

一种基于动态平滑的实时基频提取算法

胡海洋^{1,2}, 原 猛¹, 冯海泓¹

(中国科学院声学研究所东海研究站, 上海 200032; 2. 中国科学院研究生院, 北京 100049)

摘要: 基频是语音信号处理中的一个基本声学特征。传统的基频提取算法为了获得较好的检测效果, 需要复杂的时频域计算。对于资源受限的应用条件, 例如人工耳蜗等嵌入式实时系统, 很难应用计算量大的基频提取算法。语音信号的基频具有短时平稳性, 根据这个特点来确定基频候选值可以提高提取的准确性。据此, 提出一种基于动态平滑的基频提取算法, 使用此算法对汉语声调词库进行基频提取, 并与另外两种基频提取算法进行比较。实验结果表明, 新算法的基频绝对平均估计误差小于 3Hz, 优于另两种算法, 能够准确地提取基频, 同时算法计算量低, 适合实时应用。

关键词: 基频提取; 动态平滑; 实时处理

中图分类号: N912.3

文献标识码: A

文章编号: 1000-3630(2012)-06-0583-06

DOI 编码: 10.3969/j.issn1000-3630.2012.06.010

A dynamic smoothing based real-time pitch detection algorithm

HU Hai-yang^{1,2}, YUAN Meng¹, FENG Hai-hong¹

(1. Shanghai Acoustics Laboratory, Institute of Acoustics Chinese Academy of Sciences, Shanghai 200032, China;

2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Fundamental frequency is one of the most important features in speech signal processing. Traditional pitch detection algorithms (PDA) can hardly be applied in the resource-limited hardware system due to the computation complexity. A dynamic smoothing based pitch detection algorithm is proposed in this paper. As pitch is a physical quantity that does not change rapidly, using continuous speech frames to decide the best candidate of fundamental frequency can improve the accuracy. An objective experiment was carried out to compare the pitch detection accuracy of the DSPDA with two other algorithms. The experimental results show that the averaged pitch detection error is 3Hz lower than that of other algorithms.

Key words: pitch detection; dynamic smoothing; real-time

0 引言

基频信息是许多语音信号处理应用中的特征参数。在声调语言(例如汉语)中, 不同的基频变化模式对应不同的声调。声调对语义的理解具有重要意义。正常听力者的声调感知可以达到基本正确。人工耳蜗植入者的声调识别正确率在安静环境下只有 50%~70%, 随着噪声增加正确率还会进一步明显降低^[1]。有研究表明, 通过在人工耳蜗言语编码中增强基频信息有助于提高人工耳蜗植入者的声调感知能力^[2]。因此为人工耳蜗言语编码策略提供快速准确的基频估计对改善植入者的声调感知能力具有重要意义。

基频与语音信号的周期性相关, 根据基频提取算法计算方法的不同可以分为三类: 时域算法, 例如 YIN(YIN Estimator)算法^[3]和 RAPT(a Robust Algorithm for Pitch Tracking)算法^[4]; 频域算法, 例如基于倒谱的算法^[5]; 以及时频域结合的方法^[6]。这些算法已经可以获得较好的基频提取效果, 但对于人工耳蜗或助听器等低功耗嵌入式设备, 计算复杂度过高。

本文提出了一种动态平滑方法, 基于此方法实现了一种计算量低的实时基频提取算法(Dynamic Smoothing based Pitch Detection Algorithm, DSPDA)。动态平滑利用语音信号的音高不能突变的特点, 根据前一时刻语音信号的基频值来确定新的基频值。

1 动态平滑原理

在语音信号中, 语音的清音和浊音过渡段多出现在音节的开始和结尾等能量较小的部分, 语音信

收稿日期: 2012-04-06; 修回日期: 2012-07-04

基金项目: 国家自然科学基金资助项目(11104316); 上海自然科学基金资助项目(11ZR1446000)

