

## 多媒体编码及其信息安全应用

*Multimedia Coding and Its Application to  
Information Security*

# 第四讲

## AAC和ACELP编码原理

授课时间：2022年3月14日

## 内 容 提 纲

1. AAC编码原理
2. ACELP编码原理
3. 本讲小结

## 内 容 提 纲

1. AAC编码原理
2. ACELP编码原理
3. 本讲小结

# 1.1 概述

## ○AAC (Advanced Audio Coding) 高级音频编码

- 出现于1997年，基于MPEG-2的音频编码技术。由Fraunhofer IIS、杜比实验室、AT&T、索尼等公司共同开发，目的是取代MP3格式
- 2000年，MPEG-4标准出现后，AAC重新集成了其特性，加入了SBR技术和PS技术，为了区别于传统的MPEG-2 AAC又称为MPEG-4 AAC

扩展名	<b>MPEG/3GPP container</b> <ul style="list-style-type: none"><li>.m4a, .mp4, .3gp</li></ul> <b>Apple container</b> <ul style="list-style-type: none"><li>.m4a, .m4b, .m4p, .m4r, .m4v</li></ul> <b>Raw stream</b> <ul style="list-style-type: none"><li>.aac</li></ul>
互联网媒体类型	audio/aac, audio/aacp, audio/3gpp, audio/3gpp2, audio/mp4, audio/mp4a-latm, audio/mpeg4-generic
开发者	Bell Labs, Fraunhofer Institute, Dolby Labs, Sony and Nokia
初始版本	1997年, 23年前 <sup>[1]</sup>
格式类型	音频压缩, 破坏性数据压缩
专门由	MPEG-4 Part 14, 3GP and 3G2, ISO base media file format and Audio Data Interchange Format (ADIF)
标准	ISO/IEC 13818-7, ISO/IEC 14496-3

1997年制订不兼容MPEG-1的音频标准MPEG-2 AAC, 即MPEG-2 AAC

1999年MPEG-2 AAC增加LTP和PNS工具, 形成MPEG-4 AAC v1

2002年MPEG-4 AAC v1增加了SBR和错误鲁棒性工具, 形成MPEG-4 HE-AAC

2004年MPEG-4 HE-AAC引入PS模块, 提升低码率性能, 形成EAAC+

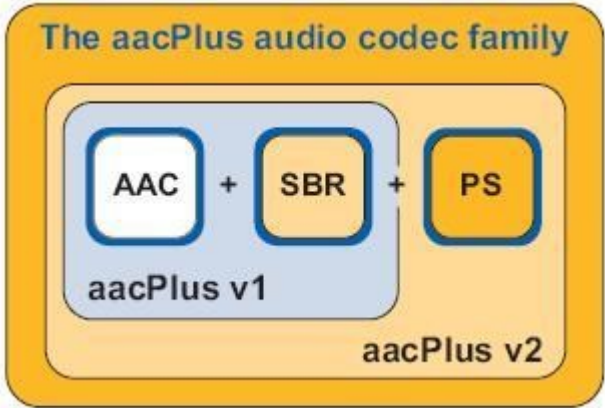
# 1.1 概述

## ○MPEG-2/4 AAC编码规格

MPEG-4 AAC LC (Low Complexity) “低复杂度规格 (LC-AAC)” 是最常用的规格，这种规格在中等码率 (96kbps-192kbps) 的编码效率以及音质方面，都能找到平衡点

层 次	规 格
MPEG - 2AAC LC	低复杂度 (Low Complexity)
MPEG - 2AAC Main	主要层次
MPEG - 2AAC SSR	可变采样率层次 (Scalable Sampling Rate)
MPEG - 4AAC Main	表 2.1: MPEG-4 AAC 编码规格说明
MPEG - 4AAC LC	
MPEG - 4AAC SSR	
MPEG - 4AAC LTP	
MPEG - 4AAC LD	
MPEG - 4AAC HE	

规格名称	规格说明
AAC Main	主规格, MPEG-2 AAC Main+PNS
AAC LC <sup>1</sup>	低复杂度规格 (Low Complexity), MPEG-2 AAC LC+PNS
AAC SSR	可变采样率规格 (Scaleable Sample Rate), MPEG-2 AAC SSR+PNS
AAC LTP	长时期预测规格 (Long Term Prediction), MPEG-2 AAC LC+PNS+LTP
AAC LD	低延迟规格 (Low Delay)
AAC HE	高效率规格 (High Efficiency), HE-AAC 包含 2 个版本: HE-AAC v1 (AAC+) 使用频段复制 (Spectral Band Replication, SBR) 提高频域的压缩效率, 适用于低码率 (64kbps 以下); HE-AAC v2 (Enhanced AAC+) 结合使用 SBR 和参数立体声 (Parametric Stereo, PS) 提高立体声信号的压缩效率, 它进一步降低了对码率的需求 (接近于 50%)。



# 1.1 概述

- AAC 作为一种改进的编码方案，它**相比于MP3 的优点**主要体现在：
  - 更多的**采样率**选择（8kHz 至96kHz）；
  - 更高的**声道数**上限（最多支持48 个主声道，16 个低频增强声道）；
  - 任意的**比特率**（8kbps至576kbps）和可变的帧长度；
  - 更高效率及更单纯的**滤波器组**（AAC 使用纯粹的MDCT，MP3 则使用较复杂的混和滤波器组）；
  - **对平稳的信号**有更高的编码效率（AAC 使用**较长的**1024/960 点区块长度，MP3则为576 点）；
  - **对暂态变化的信号**有更高的编码准确度（AAC 使用**较短的**128/120点区块长度，MP3 则为192 点）；
  - 可选择使用Kaiser 窗函数，以较大的主瓣为代价，**消除频谱泄漏效应**；
  - 对16kHz 声音频率成分的处理更优；
  - 有**额外的模块**如噪声移频（Noise shaping）、反向预测（Backward prediction）、感知噪声替代（PNS）等，可结合这些模块建构出各种不同的编码规格

# 1.1 概述

## ○常见的AAC编码器

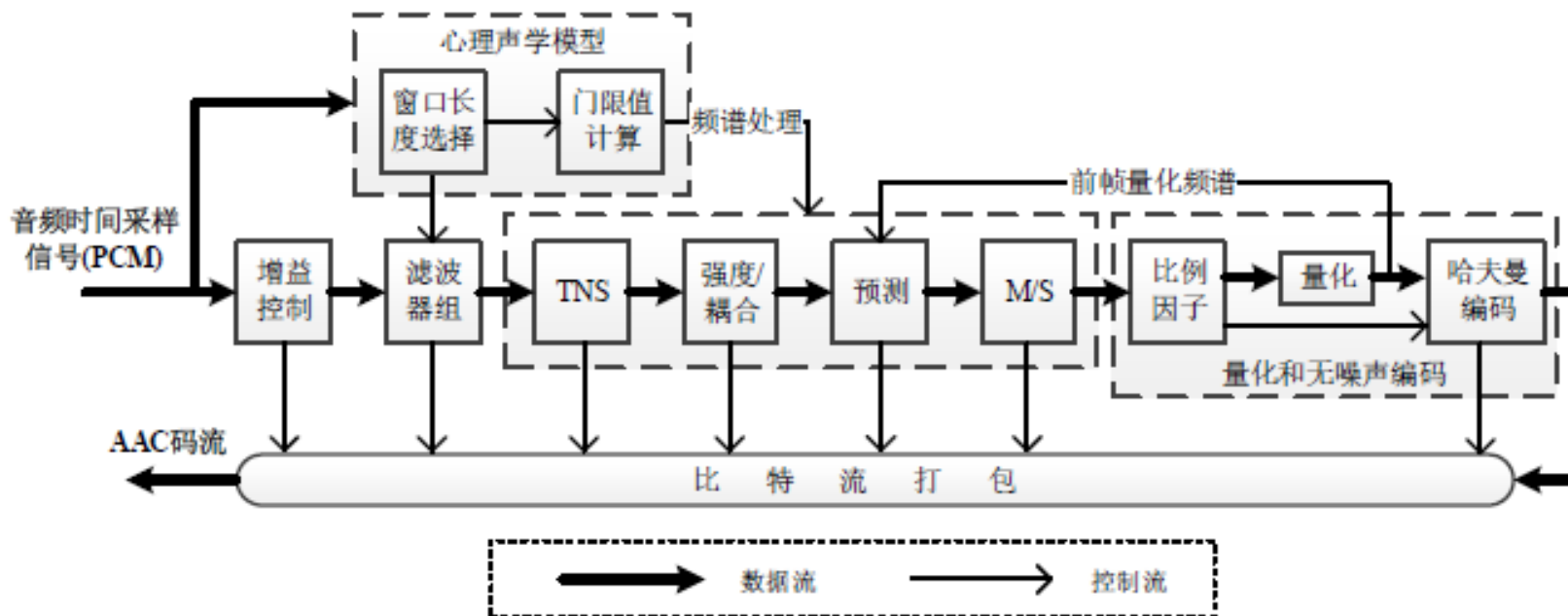
- (1) **FhG**, 它是Fraunhofer IIS 研发的权威编码器;
- (2) **Nero AAC**, 它由Nero 公司免费发布, 同时支持LC-AAC/HE-AAC 规格;
- (3) **QuickTime/iTunes**, 它们是Apple 公司的两款软件都提供了AAC 编码功能, 其编码技术来自Dolby 实验室;
- (4) **FAAC** (Freeware Advanced Audio Coder) , 是一款免费的开源软件, 支持LC/Main/LTP规格;
- (5) **DivX AAC**, 是2009 年DivX 开发的新AAC 编码器, 支持LC/HE/HEv2规格。

○在专利许可方面, **用户不需要许可或付款来传输或分发AAC 格式的音频数据流, 但是AAC 编解码器的所有制造商或开发商都需要专利许可。**

## 1.2 AAC编码总体框架

### ○MPEG-2 AAC编码框架可分为三个部分：

- 第一部分利用心理声学模型控制滤波器组模块、TNS（Temporal Noise Shaping）模块、强度/耦合模块、预测编码模块和M/S（Mid/Side）编码模块
- 第二部分通过码率/失真控制比例因子、量化和哈夫曼编码
- 第三部分是比特流打包





## 1.3 编码流程详解

### ○ (1) 心理声学模型2

### ○ (2) 增益控制(Gain Control)

- 该模块只用于SSR 框架中，它对输入信号进行增益控制，将信号作某个程度的衰减，降低其峰值大小，以减少前回声的发生。它由一个多相正交滤波器组、若干个增益检测器和增益调节器组成。

