

# 一种基于联合源-滤波器模型优化的语音声门源模型估计方法

付 强<sup>1</sup>, Peter Murphy<sup>2</sup>, 颜永红<sup>1</sup>

(1. 中科信利语音实验室, 中国科学院声学研究所, 北京 100080; 2. 爱尔兰里默瑞克大学电子与计算机工程系, 里默瑞克, 爱尔兰)

**摘 要:** 本文论述了一种基于联合源-滤波器分离的稳健声门源模型估计方法. 此方法利用 LF(Liljencrants-Fant)模型对声门波导数(glottal flow derivative)进行建模, 而声道被描述为一个时变的 ARX模型. 由于联合估计问题是一个多变量非线性优化过程, 本文采用了一个两阶段(two-pass)的实现策略来解决这一问题. 第一阶段初始化声门源和声道模型, 并为其后的联合优化过程提供稳健的初始参数. 第二阶段的联合估计则最终决定模型估计的精度, 由信任域下降优化算法实现. 通过分别对合成和真实语音的实验, 表明该方法是一种具有一定精度和较好的稳健性的声门源模型估计算法.

**关键词:** 语音产生模型; 源-滤波器; 声门波导数; ARX; 声门波估计; 联合优化

**中图分类号:** O424 **文献标识码:** A **文章编号:** 0372-2112 (2007) 05-0982-05

## A Robust Glottal Source Estimation Based on Joint Source-Filter Model Optimization

FU Qiang<sup>1</sup>, Peter Murphy<sup>2</sup>, YAN Yong-hong<sup>1</sup>

(1. ThinkIT Speech Laboratory, Institute of Acoustic, Chinese Academy of Science, Beijing 100080, China;

2. Department of ECE, University of Limerick, Limerick, Ireland)

**Abstract:** This paper describes a robust glottal source estimation method based on a joint source-filter separation technique. In this method, the glottal flow derivative is modelled as the Liljencrants-Fant (LF) model and the vocal tract is described as a time-varying ARX model. Since the joint estimation problem is a multi-parameter nonlinear optimization procedure, we separate the optimization procedure into two passes. The first pass initializes the glottal source and vocal tract models providing robust initial parameters to the following joint optimization procedure. The joint estimation determines the accuracy of model estimation, which is implemented with a trust-region descent optimization algorithm. Experiments with synthetic and real voices show the proposed method is a robust glottal source parameter estimation method with a considerable degree of accuracy.

**Key words:** speech production model; source-filter model; glottal flow derivative; ARX; glottal source estimation; joint optimization

## 1 引言

声门源估计(glottal source estimation)技术可被应用于语音分析、合成、编码、临床喉部病理分析和嗓音障碍诊断等. 传统上, 声门波形由逆滤波(inverse filtering)技术得到<sup>[1-4]</sup>. 近多年来, 一些基于联合源滤波器分离(joint source-filter separation)的方法也逐渐成为研究重点<sup>[5-8]</sup>. 在此类方法中, 声门和声道模型参数在一个联合优化准则下同时进行估计. 与基于逆滤波的时不变源滤波器相比较, 这种方法允许我们把声道看作时变系统, 即源与声道相互独立<sup>[3]</sup>. 因此, 声门源可被描述为没有共振峰纹波(formant ripples)的声门波导数模型, 如

Rosenberg<sup>[10]</sup>和 LF<sup>[11]</sup>等时间多变量模型, 它们有充足的时域细节来描述声门波信号.

本文将阐述一种准确且稳健的声门源联合估计方法. 在此方法中, 选择被广泛应用且具有一定灵活性的 LF 模型来描述声门源<sup>[7,9,12]</sup>, 声道则用时变 ARX(Auto-regressive with exogenous)模型来建模. 声道滤波器系数由 Kalman 滤波过程<sup>[13]</sup>自适应地估计.

联合估计本质上是一个多维非线性优化问题, 它的误差函数并非一个简单凸函数<sup>[14]</sup>. 其优化过程很容易陷入局部最小点<sup>[12]</sup>. 对于一个给定的优化问题, 良好的初始化过程对全局最小点的搜索至关重要<sup>[9]</sup>. 本文的策略是将优化过程分为两个阶段来实现, 图 1 给出了该方

法的结构图. 第一阶段对声门源和声道模型初始化, 确保能够为其后的联合优化提供稳健的初始化参量. 第二阶段的联合优化则主要关注模型估计的准确性.