作者简介: 胡海洋(1987-), 男, 湖北广水人, 硕士研究生, 研究方向为语音信号处理。

通讯作者: 原猛, Email: ym@mail.ioa.ac.cn

号一个音节中的基频连续变化而不会有突变。根据信号能量的不同,可以设置不同的清浊音检测阈值以提高清浊判断准确性,根据基频值连续变化的特点提高基频提取的准确性。动态平滑方法利用以上特性进行计算。

对于第 n 帧语音信号 $S(i)$ 可以获得 m 个候选值对 $\{C_{A,n}(i), C_{f,n}(i)\}$, 其中 $C_{A,n}(i)$ 为第 i 点的幅度, $C_{f,n}(i)$ 为第 i 点的频率。可以计算第 i 候选点的频率与第 $n-1$ 帧语音信号基频值 $F_0(n-1)$ 的“距离” $D_n(i)$, 如式(1)所示。

$$D_n(i) = \left| \frac{F_0(n-1) - C_{f,n}(i)}{F_0(n-1)} \right| \quad (1)$$

较大的 $D_n(i)$ 值说明候选点 i 的频率离上一帧的基频值较远,很有可能是偏离基频曲线的野点。限定候选点频率与上一帧基频值的距离在阈值 d 以内可以减少大量候选点,减小候选基频值的波动范围。对符合条件的候选点计算强度因子 $E_n(i)$, 以进一步筛选最佳候选值。 $E_n(i)$ 的定义如下:

$$E_n(i) = \begin{cases} \frac{C_{A,n}(i)}{D_n(i)} & D_n(i) < d \\ 0 & D_n(i) \geq d \end{cases} \quad (2)$$

强度因子本质上对基频曲线起到了一种动态平滑的作用,减少了基频曲线在局部范围内强烈变化的错误,提高了基频值估计的准确性。同时计算 $E_n(i)$ 时使用候选点的自相关幅度值 $C_{A,n}(i)$, 保证周期性较强的候选点能够被选择。

在基频信号的起点处,语音信号幅值较小。针对基频信号信号不同段,采取不同幅值有利于清浊音判断。

当上一帧信号为清音,即 $F_0(n-1)=0$, 若候选点自相关函数幅度值大于阈值 δ_1 , 判定信号为浊音,并选择使得自相关函数幅度值 $C_{A,n}(i)$ 最大的候选点作为最佳基频值,如式(3a)所示。

$$F_0(n) = \begin{cases} 0 & F_0(n-1)=0, \\ & C_{A,n}(i) < \delta_1; \\ C_{n,f}(k) & F_0(n-1)=0, C_{A,n}(i) \geq \delta_1, \\ & C_{A,n}(k) = \text{MAX}(C_{A,n}(i)); \end{cases} \quad (3a)$$

当上一帧信号为浊音,即 $F_0(n-1)>0$, 如候选点自相关函数幅度值大于阈值 δ_2 , 判定信号为浊音,使用动态平滑的方法计算候选值强度因子,并选取强度因子最大的候选点为最佳基频值,如式(3b)所示。

$$F_0(n) = \begin{cases} 0 & F_0(n-1)>0, \\ & C_{A,n}(i) < \delta_2; \\ C_{n,f}(k) & F_0(n-1)>0, C_{A,n}(i) \geq \delta_2, \\ & E_n(k) = \text{MAX}(E_n(i)); \end{cases} \quad (3b)$$

与动态平滑方法类似的一个方法是动态规划技术^[7](Dynamic Programming Technique)。一些基频提取算法使用动态规划的方法获得了较好的鲁棒性,例如 PRAAT^[8]、RAPT、YIN、AMPEX^[9]等。动态规划技术计算每一帧信号的所有候选值的代价因子,最后根据全局的代价因子来选择最佳基频曲线。本文动态平滑方法与动态规划的方法有明显不同,主要体现在以下两个方面:

