

十月份过程性研究记录

陶理

日期: October 28, 2023

摘要

该份过程性研究记录（截止至撰写日期）主要记录了十月上旬所完成的对于语音特征提取算法的学习以及对原理的剖析。截止至目前，我完成了对于傅里叶级数和傅里叶变换的理解，了解到了傅里叶变换在语音识别领域中的重要性，并对于傅里叶变换的计算有了初步的了解。

1 傅里叶级数 & 傅里叶变换

目前许多音频特征提取算法是基于傅里叶变换的：如通过短时傅里叶变换 (STFT)，将音频信号分成若干小段并将每一段进行傅里叶变换；通过梅尔频率倒谱系数 (MFCCs)，使用STFT将信号转化为频谱图，再对频谱图进行处理。因此对于傅里叶变换的充分理解是完成一个好的音频特征提取的基础。

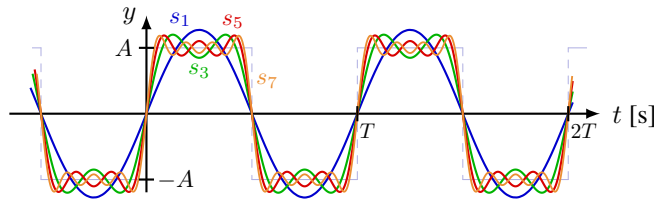
1.1 傅里叶级数

对于一个周期信号 $f(t)$ ，有：

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega t) + b_n \sin(n\omega t)] \quad (1)$$

通俗来讲，傅里叶级数就是将一个周期函数 $f(t)$ 表示为无数个余（正）弦函数的和。

e.g. 假设有 $g(t)$ 为一个方波，将其表示成无数个正弦函数的叠加，即下图所呈现。



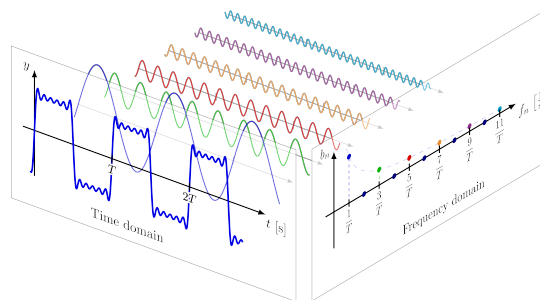
其中， $g(t)$ 就可以表达为：

$$g(t) = \frac{4A_{\max}}{\pi} \sum_{n=1}^{\infty} \frac{\sin((2n-1)\omega t)}{2n-1} \quad (2)$$

1.2 傅里叶变换

$$F(\omega) = \int f(t) \cdot e^{-i\omega t} \cdot dt \quad (3)$$

简单来讲，傅里叶变换是傅里叶级数的拓展，可以将非周期函数转化为多个余（正）弦函数的和。实际上，可以将傅里叶级数是傅里叶变换在周期函数上的特例。



傅里叶变换的意义：为研究信号的频谱特性提供了方案，提供了一种将时域信号转化为频域信号的方法。

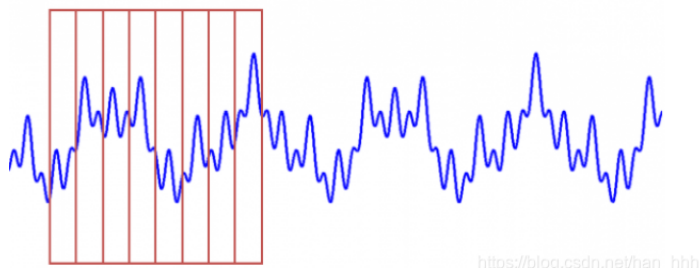
傅里叶变换在语音识别领域的意义：傅里叶变换使得时域上的音频信号转化为频域上的信号，从而对于研究音频的特征提供了方案。

2 语音特征提取算法

2.1 短时傅里叶变换

短时傅里叶变换，即Short Time Fourier Transform，简称STFT。

由于傅里叶变换将时域信号转化为频域，所以无法从傅里叶变化后的信号观察出特定时刻的频率，所以通过“加窗”的手段，将音频分为多个小的片段，如此即能将每一个小片段看作平滑的曲线，再进行傅里叶变换，就可以知道在每个时刻上的频率。



https://blog.csdn.net/han_hhh

2.2 梅尔频率倒谱系数

梅尔频率倒谱系数，即Mel-scale Frequency Cepstral Coefficients，简称MFCCs。梅氏系数也是在语音识别领域中运用的最为广泛的特征提取算法之一。梅尔频率倒谱系数通常有以下几个步骤以下引自百度百科(<https://baike.baidu.com/梅尔频率倒谱系数/2916505>)：

1. 将一段语音信号分解为多个讯框
2. 将语音信号预强化，通过一个高通滤波器。
3. 进行傅立叶变换，将信号变换至频域。
4. 将每个讯框获得的频谱通过梅尔滤波器(三角重叠窗口)，得到梅尔刻度。
5. 在每个梅尔刻度上提取对数能量。
6. 对上面获得的结果进行离散傅里叶反变换，变换到倒频谱域。

7. MFCC就是这个倒频谱图的幅度(amplitudes)。一般使用12个系数，与讯框能量叠加得13维的系数。
以下为我通过Python语言编程所完成的程序结果:

