5. 动态调整 6. 支持单声道和双声道，最多支持255个音轨 7. 抗丢包
	宽带单声语音 (SILK)	**16kHz （256kbps）**	**1chs**	**变码率，最佳码率：20kbps**	**7.8%**	
	全带单声语音 (CELT)	48kHz (768kbps)	1chs	变码率，最佳码率：32kbps	4.1%	
iLBC	20ms Frame	8kHz (128kbps)	1chs	15.2kbps	11.8%	iLBC特点： 1. 抗丢包 2. 音质较好，15.2kbps Mos 4.14
	30ms Frame	8kHz (128kbps)	1chs	13.3kbps	10.3%	
GSM	FR (LPC-PRE)	8kHz (128kbps)	1chs	13kbps	10.1%	帧块大小20ms, 蜂窝电话用的编码器
	HR (VSELP)	8kHz (128kbps)	1chs	5.6kbps	4.3%	

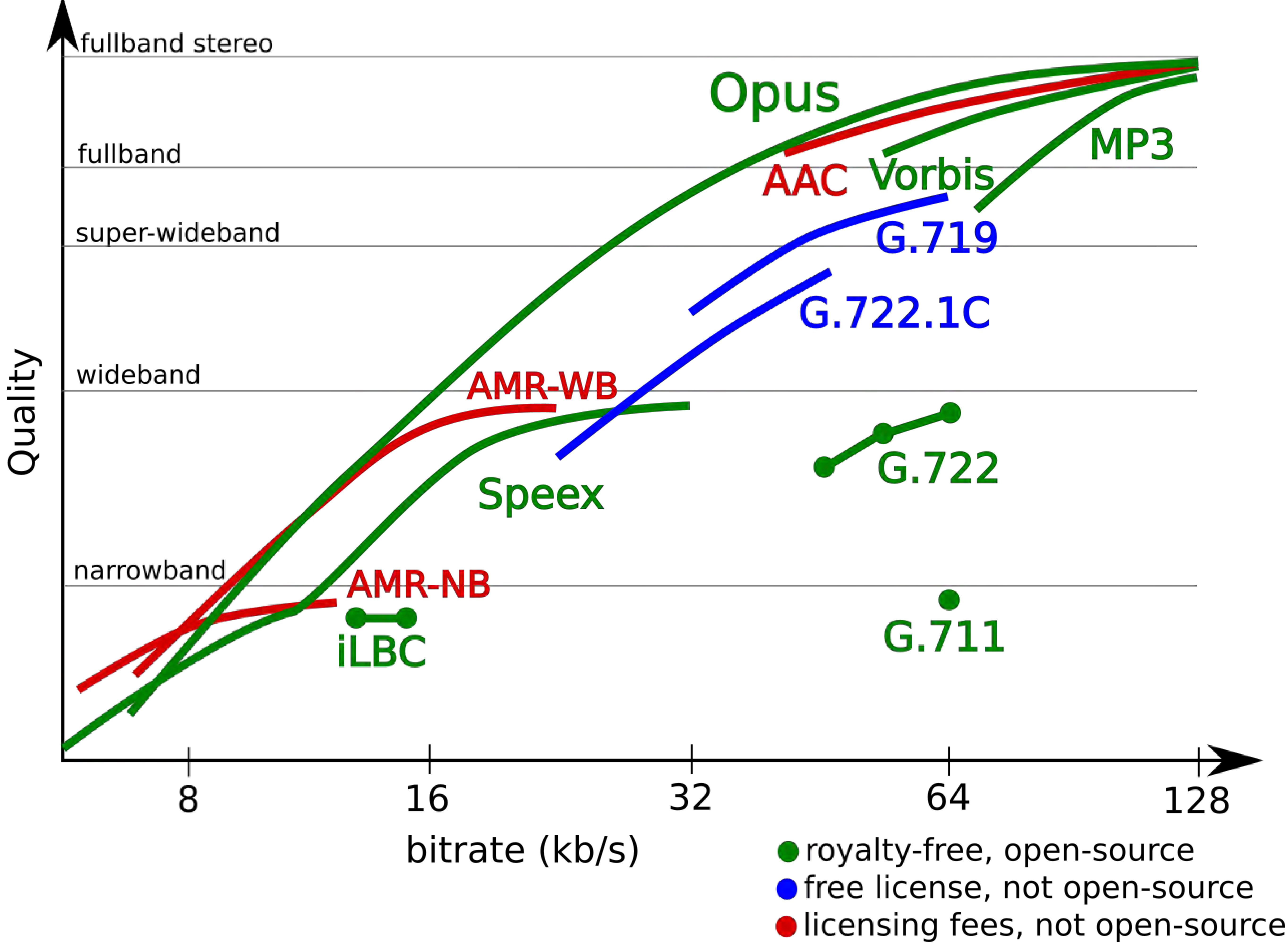
AMR系列

	原始码率	压缩码率	采样率	通道数	最大压缩比	备注
AMR-NB	128kbps	4.75kbps, 5.15kbps 5.9kbps, 6.7kbps 7.4kbps, 7.95kbps 10.2kbps, 12.2kbps	8kHz	1	3.7%	AMR-NB 是最流行的语音编码器，因为压缩效果好，且支持多种码率形式，同时满足GSM和3G网，且是Android系统最早支持语音编码器
AMR-WB	256kbps	6.6kbps, 8.85kbps 12.65kbps, 14.25kbps 18.25kbps,19.85kbps 23.05kbps, 23.85kbps	16kHz	1	2.5%	AMR-WB 是AMR-NB 向宽带的扩展版，主要用在3G和4G通话标准协议中。
AMR-WB+		融合AMR-NB, AMR-WB 外，还有扩展	支持到48kHz	2		第一个支持语音和音频的混合编码，原本3GPP是想通过AMR-WB+ 来对抗AAC，但因为各种原因，没有推广开来

ITU-T G系列

G系列	原始码率	压缩码率	压缩比	备 注	SR/Stereo
G711	128kbps	64kbps	2:1(50%)	最早电话语音压缩算法，目前基本上不用，但会作为VOIP项目互连互通必选项	8kHz/No
G721	128kbps	32kbps	4:1(25%)	波形编码ADPCM，音质略次于PCM编码，可懂度和自然度都很不错	8kHz/No
G722	256kbps	64/56/48kbps	5.3:1(18.75%)	G.722最后被G.726收编	16kHz/No
G.722.1 Siren7	256kbps	32/24/16kbps	16:1(6.25%)	Polycom 第三代Siren7 压缩技术，20ms 封帧，提供40ms的算法延迟 Microsoft office communications Server(audio conferencing)	16kHz/No
