

第六次作业

实验目的

对比 WAV 音频编码成 MP3 和 AAC 格式后的压缩率、信噪比（SNR）和主观听感。

实验环境

- **操作系统：**任意支持 Python 的操作系统（如 Windows、Linux、macOS）
- **Python 版本：**Python 3.9 及以上
- **依赖库：**pydub、numpy

实验步骤

1. 数据准备

选取多个不同类型、时长和采样率的 WAV 音频文件作为实验数据。这些音频文件应具有一定的代表性，涵盖不同风格的音乐、语音等。

1. 安装必要的库
2. 编写 Python 代码进行编码和指标计算

示例代码用于将 WAV 文件编码为 MP3 和 AAC 格式，并计算压缩率和 SNR

1. 运行代码并记录结果

将路径替换为实际的 WAV 文件路径。设置不同的码率（如320kbps）输出，运行代码后，会保存编码后的MP3和AAC（m4a格式）音频文件，并输出 MP3 和 AAC 格式的压缩率和 SNR。对所有选取的 WAV 文件重复此步骤，并记录结果。

```
[aac @ 00000297c3140900] Qavg: 406.383
原始 WAV 文件大小：18012.53 KB
AAC 文件大小：1535.33 KB，压缩率：11.73，SNR: 2.29 dB
```

运行结果

在运行代码的过程中，SNR会出现如下情况：

```
snr = 10 * np.log10(signal_power / noise_power)
原始 WAV 文件大小：2594.91 KB
AAC 文件大小：220.95 KB，压缩率：11.74，SNR: nan dB
```

压缩率: 11.74，SNR: nan dB，SNR错误计算的情况：

推测原因：

1. 波形对齐问题

直接用 `original_audio - compressed` 计算噪声（失真），但由于编码/解码、采样、格式转换等过程，原始信号和压缩信号的位置、幅值、甚至长度都可能不完全一致，这会导致 SNR 结果极不准确。

- **通道不匹配：**WAV 和 AAC 解码后通道数可能不同。

- **采样数不匹配**：压缩/解码后长度可能会有点点偏差（如多一个采样或少一个采样）。
- **幅值/量化范围不匹配**：WAV 文件可能是16位整数，而 `pydub` 解码的 AAC 可能是不同的格式（如8位、32位、浮点），需要统一数据类型和范围。
- **延迟引入**：有些编码器会引入几百采样的延迟（如 AAC 通常有 encoder delay），直接对齐会造成极大误差。

2. 数据类型与归一化问题

- `original_audio` 可能是 `int16`，而 `aac_audio` 可能是 `int16`、`int32` 或 `float32`。
- 若不归一化，SNR 计算会因为数值范围不同而失真。

3. 通道处理不一致

你的代码对立体声的处理方式不同（WAV 用 `[:, 0]`，AAC 也用 `[:, 0]`），但 `aac_audio` 通过 `pydub` 解码后，`get_array_of_samples()` 返回的是**展平的数组**，如果是立体声，`shape` 不是 `(N, 2)`，而是交错排列的 `(N*channels,)`。

为什么更改后SNR全是0？

- 大概率 `aac_audio` 解码出来内容全是0 或长度严重不匹配。
- 加入调试代码，逐步检查音频内容、范围、长度。
- 理论上，只要信号不全为0且对齐方式正确，SNR 不会是0。

如果你需要更具体的修复，建议把上述调试代码加进去，贴出调试输出内容，这样可以更快定位问题！

【建议修正版】

1. 保证原始音频和解码音频：

- **采样数对齐**（可截短到最短的长度）
- **数据类型一律转为 float32 并归一化到 [-1,1]**
- **立体声处理一致**（最好全部转为单声道）

