目录

摘要 - 2 -
Abstract - 2 -
第一章 绪论 - 3 -
1.1 背景 - 3 -
1.2 语音特性提取的重要性 - 3 -
第二章 倒谱的相关知识 - 4 -
2.1.倒谱和复倒谱 - 4 -
2.1.1 倒谱和复倒谱的定义 - 4 -
2.1.2 倒谱和复倒谱的关系 - 4 -
2.2.倒谱的特点 - 5 -
2.3.求倒谱的算法 - 7 -
第三章 MFCC 参数的提取 - 9 -
3.1.MFCC 的原理 - 9 -
3.2.MFCC 算法流程 - 11 -
3.3.差分特征参数的提取 - 11 -
3.4.MATLAB 中的设计与实现 - 11 -
第四章 倒谱法提取基音频率 - 13 -
4.1.基音的相关知识 - 13 -
4.1.1.基音的周期 - 13 -
4.1.2. 基音检测的难点 - 13 -
4.2.提取基音的方法 - 14 -
4.3.倒谱分析算法的原理 - 14 -
4.4.MATLAB 中的设计与实现 - 15 -
第五章 倒谱法提取共振峰 - 16 -
5.1.共振峰的概念 - 16 -
5.2.提取共振峰的方法 - 16 -
5.3.倒谱法的原理 - 17 -
5.4.MATLAB 中的设计与实现 - 17 -
第六章 结束语 - 20 -
附录 - 21 -
1 提取 MFCC 参数的相关程序 - 21 -
1.1 mfcc.m - 21 -
1.2 enframe.m - 21 -
1.3 mel.m - 23 -
2 提取基因和共振峰的程序 - 25 -
致谢 - 26 -

基于倒谱的语音特性提取算法设计及其实现

赵丽君

西南大学 电子信息工程学院，重庆 400715

摘要：在语音信号处理中，常用的语音特性是基于 Mel 频率的倒谱系数（MFCC）以及一些语音信号的固有特征，如共振峰和基音频率等。倒谱可以较好地将语音信号中的激励信号和声道响应分离，并只需要用十几个倒谱系数就能较好地描述语言信号的声道响应，在语音信号处理中占有很重要的位置。本论文设计了基于倒谱的语音特性参数提取算法，并在 Matlab 中予以实现。

关键词：倒谱；MFCC；基音；共振峰

The Design and Implementation of Cepstrum-based Algorithm in Voice Characteristic Extraction

Zhao Lijun

School of Electronic & Information Engineering, Southwest University, Chongqing 400715, China

Abstract: In voice signal processing, MFCC and some inherent characteristics of voice signals, such as formants and the frequency of pitch. Cepstrum can be used to separate the excitation signal and channel response, and can represent channel response with only a dozen cepstral coefficients. As a result, it has been a very important role in voice signal processing. In this paper, the cepstrum-based algorithm to extract above-mentioned voice characteristics and its implementation in MATLAB are described in detail.

Key word: Cepstrum; MFCC; pitch ; formant

第一章 绪论

1.1 背景

由于语言是人们在日常生活中的主要交流手段，因此语音信号处理在现代信息社会中占用重要地位。随着语音信号处理技术在实际生活中的应用的不断发展，语音信号处理技术已经被广泛地接受和使用。由于语音比其他形式的交互方具有更多的优势，因此这项技术已经越来越贴近人们的生活。目前，语音信号处理技术处于蓬勃发展时期，不断有新的产品被研制开发，市场需求逐渐增加，具有良好的应用前景。

1.2 语音特性提取的重要性

语音信号处理虽然包括语音通信，语音合成和语音识别等方面的内容，但其前提和基础是对语言信号进行分析。语音的压缩与恢复是语音信号处理的关键技术。近年来有关这方面的研究不断发展成熟，并形成一系列的标准。在语音信号的各种分析合成系统中，需要提取频谱包络参数，推测音源参数（清浊音的判定以及浊音周期等）。只有将语音信号分析表示成其本质特性的参数，才有可能利用这些参数进行高效的语音通信，才能建立用于语音合成的语音库，也才能建立用于识别的模板或知识库。

根据所分析的参数不同，语音信号分析可分为时域，频域，倒谱域等方法。进行语音信号分析时，最先接触到的，也是最直观的是它的时域波形。时域分析具有简单直观，清晰易懂，运算量小，物理意义明确等优点；但更为有效的分析多是围绕频域进行的，因为语音中最重要的感知特性反映在其功率谱中，而相位变化只起着很小的作用[1]。

频谱分析具有如下优点：时域波形较易随外界环境变化，但语音信号的频谱对外界环境变化具有一定的顽健性。另外，语音信号的频谱具有非常明显的声学特性，利用频域分析获得的语音特征具有实际的物理意义。如 MFCC,共振峰，基音周期等参数。

倒谱域是将对数功率谱进行反傅立叶变换后得到的，它可以进一步将声道特性和激励特性有效地分开，因此可以更好地揭示语音信号的本质特性。本文给出语音特性的提取中基于倒谱的算法设计及其实现。使读者对相关技术的基本理论，方法和基本应用有一个系统的了解。

第二章 倒谱的相关知识

2.1.倒谱和复倒谱

2.1.1 倒谱和复倒谱的定义

语音信号不是加性信号，而是卷积信号。为了能用线性系统对其进行处理，可以先采用卷积同态系统处理。经过卷积同态系统后输出的伪时序序列称为原序列的“复倒频谱”。它的定义式可以表示为：

$$\hat{x}(n) = IFT\{\ln[FT\{x(n)\}]\} \tag{2-1}$$

倒谱或称“倒频谱”的定义为：

$$c(n) = IFT\{\ln|FT[x(n)]|\} \tag{2-2}$$

它和复倒谱的主要区别是对序列对数幅度谱的傅立叶逆变换，它是复倒谱中的偶对称分量。它们都将卷积运算，变为伪时域中的加法运算，使得信号可以运用满足叠加性的线性系统进行处理。复倒谱涉及复对数运算，而倒谱只进行实数的对数运算，较复倒谱的运算量大大减少[2]。

如果$c_1(n)$和$c_2(n)$分别是$x_1(n)$和$x_2(n)$的倒谱，x(n)= $x_1(n)*x_2(n)$,那么 x(n)的倒谱 c(n)= $c_1(n)+c_2(n)$。

2.1.2 倒谱和复倒谱的关系

如果已知一个实序列 x(n)的复倒谱$\hat{x}(n)$，那么可以由$\hat{x}(n)$求出它的倒谱 c(n)。为此首先将$\hat{x}(n)$表示为一个偶对此序列$\hat{x}_c(n)$和一个奇对称序列$\hat{x}_o(n)$之和：

$\hat{x}(n)=\hat{x}_c(n)+\hat{x}_o(n)$ 其中 $\hat{x}_c(n)=\hat{x}_c(-n)$，$\hat{x}_o(n)=-\hat{x}_o(-n)$