G.722.1C Siren14	512kbps	48/32/24kbps Mono	21:1(4.68%)	2005年单声道Siren14变成ITU-T G.722.1C, 20ms封装，提供40ms的算法延迟	32kHz/Yes
G.719 Siren22	768kbps	64/48/32kbps Mono	24:1(4.16%)	2008年Siren22变成ITU-T G.719 全频带Codec	48kHz/Yes
G.723.1	128kbps	6.3/5.3kbps	24:1(4.16%)	H.324 语音编码器，30ms 封帧，37.5ms 算法延迟	8kHz/No
G.726	128kbps	40/32/24/16kbps	8:1(12.5%)	在G.711基础上进一步压缩 早期的视频会议用得较多	8kHz/No
G.729/G.729A/ G.729AB	128kbps	8kbps	16:1(6.25%)	高效的窄带语音编码器，CS-ACELP， 10ms 封装，Algorithmic delay is 15ms per frame, with 5ms look-ahead delay	8kHz/No

注：平时视频会议Polycom系统一般是用Siren14或22


fullband stereo
fullband
super-wideband
wideband
narrowband
Quality
Opus
AAC
Vorbis
MP3
G.719
G.722.1C
AMR-WB
Speex
G.722
AMR-NB
iLBC
G.711
8
16
32
64
128
bitrate (kb/s)
royalty-free, open-source
free license, not open-source
licensing fees, not open-source

ISO 系列

	子系列	支持采样率	通道数	支持码率	甜点压缩比	备注
MP3	MPEG-1/MPEG-2 audio Layer III	32kHz 44.1kHz 48kHz	2chs	32,40,48,56,64,80,96,112,128,160,192,224,256,320kbps	11:1	128kbps 达到CD音质
	MP3 Pro (MP3+SBR)	32kHz 44.1kHz 48kHz	2chs	码率能达到上面的一半	22:1	因为高频是通过算法生成，会在听觉上产生不自然。另外因为专利和应用上推广的限制，MP3 Pro实际上没有大规模应用。
AAC	AAC-LC	8～96kHz	最多支持48	48kbps ～ 400kbps	甜点码率：96kbps 6.8%	96kbps aac 质量相当于MP3 128kbps, 同样达到CD音质
	HE-AAC	8～96kHz	最多支持48	20～192kbps	甜点码率：32kbps 2.3%	主要用于<80kbps 的编码，由于采用SBR+PS, 在同质量情况下，压缩率可以做到最低
	AAC-LD/ELD	8～96kHz	最多支持48	32~64kbps	4.6%	主要是做到了AAC的低延时，最小延时做到7.5ms, 达到了延时和压缩比的最佳平台