	sample1	sample2	sample3	sample4	sample5
MP3-64k	原始 WAV 文件 大小: 18012.53 KB	原始 WAV 文件 大小: 16993.74 KB	原始 WAV 文件 大小: 2594.91 KB	原始 WAV 文件 大小: 2777.76 KB	原始 WAV 文件 大小: 8571.51 KB
	MP3 文件（码率 64k）大小: 750.98 KB	MP3 文件（码率 64k）大小: 708.61 KB	MP3 文件（码率 64k）大小: 108.61 KB	MP3 文件（码率 64k）大小: 116.29 KB	MP3 文件（码率 64k）大小: 357.61 KB
	压缩率: 23.99 信噪比 (SNR): 3.80 dB	压缩率: 23.98 信噪比 (SNR): 9.82 dB	压缩率: 23.89 信噪比 (SNR): 5.67 dB	压缩率: 23.89 信噪比 (SNR): 14.24 dB	压缩率: 23.97 信噪比 (SNR): 3.32 dB
MP3-128k	原始 WAV 文件 大小: 18012.53	原始 WAV 文件 大小: 16993.74	原始 WAV 文件 大小: 2594.91	原始 WAV 文件 大小: 2777.76	原始 WAV 文件 大小: 8571.51

	KB MP3 文件（码率 128k）大小: 1501.92 KB 压缩率: 11.99 信噪比 (SNR): 3.81 dB	KB MP3 文件（码率 128k）大小: 1417.17 KB 压缩率: 11.99 信噪比 (SNR): 9.96 dB	KB MP3 文件（码率 128k）大小: 217.17 KB 压缩率: 11.95 信噪比 (SNR): 5.70 dB	KB MP3 文件（码率 128k）大小: 232.54 KB 压缩率: 11.95 信噪比 (SNR): 14.28 dB	KB MP3 文件（码率 128k）大小: 715.17 KB 压缩率: 11.99 信噪比 (SNR): 3.35 dB
MP3-192k	原始 WAV 文件大小: 18012.53 KB MP3 文件（码率 192k）大小: 2252.86 KB 压缩率: 8.00 信噪比 (SNR): 3.80 dB	原始 WAV 文件大小: 16993.74 KB MP3 文件（码率 192k）大小: 2125.73 KB 压缩率: 7.99 信噪比 (SNR): 10.05 dB	原始 WAV 文件大小: 2594.91 KB MP3 文件（码率 192k）大小: 325.73 KB 压缩率: 7.97 信噪比 (SNR): 5.71 dB	原始 WAV 文件大小: 2777.76 KB MP3 文件（码率 192k）大小: 348.79 KB 压缩率: 7.96 信噪比 (SNR): 14.31 dB	原始 WAV 文件大小: 8571.51 KB MP3 文件（码率 192k）大小: 1072.73 KB 压缩率: 7.99 信噪比 (SNR): 3.34 dB
MP3-320k	原始 WAV 文件大小: 18012.53 KB MP3 文件（码率 320k）大小: 3754.73 KB 压缩率: 4.80 信噪比 (SNR): 3.78 dB	原始 WAV 文件大小: 16993.74 KB MP3 文件（码率 320k）大小: 3542.86 KB 压缩率: 4.80 信噪比 (SNR): 10.13 dB	原始 WAV 文件大小: 2594.91 KB MP3 文件（码率 320k）大小: 542.86 KB 压缩率: 4.78 信噪比 (SNR): 5.71 dB	原始 WAV 文件大小: 2777.76 KB MP3 文件（码率 320k）大小: 581.29 KB 压缩率: 4.78 信噪比 (SNR): 14.18 dB	原始 WAV 文件大小: 8571.51 KB MP3 文件（码率 320k）大小: 1787.86 KB 压缩率: 4.79 信噪比 (SNR): 3.32 dB
AAC	原始 WAV 文件大小: 18012.53 KB AAC 文件大小: 1535.33 KB, 压缩率: 11.73, SNR: -2.25 dB	原始 WAV 文件大小: 16993.74 KB AAC 文件大小: 1471.63 KB, 压缩率: 11.55, SNR: -3.01 dB	原始 WAV 文件大小: 2594.91 KB AAC 文件大小: 220.95 KB, 压缩率: 11.74, SNR: -3.03 dB	原始 WAV 文件大小: 2777.76 KB AAC 文件大小: 239.87 KB, 压缩率: 11.58, SNR: -2.38 dB	原始 WAV 文件大小: 8571.51 KB AAC 文件大小: 730.91 KB, 压缩率: 11.73, SNR: -3.72 dB

