

# 基础信息论

## 有损信源编码

华中科技大学电信学院

# 有损信源编码

- 目标
  - 理解MP3编码中的关键步骤
- 内容
  - 感知编码
  - 时频分析
  - 加掩码
  - 非均匀量化

# 感知编码

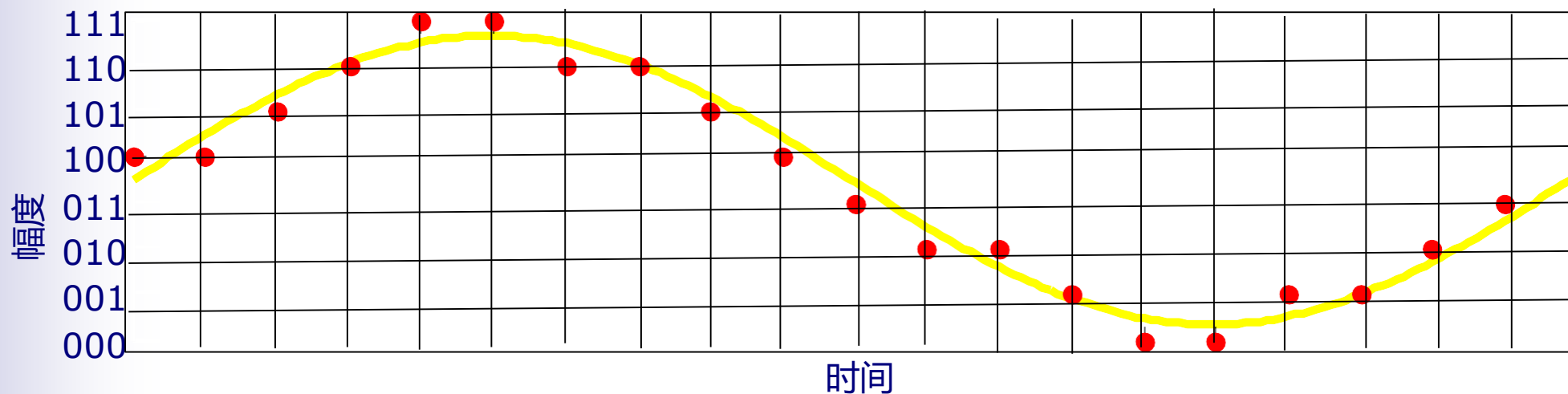
# 有损信源编码



- 输出和输入相近或者**类似**
- 通常用于面向人际通信的数据流，例如音频和视频数据
  - 例子: MP3, JPEG, MPEG
  - 通过丢弃信息实现压缩

# 脉冲编码调制

- 无压缩音频信号的标准格式 (例如: 音频CD)
- 音频信号特点:
  - 在时域采样
  - 幅度量化



**n = 3 bits**

**输出比特流: 100 100 101 110 111 111 110 110 101 100 011 010 010...**

# 移除什么样的信息？

- 减小分辨率？
  - 减少每个样本的比特数
    - 例如：16比特以2为因子可以压缩为8比特
  - 然而，引入了不可忽视的失真
- 减少采样速率？
  - 在给定的时间窗口内减少采样的数目
    - 例如：44kHz以2为因子可以压缩为22kHz
  - 然而，引入了不可忽视的高频信息的丢失
- 这些方法导致信号产生较明显的可感知变化，但是相对而言信息压缩量下降较小

# 感知编码

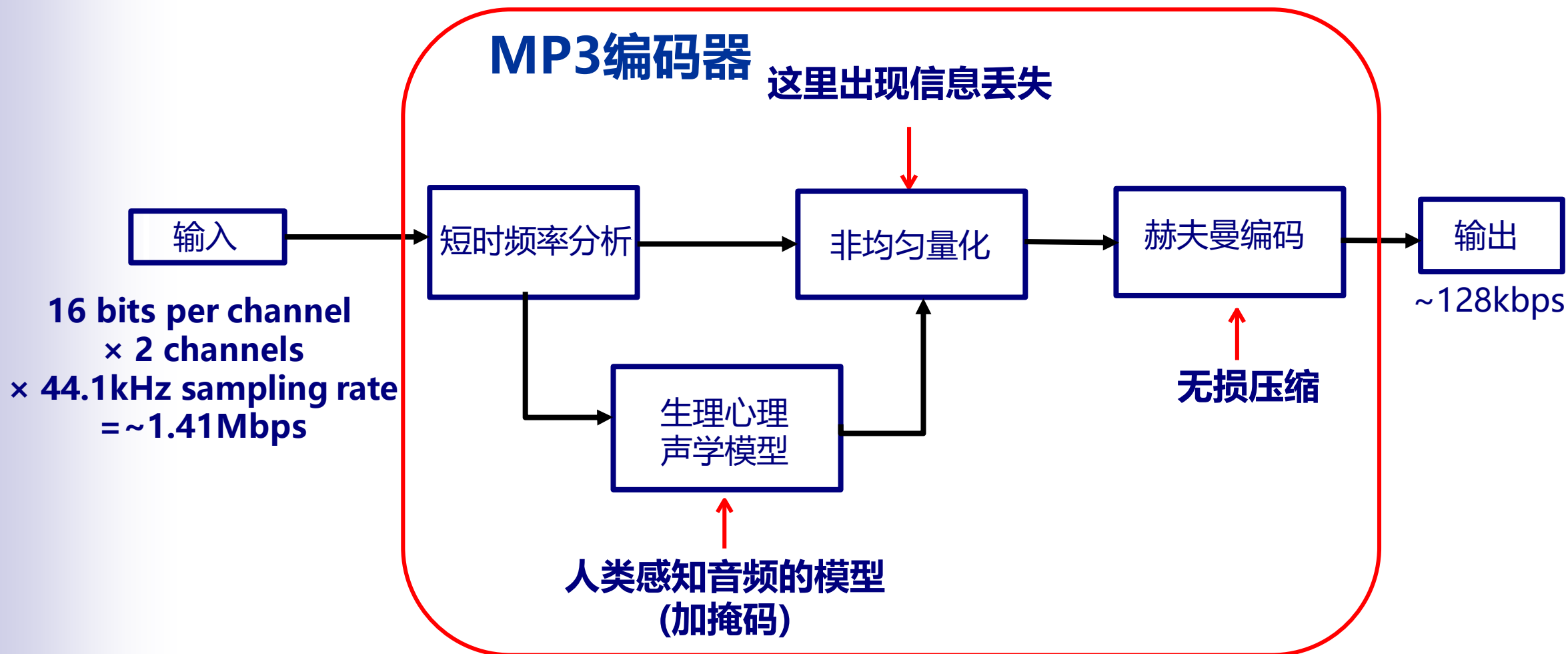
- 重要的是接收者（人的眼睛或者耳朵）如何理解输入
- 感知编码致力于寻求在不会产生可感知的影响的情况下可以被从比特流中移除的特定信息
- 这个过程基于人的感知模型

# MP3

- MPEG = Moving Pictures Experts Group (运动图像专家组)
  - 由ISO (International Standards Organization)建立
- MP3 = MPEG audio layer III
- MP3 实现了10:1 的压缩率!
- 可以实现
  - 比特流传输
  - 压缩的音频存储



