语音信号处理中基频提取算法综述

张 杰，龙子夜，张 博，陈咏丽，秦玉英

(海军装备研究院 北京 昌平区 102249)

【摘要】基频提取在语音信号处理领域是一个基础的课题。经过多年的研究，现在的技术在准确率和鲁棒性方面还没有达到一个十分令人满意的水平。当语音是一个干净的语音时，大部分的基频提取算法的结果都很好，但是当语音中混有较强的噪声，或者语音是多个语音的混合，从而同时含有多个基频的时候，很多现在的技术都表现得不好。该文介绍了若干主要的基频提取算法，并对这些算法的改进进行了讨论。

关 键 词 滤波器； 基频提取； 语音信号处理； 小波变换

中图分类号 TN912.3 **文献标识码** A doi:10.3969/j.issn.1001-0548.2010.z1.024

A Summarize of Pitch Detection Algorithmic in Speech Signals Processing

ZHANG Jie, LONG Zi-ye, ZHANG Bo, CHEN Yong-li, and QIN Yu-ying

(Naval Academy of Armament Changping Beijing 102249)

Abstract Pitch detection is a basic topic in speech signals processing. Through many years' research, there is yet a satisfied technology in terms of accuracy and robust. When the speech is clean, most of the pitch detection algorithms are fine, but many technologies are not satisfied when the speech mixed with noise or the speech is a mixture of multi-speech. This paper introduces some mainstream pitch detection algorithms, and discusses the improvement of these algorithms.

Key words filter; pitch detection; speech signals processing; wavelet transformation

语音分辩的一个主要特征是激励的类型，根据激励类型的不同，可以将语音信号分为浊音和清音两大类。语音中只有浊音才有基频，浊音的激励是周期性的脉冲串，脉冲串的频率就是基音频率，简称基频。由于发声器官生理方面的差异，男性和女性的基频范围不同，一般地，男性的基频范围为50～250 Hz；女性的基频范围为120～500 Hz；婴儿的基频范围大约为250～800 Hz；新生婴儿的哭声基频范围则更高[1-4]。

语音信号是非平稳的，因此语音信号处理也必须是短时的，即在一个短的时间窗内处理语音信号。窗长取决于语音信号的特征，通常至少要包括两个基音周期。

语音基频提取在语音信号处理领域有很多应用，如语音分离、语音合成等。对于汉语的语音识别来说，在没有考虑韵律的情况下，当前的主流语音识别技术没有用到基频。基频是韵律的重要部分，所以如果要把韵律信息加入语音识别系统，基频提取是必须的。如何加入韵律信息也是汉语语音识别系统的重要研究方向。

根据处理域的不同，可将基频提取算法分为时域的算法、频域的算法、统计的算法3类。

1 时域的算法

因为语音信号的时域波形代表了随时间变化的声音激励的变化，基频提取的最基本方法就是通过观察语音信号的波形，并从波形中检测出基频。

1.1 时域的事件发生率检测

基频提取方法中的一些方法是试图通过观察语音信号的波形重复自己的频率，估计基频。这些方法的理论依据是，如果语音信号是周期的，那么就会有随着时间不断重复出现的事件发生，统计这些事件在单位时间内的发生次数，就能估计出基频。

1.1.1 过零率

简单地说，过零率就是单位时间内波形通过零点的次数。开始对过零率的性能存有怀疑，但最近

收稿日期： 2009 ↖ 11 ↖ 15

作者简介：张 杰(1976 ↖)，女，工程师，主要从事信号处理和计算机应用等方面的研究.

过零率方法由于文献[5]而变得流行和活跃起来。使用过零率的一个重要目的就是提取基频，研究者曾经认为过零率与波形在单位时间内重复的次数有直接关系。但是不久人们就发现了以这样的思路使用过零率提取基频的方法有问题[6]。如果信号的能量都集中在基频附近，那么一个周期内它将两次穿过零。但是如果信号包含了高频能量，在一个周期内它穿过零的次数将大于2。所以如果使用过零率检测基频，要先滤掉高频成分。确定滤波器的截止频率，既要尽可能多地去掉高频成分，又要防止基频被滤掉。另一个可能的使用过零率提取基频的方法是先识别出过零率的模式，然后基于信号的模式估计基频。