### ○ (3) 滤波器组(Filterbank)

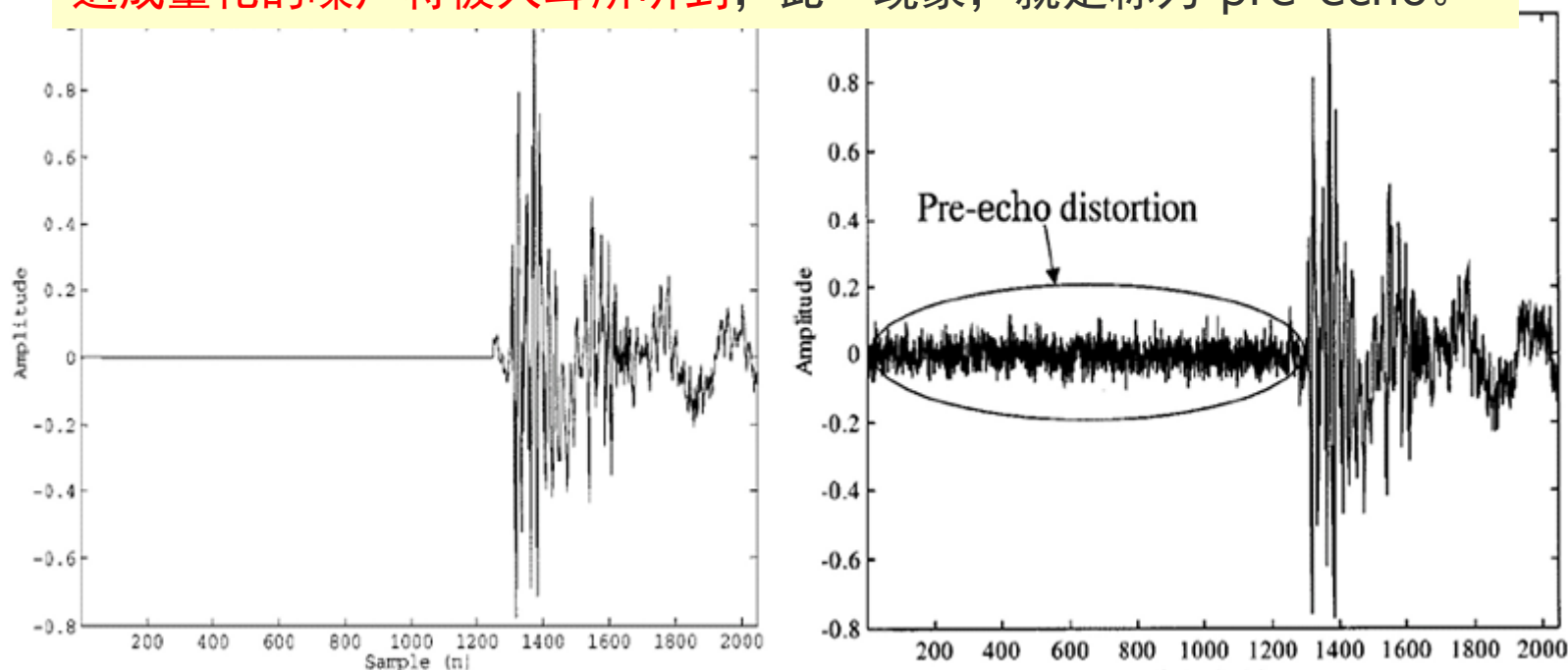
- 使用加窗的MDCT 变换将输入的时域信号转换到频域，窗的形状以及变换块的长度自适应于输入信号。为了克服在块边界处由量化精度的不均匀而产生的不连续噪声，相邻的变换块在时间上做50%重叠，即时域混叠抵消技术。

### ○ (4) 瞬时噪声整形TNS (Temporal Noise Shapping)

- 用于抑制前回声失真，它能够控制每个变换窗口内量化噪声的时域形状，把量化噪声置于实际信号之下。

## 1.3 编码流程详解

从时间域上的遮蔽效应可发现，若一较高的能量是在转换长块的前半部时，其经由量化所产生的噪声，可能被postmasking 所遮蔽，但是若较高的能量是在长块的后半部时，则散布到前半部的噪声将无法被 premasking 所遮蔽，这就是由于对长块而言，其在时间域上的分辨率较低，因此噪声分布范围超过 premasking 的遮蔽范围，而造成量化的噪声将被人耳所听到，此一现象，就是称为 pre-echo。



## 1.3 编码流程详解

### ○ (5) 联合立体声编码

○联合立体声编码(Joint Stereo Coding)是一种**空间编码技术**，其目的是去掉空间的冗余信息。MPEG-2 AAC 系统包含两种空间编码技术：**M/S编码(Middle/Side Encoding)**和**声强/耦合(Intensity/Coupling)**。

#### ○ M/S 编码

- 使用矩阵运算，因此把M/S编码称为矩阵立体声编码(Matrixed Stereo Coding)。
- 由于左右声道具有相关性，M/S 编码不传送左右声道信号，而是使用标准化的“和”信号和“差”信号，前者用于中央M (Middle) 声道，后者用于边S(Side)声道，来代替原来的左、右声道，因此M/S编码也叫做“**和/差编码**”(Sum-difference Coding)。
- 当左右声道频谱相似度大时，使用M/S 编码能减少其中一个声道编码所需的比特数。记L 和R 分别表示初始的左右声道，令 $M = (L + R)/2$ 代替L， $S = (L - R)/2$  代替R。对于每一个无噪声编码带，使用L/R 或M/S 两者中的哪一种方式进行编码，取决于谁耗费的比特数更少。

## 1.3 编码流程详解

### ○声强/耦合(Intensity/Coupling)

- 声强/耦合编码的名称也很多，如声强立体声编码(Intensity Stereo Coding)、声道耦合编码(Channel Coupling Coding)等，它们探索的基本问题是声道间的不相关性(Irrelevance)。
- 人耳听觉系统在听4 kHz以上的信号时，双耳的定位对左右声道的强度差比较敏感，而对相位差不敏感。声强/耦合就利用这一原理，在某个频带以上的各子带使用左声道代表两个声道的联合强度，右声道的谱线置为零，不再参与量化和编码。
- 做法为：将左右声道之频谱值相加，再乘上一个调整因子，最后将新的频谱系数送出。

$$\text{spec}_i[i] = (\text{spec}_l[i] + \text{spec}_r[i]) \times \sqrt{\frac{E_l[\text{sfb}]}{E_r[\text{sfb}]}}$$

## 1.3 编码流程详解

### ○ (6) 频域帧间预测(Predication)

- 在信号较平稳的情况下，利用预测编码可进一步减小信号的冗余度
- 在AAC编码器中预测是**利用前面两帧的频谱来预测当前帧的频谱**，再求预测的残差，然后**对残差进行编码**。预测使用经过量化后重建的频谱信号，具体步骤如下：
  - ①使用前两帧的重建频谱信号预测当前帧的频谱
  - ②将当前频谱与预测频谱相减得到残差信号
  - ③对残差信号量化
  - ④对残差信号反量化，利用预测残差和预测值重建当前帧频谱信号
  - ⑤更新预测器

## 1.3 编码流程详解

### ○ (7) 量化器(Quantizer)

- 上述5个模块都可以达到数据压缩的目的，但是主要压缩工作是在量化与熵编码阶段完成的。
- 对频谱数据进行非均匀量化，使得量化噪声满足心理声学模型的要求，并且量化后频谱的编码比特数必须不超过当前块可分配到的平均比特数（取决于采样率和压缩比特率）。

### ○ (8) 无噪声编码(Noiseless Coding)

- 无噪声动态范围压缩应用在Huffman编码前。在量化系数矩阵前可以放置 $\pm 1$  作为基值偏差，标示频率位置，仅应用于有足够存储空间时。此编码最多有4 种系数输入。
- 编码标准中包括1 个比例系数哈夫曼码书和11 个频谱哈夫曼码书。

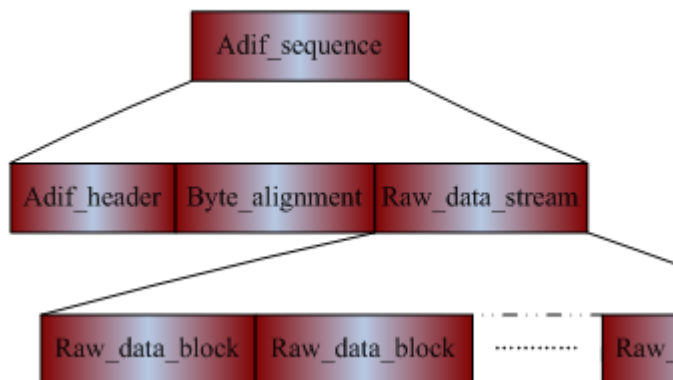
### ○ (9) 多比特转换(Bit stream Multiplexing)

- AAC 可以单纯地传送原始数据，也可利用先进的声音转换逻辑传送。

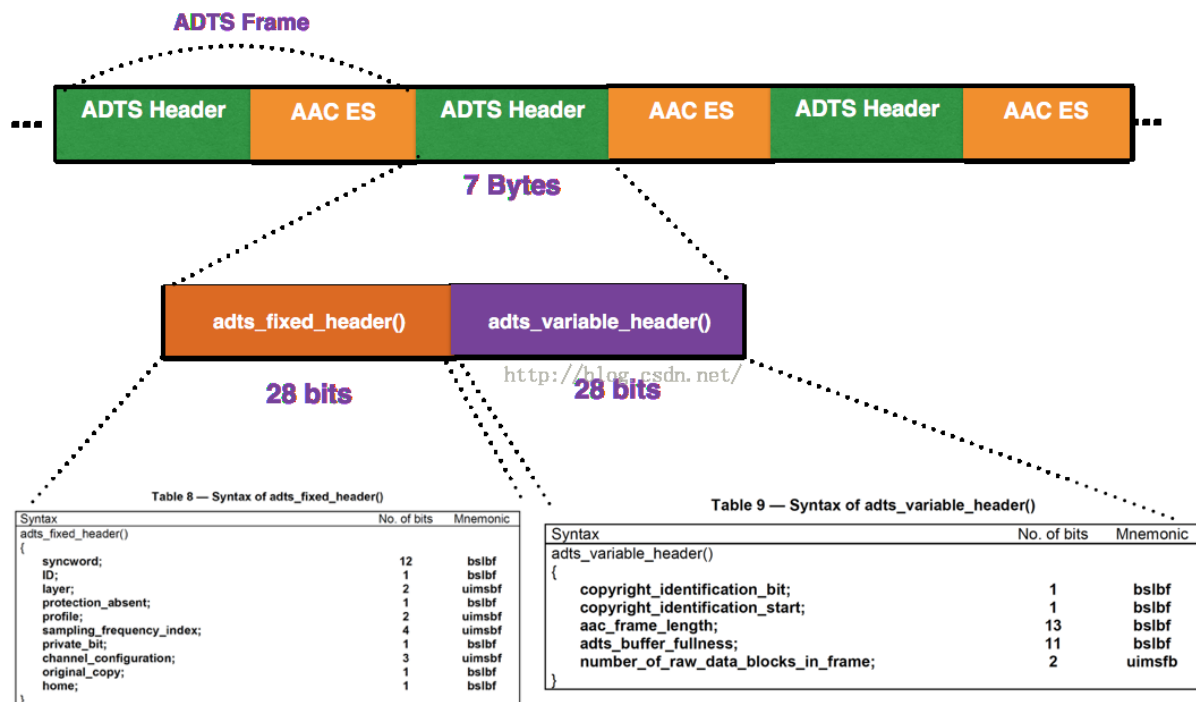
# 1.4 AAC码流结构

- MPEG-2 AAC规定了2种码流格式**ADIF** (Audio Data Interchange Format) 和**ADTS** (Audio Data Transport Stream)，前者用于属于文件格式用于存储，后者属于流格式，用于传输。