# MP3编码中的关键步骤

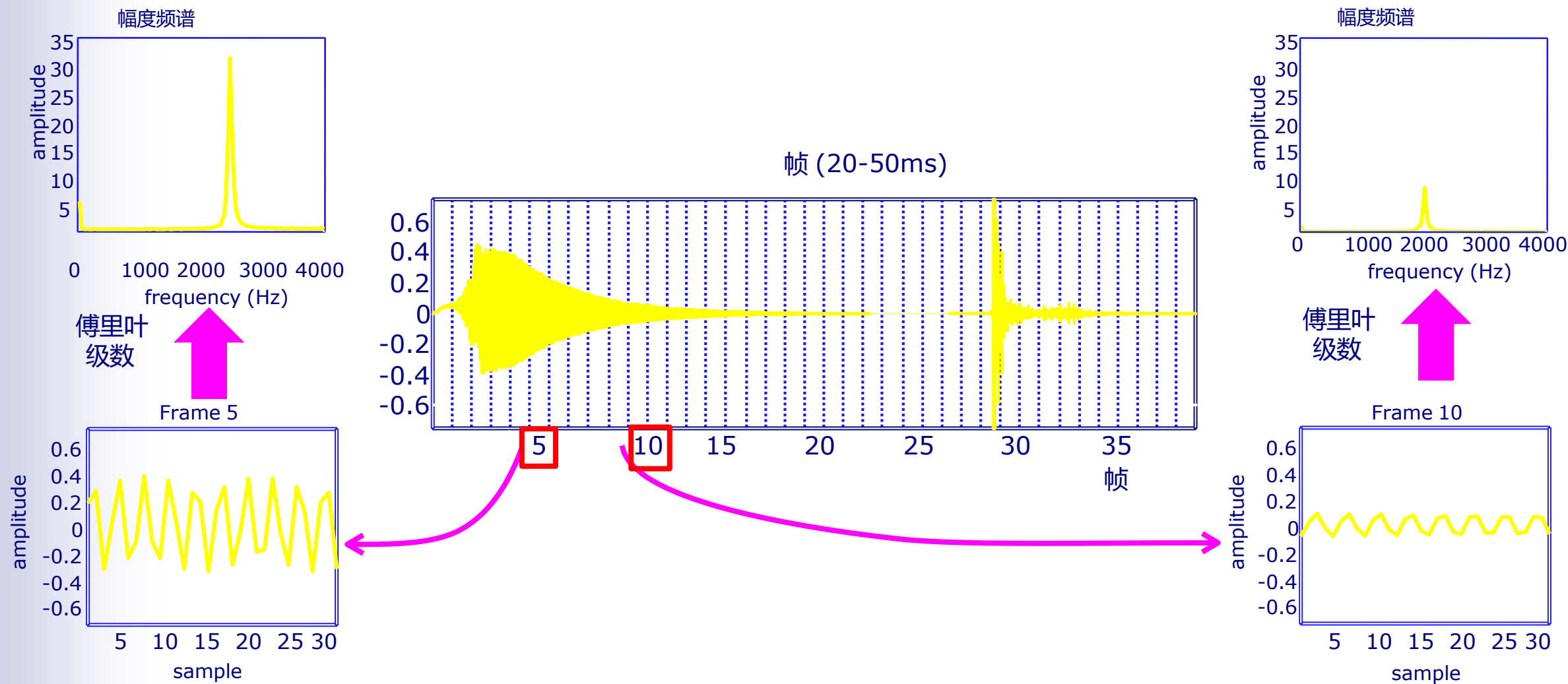


# 时频分析

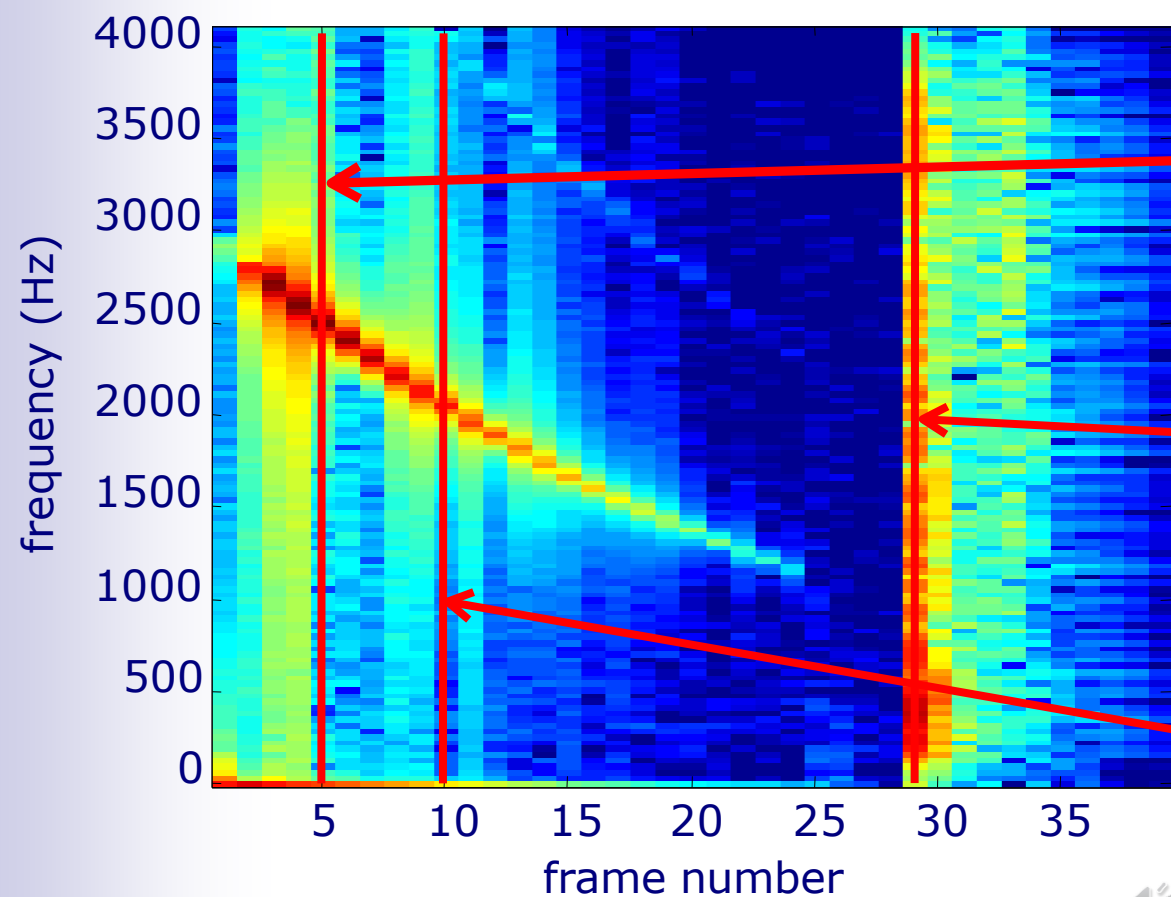
# 时频分析

- 对于复杂的信号（例如：演讲或者音乐），
  - 短时间尺度的分段可被一些重要的频率成分准确描述
  - 长时间尺度的分段则没有
- 时频分析揭示了短时间的频率成分在长时间间隔内时怎么改变的。
  - **时频图**就是时频分析的一个例子

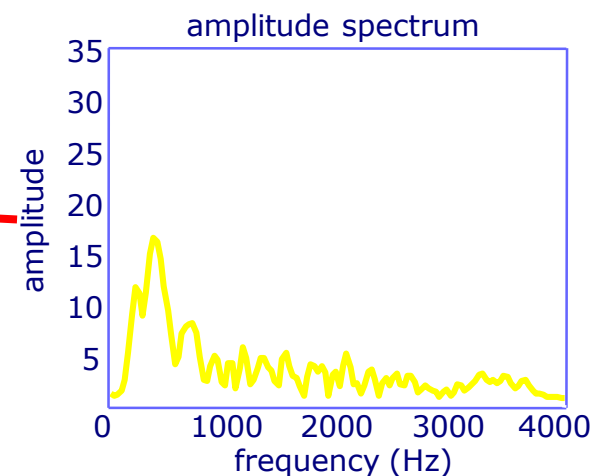
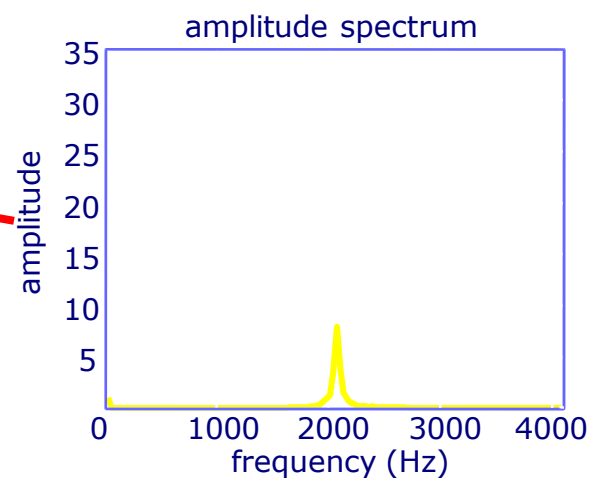
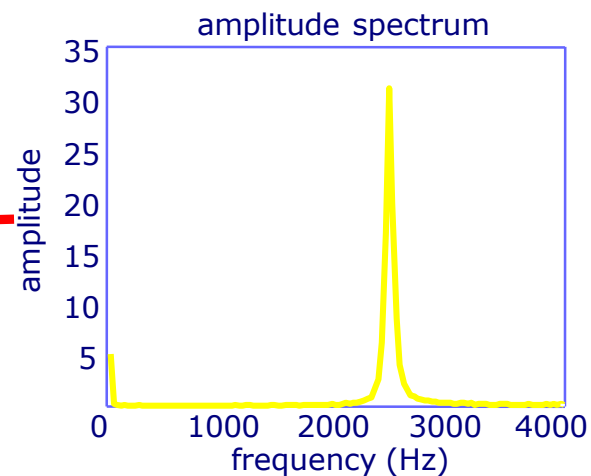
# 短期频谱