(1) 本文使用的动态平滑方法只利用前一帧已有的基频信息搜索新的基频候选值,而不依赖于任何后续信息,计算延迟低,适于实时处理。

(2) 动态平滑方法中,根据前一帧信号的基频值限定了一个基频候选值搜索范围,使基频估计值在限定范围内波动。

动态平滑方法是本文基频提取算法中的主要部分,下文将介绍算法具体的基频提取流程。

2 基频提取算法流程

动态平滑方法是本文基频提取算法使用的主要方法,为了提高算法的鲁棒性和准确性,还需要使用其他处理方法,如增加预处理阶段可以削弱干扰成分,减少动态平滑选择时的候选值个数。对自相关函数进行修正可以减少处理过程的不利影响。基频提取算法流程如图1所示。

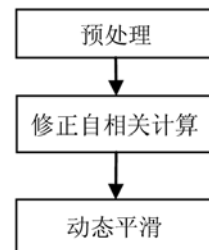


图1 基频提取算法流程

Fig 1 Flow chart of pitch detection algorithm

2.1 预处理

语音信号中包含噪声、清音、共振峰等复杂成份,由于这些成份的影响,在变换域中表征语音信号基频的峰值存在偏移或被其他成份的峰值所淹没,使得基频提取结果产生错误和误差。在提取基频之前,削弱与基频无关的成份,可以提高基频提取算法的鲁棒性。

考虑到需要实现一个计算复杂度低的实时基频提取算法,本文选择中心削波的方法进行预处理。中心削波的方法只需要常数量级的加法运算即可实现,相比于LPC滤波波等需要大量乘法运算的方法,计算量要少得多。中心削波函数如公式(4),

式中 C_L 为削波阈值。

$$y(n) = \begin{cases} x(n) - C_L, & x(n) > C_L \\ 0, & |x(n)| \leq C_L \\ x(n) + C_L, & x(n) < -C_L \end{cases} \quad (4)$$

中心削波方法是基于 Licklider 和 Pollack 的研究：对语音信号进行削峰，语音信号的可懂度下降很少；如果使用中心削波，语音的可懂度则会严重下降，这是因为语音中共振峰信息对可懂度有重要作用^[10]。图 2 是一个中心削波的示例图。

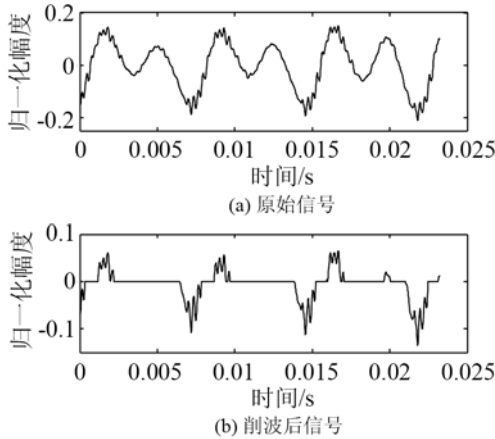


图 2 中心削波示例
Fig.2 Example of center clipping

从图 2 中可以看出，信号一个周期内由声道共振峰调制引起的衰减振荡被削减，而信号波形的周期性仍然保留。

2.2 修正自相关计算

本文基频提取算法基于自相关法。长度为 L 的加窗信号 $S_n(i)$ 的自相关函数 $R_n(k)$ 为

$$R_n(k) = \sum_{i=0}^{L-k-1} S_n(i)S_n(i+k), \quad 0 \leq k < L \quad (5)$$

自相关法通过搜索零点以外的最大峰值来确定基频值。对于加窗信号，其自相关函数的幅值随着自相关延迟时间的增加而逐渐衰减。当信号的基频值较低时，对应基频的自相关函数峰值出现在离零点较远处，这个峰值可能因窗函数衰减作用而被漏选。Paul Boersma 的研究指出，消除窗函数的这种衰减作用可以提高算法的准确性^[11]。同时他并提出如下改进方法^[11]，令