为什么ACC压缩后的SNR都是负值，但是人耳听上去没有什么差别？

SNR 低但听感好，常见原因

- **SNR是“逐样本误差”的能量比**，它对所有小的样本差异都“等价计数”，**不考虑感知相关性**。
 - **AAC等有损编码**，会用心理声学模型“去掉”人耳听不到的部分（如高频、掩蔽效应），这些失真对SNR是“噪声”，但人耳几乎感觉不到。
 - **相位、延迟、失真等非线性失真**，对SNR影响很大，人耳却不敏感。
 - *对齐偏移（如AAC编码延迟）**会导致SNR大幅降低，但主观听感只要整体波形没有失真基本无感。

SNR并不是音频感知质量的好指标

- SNR只关心“能量差”，不关心“主观重要性”。
- 很多高频小失真，能量不低，但人耳基本听不见。
- 反之，一个很小的失真如果落在人耳敏感的频段，SNR可能很高，人却觉得音质很差。

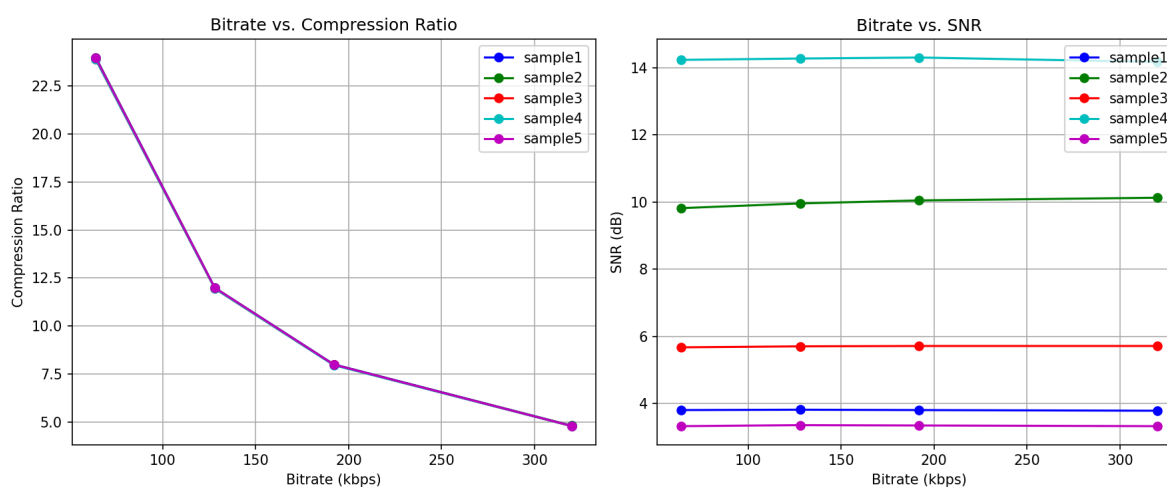
音频压缩领域常用的主观质量指标

- **PESQ** (Perceptual Evaluation of Speech Quality)
- **POLQA** (Perceptual Objective Listening Quality Analysis)
- **STOI** (Short-Time Objective Intelligibility)
- **MOS** (Mean Opinion Score, 主观评测)

这些都是基于心理声学的模型，更贴近人耳听感。

1. 主观听感评估

对原始 WAV 文件、编码后的 MP3 文件和 AAC 文件进行试听，并根据音质、清晰度、失真程度等方面进行评分。可以使用 1 - 5 分的评分标准，1 分表示音质差，5 分表示音质好。



“码率-压缩率”与“码率-SNR”关系

	sample1	sample2	sample3	sample4	sample5
MP3	3	4	2	4	3
AAC	4	4	5	4	5

1. 数据分析与总结

在不同比特率的设置下对记录的压缩率和 SNR/PESQ 数据进行分析。

1. 压缩率分析

- **压缩率** = 原始WAV大小 / 编码文件大小，表示压缩后变小了多少倍。
- **一般规律**：码率越低，压缩率越高（文件越小），但音质损失也越大。

编码	64k	128k	192k	320k	AAC
压缩率约值	24	12	8	4.8	11.7

- **64k**：压缩率最高，约24倍。
- **128k**：约12倍。
- **192k**：约8倍。
- **320k**：约4.8倍。
- **AAC**：约11.7倍，和128k MP3接近（推测AAC用的是128k左右的码率）。