ADIF 格式结构



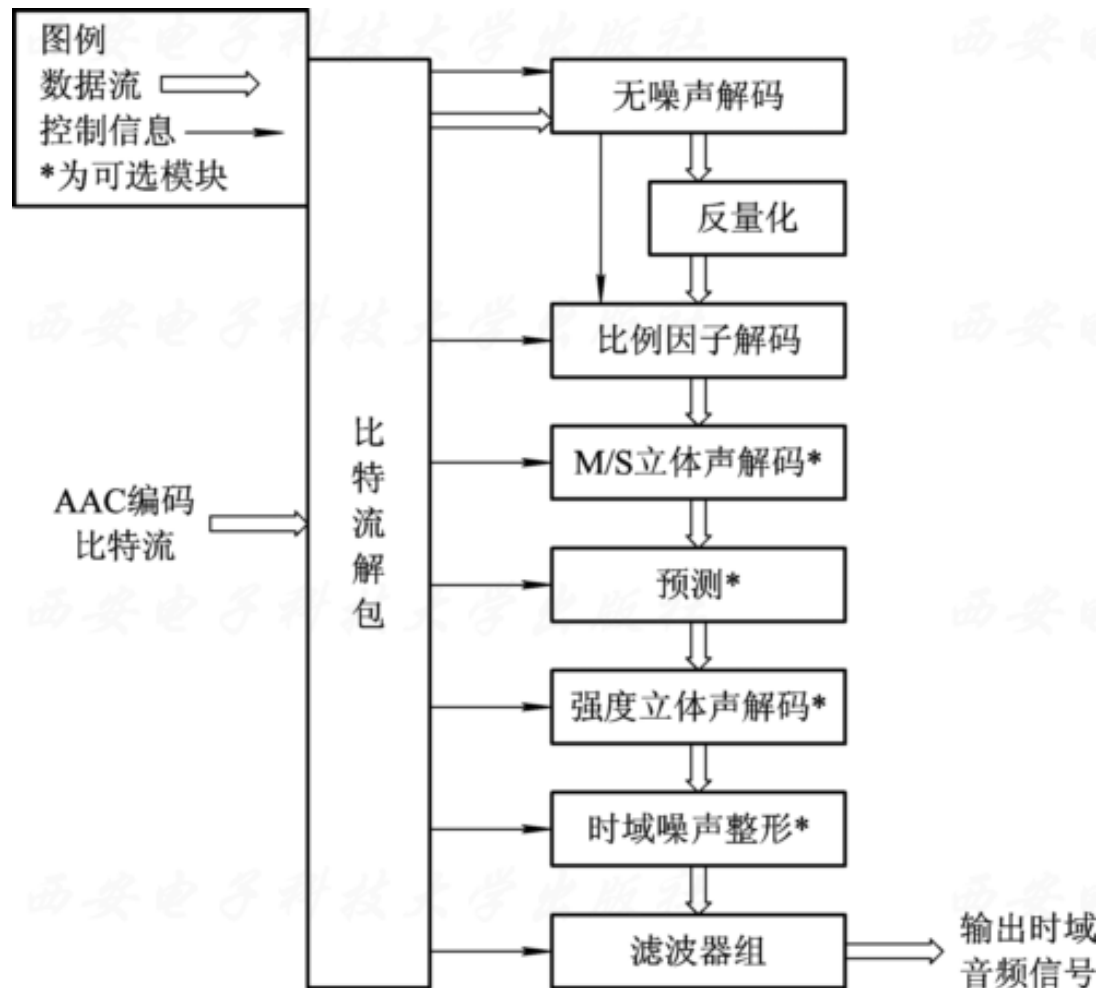
ADTS Sequence (AAC)



注:ADTS Header的长度可能为7字节或9字节. protection\_absent=0时, 9字节. protection\_absent=1时, 7字节.

## 1.5 AAC解码流程

- 解包模块从原始AAC码流中**分离出数据和控制信息**提供给各个相关工具；
- 无噪声解码模块利用霍夫曼码本将频谱的编码数据变换成量化数据；
- 编码端把每帧1024条谱线分成若干区，每个区由一种码本编码，**解码端**所需的码本和分区信息从控制信息中的编码分区数据中获得。





## 内 容 提 纲

1. AAC编码原理
2. ACELP编码原理
3. 本讲小结

## 2.1 语音的产生与感知

### (1) 人类发音系统

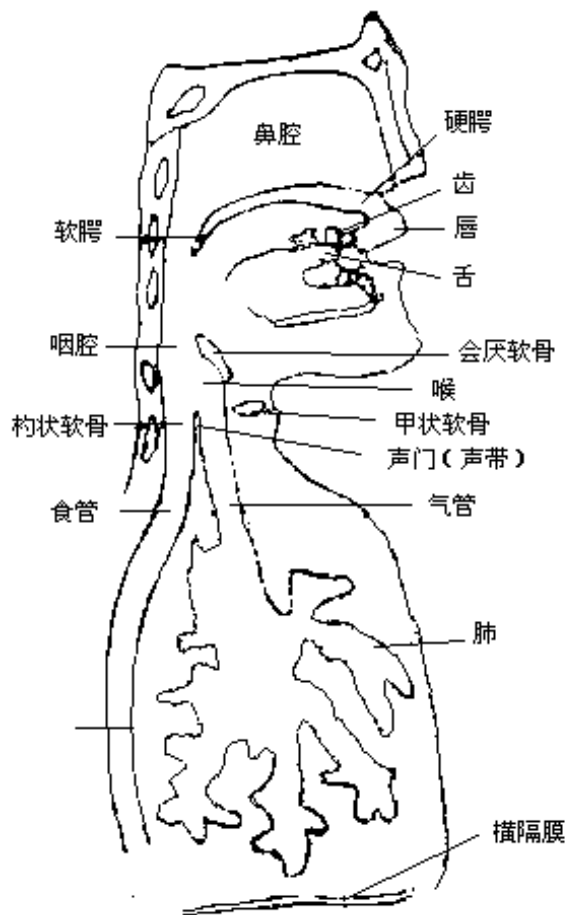
➤**语音**：由人体发音器官在大脑控制下的生理运动产生。

➤**发音器官**：由**肺**和**气管**、**喉**（包括**声带**）、**声道**（**咽腔**、**鼻腔**和**口腔**）三部分组成。

◆**肺和气管**：整个语音系统的**能源**提供者

◆**喉**：主要的声音**生成机构**

◆**声道**：则对生成的声音进行**调制**



## 2.1 语音的产生与感知

- 肺的功能：
  - ◆呼吸功能，进行气体交换
  - ◆提供能量，将压缩空气供给发音器官
- 气管：连接肺和喉，是肺与声道联系的通道
- 喉：由软骨和肌肉组成的复杂系统，含声带（发音器官）
- 声带：
  - ◆是阀门，又是振动部件
  - ◆声带紧绷在喉头的前后壁上，有折叠
  - ◆声带的长度约10 ~ 14 mm
- 声带的声学功能：为语音提供主要的激励源

## 2.1 语音的产生与感知

- 空气作用：使声带**开启/闭合**，形成**脉动气流**（**声门脉冲串**）
- 基音周期（振动周期）：声带每**开启/闭合**一次的时间
- 基音频率（基频）：基音周期的**倒数**