$$R_x(\tau) \approx \frac{R_{xw}(\tau)}{R_w(\tau)} \quad (6)$$

即用加窗后的自相关函数除以窗函数本身的自相关函数，近似等于信号的理想自相关函数。图 3 是自相关函数使用这个方法修正前后的比较。

经过这一处理之后，自相关函数中延迟 τ 较大处的自相关函数值被提升，消除了因窗函数的衰减

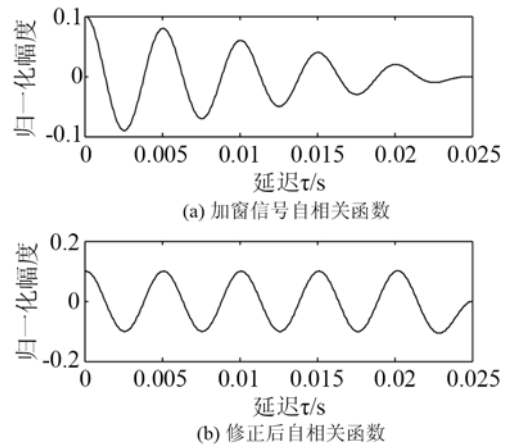


图 3 对加窗信号的自相关函数的修正
Fig.3 Modified autocorrelation function of windowed speech

作用而漏选的基频候选值的错误。但是这一近似的补偿方法中，当自相关函数值本身很大(接近于 1) 时，在 τ 较大处的补偿过大，导致修正后的自相关函数值可能会大于 1。对于这个错误，修正方法对过大的候选峰值进行如式(7)所示的处理。

$$R'_x(\tau) = \frac{0.95}{R_x(\tau)}, \quad \text{当 } R_x(\tau) > 0.95 \text{ 时} \quad (7)$$

2.3 动态平滑

利用修正后的自相关函数，使用动态平滑的方法即可得每一帧信号的基频值输出。根据公式(1)、(2)、(3)，动态平滑部分的算法流程如图 4 所示。

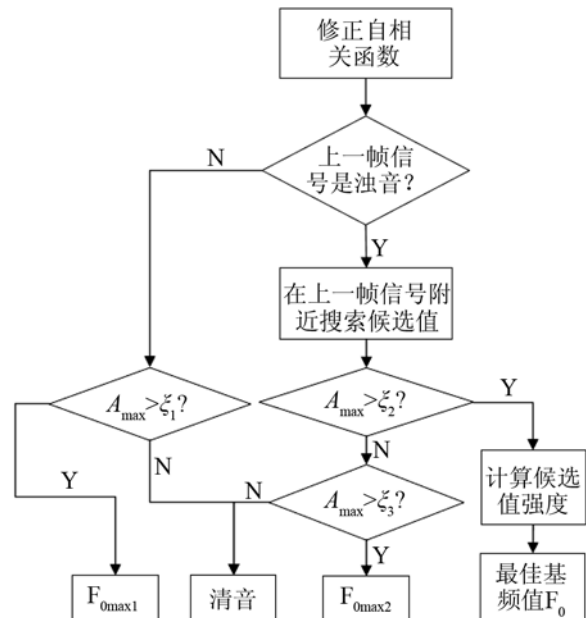


图 4 动态平滑算法流程图
Fig.4 Flow chart of dynamic smoothing algorithm

在图 4 中， A_{max} 为信号自相关函数除零点以外的最大峰值， A_{maxc} 为信号自相关函数选定区域中最大候选值的幅度值。 δ_1 、 δ_2 、 δ_3 为不同情况的清