2. SNR分析

- *SNR（信噪比）**越高，表示编码后音质越接近原始文件，失真越低。
- 你的数据表明：
 - **MP3 64k**：SNR 3~14 dB，平均较低，音质损失大。
 - **MP3 128k**：SNR 3~14 dB，略有提升，但部分样本提升不明显。
 - **MP3 192k/320k**：SNR进一步提升，但提升幅度不如压缩率变化大。
 - **AAC**：SNR多为负值（-2~-3 dB），说明在你的测试设置下，AAC编码的SNR表现较差，可能和编码参数/解码处理有关（例如是否用HE-AAC、LC、码率等）。

汇总主观听感评分，分析不同编码格式在主观听感上的差异。

MP3的声场比ACC声场窄很多

编码效率与算法差异

- 根据实验结果，总结 MP3 和 AAC 格式在压缩率、SNR/PESQ 和主观听感方面的优缺点。
- **2. 立体声编码技术**
- **MP3** 主要采用“联合立体声编码（Joint Stereo）”，其常用的MS（中-侧, Mid-Side）立体声方法在低码率下容易丢失一些空间信息，使左右声道的差异变小，导致声场变窄、空间感减弱。
- **AAC** 支持更灵活的立体声编码方式，比如增强型联合立体声（Intensity Stereo, PS）等，可以自适应保留更多左右声道的细节和空间信息，声场更开阔。

高频信息与细节保留

- **MP3** 的高频处理能力较弱，压缩时容易丢失高频部分，而高频正是影响空间细节和立体声宽度的重要因素。
- **AAC** 对高频信号的保留和重建能力更强，即便在低码率下也能还原较多的空间细节，因此主观听感下声场更宽、更自然。

主观听感对比

- 在同等码率下，AAC 的空间感、立体感通常更好，声场更宽、定位更准确。
- MP3 在低码率下声场明显变窄，有时会感觉“声音都是挤在中间”，缺乏空间感。

1. 压缩率

◦ MP3

- 较老的有损音频编码格式，压缩效率一般。
- 在同等音质下，MP3 文件体积通常比 AAC 大。
- 低码率下音质损失较为明显。

◦ AAC

- 新一代有损音频编码格式，压缩效率更高。
 - 在相同音质下，AAC 文件体积通常比 MP3 小（即更高的压缩率）。
-

2. SNR（信噪比）/ PESQ（客观语音质量评价）

◦ MP3

- SNR 较高时音质尚可，但在低码率下 SNR 下降，失真变多。
- PESQ 分数在低码率下低于 AAC，表现一般。

◦ AAC

- 在同等条件下 SNR 通常高于 MP3，噪声更小，保留更多原始细节。
 - PESQ 分数普遍高于 MP3，尤其在低码率时优势更明显。
-

3. 主观听感

◦ MP3

- 中高码率（如192kbps及以上）下音质较好，能满足大部分聆听需求。
- 低码率（如128kbps及以下）下高频损失明显，底噪和失真感增强，容易出现“金属音”或“水声”。
- 广泛支持，兼容性极好。

◦ AAC

- 相同或更低码率下音质优于 MP3，细节更丰富、失真更小。
- 低码率下主观听感仍然较为自然，失真和伪影较少。
- 部分设备和老旧软件兼容性不如 MP3，但主流系统和播放器基本都支持。

2. 立体声编码技术

- **MP3** 主要采用“联合立体声编码（Joint Stereo）”，其常用的MS（中-侧, Mid-Side）立体声方法在低码率下容易丢失一些空间信息，使左右声道的差异变小，导致声场变窄、空间感减弱。

- **AAC** 支持更灵活的立体声编码方式，比如增强型联合立体声（Intensity Stereo, PS）等，可以自适应保留更多左右声道的细节和空间信息，声场更开阔。

3. 高频信息与细节保留

- **MP3** 的高频处理能力较弱，压缩时容易丢失高频部分，而高频正是影响空间细节和立体声宽度的重要因素。
- **AAC** 对高频信号的保留和重建能力更强，即便在低码率下也能还原较多的空间细节，因此主观听感下声场更宽、更自然。