易于证明

$$\hat{x}_c(n)= \frac{1}{2}[\hat{x}(n)+\hat{x}(-n)] \tag{2-3}$$

$$\hat{x}_o(n)= \frac{1}{2}[\hat{x}(n)-\hat{x}(-n)] \tag{2-4}$$

由于一个偶对称序列的 DTFT 是一个实函数，而一个奇对称序列的 DTFT 是一个虚函数，可得

$$\hat{x}_c(n) = F^{-1}[\mathrm{Re}[\hat{X}(\exp jw)]] = F^{-1}[\ln|X(\exp jw|] \quad (2\text{-}5)$$

$$c(n) = F^{-1}[\ln|X(\exp jw)|] \quad (2\text{-}6)$$

$$c(n) = \hat{x}_c(n) = \frac{1}{2}[\hat{x}(n) + \hat{x}(-n)] \quad (2\text{-}7)$$

这样，由 $\hat{x}(n)$ 即可求得 c(n)。如果设

$$p(n) = F^{-1}[Arg|X(\exp jw)|] \quad (2\text{-}8)$$

那么可以同理导出：

$$p(n) = \hat{x}_o(n) = \frac{1}{2}[\hat{x}(n) - \hat{x}(-n)] \quad (2\text{-}9)$$

$p(n)$ 称为“相位倒谱”，不难看出，c(n)表现的是 x(n)的 DTFT $X(\exp jw)$ 的模函数的特征，$p(n)$ 表现的是 $X(\exp jw)$ 相位函数的特征，而 $\hat{x}(n)$ 包括两个方面的特征。

只有当 x(n)是一个因果最小相位序列时 $\hat{x}(n)$ 才是一个因果稳定序列。x(n)应满足两个条件。第一，x(n)=x(n)u(n)。第二，X(Z)=Z[x(n)]的零极点皆成为 $\hat{X}(N)$ 的极点。这样，只有当 X(Z)的零极点皆在单位圆内时才能使 $\hat{X}(N)$ 的极点全在单位圆内，这样才能保证 $\hat{x}(n)$ 是一个因果稳定序列。只有当 x(n)是一个反因果最大相位序列时，$\hat{x}(n)$ 才是一个反因果稳定序列。它的条件与前一情况正好完全相反[3]。

这样，只要 x(n)是因果最小相位序列或反因果最大相位序列，便可以由 c(n)算出 $\hat{x}(n)$。

2.2.倒谱的特点

假设所处理的语音信号是一个离散时域中的实序列 x(n),由于对语音信号必须进行短时分析，x(n)的非零间隔[N_1，N_2]必然是一个有限间隔，为了便于分析与计算，一般设置 N_1=0，N_2=N-1，这时间隔内共有 N 个样点。此时 x(n)的 Z 变换 X(Z)可以表示为如

下形式：

$$X(Z)=\sum_{n=0}^{N-1}x(n)z^{-n}=AZ^{-N_B}\prod_{i=1}^{N_A}(1-a_iz^{-1})\prod_{j=1}^{N_B}(1-b_jZ) \qquad (2\text{-}10)$$

其中$|\alpha_i|<1$，$|b_j|<1$。$Z=\alpha_i$，i=1～N_A，是X(Z)在单位圆内的零点。$Z=1/b_j$，j=1～N_B，是X(Z)在单位圆外的零点。N_A+N_B=N-1。A是一个实数，它可以根据下列公式计算：

$$x(0)=A\prod_{j=1}^{N_B}[-b_j] \qquad (2\text{-}11)$$

如果x(n)是最小相位序列，那么N_B=0，N_A=N-1，且x(0)=A.

借助与式（2-3）可以求得该序列的复倒谱$\hat{x}(n)$。首先求X(Z)的对数，得到$\hat{X}(Z)$如下。

$$\hat{X}(Z)=\ln A+\ln Z^{-N_B}+\sum_{i=1}^{N_A}\ln(1-a_iZ^{-1})+\sum_{j=1}^{N_B}\ln[1-b_jZ] \qquad (2\text{-}12)$$

此式右侧第二项$\ln Z^{-N_B}$是一个表示延迟量大小的项，它不包含有关序列 x(n)特征的任何有用信息，相反，可以证明，它的存在会对有用信息造成干扰。事实上，如果将间隔[N_1，N_2]的起点N_1=0改变为$N_1=N_B$，此项就消失了。如果为了方便，永远选N_1=0，那么当x(n)为非最小相位时，就需要采取措施将其消除，如果这个第二项已被消除，便对式（2-5）右侧第三，四两项和式中的每个对数在单位圆（|Z|=1）上用台劳级数展开，就可以得到下列表达式：

$$\ln(1-a_iZ^{-1})=\sum_{n=1}^{\infty}-\frac{a_i^n}{n}Z^{-n} \qquad (2\text{-}13)$$

$$\ln(1-b_iZ)=\sum_{n=1}^{\infty}-\frac{b_j^n}{n}Z^{n} \qquad (2\text{-}14)$$

这样式（2-5）可表达为下列形式（右侧第二项已去除）：

$$\hat{X}(Z)=\ln A+\sum_{n=1}^{\infty}-[\sum_{i=1}^{N_A}\frac{a_i^n}{n}Z^{-n}]Z^{-n}+\sum_{n=1}^{\infty}-[\sum_{j=1}^{N_B}\frac{b_j^n}{n}]Z^{n} \qquad (2\text{-}15)$$

对照 $\hat{X}(Z)=\sum_{n=-\infty}^{+\infty}\hat{x}(n)Z^{-n}$ 立即可以得到：

$$\hat{x}(n)=\begin{cases}-[\sum_{i=1}^{N_A}\frac{a_i^n}{n}] & ,\ n>0\\ \ln A & ,\ n=0\\ -[\sum_{j=1}^{N_B}\frac{b_j^n}{-n}] & ,\ n<0\end{cases} \tag{2-16}$$

由式（2-9）可以看到，x(n)随着|n|的增大而呈减小趋势，当各$|a_i|$，$|b_j|$越接近于零，其衰减速度越快[3]。

由于$c(n)=\frac{1}{2}[\hat{x}(n)+\hat{x}(-n)]$，倒谱 c(n)随 n 的变化规律与$\hat{x}(n)$大致相似，只是 c(n)是围绕原点对称的衰减序列，而$\hat{x}(n)$是非对称的。

2.3.求倒谱的算法

假设被处理的序列 x(n)所占的间隔是[0,N-1]。这里所用的间隔长度 N 并不一定确切等于 x(n)的实际长度，它可以选得比实际长度大一些。N 选的大一些可以达到两个目的，第一是防止求出的 c(n)中有混叠存在，第二是使它所代表的离散时域频谱有更佳的分辨率。当 N 大于 x(n)的实际长度是，可以在 x(n)的后方添若干个零来补足所需的长度，这称为“补零”。用 DFT 和 IDFT 实现的同态处理特征系统如下所列。

特征系统D^*：

$$x(k)=\sum_{n=0}^{N-1}x(n)\exp(-j\frac{2\pi}{N}nk),\quad k=0\sim(N-1) \tag{2-17}$$

$$C(k)=\ln|X(k)|,\quad k=0\sim(N-1) \tag{2-18}$$

$$C_P(n)=\frac{1}{N}\sum_{n=0}^{N-1}C(k)\exp(j\frac{2\pi}{N}nk),\quad n=0\sim(N-1) \tag{2-19}$$