## 2 联合估计问题的提出

### 2.1 基音周期的确定

由于该算法是基音同步的, 因此需要首先确定每一基音周期的起止位置. 声门关闭瞬间 (glottal closure instant, GCI) 是所有能够刻化基音周期的语音事件中, 最容易被检测到的<sup>[2]</sup>. 它对应着所有共振峰被同时激励的瞬间, 在声门波导数波形上则表现出明显的奇异性, 如图 1 所示. 为此我们选择 GCI 作为基音周期的起止位置.

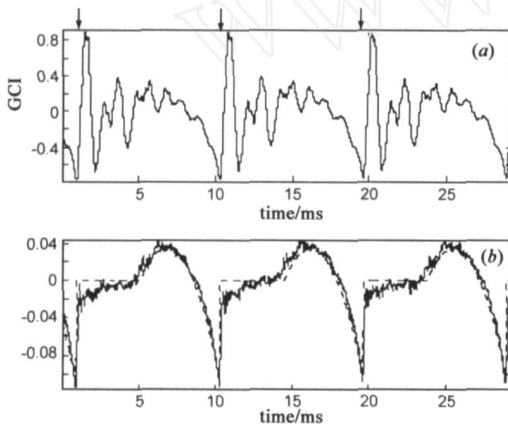


图 1 (a) 带有 GCI 标注的原始语音, (b) 用基音同步逆滤波法估计的声门波导数波形 (虚线为利用 3.1 中方法建立的最佳匹配 Rosenberg 模型)

### 2.2 时变 ARX 语音产生模型

在所提出的算法中, 语音产生模型被看作一个时变 IIR 系统<sup>[8]</sup>, 如下

$$s[n] + \sum_{i=1}^p a_i[n]s[n-i] = \sum_{j=1}^q b_j[n]v_g[n-j] + v_g[n], \quad 0 \leq n \leq N_0 - 1 \quad (1)$$

这里  $s[n]$  和  $v_g[n]$  分别表示在一个基音周期内观察到的语音信号和未知的声门波导数波形信号.  $N_0$  表示每基音周期样本长度.  $a_i[n]$  和  $b_j[n]$  分别是时变 IIR 滤波器的自回归 (AR) 和滑动平均 (MA) 系数.  $p, q$  表示滤波器的阶数, 即分别代表模型极、零点的个数, 其典型值分别为 14 和 2.

因为 LF 模型能够以较少的参数覆盖常见的现实声门波形, 以及在匹配大多数发音 (phonation) 上的灵活性<sup>[11]</sup>, 所以这里的声门波导数采用 LF 模型来描述. 设最大声门流瞬间, 最大声门波导数负值瞬间和声门闭合瞬间分别为  $N_p, N_e$  和  $N_c$ , 则离散时间的 LF 模型可被定义为:

$$v_g^{LF}[n] = \begin{cases} E_0 e^{-n} \sin(\omega_g n), & 0 \leq n < N_e \\ -\frac{E_c}{N_a} [e^{-n(N_c-N_e)} - e^{-N_c(N_c-N_e)}], & N_e \leq n < N_c \\ 0, & N_c \leq n \leq N_0 - 1 \end{cases} \quad (2)$$

除去声门闭合幅度参数  $E_c$ , LF 可以由两组相互独立的参数指定: 直接合成参数 ( $E_0, \omega_g$ ) 和时间参数 ( $N_p, N_e, N_a, N_c$ ). 如前所述, 在我们的方法中, 每个基音周期的起点和终点是 GCI, 这与  $v_g^{LF}$  不同. LF 的表达式  $v_g^{LF}$  和  $v_g^{LF}$  之间有一个时间平移关系, 可以表示为:

$$v_g^{LF} = \begin{cases} v_g^{LF}[n + N_e + 1], & 0 \leq n \leq N_0 - N_e - 2 \\ v_g^{LF}[n - (N_0 - N_e - 1)], & N_0 - N_e - 1 \leq n \leq N_0 - 1 \end{cases} \quad (3)$$

### 2.3 联合估计问题的方程

$$X[n] = \{ a_1[n], \dots, a_p[n], b_1[n], \dots, b_q[n] \}^T$$

$$H[n] = \{ -s[n-1], \dots, -s[n-p], v_g^{LF}[n-1], \dots, v_g^{LF}[n-p] \}^T$$

分别定义为 ARX 模型的状态向量和观测向量. 我们采用最小平方准则对声门源和声道进行联合估计, 即:

$$\min E(\cdot) = \min_{N_0} \frac{1}{N_0} \sum_{n=0}^{N_0-1} \{ e[n] \}^2 = \min_{N_0} \sum_{n=0}^{N_0-1} e[n]^2$$

$$= \min_{N_0} \sum_{n=0}^{N_0-1} \{ s[n] - H[n]^T X[n] - v_g^{LF} \}^2 \quad (4)$$