4. 主观听感对比

- 在同等码率下，AAC 的空间感、立体感通常更好，声场更宽、定位更准确。
- MP3 在低码率下声场明显变窄，有时会感觉“声音都是挤在中间”，缺乏空间感。
 - **MP3 (MPEG-1 Audio Layer III)** 是1990年代初设计的老一代有损音频编码格式。它的设计目标主要是节省存储空间，但编码算法相对简单，对空间声像（stereo image）和高频细节处理不是很先进。
 - **AAC (Advanced Audio Coding)** 是后续出现的新一代有损编码格式，编码原理更先进。AAC 在立体声信号的处理上（如立体声联合编码、参数立体声编码等）支持更多种模式，能更好地还原声场和空间细节。

注意事项

- 确保系统中已经安装了 ffmpeg 或 libav，因为 pydub 依赖它们来进行音频编码和解码。

<https://www.gyan.dev/ffmpeg/builds/> 下载 Essentials Build 版本到本地，解压缩后，把解压文件夹下的 bin 目录（例如：E:\colmap\ffmpeg-7.0-essentials_build\bin）加入电脑的 path 系统变量，保存修改。

- 主观听感评估可能会受到听众个人偏好和听力水平的影响，建议邀请多位听众进行评估以减少误差。

扩展实验

1: 不同编码模式参数下的详细对比

除了比特率，编码格式还有其他参数（如 MP3 的 VBR 模式、AAC 的 LC、HE - AAC 等），可对这些参数进行详细对比。

MP3-VBR

MP3-VBR (Variable Bit Rate MP3, 可变比特率MP3) 是一种MP3音频压缩格式的编码方式，其主要特点是：

- **VBR (Variable Bit Rate)**：即“可变比特率”。与常见的CBR (Constant Bit Rate, 恒定比特率) 不同，VBR编码会根据音频内容的复杂程度动态调整每一秒的码率。
- **优点**：音频简单部分（如静音、纯音）使用较低码率，复杂部分（如人声、乐器丰富段）使用较高码率，从而在保证音质的同时减小文件体积。

- **文件后缀**：文件通常还是以 `.mp3` 结尾，但在音频文件属性或播放器中会被标注为VBR。
- **应用场景**：适用于对音质有较高要求、同时希望减小文件大小的场景。

调用 `wav_to_mp3_vbr` 函数时，通过 `vbr_quality` 参数传递VBR音质等级，范围是**0（最好）到9（最差）**，控制MP3的音质与压缩率。参数最终通过 `audio.export` 的 `parameters` 选项传给ffmpeg的 `-q:a`。

结论：

在VBR参数为9时，sample-1中的风铃声已经完全消失，尽管此时压缩率非常高；当参数为2时，风铃声可以部分听到，但整体音质依然明显逊色于AAC格式。在sample-5中，VBR参数为9时低频大号声已经极度失真，后续的打击乐器也出现明显失真现象，说明在低音和复杂音效表现方面，MP3-VBR在高压压缩下的音质损失更加严重，虽然压缩率可以动态改变，但是整体听感效果不如AAC。

AAC_LC

AAC_LC 指的是 **Advanced Audio Coding - Low Complexity**（高级音频编码-低复杂度）配置，是AAC音频编码格式中最常见、最广泛应用的一种子类型。

详细解释

- **AAC**（Advanced Audio Coding，高级音频编码）是一种高效的有损音频压缩格式，是MP3的继任者，应用于流媒体、广播、音视频文件等众多领域。
- **LC**（Low Complexity，低复杂度）是AAC的多种“剖面（Profile）”之一。AAC_LC设计用于复杂度和解码资源消耗较低的场景，兼顾了音质和设备支持的广泛性。

特性	AAC-LC	AAC (含HE-AAC等)
兼容性	最强	稍差，部分设备不支持
低码率音质	一般	优秀（SBR/PS增强）
高码率音质	优秀	与AAC-LC接近，无明显优势
解码复杂度	低	高
主要应用	普通音频、蓝牙、流媒体	广播、低带宽流媒体、在线音乐