由此求出的$C_P(n)$与真实的 c(n)的关系是：

$$C_P(n)=[\sum_{r=-\infty}^{+\infty}c(n+rN)]R_N(n) \tag{2-20}$$

它的防混叠条件是 $N > 2\max\{n_a, n_b\}$。 $\max\{n_a, n_b\}$ 表示取 $|n_a|$，$|n_b|$ 中的最大值。

在语音信号处理中 x(n)的实际长度一般为100□200。而 N 值一般选为 256，512 或 1024,这时既有足够高的分辨率和避免混叠的能力又具有相应的高效 FFT 算法可资利用。

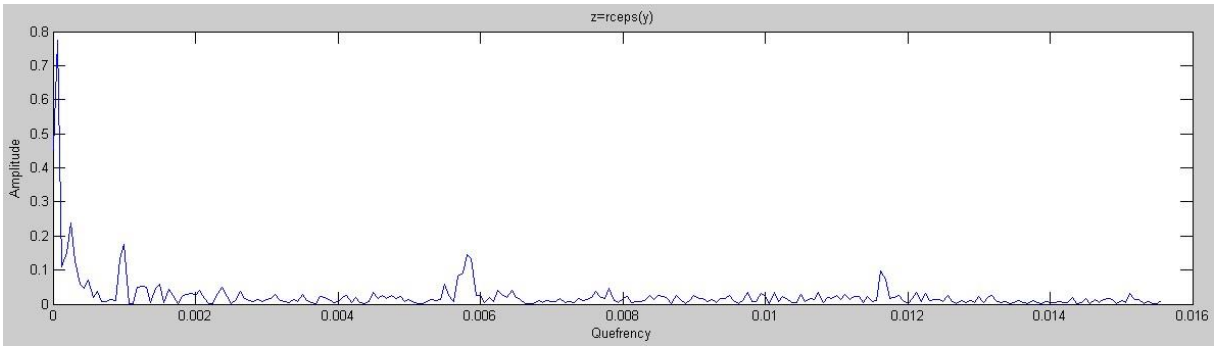

图 2.1 MATLAB 中所作的倒谱图

Figure2.1 Cepstrum Figure in MATLAB

rceps(y)为 MATLAB 提供的倒频谱函数，通过对时域信号的傅里叶变换的幅值求自然对数，然后再做傅里叶逆变换。编程时可直接运用。

第三章 MFCC 参数的提取

3.1.MFCC 的原理

在语音识别和说话人识别中，常用的语音特征是基于 Mel 频率的倒谱系数（mel frequency cepstrum coefficient,MFCC）.由于 MFCC 参数是将人耳的听觉感知特征和语音的产生机制相结合，因此目前大多数语音识别系统中广泛使用这种特征。

人的耳朵具有一些特殊的功能，这些功能使得人耳能够从嘈杂的背景噪声中，以及各种变异情况下听到语音信号，这是因为人的内耳基础膜对外来信号会产生调节作用。对不同的频率，在相应的临界带宽内的信号会引起基础膜上不同位置的振动。由此可用带通滤波器组来模仿人耳听觉，从而减少噪声对语音的影响。

耳蜗实质上相当于一个滤波器组，耳蜗的滤波作用是在对数频率尺度上进行的，在 1000Hz 以下为线性尺度，而 1000Hz 以上为对数尺度，这就使得人耳对低频信号比对高频信号更敏感。根据这一原则，研究者根据心理学实验得到了类似于耳蜗作用的一组滤波器组，就是 Mel 频率滤波器组。对频率轴的不均匀划分是 MFCC 特征的特点。将频率变换到 Mel 域后，Mel 带通滤波器组的中心频率是按照 Mel 频率刻度均匀排列的[4]。

设语音信号的 DFT 为

$$X_a(k)=\sum_{n=0}^{N-1}x(n)e^{-j2\pi nk/N}\quad 0\le k<N \tag{3-1}$$

其中 x(n)为输入的语音信号,N 表示傅立叶变换的点数。

我们定义一个 M 个滤波器组,采用的滤波为三角滤波器，中心频率为 f(m),m=1,2, , ...M,则三角滤波器的频率响应按式(3-2)定义，频率响应波形如图(3.1)所示。

$$H_m^{'}(k)=\begin{cases}0 & k<f(m-1)\\ \dfrac{2(k-f(m-1))}{(f(m+1)-f(m-1)(f(m)-f(m-1))} & f(m-1)\le k\le f(m)\\ \dfrac{2(f(m+1)-k)}{(f(m+1)-f(m-1))(f(m+1)-f(m))} & f(m)\le k\le f(m+1)\\ 0 & k>f(m+1)\end{cases} \tag{3-2}$$

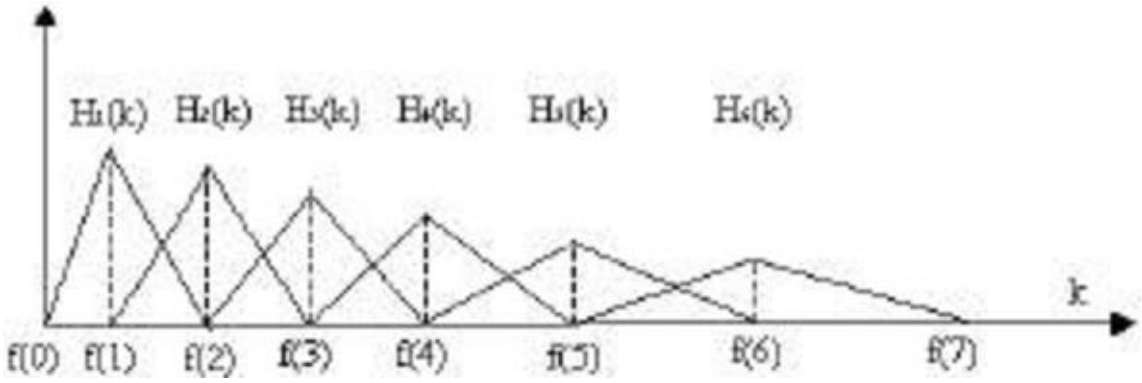

图 3.1 用于计算 Mel 倒谱的三角滤波器

Figure2.1 The Triangular Filter Used to Calculate The Mel Cepstrum

为便于计算，本文将式(3-2)的三角滤波器简化为

$$H_m^{'}(k)=\begin{cases}0 & k<f(m-1)\\ \dfrac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1)\le k\le f(m)\\ \dfrac{f(m+1)-k}{f(m+1)-f(m)} & f(m)\le k\le f(m+1)\\ 0 & k>f(m+1)\end{cases} \tag{3-3}$$

其中 $\sum_{m}^{M-1}H_m^{'}(k)=1$。

Mel 滤波器的中心频率定义为：

$$f(m)=\frac{N}{Fs}B^{-1}(B(f_l)+m\frac{B(f_h)-B(f_l)}{M+1}) \tag{3-4}$$

其中 f_h 和 f_l 分别为滤波器组的最高频率和最低频率，Fs 为采样频率，单位为 Hz 。M 是滤波器组的数目，N 为 FFT 变换的点数，式中 $B^{-1}(b)=700(e^{\frac{b}{1125}}-1)$。