1.1.2 峰值率

时域的算法统计在单位时间内波形峰值出现的次数。理论上，信号在一个周期内有一个最大值和一个最小值，所以只需要统计单位时间内最大值的个数就可估计语音的基频。局部的峰值检测器必须用于检测信号在局部的最大值，单位时间内的最大值个数就是语音的基频。从另一个角度考虑，相邻两个最大值的时间差的倒数也可以用于估计基频。

1.1.3 信号导数的事件检测

如果信号是周期性的，那么信号的导数也是周期性的，而且信号周期与原始语音信号的周期相同。所以过零率和峰值率两种算法对于信号的导数同样适用。某些情况下，在信号的导数上检测过零率或峰值率，会比在原始信号上直接检测过零率或峰值率包含更多的信息，或者更加鲁棒，这取决于信号本身的特性。

1.2 自相关函数法

以两个信号之间的相关函数度量它们之间的相似性，相关函数的结果随两个信号波形开始时间的延迟而变化。自相关函数是信号自身的相关函数，以自相关函数度量信号自身的相似性。对于无限长的离散信号$x[n]$，自相关函数的定义为：

$$R_x(v)=\sum_{n=-\infty}^{\infty}x[n]x[n+v] \tag{1}$$

式中 v为信号的延时，对于一般的8 kHz采样的语音信号，取值范围为20～150，相应的基音频率范围为60～500 Hz。

对于长度为N的离散信号$x_N[n]$，自相关函数的定义为：

$$R_{x_N}(v)=\sum_{n=0}^{N-1-v}x_N[n]x_N[n+v] \tag{2}$$

对于周期性函数，自相关函数也呈现周期性，并且在基音周期的各个整数点上有很大的峰值。只要找到第一最大峰值点的位置并计算它与v=0点的间隔，便能估计出基音周期，而基音周期的倒数就是基频。为了防止窗长过短引起计算的错误，需要使窗长至少大于两个基音周期，才可能取得较好的计算结果。

1.3 平均幅度差函数法

还可以采用平均幅度差函数法求基频，计算公式为：

$$r_x(v)=\sum_{v=0}^{N-1-v}|x(n+v)-x(v)| \tag{3}$$

与自相关函数法相同，对于周期性的函数$x(n)$，平均幅度差函数$r(v)$也呈现周期性，不同点在于自相关函数法的结果在基音周期的各个整数点有很大的峰值，而平均幅度差函数法在基音周期的各个整数点有谷值。

1.4 阴阳估计法

文献[7]根据东方阴阳平衡的哲学理论提出阴阳估计法，该方法试图在自相关函数的生成和取消之间取得平衡。自相关函数法的难点在于峰值也会出现在谐波处，所以有时很难判断哪个峰值对应基频。阴阳估计法基于差异函数，与前面的平均幅度差函数法一样，在基音周期的整数倍点取得谷值。差异函数表示为：

$$d_t(\tau)=\sum_{j=0}^{W-1}(x(j)-x(j+\tau))^2 \tag{4}$$

式中 W为窗长。为了减少高次谐波处的谷值带来的错误，可以用累计平均函数代替式(4)定义的差异函数。平均函数表示为：

$$d'_t(\tau)=\begin{cases}1, & \tau=0\\ \dfrac{d_t(\tau)}{\frac{1}{\tau}\sum_{j=1}^{\tau}d_t(j)}, & \text{其他}\end{cases} \tag{5}$$

平均函数用差异函数除以它前面所有值的均值，与差异函数不同，它在延迟为0时的值是1而不是0，并且在延迟很小时都能取得较大值，在差异函数的值小于其前面的均值时才降到1以下。使用式(5)的优点是：(1) 减小错误率。(2) 避免原来延迟为0时的谷值影响。(3) 归一化结果，为后续处理带来方便。要了解更加详细的处理过程，请参看文献[7]。

2 频域的算法

频域有更多的与基频相关的信息。具有基频的

信号是由频率具有谐波关系的信号组成的，因此有很多尝试利用频域信息提取基频的方法。

2.1 基于滤波器的算法

2.1.1 最佳梳状滤波器法

最佳梳状滤波器法[8]是具有高鲁棒性但计算代价很大的算法。一个梳状滤波器有很多等距离分布的通带，在最佳梳状滤波器算法中，通带的位置都是由第一个通带决定的，即通带的中心频率都是第一个通带中心频率的整数倍。输入信号通过多个与第一个通带中心频率不同的梳状滤波器。如果输入信号是由一组频率成谐波关系的信号组成的，那么滤波器的输出在全部谐波成分都通过滤波器时达到最大。但是如果信号只有一个基频成分，该方法就会失效，因为会有很多个梳状滤波器能让信号通过。不过，语音信号的频率具有谐波结构，所以可采用该方法提取基频。