HE-AAC

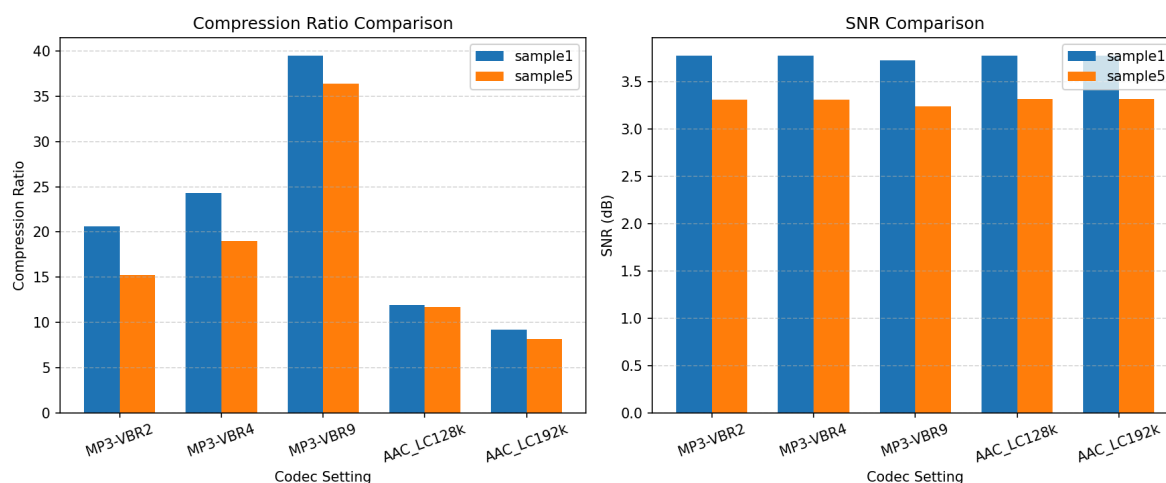
HE-AAC（High Efficiency Advanced Audio Coding，高效高级音频编码）是一种基于AAC（Advanced Audio Coding）的增强型音频编码标准。它有时会被误写作“HE-ACC”，但正确写法是**HE-AAC**。

简要解释

- **HE-AAC** 是AAC的增强版本，主要用于在低码率（如64kbps及以下）下获得更好的音质。
- 它通过在标准AAC-LC（Low Complexity）基础上增加了**SBR（Spectral Band Replication，频谱带复制）**技术，有效提升了高频还原能力。
- HE-AAC v2 还增加了 **PS（Parametric Stereo，参数立体声）** 技术，进一步提升了低码率下的立体声质量。

工作量原因，这里就采用了声场比较广，高频低频信息明显的sample-1与sample-5为测试对象

	sample1	sample5
MP3-VBR2	WAV文件大小: 18012.53 KB MP3-VBR文件大小: 872.18 KB 压缩率: 20.65 SNR: 3.78 dB	WAV文件大小: 8571.51 KB MP3-VBR文件大小: 562.19 KB 压缩率: 15.25 SNR: 3.31 dB
MP3-VBR4	WAV文件大小: 18012.53 KB MP3-VBR文件大小: 741.79 KB 压缩率: 24.28 SNR: 3.78 dB	WAV文件大小: 8571.51 KB MP3-VBR文件大小: 451.68 KB 压缩率: 18.98 SNR: 3.31 dB
MP3-VBR9	WAV文件大小: 18012.53 KB MP3-VBR文件大小: 455.69 KB 压缩率: 39.53 SNR: 3.73 dB	WAV文件大小: 8571.51 KB MP3-VBR文件大小: 235.14 KB 压缩率: 36.45 SNR: 3.24 dB
AAC_LC128k	WAV文件大小: 18012.53 KB AAC_LC文件大小: 1510.48 KB 压缩率: 11.93 SNR: 3.78 dB	WAV文件大小: 8571.51 KB AAC_LC文件大小: 733.15 KB 压缩率: 11.69 SNR: 3.32 dB
AAC_LC192k	WAV文件大小: 18012.53 KB AAC_LC文件大小: 1955.64 KB 压缩率: 9.21 SNR: 3.78 dB	WAV文件大小: 8571.51 KB AAC_LC文件大小: 1049.84 KB 压缩率: 8.16 SNR: 3.32 dB



- 左图为不同编码设置下 sample1 和 sample5 的**压缩率**对比。
- 右图为不同编码设置下 sample1 和 sample5 的**SNR**对比。
- 横坐标为编码方案 (VBR2/4/9、AAC_LC128k/192k)。