# 时频图

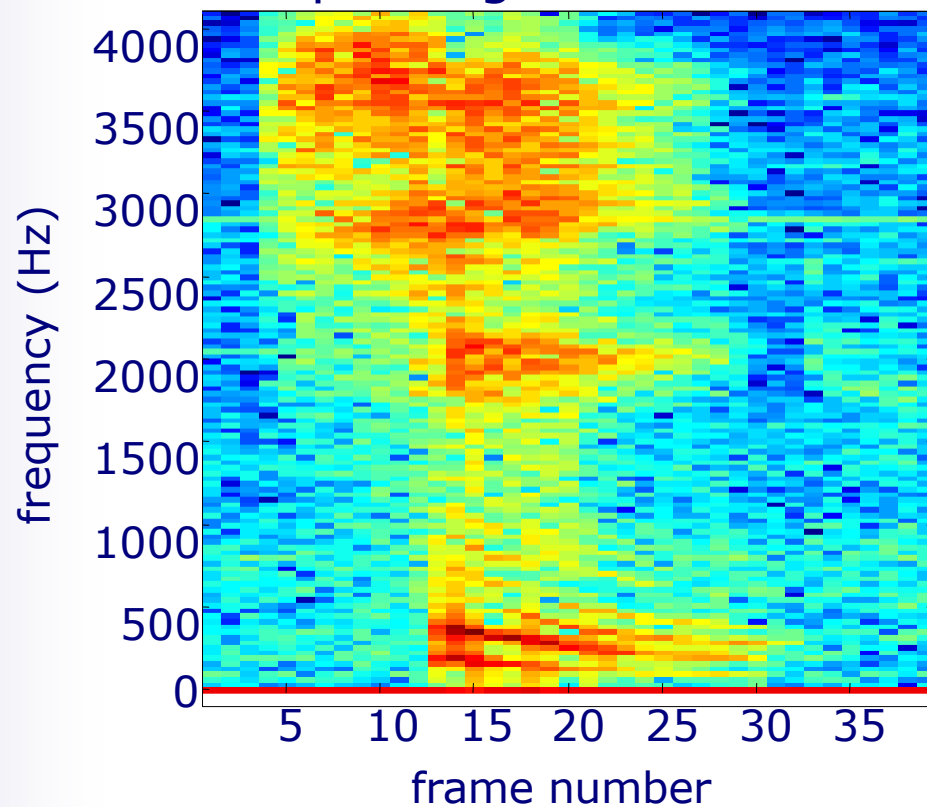


blue = quiet  
red = loud

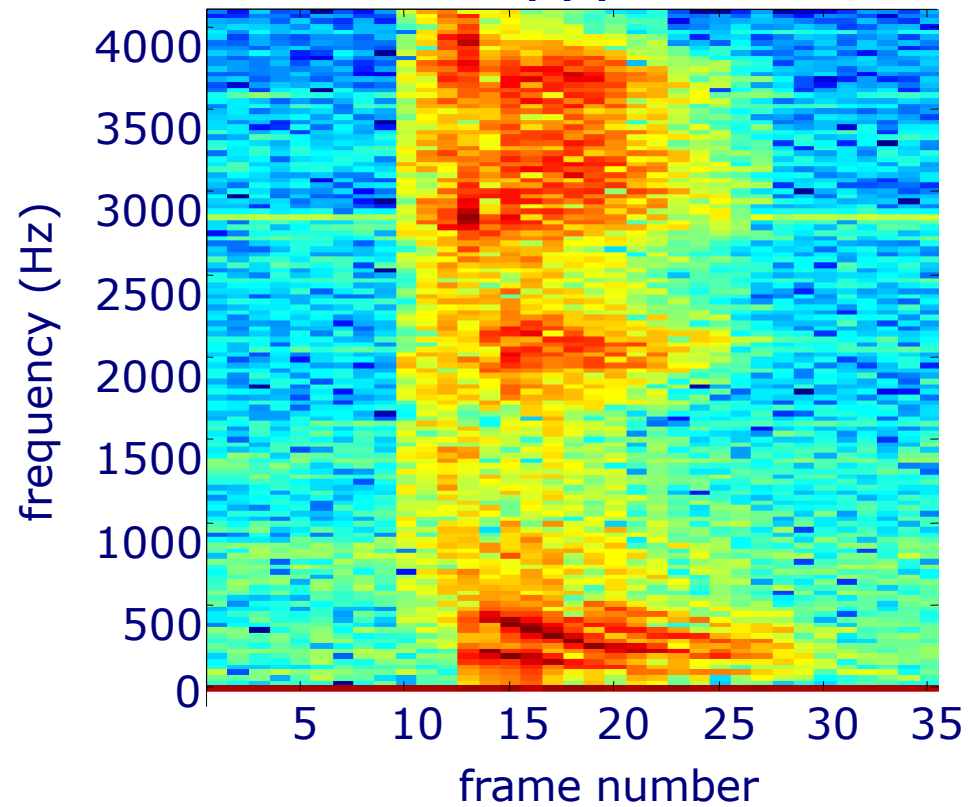


# 演讲时频图

spectrogram of "she"