当  $0 < N_p < N_e < N_c = N_0$  时成立. 这里  $\mathcal{N}$  表示离散 LF 模型参数集, 即  $\mathcal{N} = \{ N_p, N_e, N_a, N_c, E_c \}$ . 当选定一个  $\mathcal{N}$  时,  $v_g^{LF}$  也就唯一地确定了. 于是式 (4) 可以化简为

$$\min_{X[n]} E(X[n]) = \min_{X[n]} \sum_{n=0}^{N_0-1} \{ z[n] - H[n]^T X[n] \}^2 \quad (5)$$

其中,  $z[n] = s[n] - v_g^{LF}[n], 0 \leq n \leq N_0 - 1$ .

时变状态向量  $X[n]$  可由 Kalman 滤波器唯一确定, 且定义域内的每个  $\mathcal{N}$  都唯一对应的  $E(\cdot)$ , 由此形成一个以  $\mathcal{N}$  为自变量的多维误差函数. 从优化理论的角度观点看, 等式 (4) 是一个无约束有界非线性最小化问题. 很明显,  $E(\cdot)$  没有闭式解. 此外, 观察等式 (2) 和式 (4), 易见  $E(\cdot)$  的误差表面是非凸的, 这就意味着也没有有效的迭代算法能够保证收敛到全局最优解. 为了解决这一复杂的非线性优化问题, 我们的策略是首先构造一个能较容易确定全局解的近似问题, 这一近似问题全局解被用作第二阶段中的下降算法的初始点, 从而通过进一步迭代得到一个稳健且准确的方程 (4) 的最终解.

### 3 初始化

相对于准确性而言, 在初始化阶段我们更关心估计过程的收敛性和稳健性. 这里与接下来的联合估计过程主要有两个方面的不同. 第一, 为了简化问题, 声道被看作全极点 LTI 系统; 第二, 在这里声门波导数由数学上更为简单的两参量 Rosenberg 来建模, 而不是 LF 模型.

### 3.1 近似问题的建立

在 LTI 语音生成过程假定下,声道的传递函数可被典型地表达  $V(z) = 1/1 - \sum_{i=1}^p a_i z^{-i}$  这样输出的语音可以表示为:

$$s[n] = \sum_{i=1}^p a_i S[n-i] + v_g[n], 0 \leq n \leq N_0 - 1 \quad (6)$$

这里  $v_g$  由 Rosenberg 模型给出,它在声门波导数开相(open phase)是一个简单的二阶多项式,而在闭相则为 0,如图 1 所示.

$$v_g^{RS}(n) = \begin{cases} 2cn - 3dn^2, & 0 \leq n \leq OQ \cdot N_0 \\ 0, & OQ \cdot N_0 < n \leq N_0 - 1 \end{cases} \quad (7)$$

$$c = \frac{27 \cdot AV}{4 \cdot (OQ^2 \cdot N_0)} \quad (8)$$

$$d = \frac{27 \cdot AV}{4 \cdot (OQ^3 \cdot N_0^2)} \quad (9)$$

我们把 Rosenberg 参数向量表示为:  $\theta = [OQ, AV]^T$ . 其中  $OQ$  ( $0 < OQ < 1$ ) 是声门开商(open quotient),是声门波形幅度峰值与  $N_0$  之比. 与 2.1 中所述相同的原因, Rosenberg 模型中表达式  $v_g^{RS}$  和  $v_g^{LF}$  之间也存在一个时间平移关系,即:

$$v_g^{RS}(n) = \begin{cases} v_g^{LF}[n + OQ \cdot N_0 + 1], & 0 \leq n \leq N_0 - OQ \cdot N_0 - 2 \\ v_g^{LF}[N_0 - OQ \cdot N_0 - 1], & N_0 - OQ \cdot N_0 - 1 \leq n \leq N_0 - 1 \end{cases} \quad (10)$$

参照等式(4),得到构造一个近似的最小平方误差问题为:

$$\min E(\theta) = \frac{1}{N_0} \sum_{n=N_0-OQ \cdot N_0-1}^{N_0-1} \{ \hat{v}_g[n] - v_g^{RS}[n] \}^2$$