3GPP系列:EVRC

	EVRC	EVRC-B	EVRC-C	EVRC-D	EVRC-E
Bandwidth	200~3400Hz	200~3400Hz	50~7000Hz	50~7000Hz	50~7000Hz
Standardized	TIA 1997	3GPP2 2006	3GPP2 2007	3GPP2 2009	3GPP2 2011
Technology	RCELP/ACELP	CELP/PPP/NELP	CELP/LPC/PPP/NELP	CELP/LPC/PPP/NELP	CELP/LPC/PPP/2K/NELP
Bitrate(kbps)	6	4 to 8.3	4 to 7.5	4 to 7.5	2 to 7.5
Delay(ms)	23	23	23	23	23
Frame(ms)	20	20	20	20	20
Lookahead(ms)	3	3	3	3	3
Quality	Narrowband	Narrowband	Wideband	Wideband	Wideband
Complexity	30mis	40mips			
ROM(Bytes)	200k	200k			
RAM(bytes)	20k + 10k/ch	20k + 10k/ch			

EVRC 是CDMA 中使用的语音编解码器，由高通公司1995年提出目标是取代QCELP

极低码率

	原始码率	压缩码率	压缩比	备 注
MELP	128kbps	2.4kbps	1.8%	MELP 美国军方语音编码标准, 22.5ms 帧封装
MPEG-HVXC	128kbps	2kbps, 4kbps	1.5%	MPEG4 中的低极码充语音编码算法，用于对讲机
IMBE	128kbps	4.15kbps	3.2%	IMBE是MBE改进算法，个别洲的对讲机codec协议组成员
AMBE	128kbps	4.8kbps	3.75%	目前对讲机主要的codec算法，DVSI公司专利算法，算法延时32ms