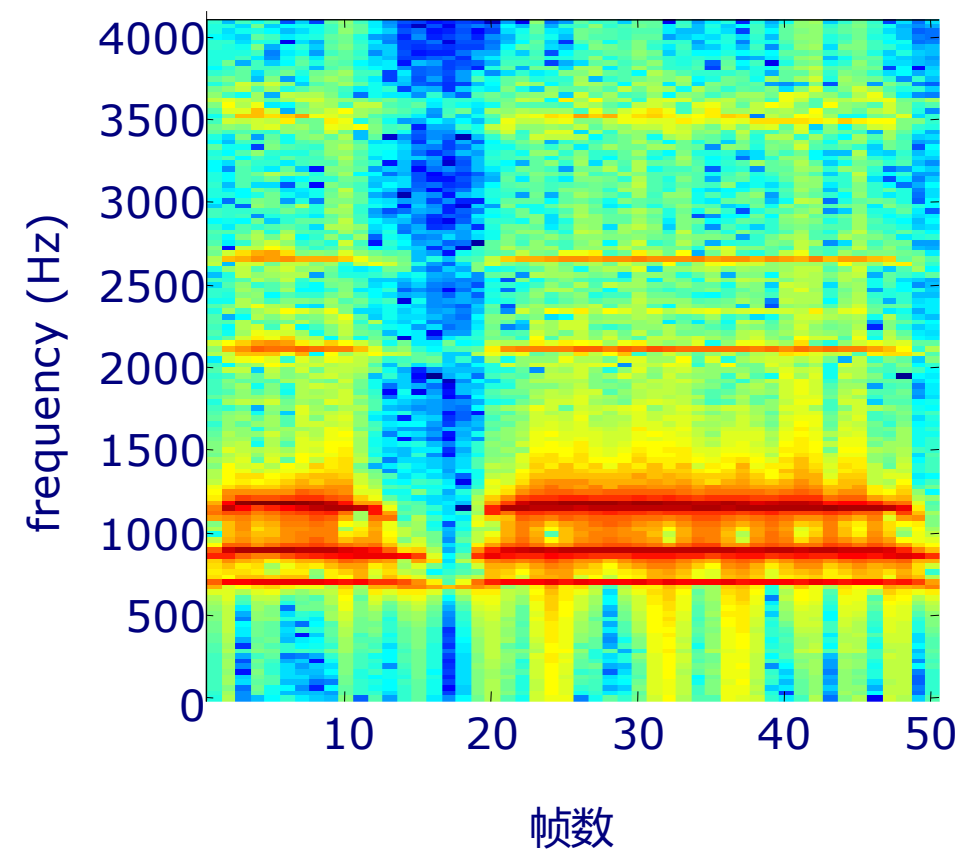
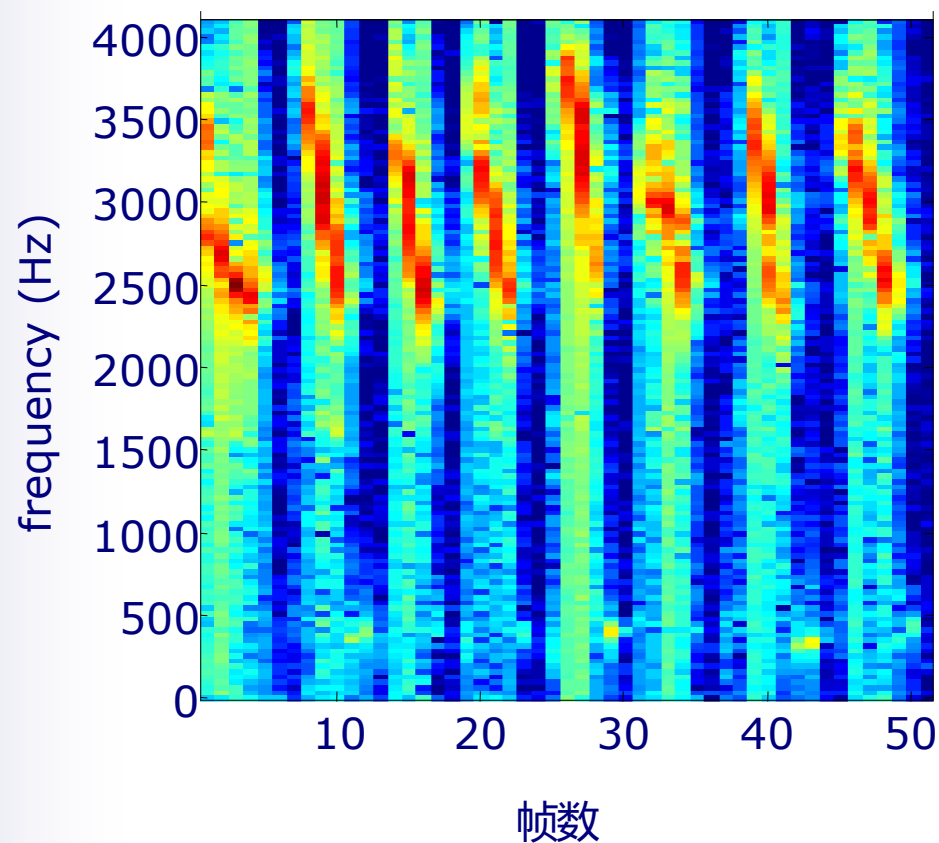


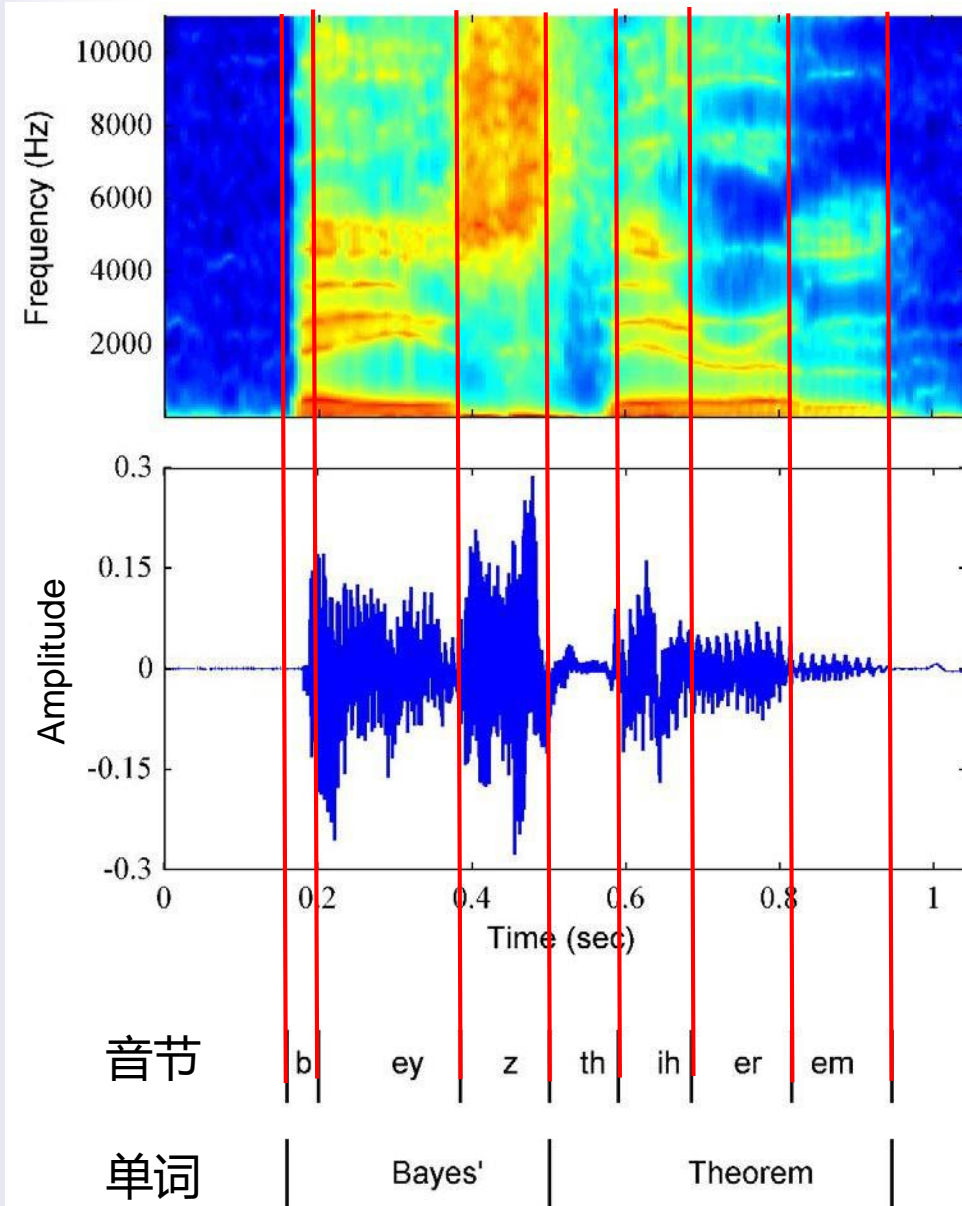
???



# 训练口哨声和鸟的叫声

上述声音分别对应哪一个时频图?





