FFT：对非平稳信号存在局限性，只能描述全局的频率特点，但是对于频域细节（各成分出现时刻并无所知），因此，时域相差很大的信号可能频谱相同。

STFT：相当于对时域信号先进行加窗，然后对每段信号进行傅立叶变换。虽然可以描述频域的细节，但是STFT的选择相对困难，窗太窄，窗内信号太短，导致频率分析不够精准，频率分辨率差；窗太宽，时域上不够精细，时间分辨率低

小波变换：将无限长的三角函数基换成了有限长的会衰减的小波基。小波基有两个变量：尺度和平移量，尺度控制小波函数伸缩——对应频率（反比），平移量控制小波函数平移——对应时间，因此可以同时反映时域信息和频域信息。低频部分具有较低的时间分辨率和较高的频率分辨率，在高频部分具有较高的时间分辨率和较低的频率分辨率，很适合于分析非平稳的信号和提取信号的局部特征

频谱混叠：采样频率不满足奈奎斯特定理后，频域周期延拓频谱出现重叠

频谱泄露：时间信号在进行运算时，需要将其截断，通常采用加窗操作，加窗相当于时域相乘，也即频域卷积，于是频谱中除了本该有的主瓣之外，还出现本不该有的旁瓣，这就是频谱泄露。整周期截断是不发生频谱泄露的充要条件，采用合适的缓变窗函数，减小旁瓣能量。

栏栅效应：DFT变换只限制在离散的频谱，就像通过栏栅看景象，只能在特定点上看到真实的景象。减少栏栅效应的方法：增加频域抽样点

因果稳定系统：输出只有当前和之前的输入有关，和之后的信息无关。充要条件：h(n) = 0,n<0

MFCC：

预加重：将音频信号通过高通滤波器，提升高频部分，突出高频的共振峰，高频包含细节信息；

分帧：音频信号在整体上是非平稳信号，但是局部是平稳的，采用分帧可以更方便观察音频特性，分帧音频会有重叠区域；

加窗：平滑信号，使用汉明窗可以减弱FFT以后旁瓣大小和频谱泄露

FFT：音频频域上特性相较于时域更容易观察

三角带通滤波器：将能量谱通过一组Mel尺度的三角滤波器组，对频谱进行平滑化，并消除谐波，突出原来语音的共振峰，另外可以降低运算量

能量值取log：由于人耳对声音感知并不是线性，用log这种非线性关系更好描述

DCT变换：能量集中在低频部分，用于压缩

图示

描述已自动生成

图表

描述已自动生成 文本, 聊天或短信

描述已自动生成

（参考：<https://www.pianshen.com/article/9582143681/>）

定点量化：浮点数转为int型，使用浮点数运算是保持高精度的简单方式，但是在当前越来越多的模型被部署到移动设备，推理的效率是一个关键问题。

传统降噪算法：

谱减法：假定噪音均为加性噪声(和语音不相关)，在假设噪声统计平稳的前提下，用无语音间隙测算到的噪声估计频谱替换有语音期间频谱，与含噪音频相减，从而获得语音频谱的估计值。特点：算法简单，运算量小，易于实现，能获得较高信噪比。

LMS准则(最小均方差)：均方差最小

维纳滤波：设计一个数字滤波器，使得带噪音频和纯净音频的误差满足LMS准则

小波降噪：音频信号在经过小波变换后，真音小波系数幅值大于噪声小波系数幅值，因此可以设置阈值进行区分。

深度学习降噪算法：

DCNN：预处理(多通道转单通道，采样率调整，密集化)——提取MFCC特征——min(带噪音频MFCC，纯净音频MFCC)

Unet：带噪音频频谱——Unet——噪音频谱

带噪音频频谱 - 噪音频谱 = 重建频谱

重建频谱 + 相位谱 = 重建音频

增强处理：

增益：根据比例对音频进行增益

短音频：音频波形复制

加噪：给音频增加高斯噪声

翻转：将音频对称翻转

分割：长音频分割为多个短音频

掩蔽：屏蔽频谱中某一组频率

深度学习音频分类算法：

VGGish：预处理(多通道转单通道，采样率调整)——提取MFCC特征——VGG对MFCC进一步特征提取——SVC(支持向量机分类)

深度学习音频增强算法：