注：1) 极低码率应用场景：对讲机 / 卫星通讯 / 军工

2) 目前最极端极端低码率可以做到300bps, 相关于原数据的：0.2%， 这时只能传达语音的内容，很多声色部分都丢掉了。

全频带

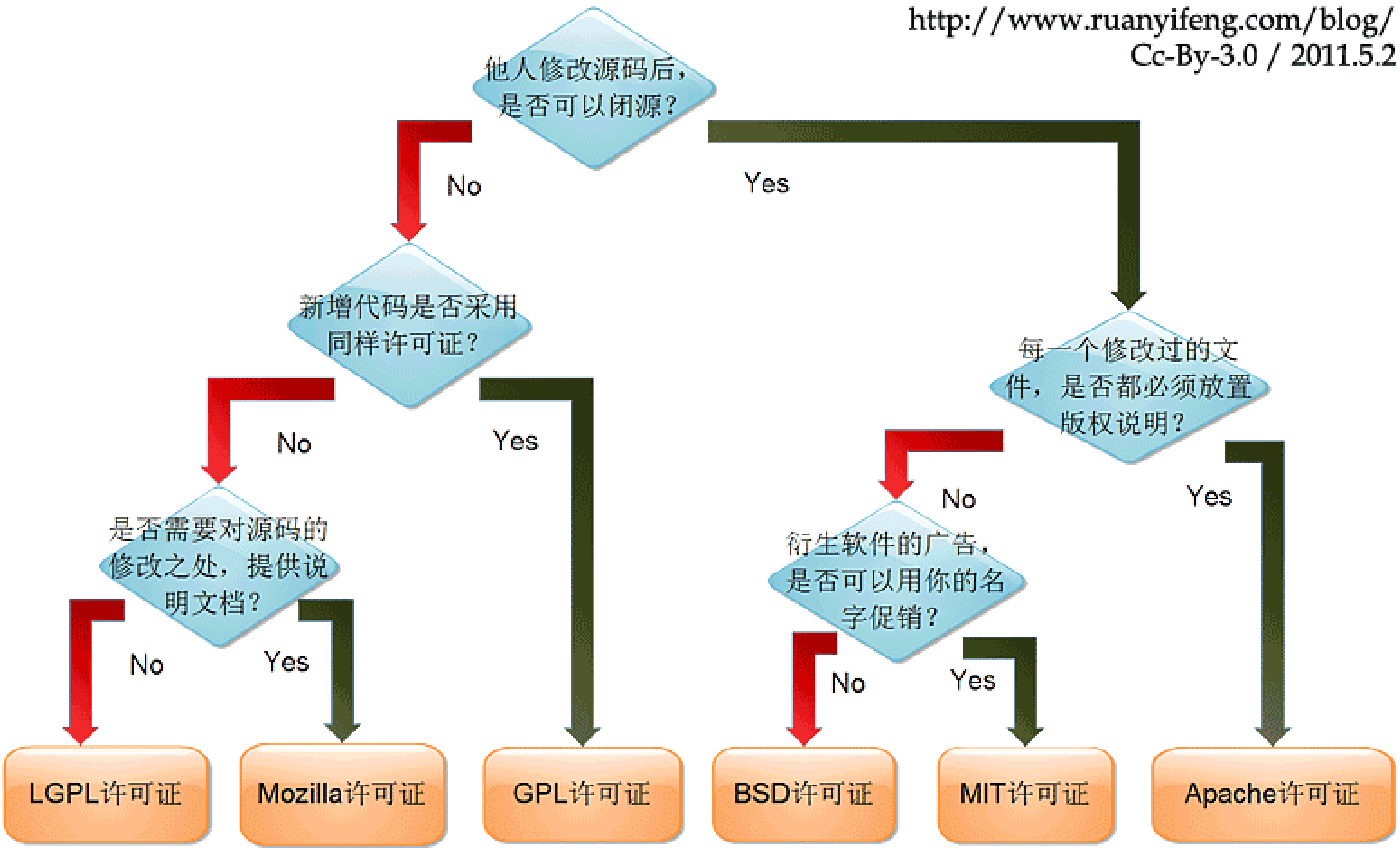
	分级	支持码率（kbps）	延时
EVS	Narrowband(NB) 8kHz	5.9, 7.2, 8, 9.6, 13.2, 16.4, 24.4	20ms 帧封装 编码延时：20ms + 0.9375ms (Resampling filter) + 8.75 (look-ahead) + 2.3125(time-domain bandwidth extension) 解码延时：30.9375ms (8kHz)
	Wideband(WB) 16kHz	5.9, 7.2, 8, 9.6, 13.2 , 13.2(通道感知) 16.4, 24.4, 32, 48, 64, 96, 128 (6.6~23.85 for AMR-WB IO)	
	Super-Wideband(SWB) 32kHz	9.6, 13.2, 13.2(通道感知) 16.4, 24.4, 32, 48, 64, 96, 128	
	Fullband (FB)	16.4, 24.4, 32, 48, 64, 96, 128	

编解码使用注意License

开源项目常用License 协议:

Top 20 Most Commonly Used Licenses in Open Source Projects

Rank	License	%
1.	GNU General Public License (GPL) 2.0	42.49%
2.	MIT License	11.49%
3.	Artistic License (Perl)	7.84%
4.	GNU Lesser General Public License (LGPL) 2.1	7.13%
5.	BSD License 2.0	6.82%
6.	GNU General Public License (GPL) 3.0	6.41%
7.	Apache License 2.0	5.47%
8.	Code Project Open 1.02 License	2.14%
9.	Microsoft Public License (Ms-PL)	1.79%
10.	Mozilla Public License (MPL) 1.1	1.03%
11.	GNU Lesser General Public License (LGPL) 3.0	0.87%
12.	Eclipse Public License (EPL)	0.71%
13.	Common Public License (CPL)	0.42%
14.	zlib/libpng License	0.36%
15.	BSD Two Clause License	0.34%
16.	Common Development and Distribution License (CDDL)	0.32%
17.	Academic Free License	0.32%
18.	Open Software License (OSL)	0.22%
19.	Microsoft Reciprocal License (Ms-RL)	0.20%
20.	Ruby License	0.20%



http://www.ruanyifeng.com/blog/
Cc-By-3.0 / 2011.5.2
他人修改源码后，是否可以闭源？
No
Yes
新增代码是否采用同样许可证？
No
Yes
每一个修改过的文件，是否都必须放置版权说明？
No
Yes
是否需要对源码的修改之处，提供说明文档？
No
Yes
衍生软件的广告，是否可以用你的名字促销？
No
Yes
LGPL许可证
Mozilla许可证
GPL许可证
BSD许可证
MIT许可证
Apache许可证