浊判断阈值, 并且 $\delta_1 < \delta_2 < \delta_3$ 。当上一帧信号为清音时, 信号幅度值可能较小, 大于较小的阈值 δ_1 的最大峰值对应的频率被认为是基频 F_{0max1} ; 当上一帧信号为浊音时, 候选值搜索区域中所有峰值都小于阈值 δ_2 , 但自相关函数的最大峰值大于较大阈值 δ_3 , 说明上一帧信号基频存在错误, 从而使搜索区域错误, 更新基频值为最大峰值对应的频率 F_{0max2} 。其他情况则使用动态平滑方法中定义的距离函数和强度因子, 选择基频值为使强度因子最大的频率 F_0 。

3 实验结果与分析

3.1 实验方法

为了对本文算法进行评价, 采用客观评价的方式按照如下步骤进行:

- (1) 准备实验用的语音文件。
- (2) 获取实验语音文件的标准基频值。
- (3) 使用本文 DSPDA 算法和其他基频提取算法对实验语音文件进行处理, 得到不同算法提取基频的结果。

(4) 根据标准基频值, 按照评价参数比较不同算法提取的结果。

在以上实验步骤中用到的实验语音文件是一个汉语声调材料库^[12]。材料库中包含 96 个双音节词和 20 组单音节字。每组单音节字包含四个汉语声调。所有材料均由以汉语普通话为母语的一名成年男性和一名成年女性的声音录制而成。对每一个录制的纯净语音添加高斯白噪声, 得到对应的 10 dB SNR 带噪语音材料。纯净语音和 10 dB SNR 带噪语音总计 704 个实验语音文件。

实验使用的标准基频值使用一个语音学常用软件 Praat^[8]得到。为了获得这些汉语声调的标准基频值, 本文先通过 praat 软件“to pitch(ac)”选项提取安静环境下语音材料的基频值, 然后再手工调整基频曲线中误差较大的点和肉眼可辨认的清浊音判断错误, 得到本文实验的标准基频值。

客观比较的对象, 本文选择与 DSPDA 算法实现原理类似的 YIN 算法和 RAPT 算法进行比较。

YIN 算法是一种基于自相关的改进算法, 具有较好的鲁棒性和准确性。许多基频提取算法在进行比较时选择了 YIN 算法^[9,13-16]。YIN 算法使用一个累积平均归一化差分方程, 将峰值搜索转化为谷值搜索, 消除了窗函数的衰减作用。在候选值选择部分, YIN 算法使用了类似于动态规划的方法。

RAPT 是另外一种时域的基于互相关的改进算

法^[5]。RAPT 算法使用归一化互相关计算, 得到基频值候选点。最后使用了类似于动态规划的方法确定最佳基频值。

在评价不同的基频提取算法时, 四种常用的评价参数^[1,13-16]为:

清音被判断为浊音的错误率, 简称清音错误率 (Unvoiced-Voiced error rate, UV), 计算方法为标准基频中清音被判断为浊音的比例。

浊音被判断为清音的错误率, 简称浊音错误率 (Voiced error rate, VU), 计算方法为标准基频中浊音被判为清音的比例。

严重错误率(Gross error rate): 与标准基频值相差 20%以上的错误率, 计算方法为被正确判断为浊音但基频值误差大于 20%的比例。

平均绝对误差(Mean absolute deviation, MAD): 浊音信号被正确判断为浊音, 并且没有发生严重错误, 使用估计基频值与标准基频值比较, 计算平均绝对误差值。假设基频估计值为 X_i , 对应的标准基频值为 Y_i , 计算方式如公式(8)所示:

$$\Delta = \frac{\sum_{i=1}^n |X_i - Y_i|}{n} \quad (8)$$

其中 X 、 Y 满足 $\frac{|X_i - Y_i|}{Y_i} \leq 0.2$ 。

3.2 实验结果

使用以上实验方法得到的实验结果列于表 1、2 中。根据信噪比的不同将实验结果分为安静环境和 10dB 两种情况, 分别使用 clean 和 10dB 标注。实验结果按照评价参数分为四类, UV 指清音错误率, VU 指浊音错误率, Gross 表示严重错误率, MAD 表示基频值的平均绝对误差。

