

语谱图

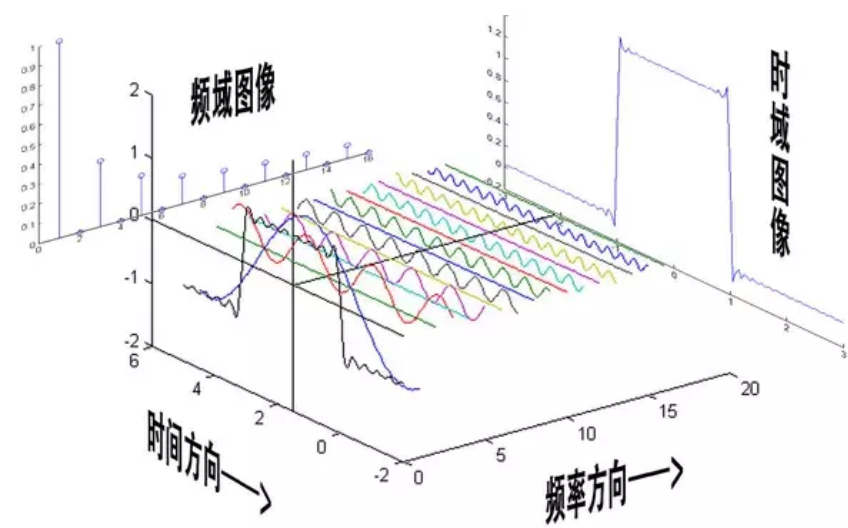
语谱图

生成语谱图的前提：

在一段时间内（如10~30ms，即所谓一帧内）可以认为频谱是不变的。【[blog链接](#)】

如何理解在一段时间内频谱是不变的？

先看图：



上图引自《傅里叶变换终极解释》

理解：对于频率方向上的若干个频率，当在时间方向上前进一段10-30ms的距离时，我们可以认为这些个频率对应的振幅不变。（时间太短，振幅还没变化。）

所以如果帧长时间太长，则这些个频率对应的振幅可能是有变化的。此时若再放在一帧里面表示则不能精确地刻画各个频率随时间变化的振幅值。

语谱图生成流程

引自blog：《[语音信号处理之（四）梅尔频率倒谱系数（MFCC）](#)》

我们处理的是语音信号，那么如何去描述它很重要。因为不同的描述方式放映它不同的信息。那怎样的描述方式才利于我们观测，利于我们理解呢？这里我们先来了解一个叫声谱图的东西。

公告

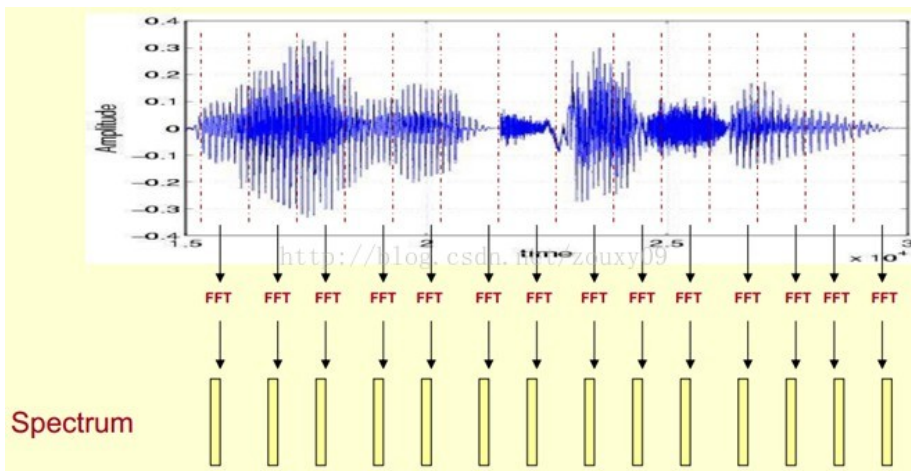
昵称：凝静志远
园龄：3年10个月
粉丝：26
关注：10
[+加关注](#)

| 2018年9月 | | | | | | |
|---------|----|----|----|----|----|----|
| 日 | 一 | 二 | 三 | 四 | 五 | 六 |
| 26 | 27 | 28 | 29 | 30 | 31 | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | 1 | 2 | 3 | 4 | 5 | 6 |