开源代码使用注意事项：

Source code files copyright and Technologies patent 区别

多媒体方面的技术专利简介:

二个多媒体技术的专利池:

MPEG-LA:

world’s leading packager of patent pools for standards and other technology platforms used in consumer electronics, as well as chemical, eCommerce, education, energy, environment, healthcare and biotechnology, manufacturing and materials, transportation and wireless technology

Via Licensing:

Via Licensing 是致力于知识产权、富于经验的专家组，它代表第三方公司组织管理授权专案。Via Licensing 公司是杜比实验室公司 (NYSE:DLB) 的独立子公司，是一家拥有 40 多年技术授权经验的公司

常见Codec Patent License

Proposed AVC License Terms, MPEG-LA Pool

Field	Internet	"Over-the-Air" Broadcast	Subscription Content	Packaged Media or Pay-per-view Content
Encoder and/or Decoder Product	$0.20/product $0.10/product above 5M units annually 100K Unit floor Paid by manufacturer of end-user product or PC operating system Caps: Total Enterprise/Entity: $3.5-5M/yr			
Use	Free for free content May be raised to local broadcast rate at license renewal	$2.5K per encoder used (one-time fee) – or – Per Market fee: Free < 100K subs. [1] $2.5K/year < 500K $5K/year < 1M $10K/year > 1M	Free <100K subs. $25K/year < 250K $50K/year < 500K $75K/year < 1M $100K/year > 1M	Lesser of $0.02 per title sold or 2% of first sale price Free for short (<12min.) content Paid by content replicator or service/content provider
	Cap: Total Enterprise/Entity: $3.5-5M/yr			

- License through 2010, grace period through 2004 (Encoding and Decoding) or 2005 (Use)

[1] In the U.S. this would exempt only the 40 smallest markets out of 210 Nielsen DMA's, the largest of these being Billings, MT or Dothan, AL.

Proposed AVC License Terms, Via Pool

Field	Internet	"OTA Broadcast"	Subscription Content	Packaged Media or Pay-per-view Content
Encoder and/or Decoder Product	$0.25/product $0.0025/product for temporary decoder provided with content Exemption for free 30-day demos of PC software Paid by manufacturer of end-user product Caps: Total Enterprise/Entity: $2.5M/yr except PC OEM (non-retail) software at $4M/yr A PC software OEM can pay product royalties that cover the PC Hardware the software is shipped on.			
Use	Free			Per title sold (Free for free content) $0.005 <30 Min $0.015 <90 Min $0.025 >90 Min $0.0025 for < 30 days, 20 plays temporary content Paid by party that completes financial transaction and authorizes content delivery
Overall Floor	Less than 50K Units and $500K revenue – no fees or reporting			

- Initial Fee of $15K at start of license, "hardship terms" for entities with <$2M annual revenues
- License term 5 years, grace period through 2004

Codec	File Format/Type
H.264	mp4, 3gp, m4v, ts, h264, 264
Mpeg4	mp4, 3gp, mpeg4
H.263 (Count together with MPEG4)	mp4, 3gp, h263, 263
WMV (Talk to Microsoft)	avi, asf, wmv
Motion Jpeg	avi
DivX	avi (DivX3 and before. For DivX4 and after, only if Mpeg4 compliant)
MP3 (Talk to Thomson)	mp3
AAC/AAC+/eAAC+	mp4, 3gp, aac, m4a, m4v
AMR NB/WB/WB+	mp4, 3gp, amr
WMA (Talk to Micrsoft)	avi, asf, wma, wmv
MIDI	mid
JPEG	jpg

Red word – Need to pay license and each license fee detail will be describe in the below.
Codec portion include H.264, Mpeg4(H.263) , AAC , AAC+ , eAAC+ , AMR
File format portion include **mp4**, **3gp**, **m4V** (Generally called **MPEG4 file format**).

Blue word – Need to pay license but Royaltek should have the experience, so I will skip this portion.

