


# 音频算法调研

音频分类算法，字面意思音频分类，与自然语言处理不同

 [https://blog.csdn.net/wudibaba21/article/details/108863431?ops\\_request\\_misc=&request\\_id=&biz\\_id=102&utm\\_term=Fbank&utm\\_medium=distribute.pc\\_search\\_result.none-task-blog-2~all~sobaiduweb~default-1-108863431.142^v93^koosearch\\_v1&spm=1018.2226.3001.4187](https://blog.csdn.net/wudibaba21/article/details/108863431?ops_request_misc=&request_id=&biz_id=102&utm_term=Fbank&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-1-108863431.142^v93^koosearch_v1&spm=1018.2226.3001.4187)

**power-normalized cepstral coefficients**相比于MFCC特征：

- 在噪声和混响场景下提升识别效果，尤其在训练语料是clean语音的时候
- 相比于MFCC，计算量提升34.6%

使用pncc相比mfcc，噪声和口音测试集可以得到10-15%的相对提升

**fbank**特征更多是希望符合声音信号的本质，拟合人耳的接收特性。

**Filter Banks**和MFCC对比：

- **计算量**：MFCC是在FBank的基础上进行的，所以MFCC的计算量更大
- **特征区分度**：FBank特征相关性较高（相邻滤波器组有重叠），MFCC具有更好的判别度，这也是在大多数语音识别论文中用的是MFCC，而不是FBank的原因
- **信息量**：FBank特征的提取更多的是希望符合声音信号的本质，拟合人耳接收的特性。MFCC做了DCT去相关处理，因此Filter Banks包含比MFCC更多的信息
- 使用对角协方差矩阵的GMM由于忽略了不同特征维度的相关性，MFCC更适合用来做特征。
- DNN/CNN可以更好的利用Filter Banks特征的相关性，降低损失。

从目前的趋势来看，因为神经网络的逐步发展，FBank特征越来越流行。

## 1. Spectrogram（频谱图）：

Spectrogram是最基本和常见的音频特征表示方法之一。它将音频信号划分为短时窗口，对每个窗口进行傅里叶变换，然后将得到的频谱信息绘制成图像。Spectrogram提供了音频信号在频域上的分布信息，可以捕捉到音频的频率特征和频谱变化。

## 2. MelSpectrogram (梅尔频谱图) :

MelSpectrogram是在Spectrogram的基础上应用了梅尔滤波器组进行转换的特征表示方法。梅尔滤波器组是一组带通滤波器，能够更好地模拟人耳对声音的感知方式。MelSpectrogram在频谱上进行横向拉伸，使较低频率的分辨率更高，以更好地捕捉音频信号的感知特性。

## 3. MFCC (梅尔频率倒谱系数) :

MFCC是一种基于MelSpectrogram的特征表示方法，通过对Mel频谱图进行离散余弦变换得到MFCC系数。MFCC首先计算Mel频谱图，然后对其进行离散余弦变换，并保留一部分低频倒谱系数。MFCC主要捕捉音频信号的声学特征，包括音高、音强和声道信息，常用于语音识别和语音相关任务。

## 4. Fbank (滤波组特征) :

Fbank是一种类似于MFCC的特征表示方法，它也使用梅尔滤波器组对Spectrogram进行转换，但在之后的处理步骤中略有不同。在Fbank中，梅尔滤波器组的响应被经过对数计算，得到每个滤波器通道的能量。Fbank特征可以捕捉到音频信号的能量分布情况。

## 5. LogMelSpectrogram (对数梅尔频谱图) :

LogMelSpectrogram是在MelSpectrogram的基础上应用了对数变换的特征表示方法。对数变换有助于减小动态范围，增加对小振幅特征的敏感性，并提高特征的区分度。

**一、音频分类主要基于将音频转变为频谱，然后使用对图像的分类实现对音频的分类，主要特征在于，机器能够识别的二维的信号，能够通过坐标转换从其他维度的坐标转换为二维坐标信息的信号都可以使用神经网络来进行分析。**

**二、因此对的分析，通过不同分析手段得到的具有不同特征的频谱是决定分类标准的重点。**

**三、使用不同的网络结构对不同的检测效果的影响。**

**四、每种频谱图凸显出来的特征是什么，这个需要有一定的了解。**

算法资料链接：

<https://github.com/tensorflow/models/tree/master/research/audioset>

\n\n Models for AudioSet: A Large Scale Dataset of Audio Events \n This repository provides models