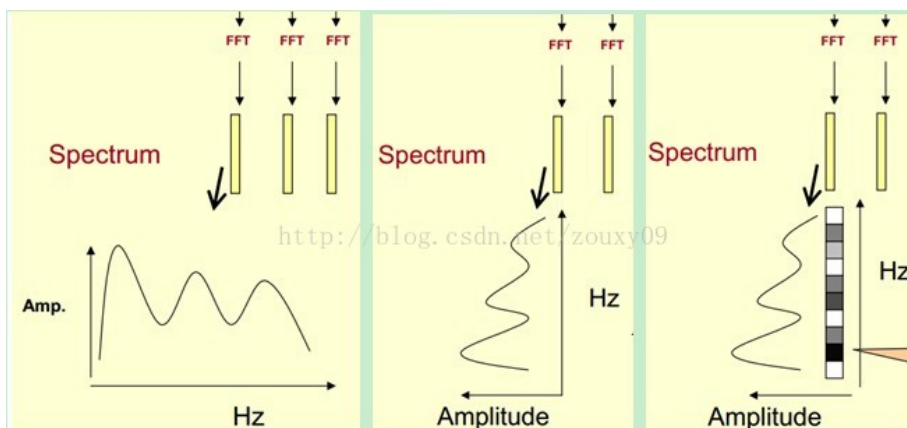
搜索

常用链接

[我的随笔](#)
[我的评论](#)
[我的参与](#)

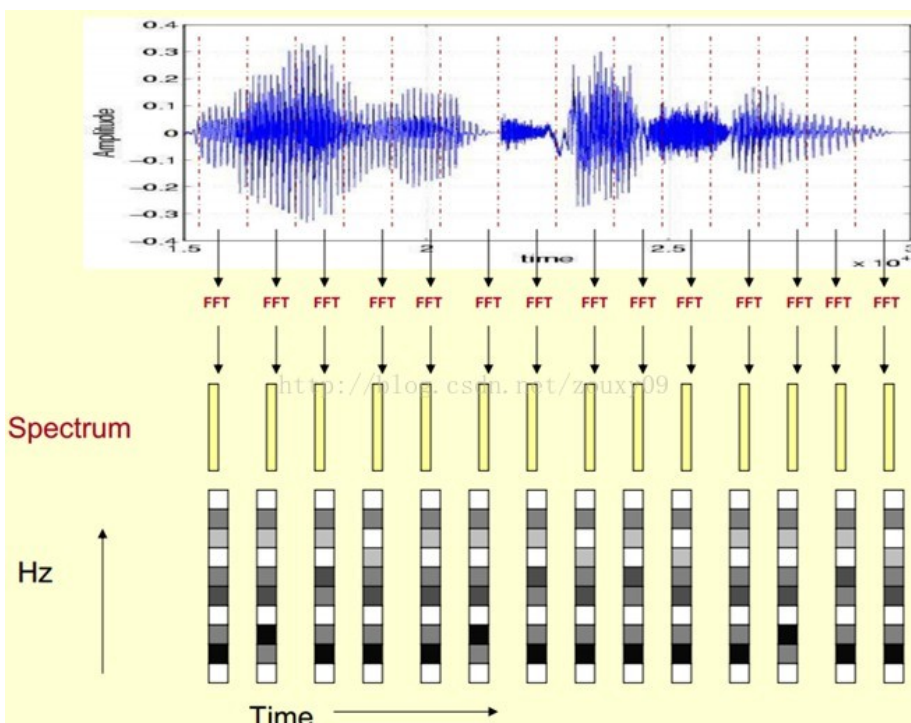


这里，这段语音被分为很多帧，每帧语音都对应于一个频谱（通过短时FFT计算），频谱表示频率与能量的关系。在实际使用中，频谱图有三种，即线性振幅谱、对数振幅谱、自功率谱（对数振幅谱中各谱线的振幅都作了对数计算，所以其纵坐标的单位是dB（分贝）【关于分贝的理解】。这个变换的目的是使那些振幅较低的成分相对高振幅成分得以拉高，以便观察掩盖在低幅噪声中的周期信号）。