常用多媒体技术Patent License :

MPEG4 (Include H.263) – Pay to MPEG LA

Set up fee: No

License fee:

Below 50K units – free

Over 50K units – USD$0.25 (Every year renew)

Yearly Cap: USD$3M (After pay over this amount, no more need to pay license fee)

(AAC) – Pay to Via licensing

Set up fee: No

License fee:

Volume (per channel/quarterly reset)	Consumer Decoder or Encoder Channels	Consumer Codec Channels	Professional Decoder Channels*	Professional Encoder Channels*
Flat Rate	n/a	n/a	$2.00	$20.00
1 to 100,000	$0.50	$1.00	-	*If greater than two channels, total per-product fees not to exceed 3% of end-user price, but not less than $50 per product or more than $2,000 per product.
100,001 to 500,000	$0.37	$0.74	-	
500,001 to 1,000,000	$0.27	$0.54	-	
1,000,001 to 5,000,000	$0.22	$0.44	-	
5,000,001 to 10,000,000	$0.17	$0.34	-	
10,000,001 or more	$0.12	$0.24	-	

(AAC+) - Pay to Via licensing

Set up fee: No

License fee:

Volume (per channel/quarterly reset)	Consumer Decoder or Encoder Channels[1]	Consumer Codec Channels[1]	Professional Decoder Channels	Professional Encoder Channels*
Flat Rate	n/a	n/a	$2.50	$25.00
1 to 100,000	$0.63	$1.25	-	*If greater than two channels, total per-product fees not to exceed 3% of end-user price, but not less than $65 per product or more than $2,000 per product.
100,001 to 500,000	$0.46	$0.92	-	
500,001 to 1,000,000	$0.34	$0.68	-	
1,000,001 to 5,000,000	$0.28	$0.56	-	
5,000,001 to 10,000,000	$0.21	$0.42	-	
10,000,001 or more	$0.15	$0.30	-	

[1]License fees for multi-channel Consumer Products (more than two channels) are calculated at a rate of three channels.

(eAAC+) – Pay to Coding Technology

P.S. Need to talk to Coding Technology for detail because they have not released any information.

(AMR) – Pay to VoiceAge

NB (Narrow Band)

Set up fee: USD$6500

Minimum amount: USD$10000 / year, even no shipping but can be deducted if shipping.

License fee:

Category 4- PDAs(1)			
Cumulative number of Codecs Sold (2)	**Decoder**	**Encoder**	**Codec**
1 to 100,000	$0.25	$0.40	$0.40
100,000+	$0.20	$0.30	$0.30
Maximum Annual Royalty	**$2,000,000 per Licensee including all its Affiliates**		

WB (Wide Band)

Setup fee: USD$6500

Mininum amount: USD$10000/ year, even no shipping but can be deducted if shipping.

License fee:

Annual Volume	Decoder Mono	Encoder Mono	Codec Mono	Decoder Stereo	Encoder Stereo	Codec Stereo
1 to 100,000	$0.40	$0.40	$0.60	$0.75	$0.75	$1.15
100,001 to 500,000	$0.30	$0.30	$0.45	$0.55	$0.55	$0.85
500,001 to 1,000,000	$0.25	$0.25	$0.40	$0.40	$0.40	$0.60
1,000,001 to 5,000,000	$0.20	$0.20	$0.30	$0.30	$0.30	$0.45
5,000,001 to 10,000,000	$0.15	$0.15	$0.25	$0.25	$0.25	$0.40
10,000,001 or more	$0.10	$0.10	$0.15	$0.15	$0.15	$0.25
Maximum Annual Royalty (non-wireless use)	$3,000,000 per Licensee including all its Affiliates					

File format portion

(Mpeg4 file format) (Decoder only) – Pay to MPEG LA

Setup fee: No

License: USD$0.15/per unit, after pay over USD$100K, no more need to pay.

参考文献

1 Stefan Meltzer and Gerald Moser, MPEG-4 audio coding for today's digital media world. Coding Technologies
2 The AAC-ELD family for high quality communication services
3 David JM Robinson, Perceptual model for assessment of coded audio
4 Design of a CELP speech Coder and Study of Complexity vs Quality Trade-offs for Different Codebooks
5 http://old.hssyxx.com/zhsj/kexue-2/zutiweb/zu17/042.htm
7 http://www.docin.com/p-724098762.html
8 http://www.cnblogs.com/tanglei-young/archive/2013/03/18/speech.html

Thanks

掌门集团
ZENMEN

LiveVideoStackCon

聚 音 视　研修不止于形

关注LiveVideoStack公众号

回复 **刘华平** 为讲师评分