其中  $\hat{v}_g$  是通过基音同步协方差线性预测逆滤波法<sup>[3]</sup>得到的声门波导数波形.  $v_g^{RS} = [v_g^{RS}[0], v_g^{RS}[1], \dots, v_g^{RS}[N_0 - 1]]^T$ . 实际上定义了一个声门波形拟合过程(glottal waveform fitting)<sup>[14]</sup>. 尽管 Rosenberg 声门模型与 LF 模型相比不够准确,适应性也不够好,但它简单的数学表达式保证了相应的声门波形拟合过程可以较容易地收敛到全局最小点,计算量也随之减少<sup>[14]</sup>. 这是一个非常重要的性质,因为这里初始化过程的主要目的就是为下一阶段的联合优化提供稳健的初始参数.

考虑到实际应用中存在 GCI 检测错误以及  $\hat{v}_g$  中存在噪声和纹波(ripples)的可能性,在实现声门适应过程时使用了 Simplex 算法<sup>[15]</sup>. 这是一种直接搜索算法,并不涉及数值或解析的梯度向量运算. 我们把第一阶段的中的最优解表示为  $OQ^1, AV^1$ .

### 3.2 直接估计

声门波拟合过程本身也需要一个初始化值. 由于

所估计的声门波导数  $\hat{v}_g[n]$  往往包含纹波(这是由未被完全逆滤波的共振峰引起的)和噪声成分,故先对它进行低通滤波以减少测量误差. 即,  $\hat{v}_g[n]$  与一个长度为 7 的 Blackman 窗函数做卷积. 直接估计中的基本规则如下<sup>[17]</sup>:  $U_m = \max(\hat{v}_g^{lp})$ :  $\hat{v}_g^{lp}$  的最大幅值,  $N_z$ : 声门闭合瞬间,  $N_o$ : 声门开启瞬间. 我们将 Rosenberg 模型的初值设为  $OQ^0 = (N_0 - N_o)/N_0$  (开商初值);  $AV^0 = U_m/N_0$  (幅度初值). 这样,声门回复期(duration of the return portion)即为  $\tau_{ce} = N_z + 1$ .

### 3.3 从 Rosenberg 到 LF 的参数转换

在联合估计过程中真正采用的声门源是 LF 模型,因此必须将所得到的 Rosenberg 模型参数估计值( $OQ^1, AV^1$ )转换为 LF 模型参数( $N_p^1, N_e^1, N_a^1, N_c^1, E_e^1$ ). 为了避免联合估计过程中误差函数出现台阶(staircase)现象<sup>[12]</sup>,我们采用时间参数与  $N_0$  的比值( $r_p^1 = N_p^1/N_0, r_e^1 = N_e^1/N_0, r_a^1 = N_a^1/N_0, r_c^1 = N_c^1/N_0$ )作为误差函数的输入变量,而不是直接应用离散时间变量. 这样,模型参数转换方程为:

$$r_e^1 = OQ^1 \quad (11)$$

$$r_p^1 = \frac{2r_c^1}{3} \quad (12)$$

$$r_c^1 = r_e^1 + \frac{\tau_{ce}}{N_0} \quad (13)$$

$$r_a^1 = \frac{2\tau_{ce}}{3} \quad (14)$$

$$E_e^1 = -v_g^{RS}[N_e^1] = -v_g^{RS}[\text{int}(r_e^1 \cdot N_0)] \quad (15)$$

## 4 联合估计的实现

### 4.1 ARX 模型辨识

用来解式(5)中的 ARX 辨识问题的状态空间模型定义如下:

$$X[n] = X[n-1] \quad (16a)$$

$$z[n] = H^T[n]X[n] + e[n] \quad (16b)$$

其中式(16a), (16b) 分别是状态转移方程和观测方程.

对每一个基音周期,  $X[n]$  的初始条件设为  $X[0] = [\hat{a}_0, \hat{a}_1, \dots, \hat{a}_p, \bar{b}_0, \bar{b}_1, \dots, \bar{b}_q]^T$ , 其中  $\bar{b}_j$  表示上一个周期 ARX 模型第  $j$  个 MA 系数的在整个周期内的平均值.