每个滤波器组的输出的对数能量为

$$S(m)=\ln(\sum_{k=0}^{N-1}|X_a(k)|^2\ H_m(k))\quad ,0\le m<M \tag{3-5}$$

经余弦变换得到 MFCC 系数

$$C(n)=\sum_{m=0}^{M-1}S(m)\cos(\pi n(m+0.5)/M)\quad ,\quad 0\le n<M \tag{3-6}$$

3.2.MFCC 算法流程

设某语音信号为 x(n) ，则算法处理流程为

1）预加重 $x_n^{'} = x_n - kx_{n-1}$，其中 k 为预加重系数，一般取 0.95；

2）加窗（hamming 窗），帧长为 N；

3）DFT 变换；

4）设计一个具有 M 个带通滤波器的滤波器组，采用三角滤波器，中心频率从 0~F/2 间按 Mel 频率分布；

5）按式 (3-5)计算每个滤波器组输出的对数能量；

6）按式 (3-6)求得 MFCC 系数。

3.3.差分特征参数的提取

在提取了 MFCC 参数后，可用式(3-7)的差分特征参数提取算法提取ΔMFCC，ΔΔMFCC 参数。

$$d_t = \begin{cases} c_{t+1} - c_t & t < \Theta \\ c_t - c_{t+1} & t \geq T - \Theta \\ \dfrac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} & \text{其它} \end{cases} \qquad (3\text{-}7)$$

式中 d_t 表示第 t 个一阶差分倒谱系数，T 是为倒谱系数的维数，Θ表示一阶导数的时间差，其值取 1 或 2，$1 \leq \theta \leq \Theta$，$c_t$ 表示第 t 个倒谱系数[5]。

3.4.MATLAB 中的设计与实现

MATLAB 中，取 Mel 滤波器的阶数为 24，fft 变换的长度为 256，采样频率为 8000Hz 预加重后，对语音信号分帧（每 256 点分为一帧），计算每帧的 MFCC 参数后，求取差分系数。合并 MFCC 参数和一阶差分 MFCC 参数，可得到如下结果。

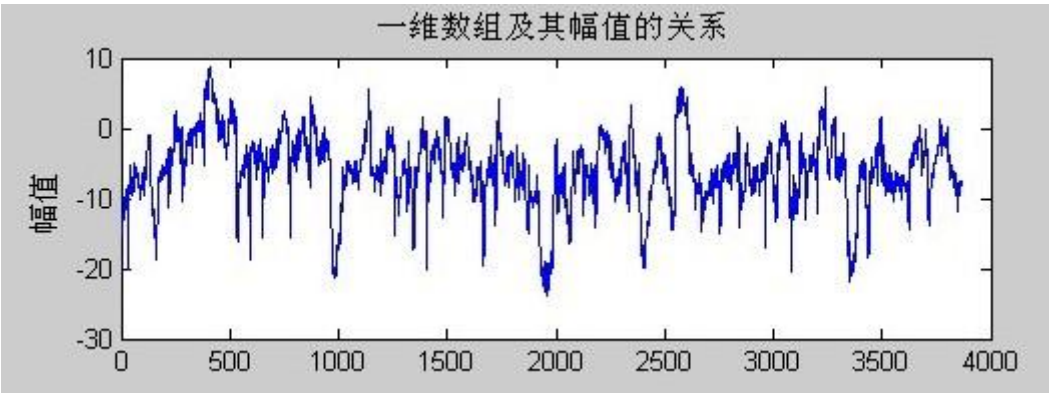

图 3.2 一维数组及其幅值的关系

Figure3.2 The Relationship Between One-dimensional Array and Amplitude

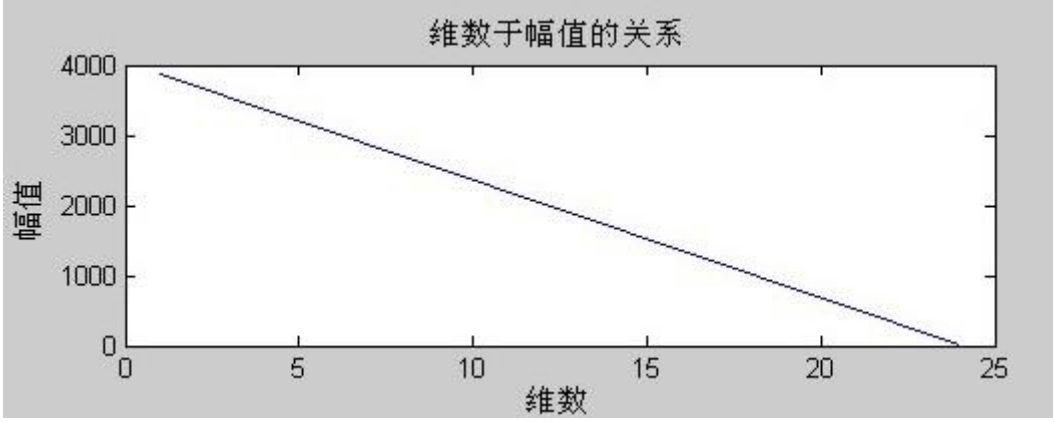

图 3.3 维数与幅值的关系

Figure3.3 The Relationship Between Dimension and Amplitude

由上图则可全面的了解 MFCC 的静态及动态特性。

第四章 倒谱法提取基音频率

4.1.基音的相关知识

4.1.1.基音的周期

基音是指发浊音时声带振动所引起的周期性，而基音周期是指声带振动频率的倒数。基音周期是语音信号最重要的参数之一，提取该参数是语音信号处理中一个十分重要的问题。对于汉语这种有调语音，基音的变化模式称为声调，它携带着非常重要的具有辨意作用的信息，有区别意义的功能。根据加窗的短时语音帧来估计基音周期，在语音编解码器，语音识别，说话人确认和辨认，对生理缺陷人的辅助系统等许多领域都是重要的一环。

4.1.2. 基音检测的难点

自进行语音信号分析研究以来，基音检测一直是一个重点研究的课题，很多方法已被提出，然而这些方法都有它们的局限性。迄今为止，尚未找到一个完善的可以适用于不同的说话人，不同的要求和环境的基音检测方法。

基音检测的主要难点表现在：

1）语音信号变化十分复杂，声门激励的波形并不是一个完全的周期序列。在语言的头，尾部并不具有声带振动那样的周期性，对有些清浊音的过渡帧很难判定它应属于周期性或非周期性，从而就无法估计出基因周期。

2）要从语音信号中去除声道的影响，直接取出仅与声带振动有关的声源信息并非易事。而声道共振峰有时会严重影响激励信号的谐波结构。

3）在浊音段很难精确地确定每个基音周期的开始和结束位置，这不仅因为语音信号本身是准周期的，也是因为波形的峰受共振峰结构，噪声等影响较大。

4）基音周期变化范围较大，从低音男声的 80Hz 直到女孩的 500Hz，这也给基音周期的检测带来了一定的困难。另外，浊音信号可能包含有三四十次谐波分量，而基波分量往往不是最强的分量。因为语音的第一共振峰通常在300~1000Hz 范围内，这就是说，2~8次谐波成分往往比基波分量还强。丰富的谐波成分使语音信号的波形变的很复杂，给基音检测带来困难，经常发生基频估计结果为实际基音频率的二三次倍频或二次分频的情况[6]。

4.2.提取基音的方法

目前基音的提取方法大致可以分为三类：

1）波形估计法。直接由语音波形来估计基音周期，分析出波形上的周期峰值。包括并行处理法，数据减少法等。

2）相关处理法。这种方法在语音信号处理中广泛使用，这是因为相关处理法抗波形的相位失真能力强，另外它在硬件处理上结构简单。包括波形自相关法，平均振幅差分函数法（AMDF），简化逆滤波法（SIFT）等。