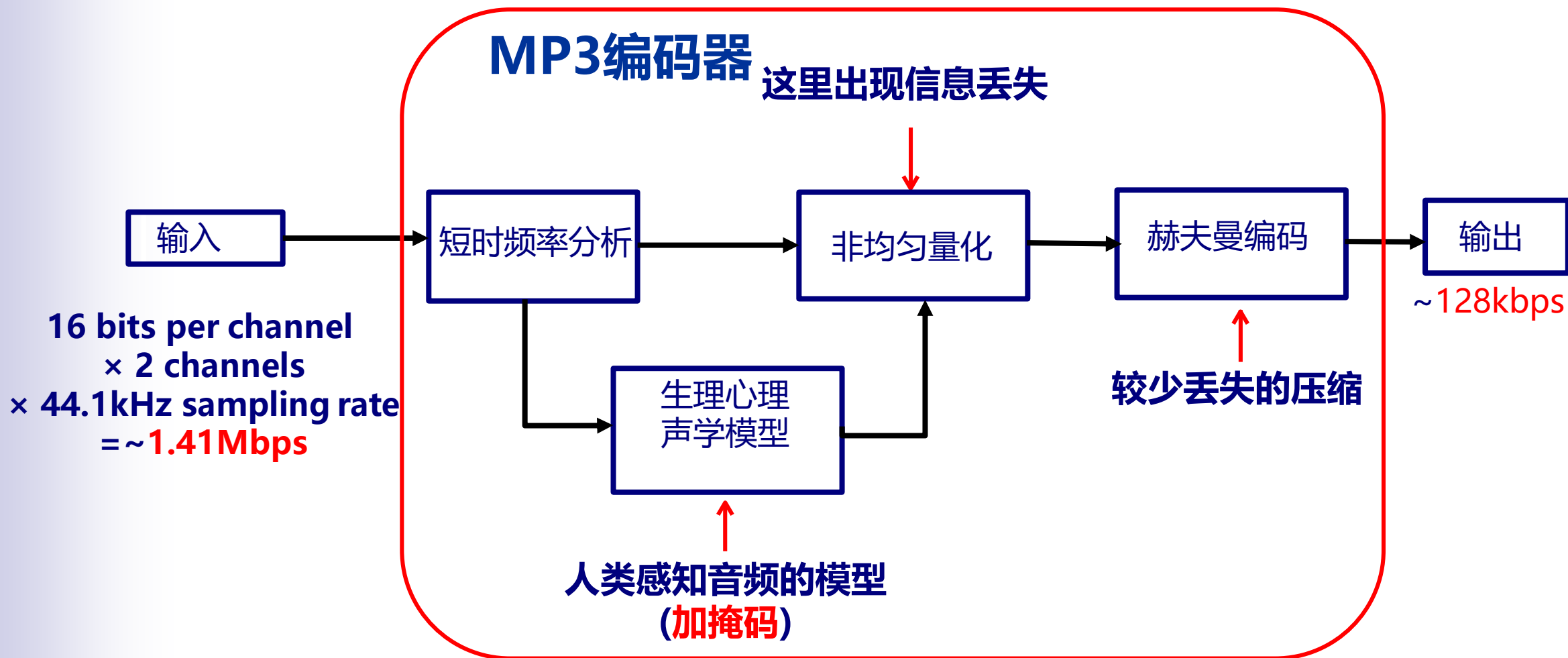
## 演讲数据

- 单词是由**音节**单元组成的
- 不同的音节有不同的频谱特性
- 当音节发生改变时，频谱特性也会变.
- **时频图**可以用来实现这一现象的可视化.