### 4.2 声门源优化

通过初始化过程,我们找到了  $\theta^1 = \{r_p^1, r_e^1, r_a^1, r_c^1, E_e^1\}$  作为等式(4)的优化问题的良好的初始点. 这样就可以用基于目标函数  $E(\theta)$  的梯度和 Hessian 矩阵的下降算法得到一个更加精确的解. 此外,式(4)是一个非线性最小平方(least-squares)问题. 我们可以从梯度和 Hessian 矩阵的特殊结构中得到比一般非约束问题更有效的收敛方式. 具体来讲,我们选择了文献[18]中描述的 interior trust region algorithm. 它能产生严格可行的迭

代,并拥有局部二次收敛特性.它还能处理如式(4)的简单边界约束问题.在信任域方法中,增量是一个有边界的二次亚问题的近似解<sup>[18]</sup>,即

$$\min_s Q_k(s) = g_k^T s + \frac{1}{2} s^T M_k s : D_k s \quad (17)$$

其中

$$M_k = B_k + C_k \quad (18)$$

$$C_k = D_k \text{diag}(g_k) J_k^v D_k \quad (19)$$

这里  $g_k = \nabla E(\cdot)_k$ ,  $B_k$  是 Hessian 矩阵  $\nabla^2 E(\cdot)_k$  的一个对称逼近,  $\Delta_k$  是一个代表信任域大小的正标量,  $D_k$  是一个尺度矩阵,定义为:

$$D(\cdot) = \text{diag}(v(\cdot)^{-1/2}) \quad (20)$$

其中  $v(\cdot)$  定义如下:对于每一个成员  $1 \leq i \leq 5$ , (i) 如果  $g_i < 0$ , 则令  $v_i = |g_i|^{-1}$ ; (ii) 如果  $g_i \geq 0$ , 则令  $v_i = |g_i|$ . 这样,当  $|v(\cdot)|$  可微时,等式(19)中的  $J_k^v$  是  $|v(\cdot)|$  的雅可比行列式.

## 5 实验及讨论

### 5.1 合成元音

合成语音按照线性时不变源-滤波器模型生成,它采用 LF 模型作为声门波导数.我们首先采用 500 个合成元音 / IY/ 来评价所提出的方法.这 500 个元音具有不同的 LF 参数,基本上覆盖了 5 种典型的噪音类型,包括 modal, vocal fry, breathy, falsetto and harsh<sup>[19]</sup>. 每种噪音类型样本有 100 种不同的 LF 时间参数设置,以典型值<sup>[19]</sup>为中心呈高斯分布.作为客观评价度量,我们计算估计值与真值的相对误差均值  $r$  如下:

$$r = \frac{1}{100} \sum_{j=1}^{100} \frac{|es_j - or_j|}{or_j} \quad (21)$$

这里  $es_j$  和  $or_j$  分别是估计的和原始的 LF 时间参数值.测试结果列于表 1. 总的来说,  $t_p$ ,  $t_e$ ,  $t_a$  和  $t_c$  的估计准确性较高而  $t_a$  和  $E_e$  相对较差.并且,不同噪音类型的元音具有不同的声门噪声强度,范围从 50dB (falsetto) 到 10dB (harsh), 而估计结果显示该算法对声门噪声强度不敏感,因此,所提方法对加性声门噪声具有较高的稳健性.

表 1 典型合成噪音的估计结果

噪音类型	Tp (%)	Te (%)	Ta (%)	Tc (%)	Ee (%)
Modal	9.97	9.56	28.39	10.80	10.33
Vocal fry	17.17	10.80	43.18	15.33	15.70
Breathy	4.68	6.23	11.97	6.99	23.65
Falsetto	5.36	7.36	10.42	4.63	13.15
Harsh	16.85	12.07	11.86	11.74	17.89

### 5.2 真实元音

为测试真实语音,我们从文献[19]所提供数据库中选取了下面的英文句子:“We were away a year ago”. 这句

话由一男声读出,音调响度都比较温和.由于这种情况下没有 LF 模型“正确”的时间参量,我们选取与该段语音同步录制的电声门 (EGG, electroglottographic) 信号的开商  $c$  参数  $OQ_{EGG}$  作为参照值.对于  $OQ_{EGG}$  的计算按照文献[3]中描述的算法,首先测量声门开启瞬间 GOI (Glottal Opening Instants) 和 GCI. 有了 GOI 和 GCI, 就能很容易的得到  $OQ_{EGG}$ .

$OQ_{EGG}$  对应于 LF 模型中的时间参数  $t_c$ . 为避免非元音信号对估计结果的影响,我们对该段语音标注出了元音段,仅对元音段进行估计结果客观评价.计算得到的相对误差  $r$  平均值为 10.47%. 这说明对于真实元音 (主要是元音),所提估计方法仍然具有较好的准确性.