3）变换法。将语音信号变换到频域或倒谱域来估计基音周期，利用同态分析方法将声道的影响消除，得到属于激励部分的信息，进一步求取基音周期，比如倒谱法。虽然倒谱分析算法比较复杂，但基音估计效果较好[4]。

4.3.倒谱分析算法的原理

对语音信号利用倒谱解卷原理，可以得出激励序列的倒谱，它具有与基音周期相同的周期，因此可以容易且精确地求出基音周期。

在发浊音时，声门激励是以基因周期为周期的冲激序列：

$$x(n)=\sum_{r=0}^{M}\alpha_r\delta(n-rN_P)$$

式中，M 是正整数；r 是正整数，且 0≤r≤M;α_r 是幅度因子；N_p 是基音周期（用样点数表示的）。根据复倒谱的定义，可以得到 x(n)的复倒谱为：

$$\hat{x}(n)=\sum_{k=0}^{\infty}\beta_k\delta(n-kN_P)$$

其中，$\beta_0=\ln\alpha_0$

$$\beta_k=-\frac{1}{k}\sum_{r=1}^{M}\alpha_r^k=-\frac{1}{k}\sum_{r=1}^{M}(\frac{\alpha_r}{\alpha_0})^k$$

从上式得出的结论为：一个周期冲激的有限长度序列，其复倒谱也是一个周期冲激序列，而且长度 N_p 不变，只是序列变为无限长度序列[5]。同时其幅度随着 k 值的增大而衰减，衰减速度比原序列要快。倒谱是复倒谱的偶对称分量，它同样具有与基音周期相同的周期，因而能容易且精确地求出基音周期。

4.4.MATLAB 中的设计与实现

当语音采样率 f_s=10kHz 时，倒谱的第一个峰值点即等于基音周期值 N_p，其变化范围在 25□ 200 之间，因而应在此范围内搜索峰值点。为了实现此搜索，语音帧数至少应该等于 200 点（即等于 20ms）。

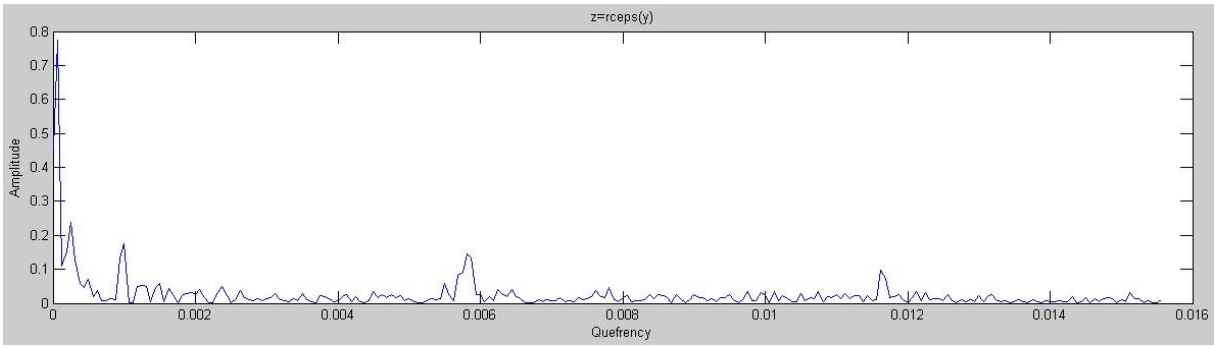

图 4.1 由 rceps 函数得到的倒谱图

Figure4.1 Cepstrum Figure from Rceps Function

图（4.1）为MATLAB中，运用rceps函数绘制的倒谱图，在图中可以清楚的发现0.006s附近的峰值点即为所求点。利用以下程序段可精确求取基因频率：

```
[Cmax Cloc]=max(abs(z(25:200)));
T0=(Cloc+40)*dt;
F0=1/T0;
fprintf('Fundamental Frequency F0=  %5.4fHz\n',F0);
```

所得结果为：

```
Fundamental Frequency F0=  170.2128Hz
```

第五章 倒谱法提取共振峰

5.1.共振峰的概念

共振峰是反映声道谐振特性的重要特征，它代表了发音信息的最直接的来源，而且人在语音感知中利用了共振峰信息。所以共振峰是语音信号处理中非常重要的特征参数，已经广泛地用作语音识别的主要特征和语音编码传输的基本信息。共振峰信息包含在频率包络之中，因此共振峰参数提取的关键是估计自然语音频谱包络，一般认为谱包络中的最大值就是共振峰[7]。与基因检测类似，共振峰估计也是表面上看起来很容易，而实际上又受很多问题困扰。这些问题包括：

1）虚假峰值。在正常情况下，频谱包络中的极大值完全是又共振峰引起的。但在线性预测分析方法出现之前的频谱包络估计器中，出现虚假峰值是相当普遍的现象。甚至在采用线性预测方法时，也并非没有虚假峰值。为了增加灵活性会给预测器增加2~3个额外的极点，有时可利用这些极点代表虚假峰值。

2）共振峰合并。相邻共振峰的频率可能会靠的太近而难以分辨。这时会产生共振峰合并现象，而探讨一种理想的能对共振峰合并进行识别的共振峰提取算法存在很多实际困难。

3）高音调语音。传统的频谱包络估计方法是利用由谐波峰值提供的样点。高音调语音（如女声和童生）的谐波间隔比较宽，因而为频谱包络估值所提供的样点比较少，所以谱包络的估计就不够精确。即使采用线性预测进行频谱包络估计也会出现这个问题。在这样的语音中，线性预测包络峰值趋向于离开真实位置，而朝着最接近的谐波峰位移动[7]。

5.2.提取共振峰的方法

提取共振峰的几种常用方法包括：

1）基于线性预测的共振峰求取方法。一种有效的频谱包络估计方法是从线性预测分析角度推导出声道滤波器，根据这个声道滤波器找出共振峰。虽然线性预测法也有一定的缺点，例如其频率灵敏度与人耳不相匹配，但对于许多应用来说，它仍然是一种行之有效的方法。线性预测共振峰通常有两种途径可供选择：一种途径是利用一种标准的寻找复根的程序计算预测误差滤波器的根，称为求根法；另一种途径

是找出由预测器导出的频谱包络中的局部极大值，称为选峰法。

2）倒谱法。声道响应的倒谱衰减很快，在[-25,25]之外的值相当小，因此可以构造一个相应的倒谱滤波器，将声道的倒谱分离，对分离出来的倒谱做相应的反变换，就可以得到声道函数的对数谱，对此做进一步处理即可求得所需的各个共振峰[8]。

5.3.倒谱法的原理

选择最普遍的极零模式来描述声道相应 x(n),其 z 变换的形式为：

$$X(z)=|A|\frac{\prod_{k=1}^{mi}(1-a_k z^{-1})\prod_{k=1}^{mo}(1-b_k z)}{\prod_{k=1}^{pi}(1-c_k z^{-1})\prod_{k=1}^{po}(1-d_k z)} \tag{5-1}$$