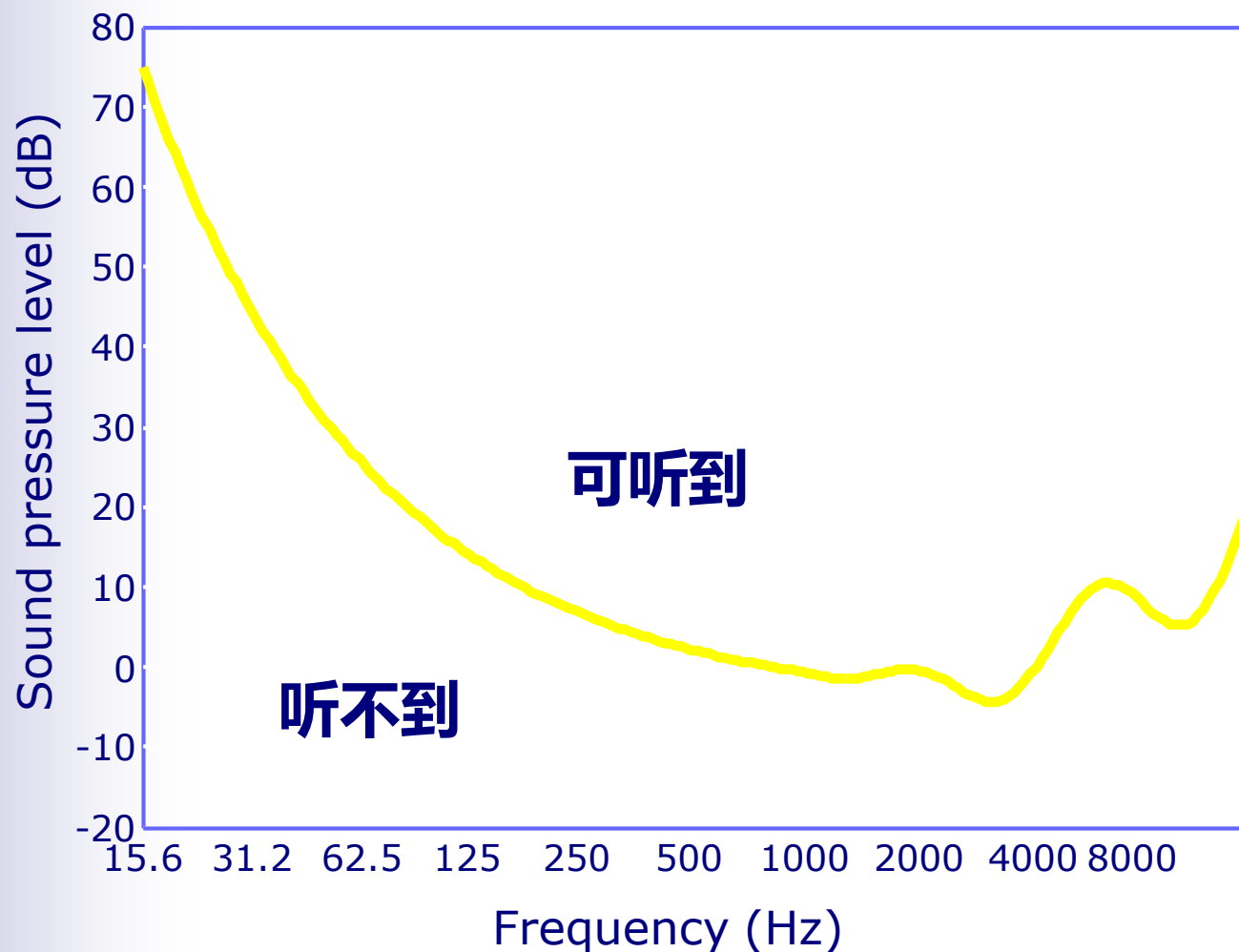


# 加掩码

# MP3 编码中的关键步骤

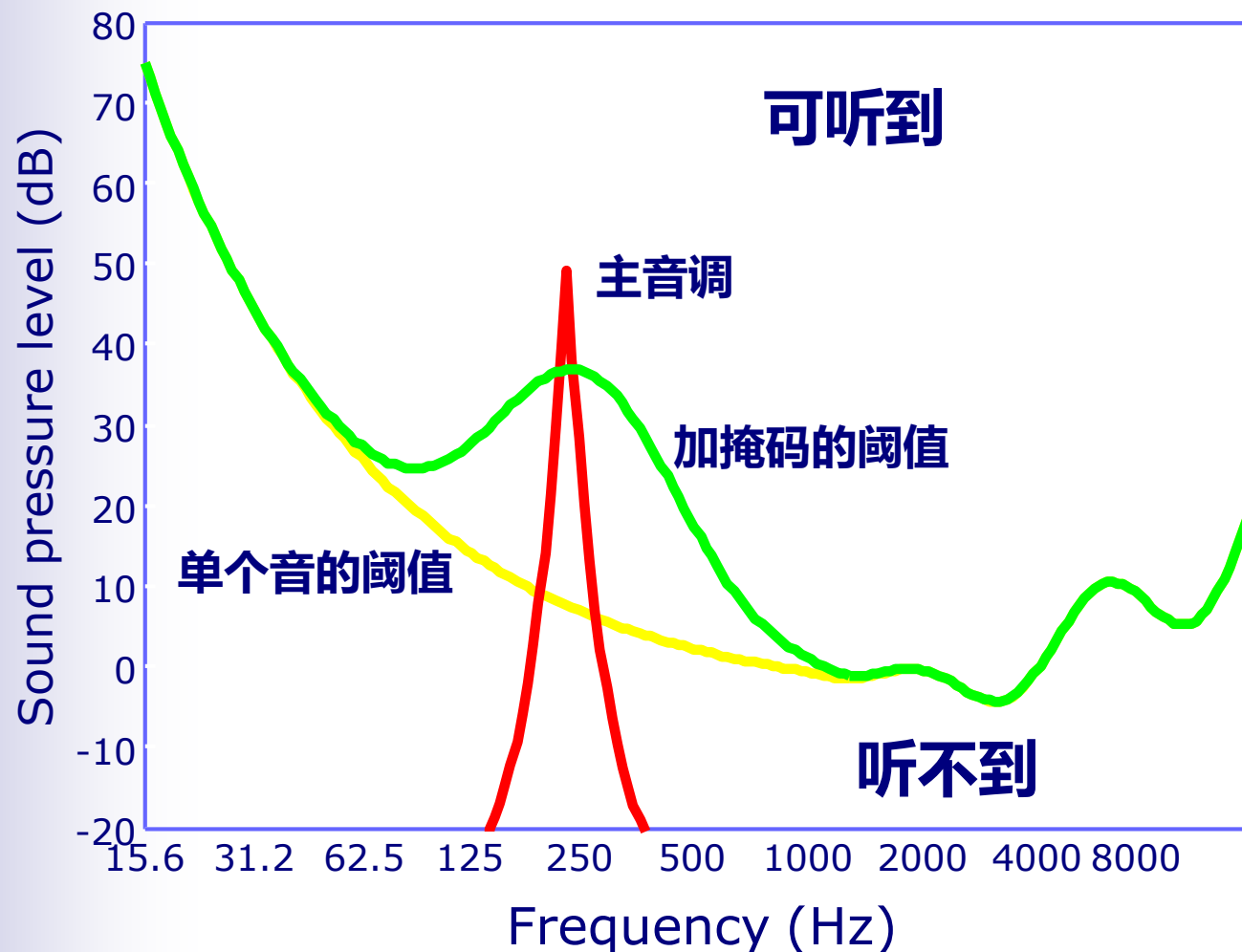


# 可听见的阈值



- 可听见的阈值是一个单音色可以被听到的**最小的声压值**.
- 丢掉在这个阈值之下的信号**不会**对声音造成可以感觉到的改变

# 加掩码

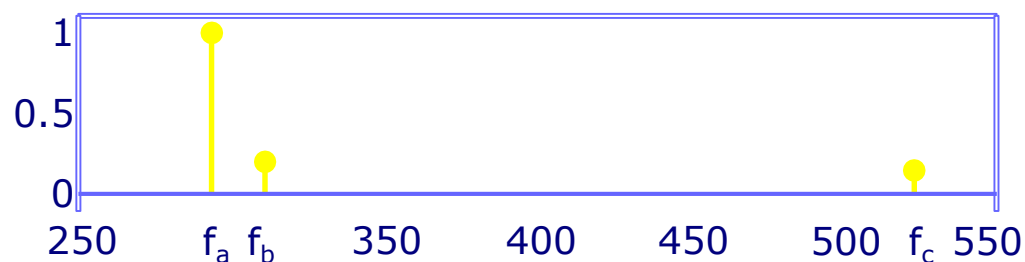


- 加掩码: 如果存在一个主音调, 那么在这个主音调附近的频率比较难以被听到.
- 在这个主音调附近频率存储信息要求的精度不高

# 加掩码例子

考虑一个有如下幅频函数的声音

声音 #1

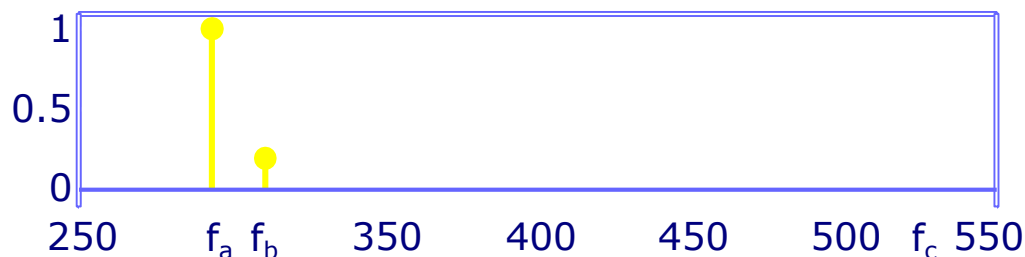


$$\begin{aligned} &\cos(2\pi f_a t) \\ &+ 0.2 \cos(2\pi f_b t) \\ &+ 0.15 \cos(2\pi f_c t) \end{aligned}$$



下面的两种声音哪一个和上面的更相似？

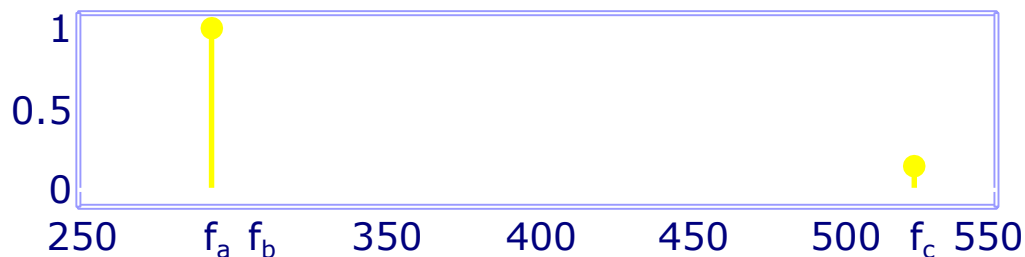
声音 #2



$$\begin{aligned} &\cos(2\pi f_a t) \\ &+ 0.2 \cos(2\pi f_b t) \end{aligned}$$



声音 #3



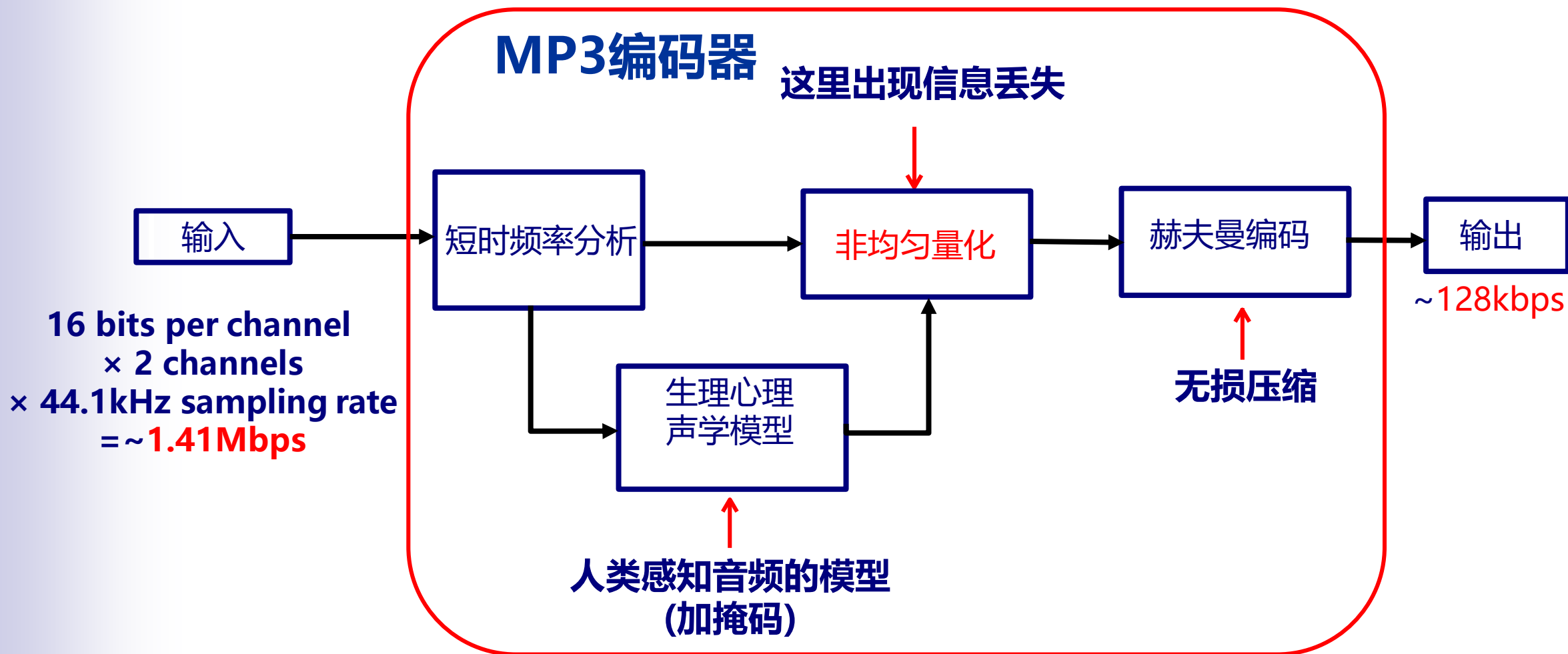
$$\begin{aligned} &\cos(2\pi f_a t) \\ &+ 0.15 \cos(2\pi f_c t) \end{aligned}$$