2.1.2 可调的IIR滤波器

文献[9]提出了一种基于中心频率可调节的带通IIR滤波器提取基频的方法，随着用户的调节，滤波器的中心频率扫过整个频域。当输入信号的一个强的频率成分在通带范围内时，滤波器会输出最大值，信号的基频就可以用此时滤波器的中心频率来估计。文献[9]提到，对于可调的IIR滤波器，有经验的用户能够识别具有一个谐波结构的信号的输出和包含多个基频信号的输出的差异。

2.2 倒谱分析法

倒谱分析是谱分析的一种方法，输出是傅里叶变换的幅度谱取对数后做傅里叶逆变换的结果。该方法所依据的理论是，一个具有基频的信号的傅立叶变换的幅度谱有一些等距离分布的峰值，代表信号中的谐波结构，当对幅度谱取对数之后，这些峰值被削弱到一个可用的范围。幅度谱取对数后得到的结果是在频域的一个周期信号，而这个频域信号的周期(是频率值)可以认为就是原始信号的基频，所以对这个信号做傅里叶逆变换就可以在原始信号的基音周期处得到一个峰值。

另外，如果对信号的傅里叶变换的幅度谱取对数后的结果直接进行分析，而不是再接着做傅里叶逆变换，就是谐波成分谱的方法。进一步，如果在求频域的变换时不使用傅里叶变换，而使用能使频谱更加精细的Chirp变换，就是基于Chirp变换的提取基频的方法，该方法具有高分辨率和高鲁棒性。

2.3 多分辨率的方法

对于任何基于傅里叶分析的频域方法都可以做的一个改进是采用多分辨率方法。该方法的思想是：如果一个特定算法在特定分辨率下的准确性是可疑的，那么采用更高或者更低的分辨率，可以进一步判断前面的基频估计是否可信。如果在全部或大部分的分辨率下求得相同的基频，那么该频率值就可以作为最终的基频估计结果。当然，在带来好处的同时，该方法也会带来计算量上的代价，因为针对每一个分辨率都需要重新计算频谱，这也是为什么多分辨率的傅里叶分析比专门的多分辨率变换(如离散小波变换)要慢的原因。

2.4 离散小波变换法

离散小波变换是一个强大的工具，它允许在连续的尺度上把信号分解为高频成分和低频成分，它是时间和频率的局部变换，能有效地从信号中提取信息。与快速傅里叶变换相比，离散小波变换的主要好处在于，在高频部分它可以取得好的时间分辨率，在低频部分可以取得好的频率分辨率。

3 统计的方法

在某种意义上，基频提取的问题可以被看作是一个统计问题。每一个输入帧都被划分给一组类中的一个，代表信号的基频估计。所以很多研究者一直试图将现代的统计方法应用于基频提取问题。

Boris和Xavier发表了一系列使用最大似然法估计基频的方法。他们的模型如下：观察集是语音信号分帧后做短时傅里叶变换的结果，每一个观察都被看作是基频激励产生的信号与其他剩余信息(包括非谐波部分和噪声)两部分的混合。该模型是由一般的语音信号产生的模型的简单化得到的，假设一个语音包括在基频及其整数倍点的值处较大的谐波成分，以及在非谐波处和噪声处的很小的值。对于一组候选的基频值，该方法计算每一个观察可能是由某一个基频产生的概率，并将概率最大的基频值作为最终的估计值。所以候选的基频值的选择是很重要的，因为从理论上讲，观察可能对应着任意的基频值。

4 算法的改进

前面提到的每种算法都有自己的改进方法，下面介绍两种对以上大部分算法均适用的改进方法。

4.1 人的听觉模型

由于基频提取本身就是听觉感知问题，所以所有的算法都可通过加入人耳的听觉模型提高性能。人耳的听觉模型将人的听觉系统对声音信号的处理

分为分析、传递和还原3个阶段。分析阶段主要考虑耳蜗的分频效应，耳蜗的外端对高频敏感，内端对低频敏感，可以用一组中心频率不同的带通滤波器来模拟。传递阶段声波振动沿基膜传播，并在听觉神经纤维内产生电流，最终传入听觉中枢。还原阶段听觉系统提取语音中诸如音质、音调、时域和位置等信息。