我们先将其中一帧语音的频谱通过坐标表示出来，如上图左。现在我们将左边的频谱旋转90度。得到中间的图。然后把这些幅度映射到一个灰度级表示（也可以理解为将连续的幅度量化为256个量化值？），0表示黑，255表示白色。幅度值越大，相应的区域越黑。这样就得到了最右边的图。那为什么要这样呢？为的是增加时间这个维度，这样就可以显示一段语音而不是一帧语音的频谱，而且可以直观的看到静态和动态的信息。优点稍后呈上。

这样我们会得到一个随着时间变化的频谱图，这个就是描述语音信号的spectrogram声谱图。



下图是一段语音的声谱图，很黑的地方就是频谱图中的峰值（共振峰formants）。

最新评论

我的标签

随笔分类

C/C++(6)

golang(4)

Java(26)

Linux/Unix(7)

操作系统(5)

分布式系统(12)

机器学习/深度学习(5)

算法/数据结构(1)

体系结构(1)

移动开发(1)

最新评论

1. Re:最容易读进去的深度学习科普贴
深度学习用FPGA和GPU各自的优缺点是什么呢？

--quray

2. Re:Java中的异常处理：何时抛出异常，何时捕获异常？
大神厉害总结的折磨详细，不过光看是没有用的，没有实际遇见过也不到怎样处理，我给大家分享一些练手的资料赶快来练习一下吧，（会以学习卡加密码的请示分享 需要打开连接右下角找人工注册免费的）...

--java大神带小白

3. Re:对分布式事务、消息队列的重新认识
1) 支付宝在扣款事务提交之前，向实时消息服务请求发送消息，实时消息服务只记录消息数据，而不真正发送，只有消息发送成功后才会提交事务；2) 当支付宝扣款事务被提交成功后，向实时消息服务确认发送。只有在得到.....

--MikeJia

4. Re:分治法与递归编程步骤

MARK

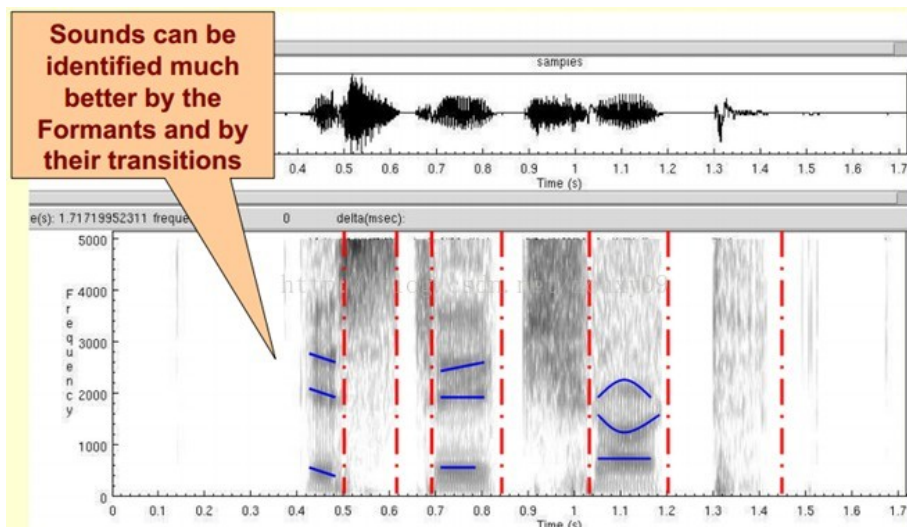
--dgdyq

5. Re:JVM学习笔记：JVM的体系结构与JVM的生命周期
有用! 谢谢

--reddit

推荐排行榜

1. 最容易读进去的深度学习科普贴(7)
2. golang学习之旅：使用go语言操作mysql数据库(1)
3. webx roadmap(1)
4. 神经网络与深度学习（5）：梯度消失问题(1)
5. 神经网络与深度学习（4）：改进神经网络的学习方法(1)