➤ 基频随人性别、年龄而不同

➤ 基频：通常为**50 ~ 450 Hz**

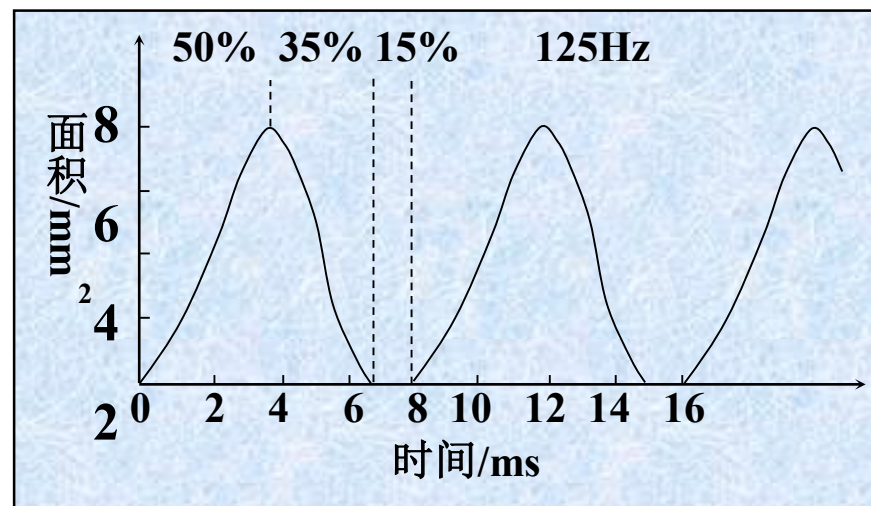
◆ 男性一般为**50 ~ 250 Hz**

◆ 女性一般为**200 ~ 450 Hz**

➤ 老年男性**偏低**，小孩和青年女性**偏高**

➤ 基频高则**音调高**，基频低则**音调低**

➤ 基频与声带的**大小、厚薄、松紧程度**以及声门上下之间的**气压差**等有关



声带开启的面积与时间的关系曲线

## 2.1 语音的产生与感知

➤ **声道**：从声门至口唇的所有**发音器官**  
包括**咽腔**、**口腔**和**鼻腔**

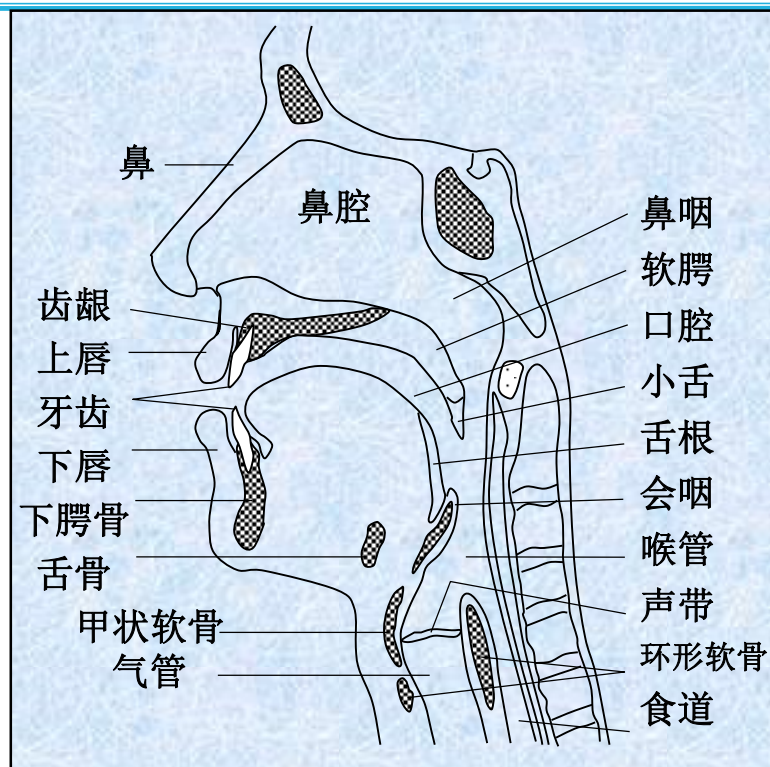
➤ **成男声道**：长**17 cm**/面积**20 cm<sup>2</sup>**

➤ **声道**可看成**非均匀截面**的**声管**，  
是**时间函数**。

➤ **口腔各器官协同动作**，空气流通过时  
形成**不同阻碍**，并产生**振颤**，发出  
**不同声音**。

➤ **口腔**是**声道**最重要的部分，其**大小**  
和**形状**可以通过调整舌、唇、齿和腭来改变。

➤ **咽腔**与**口腔**使**声道**的**形状变化增多**，能发出**较多的声音**。



声道纵剖面图



## 2.1 语音的产生与感知

➤在发音过程中，肺部与相连的肌肉相当于声道系统的激励源

➤浊音：声带处于**收紧状态**时，  
气流使**声带振动**产生的声音

➤清音：声带处于**放松状态**时，  
不伴有声带振动产生的音

➤两种清音：**摩擦音**，**爆破音**

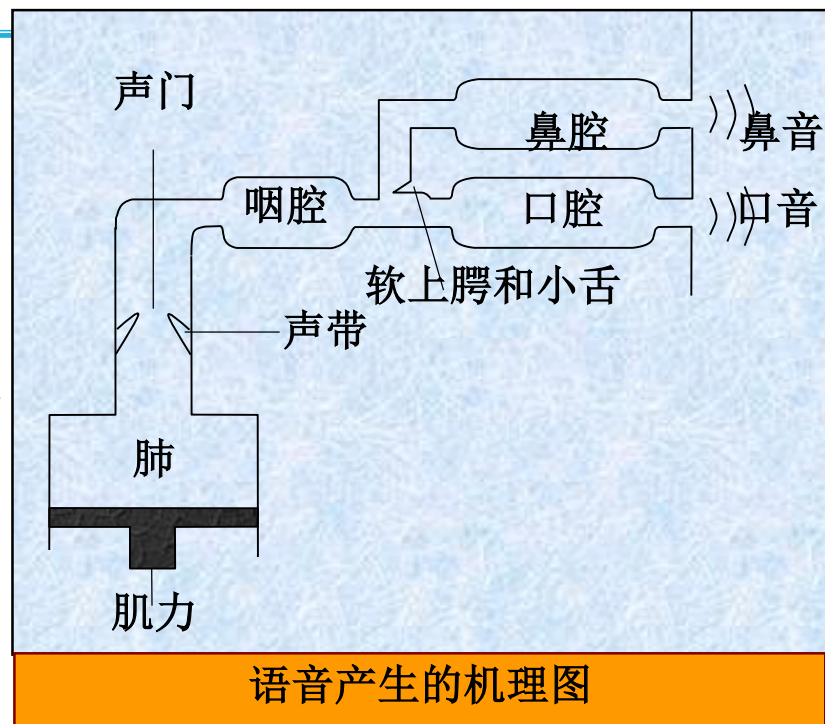
➤摩擦音：**舌**在声道的某处形成**狭窄部位（收紧点）**  
气流经过时产生**湍流**形成**噪声型**的声音