表 1 双音节词错误率比较
Table 1 Error rates of disyllabic words

	UV/%	VU/%	Gross/%	MAD/Hz
DSPDA(Clean)	6.82	4.81	3.74	2.32
YIN(Clean)	6.10	17.3	2.50	7.16
RAPT(Clean)	9.71	2.75	5.35	2.86
DSPDA(10dB)	0.67	20.2	3.22	2.64
YIN(10dB)	0.32	47.4	1.21	6.90
RAPT(10dB)	3.58	11.4	9.41	3.44

表 2 单音节字错误率比较
Table 2 Error rates of monosyllabic words

	UV/%	VU/%	Gross/%	MAD/Hz
DSPDA(Clean)	2.87	2.84	1.45	2.22
YIN(Clean)	2.31	12.4	2.20	5.33
RAPT(Clean)	5.09	3.52	3.44	1.72
DSPDA(10dB)	0.46	14.5	3.40	2.10
YIN(10dB)	0.28	35.6	1.13	5.13
RAPT(10dB)	1.35	10.9	7.61	2.05

表 1 是三种算法对实验材料中的双音节词提取基频的错误率，表 2 是三种算法对实验材料中的单音节字提取基频的错误率。

4 讨论与分析

4.1 清浊判断正确率

从表 1 和表 2 中的结果可以看到，随着噪声的增加，三种算法的浊音错误率(VU)都显著上升，而清音错误率(UV)出现下降。发生这种变化是因为随着噪声增加，浊音中在语音中的能量比例下降，有更多的语音段被判断为清音。

比较不同算法之间清浊判断错误率，YIN 算法浊音错误率显著高于另外两种算法。RAPT 算法具有最低的浊音判断错误率，并且上升比例比其他两种算法低。这说明了 RAPT 算法对语音信号的清浊判断比另外两种算法准确。

而本文 DSPDA 算法在安静环境下的清浊判断错误率与 RAPT 接近，但噪声升高之后错误率显著升高，说明 DSPDA 算法在三种算法中有较好的清浊判断正确率，但清浊判断的鲁棒性还需提高。

4.2 基频值的准确性

实验结果中，严重错误率(Gross)反映了提取的基频曲线中跳变点出现的频率，平均绝对误差(MAD)反映了基频曲线与标准基频值得偏差。

对于严重错误率(Gross)，DSPDA 算法较稳定地维持在 3.4%左右，保持了较低的水平。RAPT 算法随着噪声增加，严重错误率不断升高。YIN 算法随着噪声增加，严重错误率出现下降，可能与 YIN 算法过高的浊音判断错误率有关。

对于平均绝对误差，DSPDA 算法稳定在 3Hz 以内。而 RAPT 算法和 YIN 算法 MAD 值随着噪声增加而增加，并且 YIN 算法的误差值较大。这说明在本文实验条件下，相对其他两种算法，本文 DSPDA 算法基频值提取准确，错误率稳定。而另外两种算法基频值的准确性随噪声增加而降低。

4.3 算法实时性能分析

计算量和延时是实时算法的重要指标。本文对比的三种基频提取算法都是时域算法，但计算方法不同使得算法的这两种指标有所差异。

DSPDA 算法的主要计算量在自相关计算部分，这部分计算量可以通过限定计算范围降低。算法自相关计算以外部分的计算量与信号帧长成线性增长关系，算法总体计算量较低。并且算法运算时没

有使用后续信息，因此除计算本身消耗的时间以外，没有引入额外的延时。

RAPT 算法使用能量归一化互相关函数进行两次计算。第一次互相关计算使用降采样的方式降低计算量，第二次计算限定计算范围在第一次计算结果中的候选值附近。在能量归一化时，需要平方和运算以及开平方运算，最后的候选值选择也需要大量乘法运算和开平方、求对数等复杂运算。RAPT 算法的计算量并没有数量级上的优势。RAPT 算法同样没有引入额外的延时。