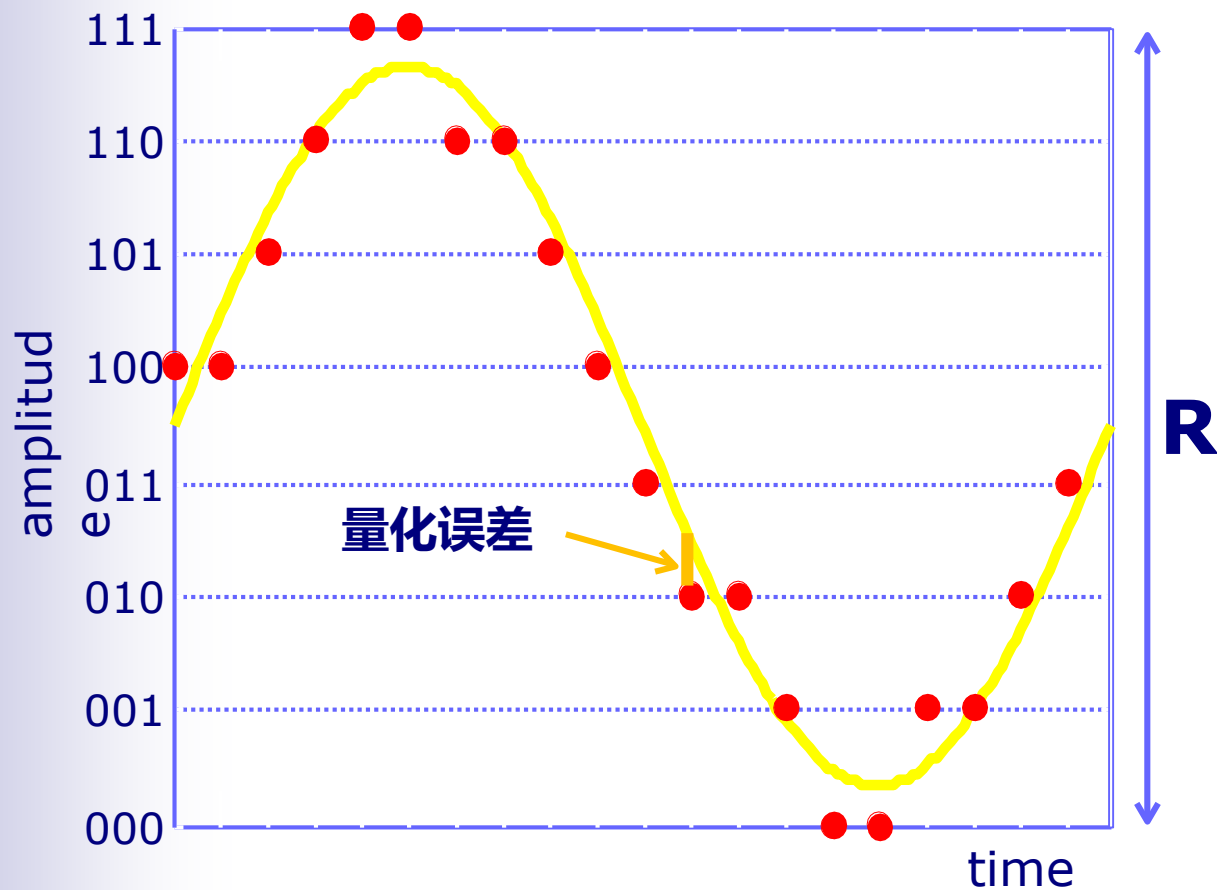
$$\begin{aligned} f_a &= 293.6\text{Hz} \\ f_b &= 311.1\text{Hz} \\ f_c &= 523.3\text{Hz} \end{aligned}$$

# 非均匀量化

# MP3 编码中的关键步骤



# 量化

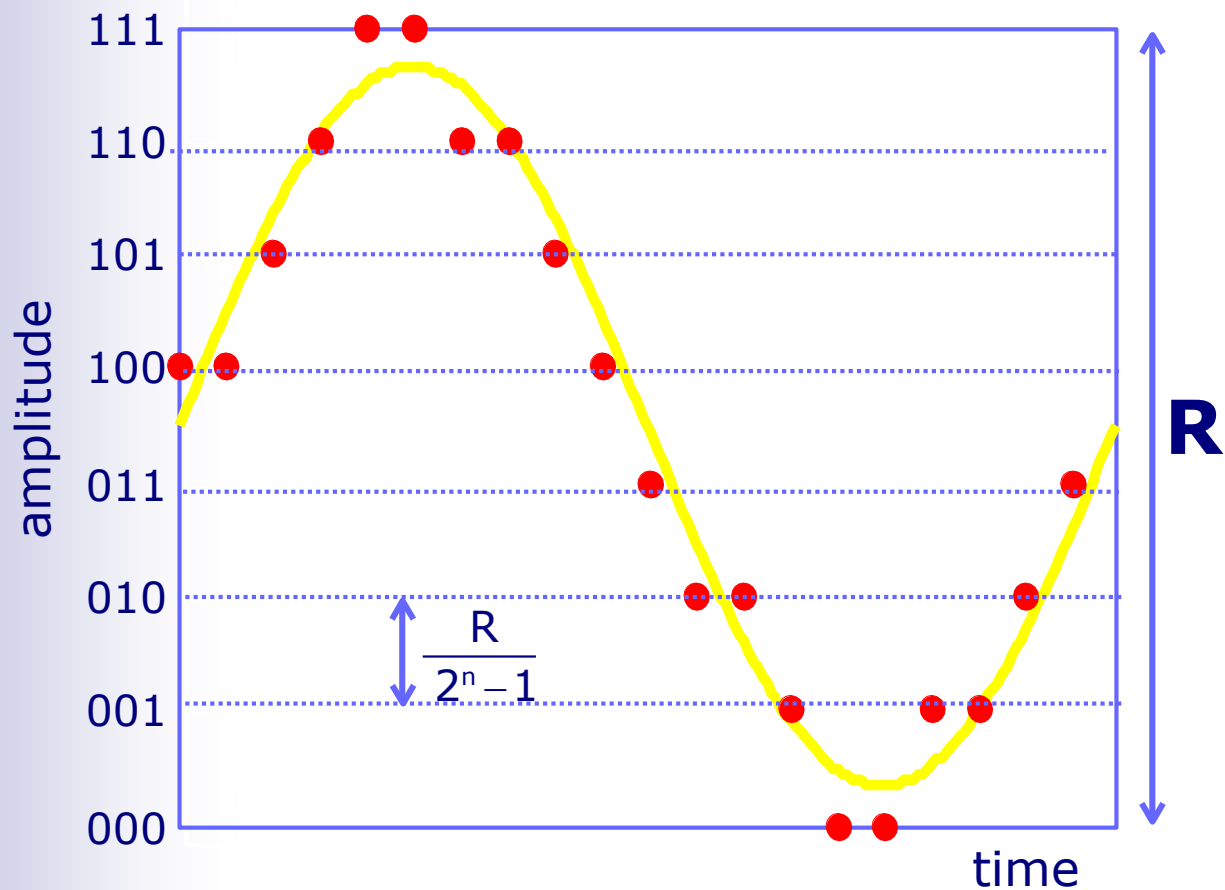


3 bits  $\rightarrow 2^3 = 8$  个量化等级  
4 bits  $\rightarrow 2^4 = 16$  个量化等级

- 二进制的数是一组有限的值:  $2^n$ ,  $n$  为比特数目.
- 量化
  - 将预期的信号范围  $R$  分为  $2^n$  个不同水平
  - 将原始的信号量化为最近的二进制值.
  - 会产生量化误差.