➤爆破音：松懈声带，用舌和嘴唇**关闭声道**，暂时**阻止气流**。

气压升高，**突然放开**舌与嘴唇，气流释放产生**短暂冲音**

◆不同的**声道收紧点**和**声道形状**，形成不同的**摩擦音**

◆不同的**声道闭紧点**和**声道形状**，形成不同的**爆破音**





# 2.1 语音的产生与感知

## (2) 人类听觉系统

### 1) 耳的结构

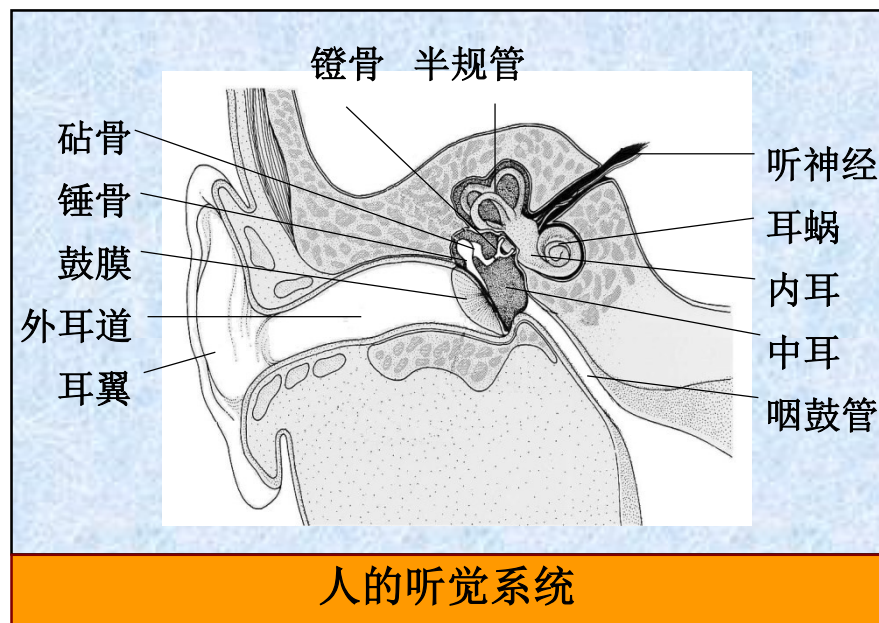
➤ 人的听觉系统。

◆ 组成：外耳、中耳和内耳。

◆ 外耳和中耳有导音的作用，  
合称为导音系；

◆ 内耳有感音作用，称感音器。

其感音作用起始于蜗神经的终端（螺旋器），故内耳的淋巴系统也属于导音系。



## 2.1 语音的产生与感知

➤外耳：由耳翼、外耳道和鼓膜组成。

◆耳翼：有保护耳孔和定向作用。

◆外耳道：是一条耳管，声音沿其传至鼓膜。

有许多共振频率，封闭时最低共振频率约为3060 Hz。

共振效应会使声音得到10 dB左右的放大。

◆鼓膜：位于外耳道内端的韧性锥形结构，  
声音的振动通过鼓膜传到内耳。

日常谈话中，鼓膜位移约为 $10^{-8}$  cm。

➤外耳的作用：有对声源定位和声放大。

➤头部的衍射效应也会增大鼓膜处的声压，

➤系统总放大：20 dB左右。



## 2.1 语音的产生与感知

### ➤中耳的结构。

➤中耳：为充气腔体，由鼓膜将其与外耳隔离，通过圆形窗和前庭窗两个小孔与内耳相通。通过咽鼓管与外界相连，以平衡气压，保护鼓膜。

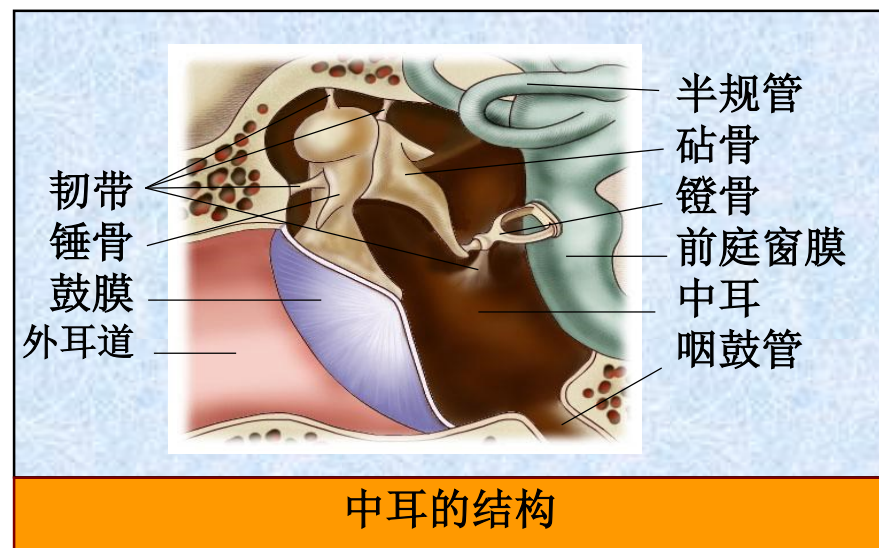
➤听骨链：由锤骨、砧骨和镫骨三块听小骨组成，由韧带悬挂在中耳的腔体内。

◆听骨链将振动传到内耳并放大，起到杠杆的作用。

◆放大30倍左右。

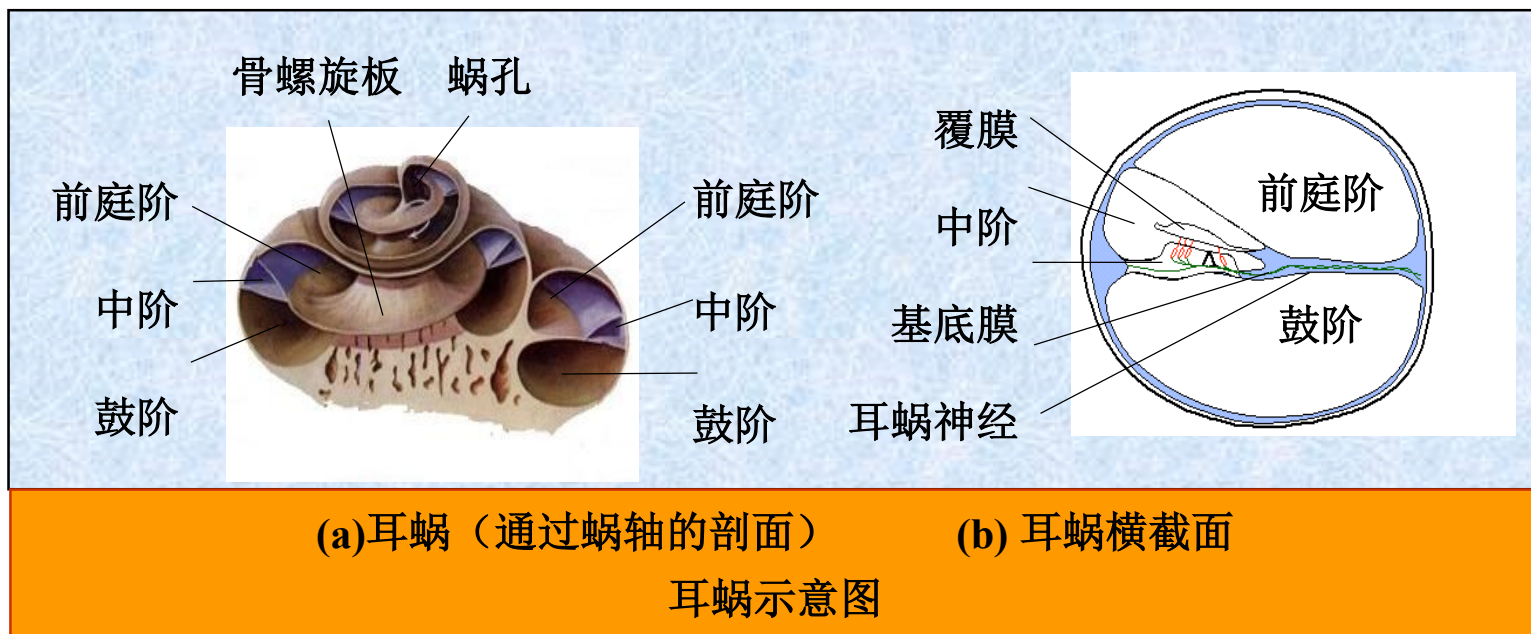
➤听小骨在不同声强范围内实现声音的线性或非线性传递。

➤中耳的作用：通过听小骨进行声阻抗变换，放大声压；保护内耳。



## 2.1 语音的产生与感知

- **内耳(迷路)**: 在颅骨腔内, 由**半规管**、**前庭窗**和**耳蜗**组成。
- **半规管**和**前庭窗**属于**本体感受器**, 与机体的**平衡机能**有关。
- ◆ 半规管内的**感受器**能感受**旋转变速运动**的刺激,
- ◆ 前庭窗内的**感受器**能感受**静止**的位置和**直线变速运动**。
- **耳蜗**: 由**鼓阶**、**中阶**和**前庭阶**三个分隔的部分组成。
- ◆ **听觉接受器**, 把声音经**机械变换**产生**神经发放信号**。



## 2.1 语音的产生与感知

◆中阶的底膜称为**基底膜**,

◆**基底膜**之上是**柯蒂氏器官**, 由**耳蜗覆膜**、**外毛细胞** (共3列, 约**20000**个) 以及**内毛细胞** (共**1**列, 约**3500**个) 构成。

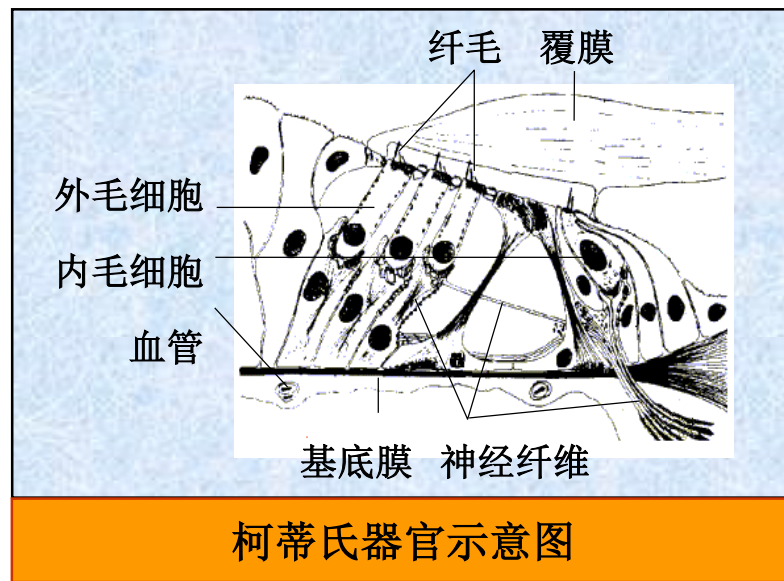
◆**柯蒂氏器官**: 是一个**传感装置**。

◆**毛细胞**上部的**微绒毛**感受耳蜗内**流体速度**的变化, 从而引起**毛细胞膜**两边**电位**的变化, 可造成**听觉神经**的**发放**或**抑制**。

➤**内耳的作用**: 感受声音。

### 听觉产生过程

声波 → 骨膜振动 → 听小骨传递  
→ 耳蜗基底膜振动 → 产生神经脉冲



柯蒂氏器官示意图

## 2.1 语音的产生与感知

➤说话过程可分五个阶段：想说阶段、说出阶段、传送阶段、接收阶段、理解阶段

### 1、想说阶段：（与大脑中枢的活动有关）

- ◆大脑决策产生说话动机

- ◆讲话神经中枢选单词、短语，按规则组合表达内容和情感

### 2、说出阶段：（与发音器官的活动有关）

- ◆大脑中枢决策，向发音器官发指令，使舌、唇、颚、声带、肺等协调动作，发出声音

- ◆大脑也发指令给其它器官，产生各种动作来配合

- ◆根据听觉系统接收的反馈语音信息，来帮助修改语音。

### 3、传送阶段：（传送声波信息的物理过程）

- ◆声波以空气为媒介传送到听者的耳中

## 2.1 语音的产生与感知

### 4、接收阶段：（与听觉系统活动有关）

- ◆外耳收集声波信息，经中耳放大，传到内耳
- ◆经内耳基底膜振动，激发柯蒂氏器官内的神经元产生脉冲
- ◆将信息以脉冲的形式传送给大脑

### 5、理解阶段：（至今尚未完全了解，机理不很清楚）

- ◆听觉神经中枢收到脉冲信息，辨认话者及所说信息

➤说话过程相当复杂，有心理、生理、物理及个人和社会因素

➤个人因素：话者口音、用词造句特色

听者的听力、理解力。

➤社会因素：话者、听者的社会基础、环境等

➤语言要素：分语言的语素、词、短语和句子等不同层次  
及词法、句法、文脉等语法和语义内容等。

➤句法的最小单位是单词，词法的最小单位是音节。

➤不同语言有不同语言规则



## 2.1 语音的产生与感知

- 语音的产生和理解：与神经系统和大脑有关，是高级活动。
- 搞清大脑产生和理解语音的机理，对语音技术有极重要意义。

特别是对语音合成与语音识别两个分支。

### ➤例，语音合成：

- ◆目前，按规则合成只能从寻找各种语言的规则入手，尽可能得出较好的人工语言。
- ◆如果发音时大脑智能活动的机理之迷揭开，就可以获得高度自然的语音合成。

### ➤例，语音识别：

- ◆目前，只能从语音信号出发，用“隐过程”（如隐马尔可夫模型）来模拟神经系统的听觉过程，不是按人的听觉过程建立处理模型。不能达到理想的识别和理解效果。
- ◆这种方法与大脑用的方法不一致。