经过傅立叶变换，取对数和逆傅立叶变换后可以得到其复倒谱：

$$\hat{x}(n)=\begin{cases}\ln|A| & (n=0)\\ \sum_{k=1}^{pi}\frac{c_k^n}{n}-\sum_{k=1}^{mi}\frac{a_k^n}{n} & (n>0)\\ \sum_{k=1}^{mo}\frac{b_k^{-n}}{n}-\sum_{k=1}^{po}\frac{d_k^{-n}}{n} & (n>0)\end{cases} \tag{5-2}$$

对于倒谱可以只考虑它的幅度特性，可以看出，它是一个衰减序列，且衰减的速度比 1/|n|快。因而它比原信号 x(n)更集中于原点附近，或者说它更具有短时性。

5.4.MATLAB 中的设计与实现

倒谱算法运用对数运算和二次变换将基音谐波和声道的频谱包络分离开来。根据其特点利用短时窗可以从语音信号倒谱 c(n)中截取出 h(n)。由 h(n)经 DFT 得到的 H(K)就是声道的离散谱曲线，由于它去除了激励引起的谐波动，因此能更精确地得到共振峰参数。

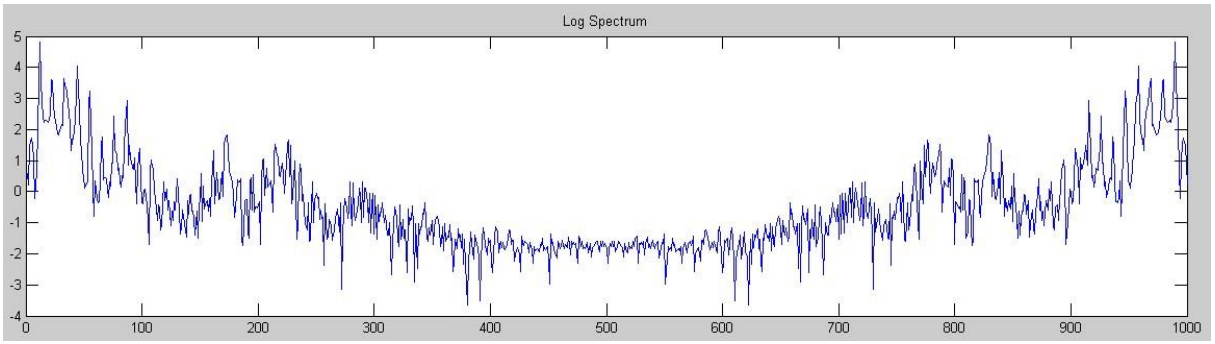

图 5.1 信号的对数频谱图

Figure5.1 The Log Spectrum Figure of Signal

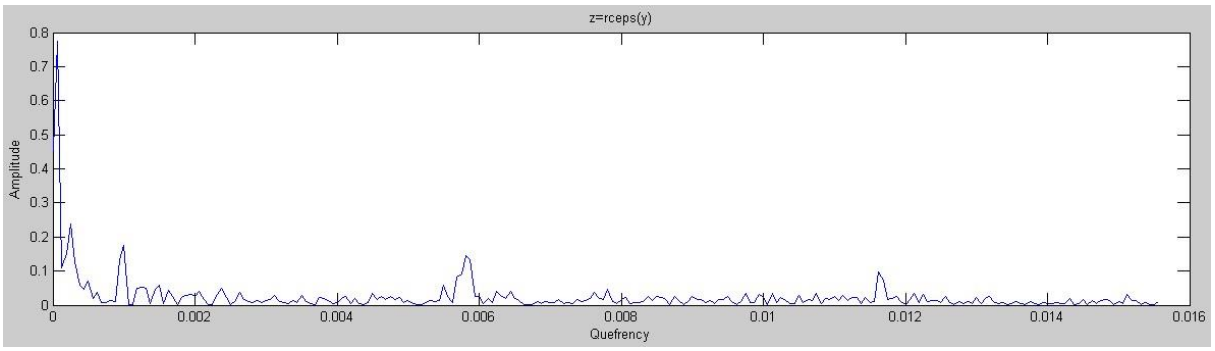

图 5.2 倒谱图

Figure5.2 Cepstrum Figure

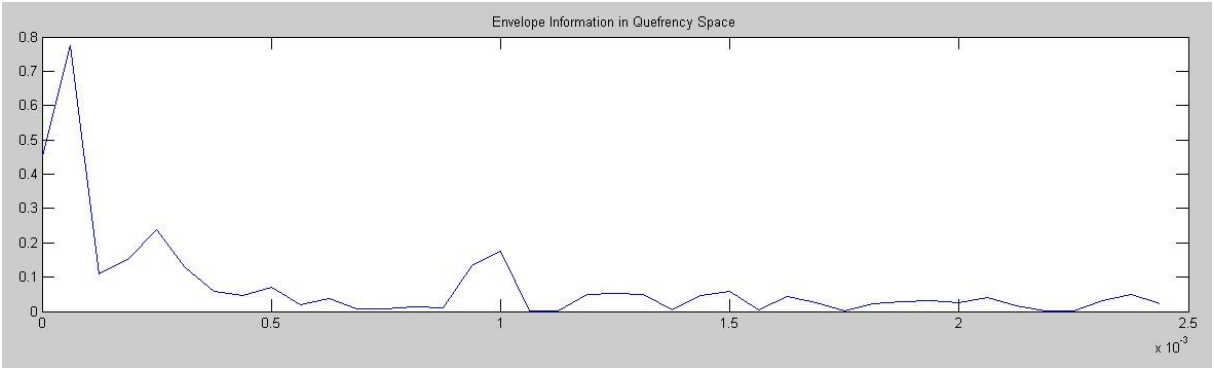

图 5.3 加窗截取部分倒谱图

Figure5.3 Cepstrum Figure after Window Interception

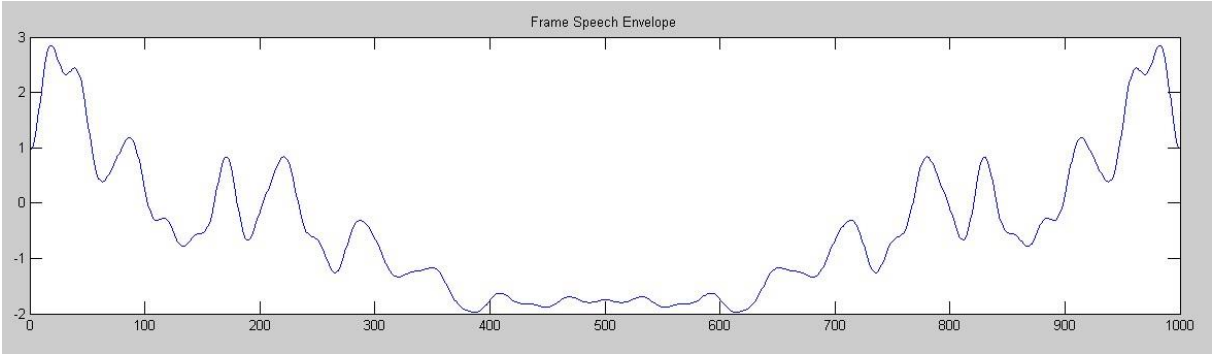

图 5.4 频谱包络图

Figure5.4 Envelope Spectrum Figure

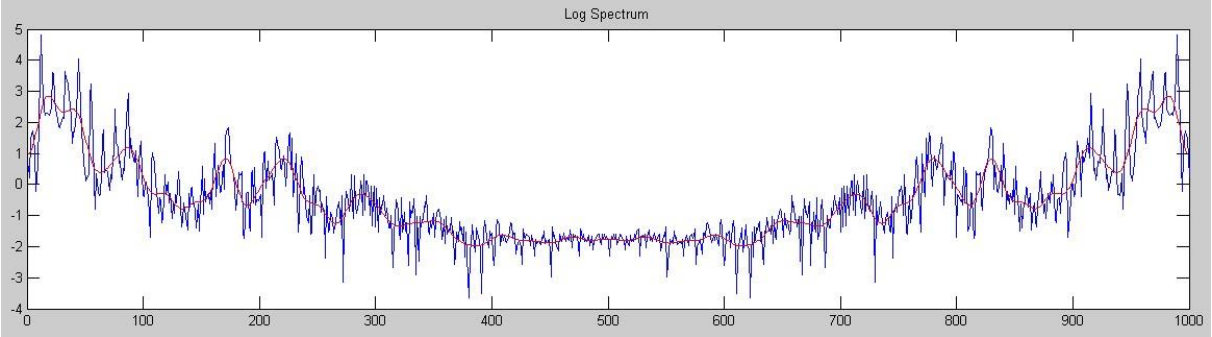

图 5.5 将对数频谱图和包络图绘制在同一个图上

Figure5.5 Drawing The Log Spectrum and Envelope in The Same Figure

MATLAB 中对信号做 fft 变换，绘制对数频谱图（5.1），运用 MATLAB 提供的倒谱函数 rceps 计算倒谱并绘制。对倒谱加窗后做 fft 变换，即得到频谱包络和共振峰信息。

第六章 结束语

本论文介绍了倒谱以及常用的语音特性参数 MFCC,基音频率和共振峰的相关知识和原理，并设计了基于倒谱的算法，在 MATLAB 中编程实现了以上参数的提取。由于作者对相关知识了解有限以及经验的不足，本文中处理的语音信号均采用简单的短时信号。各个程序虽然在 MATLAB 中得到了较好的效果，但在实际应用中，会面临很多具体的问题。需要考虑环境，说话人，应用要求等因素，去除各种影响才能取得好的分析结果。

倒谱法根据对数功率谱的逆傅立叶变换，能够分离频谱包络和细微结构，很精确地得到基音频率和共振峰信息，但它的运算量比较大。当采用无噪语音时，用倒谱进行基音提取的效果是很理想的。然而当存在加性噪声时，在对数功率谱的低电平部分会被噪声填满，从而掩盖了基音谐波的周期性。这意味着倒谱的输入不再是纯净的周期性成分，而倒谱中的基音峰值将会展宽，并受到噪声的污染从而使倒谱检测方法的灵敏度也随之下降。在基音估计中还可以使用经过中心削波或三电平削波后的自相关方法，这种方法在信噪比低的情况下可以获得良好的性能。

与基因检测类似，共振峰估计也是表面上看起来很容易，而实际上又受很多问题困扰。随着语言处理技术的发展，越来越多的语音特征提取方法被提出和完善，相信将克服各种困难，为人们的生活提供更多便利。

参考文献：

[1] 胡航.语音信号处理.哈尔滨：哈尔滨工业大学出版社，2000

[2] 陈永彬.语音信号处理.上海：上海交通大学出版社，1991

[3] M.M.Sondhi.New Methods of Pitch Extraction.IEEE Trana.AU,1968;16(1):262-266

[4] R.W.Schafer,L.R.Rabiner.System for Automatic Formant Analysis of Voiced Speed.J.A.S.A.,1997;47(2);634-648

[5] 王晓亚.倒谱在语音的基音和共振峰提取中的应用.无线电工程 2004（34--1）

[6] 杨行峻 ,迟惠生.数字语音信号处理.北京:电子工业出版社 ,1995

[7] RabinerL ,JuangB H. Fundamental of Speech Recognition.New York:Prentice Hall ,1993

[8] Furui S.Speaker Independent Isolated Word Recognition Using Dynamic Feature of Speech Spectrum. IEEE Trans on Acoustics,Speech,Signal Processing,1986,34 (1):52～59

附录

1 提取 MFCC 参数的相关程序

1.1 mfcc.m

```
close all
clear
clc
[x fs]=wavread('speech.wav');
bank=mel(24,256,fs,0,0.4,'m');%Mel 滤波器的阶数为 24，
fft 变换的长度为 256，采样频率为 8000Hz
% 归一化 mel 滤波器组系数
bank=full(bank);
bank=bank/max(bank(:));
% DCT 系数,12*24
for k=1:12
   n=0:23; dctcoef(k,:)=cos((2*n+1)*k*pi/(2*24));
end
% 归一化倒谱提升窗口
w = 1 + 6 * sin(pi * [1:12] ./ 12);
w = w/max(w);
% 预加重滤波器
xx=double(x);
xx=filter([1 -0.9375],1,xx);
% 语音信号分帧
xx=enframe(xx,256,80);%对 x 256 点分为一帧