在声学中，声强是指单位时间内通过垂直于声波传播方向的单位面积的声波能量，用I表示。当声波的频率在20～20 000 Hz(可闻频率)之间，而声强达到一定的强度(听阈)，就能被人耳感知。前人大量的实验测试结果表明，人耳对不同频率的声波感受到相同响度时的声强是不同的。人耳对两端频段的声波反应较为迟钝，而对中间频段的声波反应相对较为敏感。

对于任意的频域方法，简单的改进是用Q值恒定的谱变换方法代替傅里叶变换。恒Q的变换方法计算代价更大，但更接近于人的听觉感知系统。

在决定是否使用人的听觉模型时必须考虑两个因素：(1) 基频提取的用途。如果应用的目的很简单，要求也不是太高，那么人的听觉感知因素也许不是很必要。(2) 计算的复杂度。使用人的听觉感知模型会使计算复杂度大大增加，如果原来算法的复杂度已经很大，再加入人的听觉感知模型可能会使算法的复杂度过高。

4.2 基频的跟踪

另一种对基频提取的改进是基频跟踪。前面提到的基频提取都是在一个单独的时间窗内进行的。人的听觉系统是能够跟踪输入信号的基频的。一个只包含有限个基音周期的时间窗内的基频是很难提取的。但是，如果输入是连续的语音信号，相当于很多时间窗一个接一个输入，基频的提取反而变得很容易。研究发现，语音信号的基频具有连续性，即前后两帧的基频是连续的，不出现跳变。一帧内的基频提取常见的问题是得到的估计值是正确值的整数倍或者整数倍分之一。针对该问题，利用语音信号基频的连续性，可对基频提取算法做一个简单的改进：在计算某一帧的基频时对于它前面一帧的基频附近的值给予更大的可能性，即一帧语音信号中基频的值不可能出现跳变的情况。这就是简单的基频跟踪思想，并且不会在计算上增加任何复杂度。

另外一种比较复杂的基频跟踪方法是使用隐马尔科夫模型。

5 经典的基频检测方法

自从有了语音信号分析研究这门学科以来，基频的检测一直是一个重点研究的课题。经典的基频检测方法可以大致分为3类，如表1所示。

表1 经典的基音检测方法以及特点

分类	基因检测方法	特点
波形估计法	并行处理法	由多种简单的波形峰值检测器提取基音周期
	数据减少法	根据各种理论操作，从波形中去掉修正基音以外的数据
	过零率法	利用波形的过零率，着眼于重复图形
相关处理法	自相关法及其改进	利用语音波形的自相关函数提取基音，采用中心削波平坦处理频谱，采用峰值削波可以简化运算
	SIFT法	语音波形降低采样率后，进行LPC分析，用逆滤波器平坦处理频谱，通过预测误差的自相关函数恢复时间精度
	AMDF法	采用平均幅度差函数(AMDF)检测周期性，也可以根据残差信号的AMDF进行提取
变换法	倒谱法	根据对数功率谱的傅立叶反变换，分离频谱包络和微细结构
	循环直方图法	在频谱上求出基频高次谐波成分的直方图，根据高次谐波的公约数决定基音

(1) 波形估计法。直接由语音波形估计、分析波形上的周期峰值。

(2) 相关处理法。时域中周期信号最明显的特征是波形的类似性，因而可以通过比较原始信号和它位移后的信号之间的相似性确定基音周期。该类方法抗波形的相位失真能力强，且硬件处理结构简单。

(3) 变换法。将语音信号变换至频域或倒谱域估计基音周期

6 总 结

本文列出了若干基频提取的主要方法，对它们分别进行了简单的介绍，并讨论了对算法的改进。需要注意的是，所介绍的方法都是针对一个语音信号而言的，对于混合的语音信号的基频提取，如果可以先将混合的语音信号分离开，那么基频提取就会变得很简单。同样地，在一些基于时频分析的语音分离算法中，如果知道了各个语音的基频，那么语音分离也就变得很容易解决了。

(下转第126页)