## 2.1 语音的产生与感知

### (3) 听觉感知特性

➤ 听觉系统的两个重要特性：耳蜗对于声信号的时频分析特性；听觉掩蔽效应。

➤ 耳蜗的时频分析特性：

◆ 声音使镫骨运动，使耳蜗内流体压强变化，引起行波沿基底膜的传播。

◆ 声频不同，产生的行波不同，峰值出现在基底膜的位置不同。

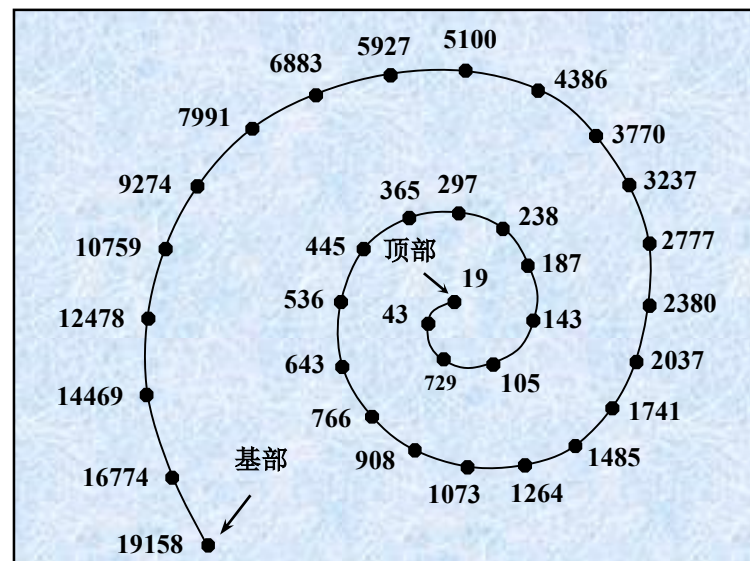
为对数型分布。

◆ 频率低，峰值出现在基底膜的顶附近；

◆ 频率高，峰值出现在基底膜的底附近。

◆ 振动强度增加，基底膜运动幅度加大，

➤ 耳蜗：有频谱分析作用。



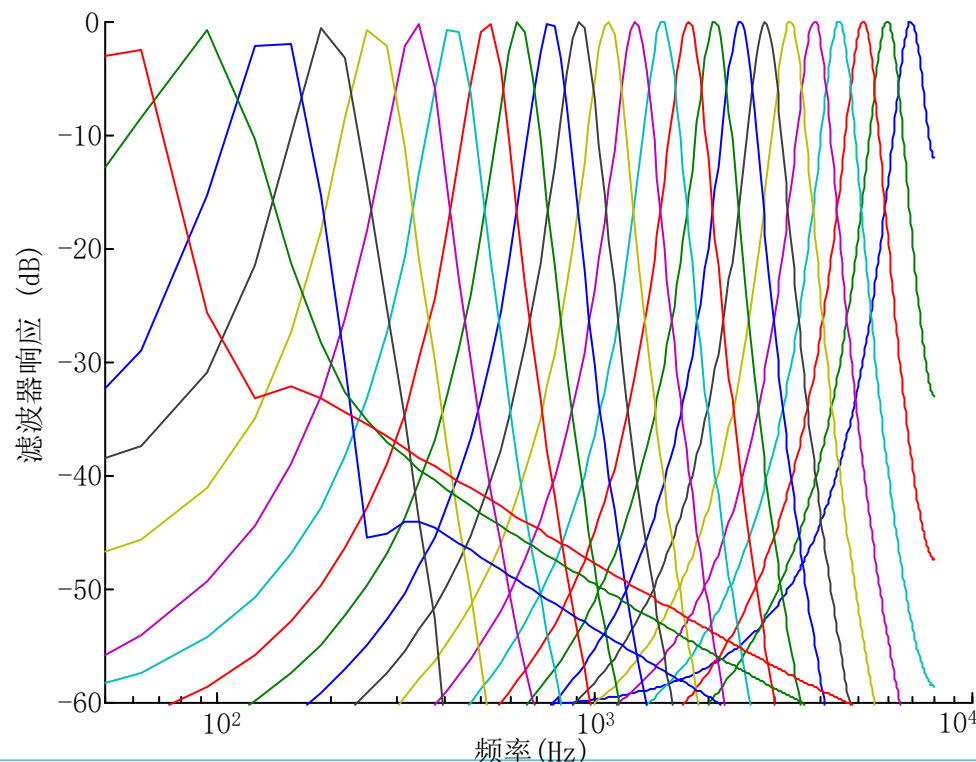
基底膜的频率响应分布图

## 2.1 语音的产生与感知

➤ **Gammatone滤波器组**：听觉选择性通常由一组基于等效矩形带宽刻度的Gammatone滤波器实现，每个滤波器模拟基底膜不同部位最大位移处的响应。其冲激响应函数可表示为：

$$g_m(t) = t^{n-1} e^{-2\pi B_m t} \cos(2\pi f_m t + \phi_m) \mu(t), 1 \leq m \leq N$$

➤ **四阶**的Gammatone滤波器能够很好地模拟基底膜的滤波特性。





## 2.1 语音的产生与感知

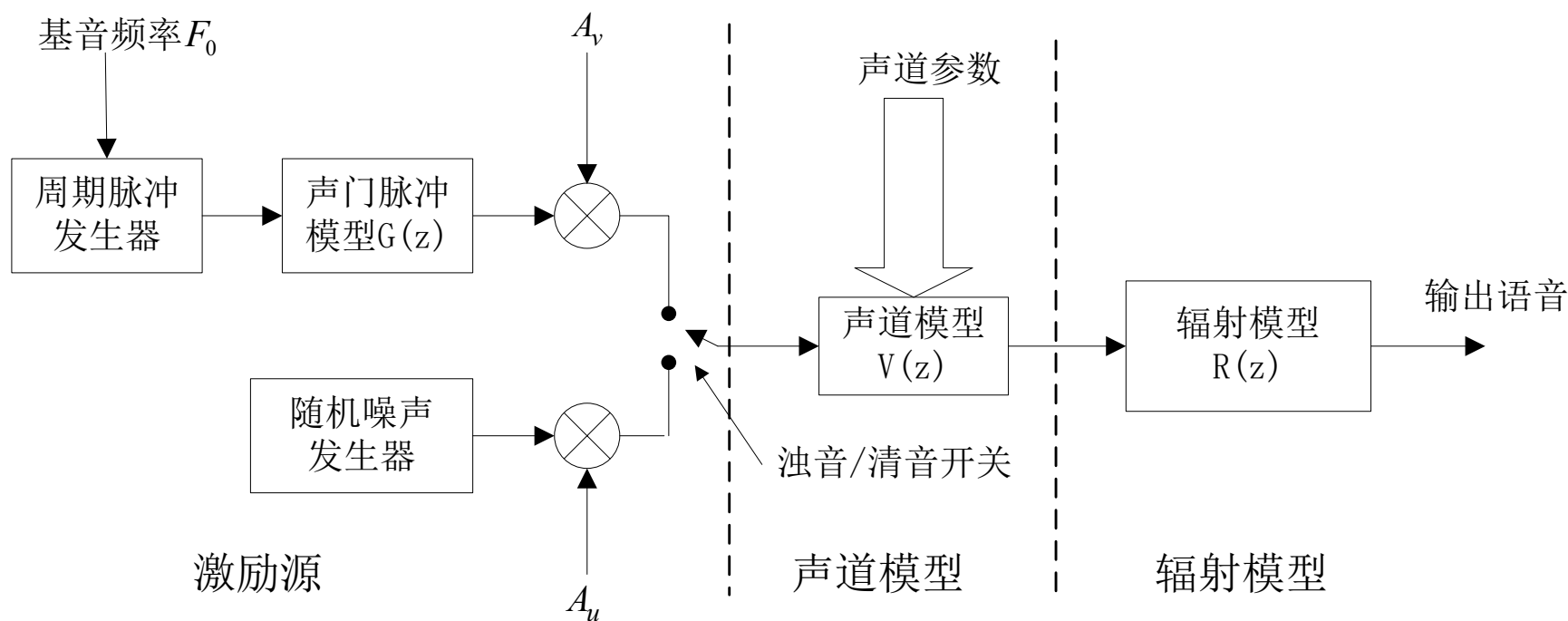
### ➤基底膜上的绒毛细胞的特性:

- ◆振动使基底膜和耳蜗覆膜之间的毛细胞上的绒毛发生弯曲。  
绒毛弯向一边，引起毛细胞的去极化，加强传入神经的作用；  
绒毛弯向另一边，引起毛细胞的超极化，导致抑制效应。
- ◆基底膜上不同部位的毛细胞具有不同的电学和力学特性。  
在基部，基底膜窄而劲度强，毛细胞及其绒毛短而有劲度；  
在顶部，基底膜宽而柔和，毛细胞及其绒毛较长而柔和。
- ◆这种差异是基底膜有频率选择性和对数分布性的重要因素。
- 人的听觉范围：20 Hz ~ 20 kHz，0 ~ 130 dB的声音信号。
- ◆听觉范围外的信号分量可忽略掉，以节省处理成本。
- ◆人耳的感觉不是绝对的，随着信号特性的不同而不同。

## 2.2 语音产生的数学模型

语音生成系统分成三个部分：

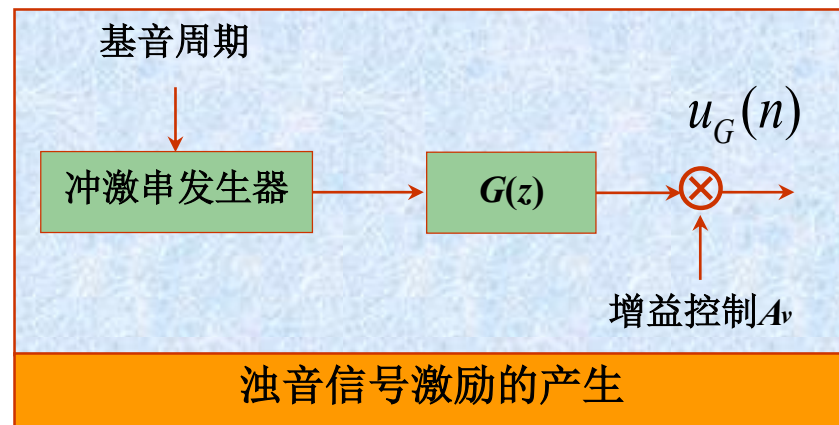
- 1) 在声门（声带）以下，称为“声门子系统”，它负责产生激励振动，是“**激励系统**”；
- 2) 从声门到嘴唇的呼气通道是声道，是“**声道系统**”；
- 3) 语音从嘴唇辐射出去，所以嘴唇以外是“**辐射系统**”。



## 2.2 语音产生的数学模型——激励模型

- ▶ 语音分成清音和浊音，清音由随机噪声激励产生  
浊音由准周期脉冲串激励产生，其周期称为基音周期。
- ▶ 浊音情况下，激励信号的产生示意图如下。
- ◆ 冲激串发生器输出的单位冲激序列（冲激间隔为基音周期）。
- ◆ 线性激励系统函数为  $G(z)$ ，经幅度控制后输出为浊音激励。
- ◆  $G(z)$  的反变换  $g(n)$  可以用 Rosenberg 函数近似表示：

$$g(n) = \begin{cases} 0.5(1 - \cos \frac{\pi n}{N}); & 0 \leq n \leq N_1 \\ \cos \frac{\pi(n - N_1)}{2N_2}; & N_1 \leq n \leq N_1 + N_2 \\ 0; & \text{其它} \end{cases}$$