% 计算每帧的 MFCC 参数
for i=1:size(xx,1)
   y = xx(i,:);
   s = y' .* hamming(256);
   t = abs(fft(s));%fft 快速傅立叶变换
   t = t.^2;
   c1=dctcoef * log(bank * t(1:129));
   c2 = c1.*w';
   m(i,:)=c2';
end
%求取差分系数
dtm = zeros(size(m));
for i=3:size(m,1)-2
   dtm(i,:) = -2*m(i-2,:) - m(i-1,:) + m(i+1,:) + 2*m(i+2,:);
end
dtm = dtm / 3;
%合并 mfcc 参数和一阶差分 mfcc 参数
ccc = [m dtm];
%去除首尾两帧，因为这两帧的一阶差分参数为 0
ccc = ccc(3:size(m,1)-2,:);
subplot(211)
ccc_1=ccc(:,1);
plot(ccc_1);title('MFCC');
ylabel('幅值');
title('一维数组及其幅值的关系')
[h,w]=size(ccc);
A=size(ccc);
subplot(212)
plot([1,w],A);
xlabel('维数');
ylabel('幅值');
title('维数于幅值的关系')
```

1.2 enframe.m

```
function f=enframe(x,win,inc)
%ENFRAME split signal up into (overlapping) frames: one per row. F=(X,WIN,INC)
%
%     F = ENFRAME(X,LEN) splits the vector X(:) up into
%     frames. Each frame is of length LEN and occupies
%     one row of the output matrix. The last few frames of X
```

```
%     will be ignored if its length is not divisible by LEN.
%     It is an error if X is shorter than LEN.
%
%     F = ENFRAME(X,LEN,INC) has frames beginning at increments of INC
%     The centre of frame I is X((I-1)*INC+(LEN+1)/2) for I=1,2,...
%     The number of frames is fix((length(X)-LEN+INC)/INC)
%
%     F = ENFRAME(X,WINDOW) or ENFRAME(X,WINDOW,INC) multiplies
%     each frame by WINDOW(:)

%       Copyright (C) Mike Brookes 1997
%       Version: $Id: enframe.m,v 1.4 2006/06/22 19:07:50 dmb Exp $
%
%     VOICEBOX is a MATLAB toolbox for speech processing.
%     Home page: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%     This program is free software; you can redistribute it and/or modify
%     it under the terms of the GNU General Public License as published by
%     the Free Software Foundation; either version 2 of the License, or
%     (at your option) any later version.
%
%     This program is distributed in the hope that it will be useful,
%     but WITHOUT ANY WARRANTY; without even the implied warranty of
%     MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.    See the
%     GNU General Public License for more details.
%
%     You can obtain a copy of the GNU General Public License from
%     ftp://prep.ai.mit.edu/pub/gnu/COPYING-2.0 or by writing to
%     Free Software Foundation, Inc.,675 Mass Ave, Cambridge, MA 02139, USA.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

nx=length(x(:));
nwin=length(win);
if (nwin == 1)
    len = win;
else
    len = nwin;
end
if (nargin < 3)
    inc = len;
end
nf = fix((nx-len+inc)/inc);
f=zeros(nf,len);
indf= inc*(0:(nf-1)).';
inds = (1:len);
f(:) = x(indf(:,ones(1,len))+inds(ones(nf,1),:));
if (nwin > 1)
      w = win(:)';
      f = f .* w(ones(nf,1),:);
end
```