# 分辨率



■ 分辨率可以用以下两种方式表示

□ 比特数目

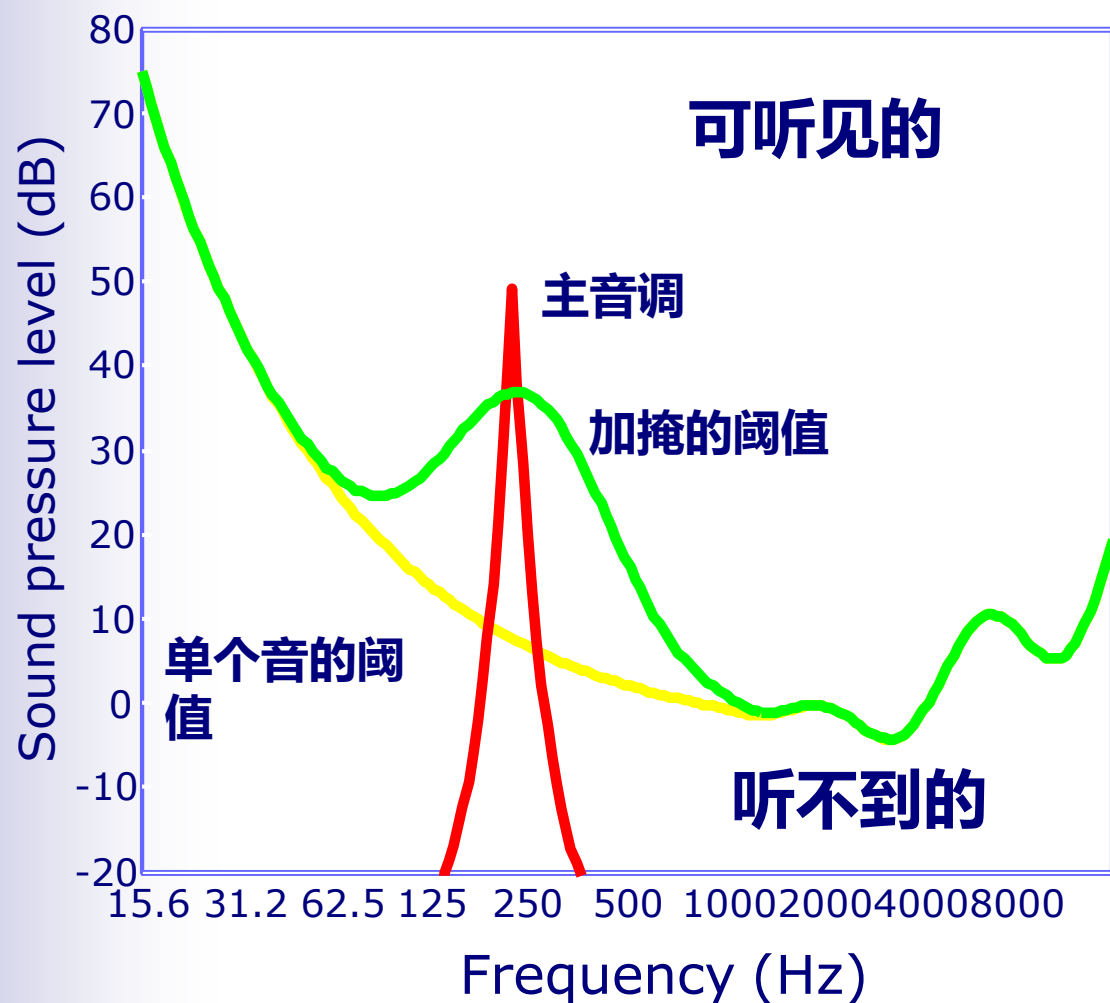
e.g.,  $n = 3$  比特

□ 不同量化等级之间的距离

e.g.,

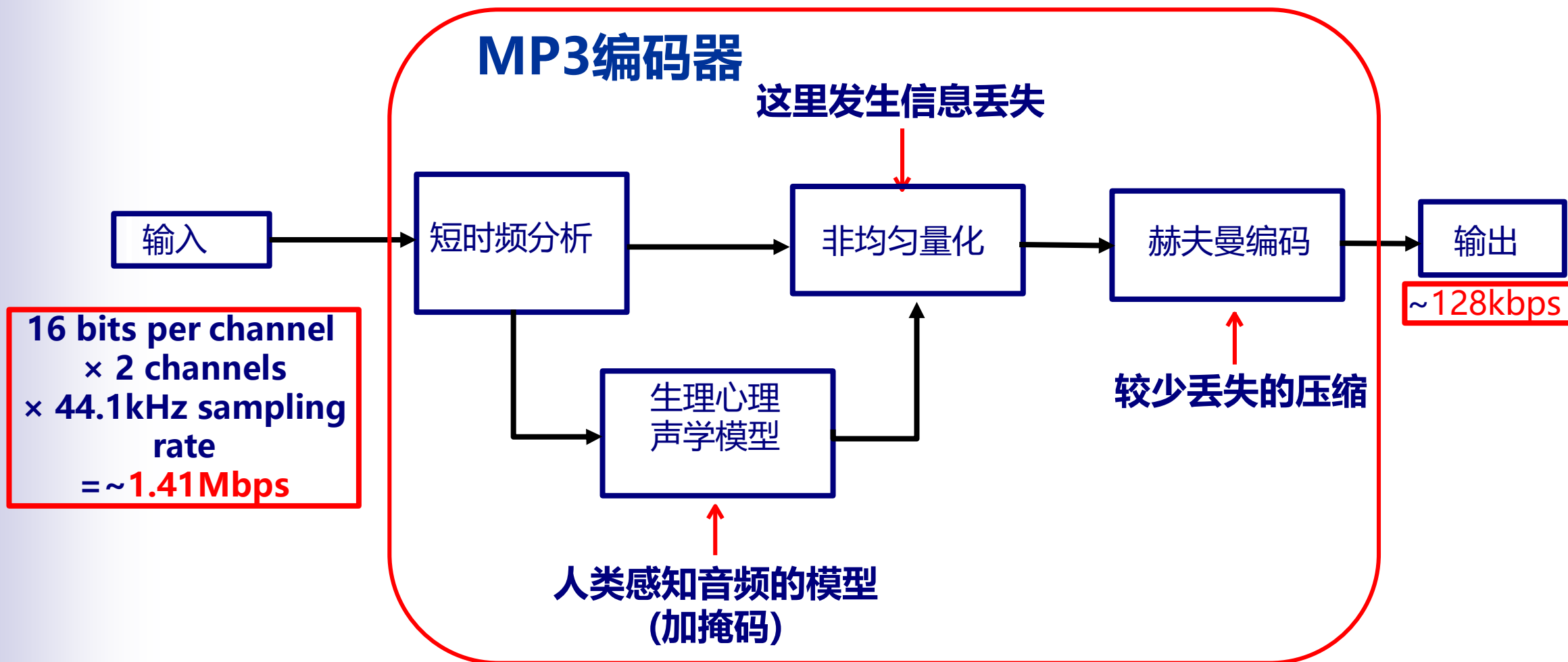
$$\text{分辨率} = \frac{R}{2^n - 1}$$

# 非均匀量化



- MP3 中将频率的幅度用各种分辨率量化:
  - 接近主要的掩码附近的频率成分采用更少的量化比特

# 总结



# 作业：有损信源编码测验

登录微助教

<http://portal.teachermate.com.cn/>

## 补充阅读

- Ref. Link: How MP3 Compression Works  
<http://www.soundonsound.com/sos/may00/articles/mp3.htm>
- Ref. Wiki: MP3 <https://en.wikipedia.org/wiki/MP3>

谢谢!

黑晓军

华中科技大学

电子信息与通信学院

Email: [heixj@hust.edu.cn](mailto:heixj@hust.edu.cn)

网址: <http://eic.hust.edu.cn/aprofessor/heixiaojun>

## 参考资料

- *A System View of Communications: From Signals to Packets (Part 2)*

<https://www.edx.org/course/a-system-view-of-communications-from-signals-to-packets-part-2-2>