式中， $N_1$ ——斜三角波上升部分的时间，约占基音周期的50%；  
 $N_2$ ——斜三角波下降部分的时间，约占基音周期的35%。

## 2.2 语音产生的数学模型——激励模型

➤斜三角波的占时比例关系与声带开启面积的与时间关系对应。

➤图示为单斜三角波波形及频谱。

◆是低通滤波器。

◆其 $z$ 变换的全极点（二极点）模型：

$$G(z) = \frac{1}{(1 - e^{-CT} z^{-1})^2}$$

式中， $C$ 是一个常数。

➤斜三角波串可看成加权单位脉冲激励单斜三角波模型的结果。

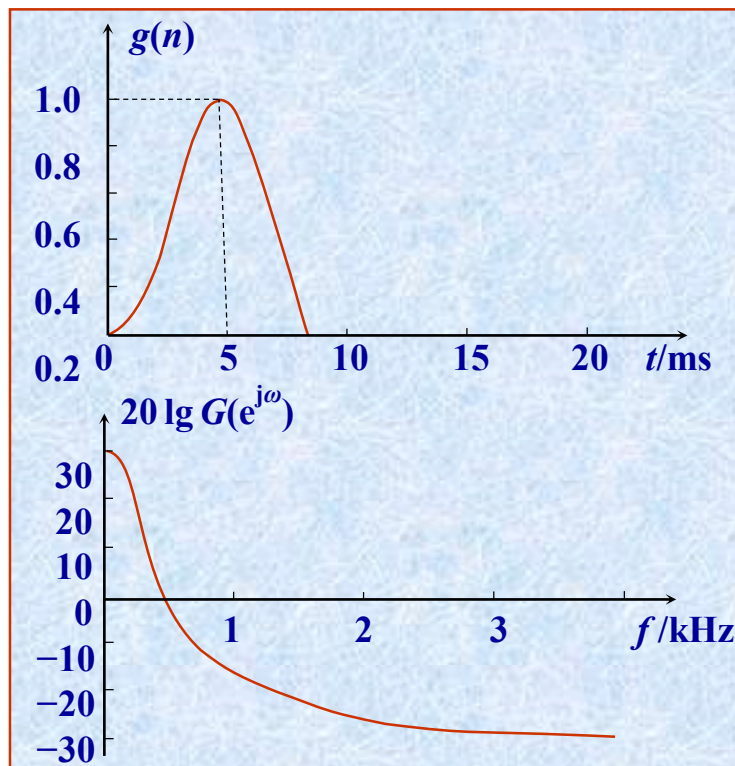
◆ $A_v$ 是单位脉冲串的幅度因子；

◆单位脉冲串的 $z$ 变换为：

$$E(z) = \frac{A_v}{1 - z^{-1}}$$

➤完整的激励模型为：

$$U(z) = \frac{A_v}{1 - z^{-1}} \frac{1}{(1 - e^{-CT} z^{-1})^2}$$

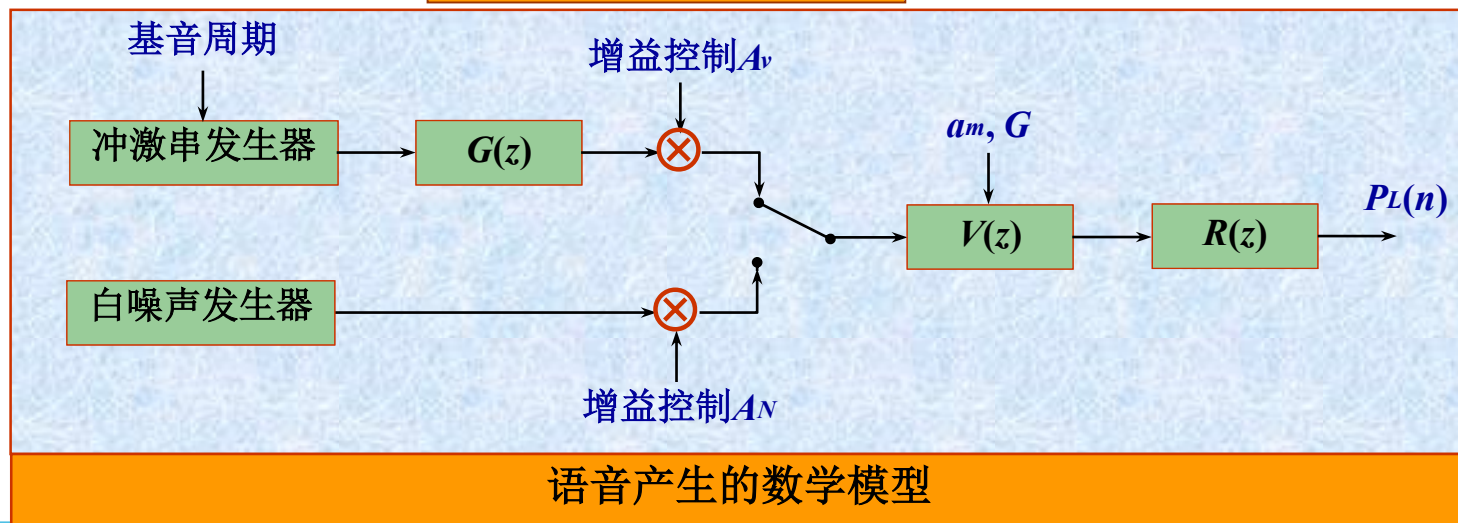


单个斜三角波及其频谱

## 2.2 语音产生的数学模型——激励模型

- 清音情况下，发塞音或摩擦音，声道被阻形成湍流。
  - ◆ 激励可模拟成随机白噪声，
  - ◆ 用均值为0、方差为1，时间或/和幅值为白色分布的序列。
- 图示为考虑所有的激励因素，语音产生的数字模型。
  - ◆ 特点：二元激励，浊音、清音激励交替进行。
  - ◆ 声道可以用多种滤波器来模拟，
  - ◆ 通常，把辐射和声道等因素全部结合，表示为全极点函数：

$$H(z) = V(z)R(z)$$



## 2.2 语音产生的数学模型——激励模型

### ➤结论:

- ◆**优点:** 该模型对大多数语音是一个好模型，  
能合成出**较满意的语音**，是分析语音最重要的基础。
- ◆**缺点:** 二元激励模型有**局限性**。  
模型建立“**短时**”平衡为前提，**不完全符合实际**；  
理论上**鼻音**和**擦音**需有**零点**，**浊擦音**不是简单的  
浊音和清音的叠加，因此不能用该模型模拟。

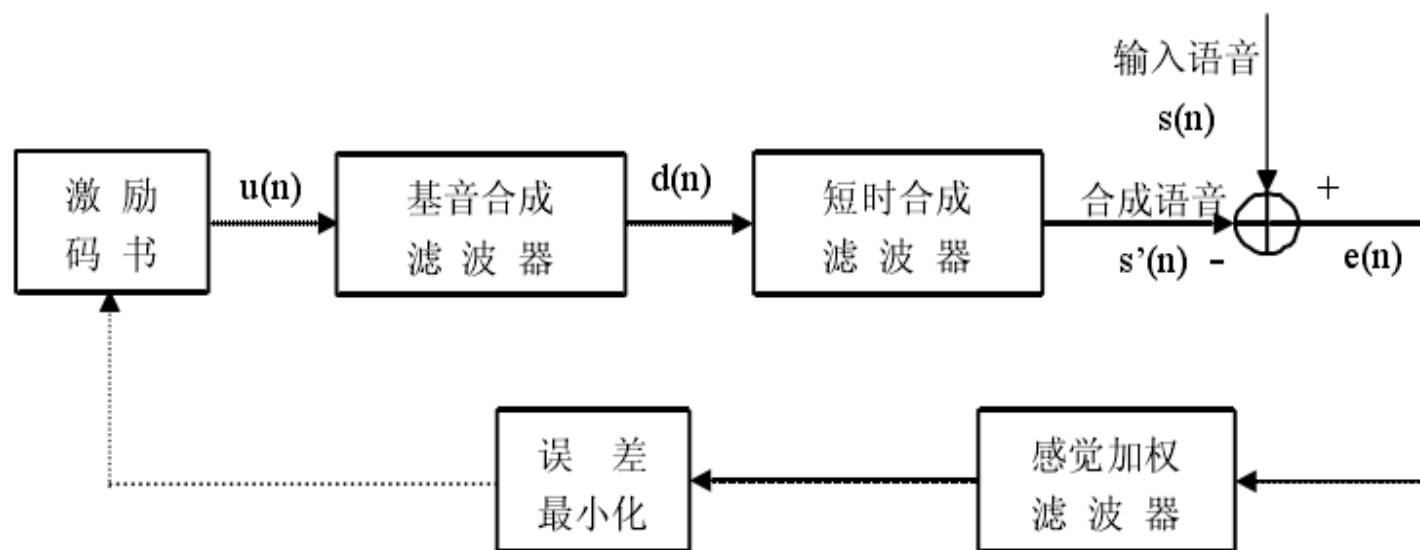
## 2.3 码激励线性预测语音编码（CELP）算法

- **码激励线性预测CELP**（Code Exited linear prediction）编码技术是一种有效的**中低速率**语音压缩编码技术，它**以码本作为激励源**，具有速率低、合成语音质量高、抗噪性强及多次音频转接性能良好等优点。在**4.8~16Kb/s**速率上得到广泛的应用。
- 采用CELP算法的低速率语音压缩编码很多，如北美的IS-95、IS-96，日本的JDC半速率标准，ITU的G.723、G.728和**G.729**采用的都是CELP算法。

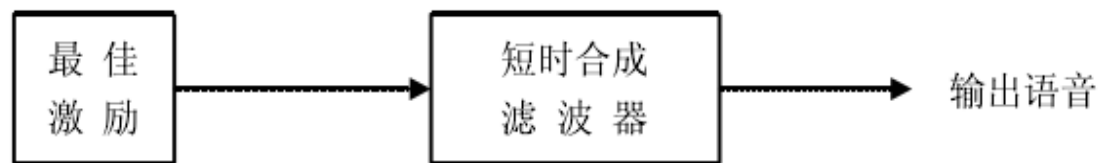


## 2.3 码激励线性预测语音编码（CELP）算法

### ○基于合成分析过程（Analysis-by-Synthesis, **AbS**）的 CELP语音编解码模型



(a) 编码器



(b) 解码器

以感觉加权均方误差最小为判决准则, 采用闭环搜索的方法—**合成分析法** (AbS) 来选取最佳激励矢量, 以得到最佳逼近原始语音的效果。



## 2.3 码激励线性预测语音编码（CELP）算法

- 为了获得与原始语音信号的最佳匹配，CELP编码模型需要频繁地修正**时变滤波器参数和激励参数**。系统的分析过程是**按帧分序**进行的，即首先确定时变滤波器的参数，然后确定固定激励参数。分析帧的长度和修正速率决定了编码方案的比特率。
- 此编码方案的**基本步骤**如下：
  - 1) **初始化**短时合成滤波器和基音合成滤波器历史（通常初始化为零值或低电平随机噪声）。
  - 2) 缓存一帧语音信号，然后对这帧语音信号进行**线性预测分析**，确定一组LPA系数。
  - 3) 利用已经确定的LPA系数和线性预测误差滤波器 $A(z)$ ，**计算未量化的残差信号**。
  - 4) 为了有效的确定激励参数，**将LPA帧分为几个子帧**。