那我们为什么要在声谱图中表示语音呢？

首先，音素（Phones）的属性可以更好的在这里面观察出来。另外，通过观察共振峰和它们的转变可以更好的识别声音。隐马尔科夫模型（Hidden Markov Models）就是隐含地对声谱图进行建模以达到好的识别性能。还有一个作用就是它可以直观的评估TTS系统（text to speech）的好坏，直接对比合成的语音和自然的语音声谱图的匹配度即可。

总结：

对于一段语音信号 $x(t)$,

- (1) 首先分帧，变为 $x(m,n)$ (n 为帧长， m 为帧的个数)；
- (2) 然后做FFT变换，得到 $X(m,n)$ ，做周期图 $Y(m,n)$ ($Y(m,n) = X(m,n) * X(m,n)'$) ；
- (3) 接着取 $10 * \log_{10}(Y(m,n))$ ，把 m 根据时间变换一下刻度 M ， n 根据频率变化一下刻度 N ；
- (4) 最后将 $(M,N, 10 * \log_{10}(Y(m,n)))$ 画成二维图就是语谱图了。

共轭复数，两个实部相等，虚部互为相反数的复数互为共轭复数(conjugate complex number)。复数 z 的共轭复数记作 z' 。

$$z \cdot z' = |z|^2 = a^2 + b^2 \quad (\text{实数}) ;$$

在matlab中 $\text{conj}(x)$ 是求 x 的共轭数。

问题：

序列的DFT的输出是什么？

一个序列可以看成是一个函数。比如序列 $[a,b,c,d]$ ，可以看成是一个函数 $f(x)$ 在 $x=1,2,3,4$ 出分别取值 a,b,c,d 。

由于任何周期函数都可以用正弦函数和余弦函数构成的无穷级数来表示。

那么对该序列做DFT的意义就是找出这个函数 $f(x)$ 是由哪些振幅和频率的正弦余弦函数构成的。

参考链接

[傅里叶变换终极解释](#)

[语音信号处理之（四）梅尔频率倒谱系数（MFCC）](#)

[信号与频谱](#)

[语音信号语谱图分析](#)

<http://www.frank-zalkow.de/en/code-snippets/create-audio-spectrograms-with-python.html?ckattempt=1&i=1>

http://old.sebug.net/paper/books/scipydoc/frequency_process.html

参考书籍：

王永琦. MATLAB与音视频技术[M]. 清华大学出版社, 2013. P125

好文要顶

关注我

收藏该文







凝静志远

关注 - 10

粉丝 - 26

+加关注

« 上一篇: [Ubuntu14.10安装TensorFlow1.0.1](#)
» 下一篇: [《Maven实战》笔记](#)

posted @ 2017-01-07 13:51 凝静志远 阅读(1974) 评论(0) 编辑 收藏

0 0

[刷新评论](#) [刷新页面](#) [返回顶部](#)

- 注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。
- 【推荐】超50万VC++源码: 大型组态工控、电力仿真CAD与GIS源码库！
- 【免费】要想入门学习Linux系统技术，你应该先选择一本适合自己的书籍
- 【前端】SpreadJS表格控件，可嵌入应用开发的在线Excel
- 【直播】如何快速接入微信支付功能



最新IT新闻:

- 把13亿中国人拉到一个微信群会发生什么？腾讯给出了回答
 - Adobe与云营销软件开发商Marketo谈判 商讨收购事宜
 - Twitter法务主管：在适当情形 Twitter会屏蔽特朗普
 - 杭州一科技企业硅谷买楼建创新中心
 - 微软悄悄地撤下有争议的Edge弹出广告 称这是Insider测试的一部分
- » 更多新闻...



华为全联接大会 | 上海 | 2018.10.10-12

「大会门票+云服务器」专属套餐0.35折起



最新知识库文章:

- 为什么说 Java 程序员必须掌握 Spring Boot ?
 - 在学习中，有一个比掌握知识更重要的能力
 - 如何招到一个靠谱的程序员
 - 一个故事看懂“区块链”
 - 被踢出去的用户
- » 更多知识库文章...

历史上的今天:

2015-01-07 Hadoop版本变迁