YIN 算法对信号自相关函数使用一个累积平均归一化差分方程进行改进，因此算法主要部分需要两次自相关计算的计算量。在选择候选值时，YIN 算法使用了当前信号前后各两帧信号来计算，因此会引入两帧信号帧移的延时。

从上述分析可知，DSPDA 算法相较于传统自相关法只增加了少量的计算量。与其他时域改进算法相比，DSPDA 算法具有较低的计算量，算法具有实时计算的特点，没有引入额外延时。

5 结论

本文对自相关法提取基频的算法进行了改进，利用动态平滑的技术，提出一种计算量低的实时基频提取算法。动态平滑技术利用上一帧信号的基频值在小范围内搜索候选基频值，并根据候选值的自相关函数幅度以及与上一帧信号基频值的距离确定最佳基频值。对 DSPDA 算法基频提取效果评价的结果显示，DSPDA 算法基频值提取准确，平均绝对误差值小于 3Hz，对于白噪声具有一定鲁棒性。DSPDA 算法在计算新的基频值时，基频提取的计算量低，能够实时计算，适合于需要低计算量实时基频提取的场合。

参 考 文 献

- [1] Milczynski Matthias, et al. Perception of Mandarin Chinese with cochlear implants using enhanced temporal pitch cues[J]. *Hearing Research*, 2012, 285(1-2): 1-12.
- [2] Yuan Meng, Lee Tan. Cantonese tone recognition with enhanced temporal periodicity cues[J]. *J. Acoust. Soc. Am.*, 2009, 126(1): 327-337.
- [3] Alain de Cheveigne, Hideki K. YIN, a fundamental frequency estimator for speech and music[J]. *J. Acoust. Soc. Am.*, 2002, 111(4): 1917-1930.
- [4] Talkin David. A robust algorithm for pitch tracking (RAPT)[A]. Kleijn W. B., Paliwal K. K. *Speech Coding and Synthesis[C]*// Elsevier Science B.V. 1995: 495-518.
- [5] Noll A. M. Cepstrum Pitch Determination[J]. *J. Acoust. Soc. Am.* 1967, 41(2): 293-309.

- [6] Klapuri Anssi. Pitch estimation using multiple independent time-frequency windows[C]// New Paltz, New York: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Oct. 17-20, 1999: 115-118.
- [7] Ney Hermann. A dynamic programming technique for nonlinear smoothing[C]// Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP 81. Apr 1981, 6: 62-65.
- [8] Boersma Paul. Praat, a system for doing phonetics by computer[C]// Glot International 5:9/10, 2001: 341-345.
- [9] Van Immerseel LucM., Martens Jean P. Pitch and voiced/unvoiced determination with an auditory model[J]. J. Acoust. Soc. Am, 1992, 91(6): 3511-3526
- [10] Licklider J.C.R., Pollack I. Effects of differentiation, Intergration, and Infinite Peak Clipping upon the Intelligibility of Speech[J]. J. Acoust. Soc. Am, 1948, 20(1): 42-50.
- [11] Boersma Paul. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound[J]. Proceedings of the Institute of Phonetic Sciences, 1993, 17(1): 97-110.
- [12] Yuan Meng, et al. Mandarin tone perception with temporal envelope and periodicity cues from different frequency regions[C]// Chinese Spoken Language Processing, 2008. ISCSLP '08. 6th International Symposium on. 16-19 Dec. 2008: 1-4
- [13] Rabiner L. R. On the use of autocorrelation analysis for pitch detection[J]. IEEE transactions on Acoustics, Speech and Signal Processing, 1977, 25(1): 24-33.
- [14] Alain de Cheveigne, Hideki K. Comparative evaluation of F_0 estimate algorithms[C]// Eurospeech 2001 Scandinavia, 2451-2454.
- [15] Rabiner L. R. A comparative performance study of several pitch detection algorithms[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1976, 24(5): 399-418.
- [16] Luengo Iker, et al. Evaluation of pitch detection algorithms under real conditions[C]// ICASSP 2007.