## 2.3 码激励线性预测语音编码（CELP）算法

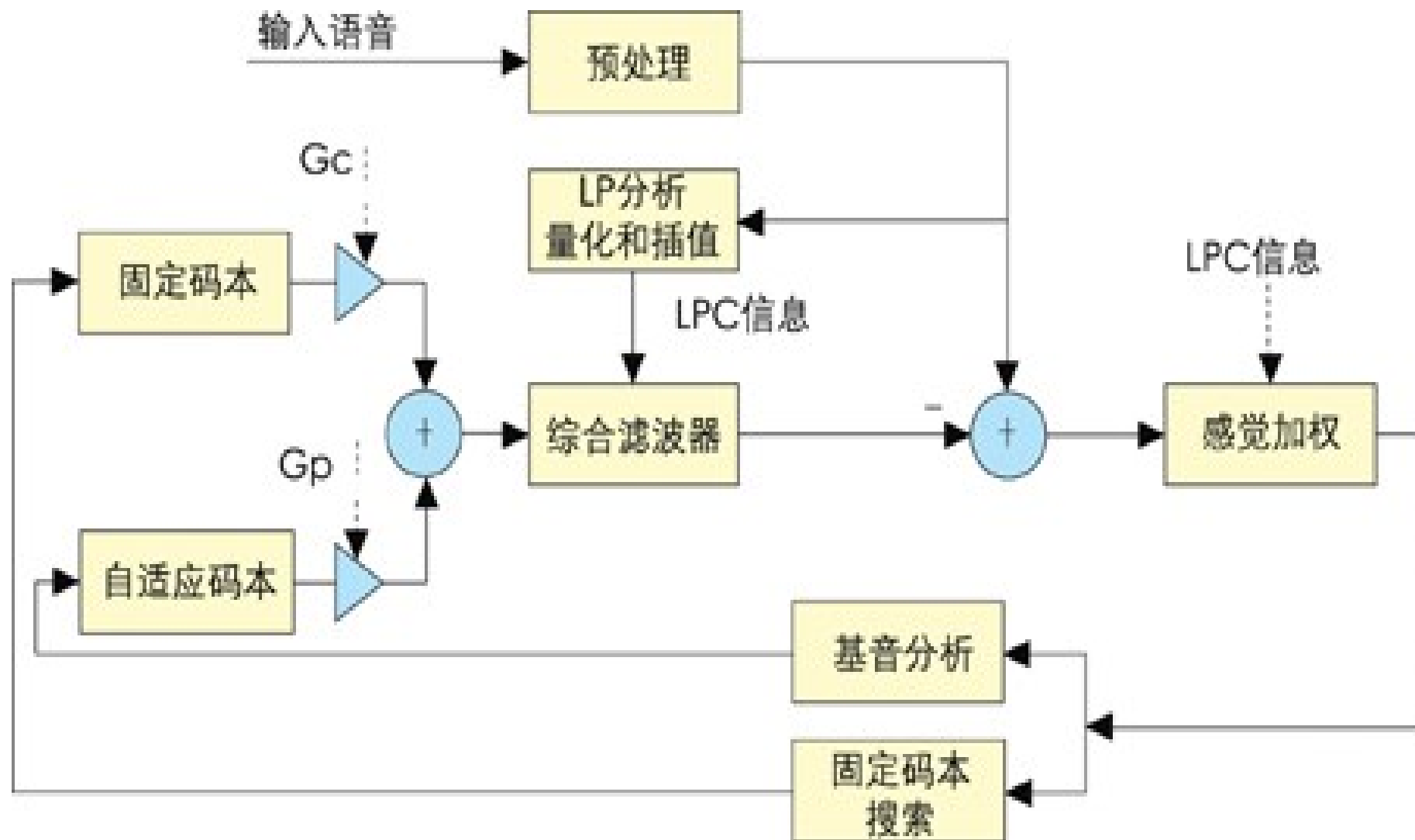
- 5) 对于每个子帧，首先用开环方法或闭环方法**确定基音预测参数**。一旦确定了基音预测器参数，则将基音合成滤波器和短时合成滤波器组合在一起形成一个级联的滤波器。其次**用激励码书中的某一矢量去激励这个级联滤波器，得到合成语音**，再计算合成语音和原始语音之间的误差，经感知特性进行加权后，选取均方误差最小的激励矢量作为**最佳矢量**。
- 6) 借助于滤波器的初始记忆内容，将最佳激励信号通过级联滤波器产生合成语音。
- 7) 对每个子帧重复第（2）至第（6）步。
- 需要注意的是，在上述CELP语音编码方案中，基音预测器对产生高质量的浊音语音发挥着十分重要的作用，这是因为**浊音语音被表征为由基音周期分隔的样点间大量相关的准周期信号**。而短时预测器在很大程度上排除了近样点的相关，得到的短时预测残差信号是一主要由基音尖峰信号构成的**低密度准周期信号**，如果没有基音预测器，只靠随机码书就不能有效地产生类似于短时预测残差信号的准周期激励信号。

## 2.4 ITU-T G.729——CS-ACELP算法

- G.729是由美国、法国、日本和加拿大的几家著名国际电信实体联合开发的。它需要符合一些严格的要求，比如**在良好的信道条件下要达到长话质量，在有随机比特误码、发生帧丢失和多次转接等情况下要有很好的稳健性等**。这种语音压缩算法可以应用在很广泛的领域中，包括IP电话、无线通信、数字卫星系统和数字专用线路
- G.729标准采用“**共轭结构代数码本激励线性预测编码算法**”(Conjugate-Structure Algebraic CELP, **CS-ACELP**)算法。这种算法综合了波形编码和参数编码的优点，以自适应预测编码技术为基础，采用了**矢量量化、合成分析和感觉加权**等技术

## 2.4 ITU-T G.729——CS-ACELP算法

### CS-ACELP算法原理图

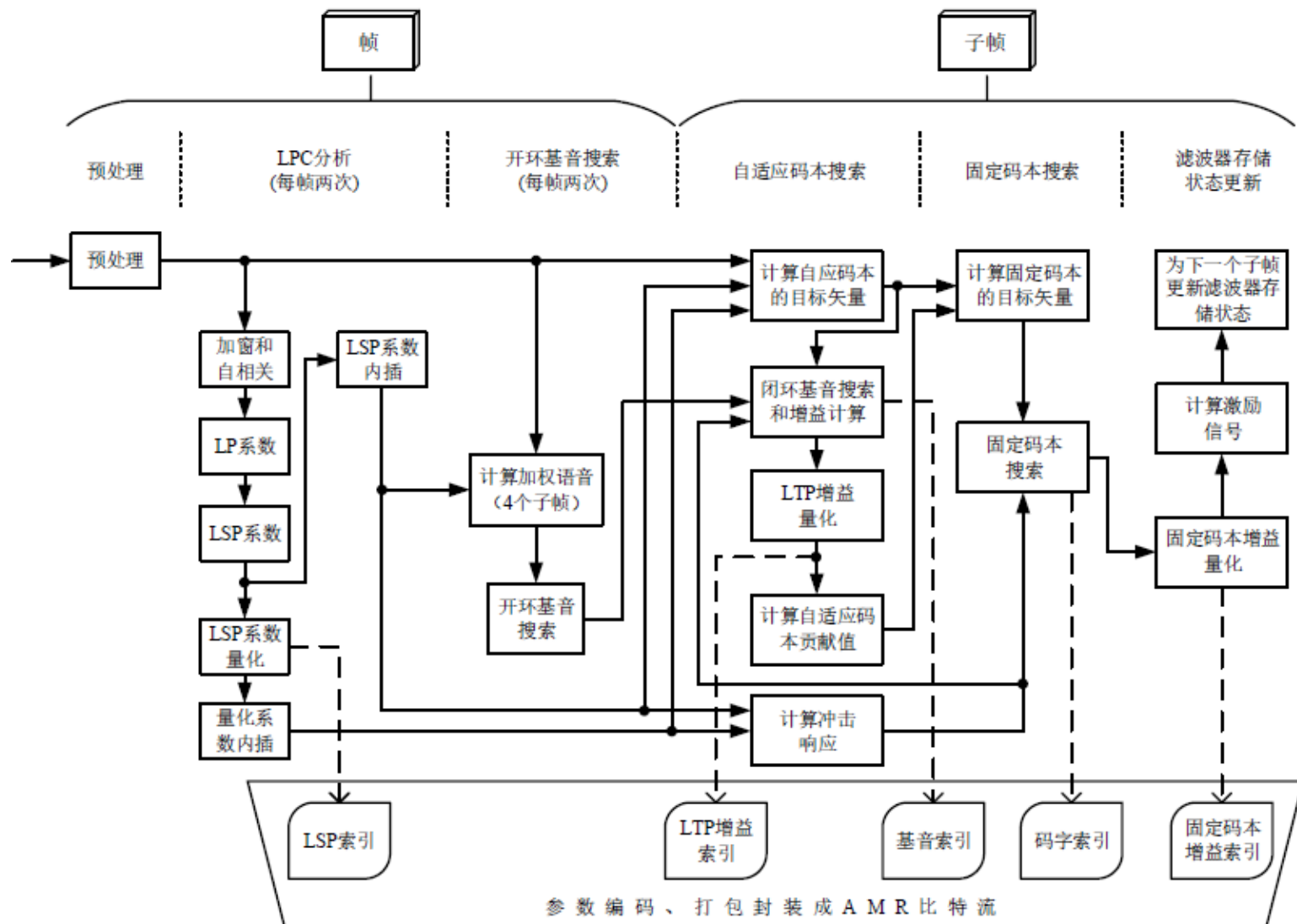


## 2.4 ITU-T G.729——CS-ACELP算法

- 编码器对**10ms**长的**语音帧**进行处理，**每帧分为两个子帧**。输入语音首先要在预处理模块中经过**高通滤波和幅度压缩变换**，**以去除低频干扰及防止在后面运算中出现溢出**。每帧进行一次线性预测（LP）分析，并将LPC参数转换到线谱，对（LSP）形式进行预测式二阶段矢量量化（VQ）。然后使用分析合成法，按照合成信号和原始信号间感觉加权失真最小的准则来提取激励参数
- 激励参数（包括固定码本和自适应码本参数）要每子帧（5ms）计算一次。每帧要利用感觉加权语音进行一次**整数基音延时估计**，然后进行的**分数基音延时分析**，**确定自适应码本的延时和增益**，下面再进行**固定码本的搜索**。固定码本是使用交织单脉冲排列设计的代数码本。在搜索时使用**迭代式深度优先树型搜索算法**。这种算法的运算量比较小，并且具有固定的运算复杂度，比较有利于使用硬件实现。自适应码本和固定码本的增益使用预测式二阶段共轭结构码本进行**矢量量化**

## 2.4 ACELP算法

### 基于ACELP算法的AMR格式编码



## 内 容 提 纲

1. AAC编码原理
2. ACELP编码原理
3. 本讲小结



# 3.本讲小结

## ○AAC编码原理

- AAC标准的发展、编码规格、比MP3的优势、开源编码器
- AAC编码总体框架及编码流程详解
- AAC码流结构
- AAC解码流程

## ○ACELP编码原理

- 语音的产生与感知：人的发音系统、人耳听觉系统、听觉感知特性
- 语音产生的数学模型——激励模型
- CELP算法、ACELP算法、ITU-T G.729 (CS-ACELP) 算法

## 第四讲 AAC和ACELP编码原理

**谢 谢**  
**Q&A**

欢迎电子邮件、QQ与微信交流问题！