1.3 mel.m

```
function [x,mn,mx]=mel(p,n,fs,fl,fh,w)
%MELBANKM determine matrix for a mel-spaced filterbank [X,MN,MX]=(P,N,FS,FL,FH,W)
%
% Inputs:   p      number of filters in filterbank
%           n      length of fft
%           fs     sample rate in Hz
%           fl     low end of the lowest filter as a fraction of fs (default = 0)
%           fh     high end of highest filter as a fraction of fs (default = 0.5)
%           w      any sensible combination of the following:
%                    't'    triangular shaped filters in mel domain (default)
%                    'n'    hanning shaped filters in mel domain
%                    'm'    hamming shaped filters in mel domain
%
%                    'z'    highest and lowest filters taper down to zero (default)
%                    'y'    lowest filter remains at 1 down to 0 frequency and
%                      highest filter remains at 1 up to nyquist freqency
%
%                     If 'ty' or 'ny' is specified, the total power in the fft is preserved.
%
% Outputs: x         a sparse matrix containing the filterbank amplitudes
%                    If x is the only output argument then size(x)=[p,1+floor(n/2)]
%                    otherwise size(x)=[p,mx-mn+1]
%           mn      the lowest fft bin with a non-zero coefficient
%           mx      the highest fft bin with a non-zero coefficient
%
% Usage:    f=fft(s);                    f=fft(s);
%           x=melbankm(p,n,fs);          [x,na,nb]=melbankm(p,n,fs);
%           n2=1+floor(n/2);             z=log(x*(f(na:nb)).*conj(f(na:nb)));
%           z=log(x*abs(f(1:n2)).^2);
%           c=dct(z); c(1)=[];
%
% To plot filterbanks e.g.      plot(melbankm(20,256,8000)')
%


%           Copyright (C) Mike Brookes 1997
%           Version: $Id: melbankm.m,v 1.3 2005/02/21 15:22:13 dmb Exp $
%
%     VOICEBOX is a MATLAB toolbox for speech processing.
%     Home page: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html
%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%       This program is free software; you can redistribute it and/or modify
%       it under the terms of the GNU General Public License as published by
%       the Free Software Foundation; either version 2 of the License, or
%       (at your option) any later version.
%
%       This program is distributed in the hope that it will be useful,
%       but WITHOUT ANY WARRANTY; without even the implied warranty of
%       MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.    See the
%       GNU General Public License for more details.
%
%       You can obtain a copy of the GNU General Public License from
%       ftp://prep.ai.mit.edu/pub/gnu/COPYING-2.0 or by writing to
%       Free Software Foundation, Inc.,675 Mass Ave, Cambridge, MA 02139, USA.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

if nargin < 6
   w='tz';
   if nargin < 5
      fh=0.5;
      if nargin < 4
         fl=0;
      end
   end
end
f0=700/fs;
fn2=floor(n/2);
lr=log((f0+fh)/(f0+fl))/(p+1);
% convert to fft bin numbers with 0 for DC term
bl=n*((f0+fl)*exp([0 1 p p+1]*lr)-f0);
b2=ceil(bl(2));
b3=floor(bl(3));
if any(w=='y')
   pf=log((f0+(b2:b3)/n)/(f0+fl))/lr;
   fp=floor(pf);
   r=[ones(1,b2) fp fp+1 p*ones(1,fn2-b3)];
   c=[1:b3+1 b2+1:fn2+1];
   v=2*[0.5  ones(1,b2-1)  1-pf+fp  pf-fp  ones(1,fn2-b3-1)
0.5];
   mn=1;
   mx=fn2+1;
else
   b1=floor(bl(1))+1;
   b4=min(fn2,ceil(bl(4)))-1;
   pf=log((f0+(b1:b4)/n)/(f0+fl))/lr;
   fp=floor(pf);
   pm=pf-fp;
   k2=b2-b1+1;
   k3=b3-b1+1;
   k4=b4-b1+1;
   r=[fp(k2:k4) 1+fp(1:k3)];
   c=[k2:k4 1:k3];
   v=2*[1-pm(k2:k4) pm(1:k3)];
   mn=b1+1;
   mx=b4+1;
end
if any(w=='n')
   v=1-cos(v*pi/2);
elseif any(w=='m')
   v=1-0.92/1.08*cos(v*pi/2);
end
if nargout > 1
   x=sparse(r,c,v);
else
   x=sparse(r,c+mn-1,v,p,1+fn2);
end
```

2 提取基因和共振峰的程序

dp.m

```
%
% dp
clear all;
clc
close all hidden
format long
waveFile='sunday.wav';
[speech, sf, nbits]=wavread(waveFile);
index1=4000;
nfft=1000;
index2=index1+nfft-1;
y=speech(index1:index2);
t=0:1/sf:(nfft-1)/sf;
nn=1:nfft/4;
dt=1/sf;

subplot(2,1,1)
z=rceps(y);          %MATLAB 提供的倒频谱函数
plot(t(nn),abs(z(nn)));
title('z=rceps(y)')
ylabel('Amplitude');
xlabel('Quefrency')

Y=fft(y);
subplot(2,1,2)
plot(log(abs(Y))); hold on
title('Log Spectrum');

figure(2)
mcep=40;

%加窗
subplot(2,1,1)
nn=1:mcep;
z1=z(nn);
plot(t(nn),abs(z1(nn)));
title('Envelope Information in Quefrency Space')
z11=[z1' zeros(1,1000-2*mcep) z1(mcep:-1:2)'];

%fft 变换
subplot(2,1,2)
z2=fft(z11);
plot(real(z2));
title('Frame Speech Envelope');
figure(1);
subplot(2,1,2); plot(real(z2),'r');

[Cmax Cloc]=max(abs(z(41:320)));
T0=(Cloc+40)*dt;
F0=1/T0;
fprintf('Fundamental Frequency F0=    %5.4fHz\n',F0);
```

致谢：

本论文的指导老师是西南大学电子信息工程学院的李太华老师。从论文的选题到深入研究的过程中，李老师都给了我很多启发性的指导，让我受益匪浅，得以顺利完成毕业论文。这期间身边的同学也帮我一起解决了很多相关知识和疑问。

在此，对李老师和帮助我的同学们致以诚挚的谢意！