## 6 结论

本文介绍了一种无需闭相位置信息的基音同步声门源估计算法.在每个基音周期内,该算法用自适应迭代方式对声门源和声道参数进行同步估计.通过 Kalman 滤波器的使用,将声门源和声道都看作时变的,这样就能够描述更为复杂的声道特性.选择 LF 声门源模型,使得该算法能够适应较广泛的噪音类型.所提算法的一个重要特点是,将多维联合估计过程分为两个阶段实现,能够有效地提高优化过程的收敛性和准确性.通过对合成声音和真实语音的实验表明,该算法对声门源模型的估计具有较好的稳健性和准确性.

### 参考文献:

- [1] Wong D Y, Markel J D, Gray Jr. A H. Least squares glottal inverse filtering from the acoustic speech waveform [J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1979, ASSP-27(4): 350 - 355.
- [2] Hess W, Ed by K-S Fu, et. al. Pitch Determination of Speech Signals- Algorithms and Devices [M]. Springer-Verlag, Berlin, 1983.
- [3] Krishnamurthy A K, Childers D G. Two-channel speech analysis [J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1986, ASSP-34(4): 730 - 743.
- [4] Alku P. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering [J]. Speech Communication, 1992, 11(2-3): 109 - 118.
- [5] Milenkovic P. Glottal inverse filtering by joint estimation of an AR system with a linear input model [J]. IEEE Trans on Acoustics, Speech and Signal Processing, 1986, ASSP-34(1): 28 - 41.
- [6] Fujisaki H, Ljungqvist M. Estimation of voice source and vocal tract parameters based on ARMA analysis and a model for the glottal source waveform [A]. Proc IEEE ICASSP[C]. Dallas, TX: IEEE, 1987. 637 - 640.

- [7] Krishnamurthy A K. Glottal source estimation using a sum-of-exponentials model [J]. IEEE Trans on Signal Processing, 1992, 40(3) :682 - 686.
- [8] Ding W, Kasuya H, Adachi S. Simultaneous estimation of vocal tract and voice source parameters based on an ARX model[J]. IEICE Trans Inf & Syst, 1995, E78-D(6) :738 - 743.
- [9] Frohlich M, Michaelis D, Strube H W. SIM -simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals[J]. J Acoust Soc Am, 2001, 110(1) :479 - 488.
- [10] Rosenberg A E. Effect of glottal pulse shape on the quality of natural vowels[J]. J Acoust Soc Am, 1971, 49(2) :583 - 590.
- [11] Fant G. A four-parameter model of glottal flow [R]. Speech Transmission Lab Quart Progr, 1985, Rep 4/ 85, KTH:1 - 3.
- [12] Strik H. Automatic parameterization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses [J]. J Acoust Soc Am, 1998, 103(5) :2659 - 2669.
- [13] Haykin S. Adaptive Filter Theory, 4th ed[M]. Prentice Hall, 2001.
- [14] Lu H L, Smith III J O. Joint estimation of vocal tract filter and glottal source waveform via convex optimization [A]. Proc IEEE Workshop on Application of Signal Processing to Audio and Acoustics [C]. New York: New Paltz, 1999. 17 - 20.
- [15] Rao S S. Engineering Optimization: Theory and Practice, 3rd Ed[M]. Interscience, 1996.
- [16] Nelder J A, Mead R. A simplex method for function minimization[J]. Comput J (Switzerland), 1964, 7(2) :308 - 313.
- [17] Alku P, Vilkman E. Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering [J]. J Acoust Soc Am, 1995, 98(2) :763 - 767.
- [18] Coleman T F, Li Y. An interior, trust region approach for nonlinear minimization subject to bounds [J]. SIAM Journal on Optimization, 1996, 6(2) :418 - 445.
- [19] Childers D G. Speech processing and synthesis toolboxes[M]. Danvers, USA: John Wiley & Sons, Inc 2000.

#### 作者简介:

**付强** 男, 1972 年生于西安. 中国科学院声学研究所副研究员. 研究方向为语音/ 音频处理中的信号分析、编码、噪声消除、客观质量评价和语音识别系统的稳健前端处理等.

E-mail: qfu@hcl. ioa. ac. cn

**Peter Murphy** 男, 爱尔兰里莫瑞克大学电子工程与计算机系 Lecturer. 研究方向为物理声学、语音处理等.

**颜永红** 男, 2002 年入选中科院百人计划, 任声学所研究员、博士生导师, 中科信利语音实验室主任和所长助理. 研究方向为大词表非特定人连续语音识别、多模口语系统、嵌入式系统、多媒体数据检索、系统自适应和快速搜索算法.