2012 年度章奎生声学设计研究所大事记

1. 2012 年全年承接新签工程合同项目共 19 项，全部为自主承接项目，合同总额达 1334 万元，与去年基本相同，其中建声厅堂音质项目占 85%，外地项目占 85%以上，50 万以上项目占 50%，最大单项建声设计合同额达 218 万，为建所以来的最大单项合同额。全年实收产值与去年基本持平；

2. 今年年初在长沙梅溪湖国际文化艺术中心项目国内外建声专业设计比选中胜过，德国 BBM、英国 TPC、新西兰 MAD 三国著名声学公司，中选并签约建所 12 年来最大建声专业设计合同额项目，为中国人争了光，也为我国建声设计业争了光；

3. 本声学所参与华东院合作设计并负责建声设计的上海梅赛德斯-奔驰文化中心的工程 2012 年度先后获全国和上海的优秀设计工程一等奖，并在集团华东院立项本所参加合作项目“超大空间建筑设计技术研究”成果中获集团科技成果一等奖；

4. 声学所宋拥民博士六月上旬因公赴伦敦到著名的扎哈·哈迪德建筑师事务所参访并洽谈梅溪湖工程合作设计协调配合工作事宜；杨志刚高工于六月上旬应邀赴日本参访日本歌剧院和音乐厅等演艺建筑工程项目；章总均因病住院未能同行。

5. 声学所在 2012 年结合项目设计独立承担完成了哈尔滨大剧院和长沙滨江文化园工程等多个剧院和音乐厅的缩尺模型试验工作，并制作了模型 1:10 混响室和微型球面声源，为本所独立对外承接厅堂音质模型试验提供了技术保证和试验条件；

6. 声学所王静波主任和宋拥民博士于 5 月下旬赴香港会展中心出席由美国声学学会、中国声学学会、香港声学学会及西太平洋声学学会联合举办的国际声学会议并作学术论文报告。章奎生所长因病休未能赴会，但他报送的学术论文也正式编入了会议论文集并已被推荐刊发于美国声学学报(J Acoustic Soc Am)2012 年第 4 期中；

7. 声学所走产学研结合发展道路，与上海中华声学装备公司合作设计建设的联合声学实验室暨测试中心今年五月正式建成使用并通过专家评审和验收，达到国内同类声学实验室领先水平，为声学所开展课题试验研究工作提供了测试基地，受到同行业专家的关注和好评；

8. 声学所与上海建筑设计研究院合作承担的无锡大剧院建筑设计和声学设计工程年内正式建成使用并通过了现场主、客观音质测试和评价，获得了好评。声学所还获得了无锡大剧院工程设计立功竞赛奖，章奎生所长应邀专程赴无锡接受无锡市长颁发的立功奖牌；

9. 年内声学所承担了集团立项课题“剧院与音乐厅音质参量的现场检测和分析评价研究”，集团和都市院 2012 年课题“声学缩尺模型测试技术研究”和“材料吸声性能现场测试技术的研究”共 3 个课题研究工作，都取得了较大的进展。预计 2013 年上半年可先后完成该 3 项课题，通过结题评审工作；

10. 声学所章奎生所长率 7 人作为一个学术团队于今年 11 月下旬应香港声学学会之邀赴香港出席第五届海峡两岸声学技术交流研讨会，并在会上提出了共 8 篇学术论文报告，受到港、澳、台共百余名与会代表的欢迎和关注。章奎生教授还应会议主席香港声学学会主席郑建汇博士之邀，作为嘉宾在会议开幕式和闭幕式上先后致辞，表示祝贺及感谢，扩大了声学所在港、澳、台的影响。章奎生所长还代表上海市声学学会与香港、台湾声学学会商定了明年秋在台北召开第六届海峡两岸声学技术交流研讨会的有关事宜。