

DOI: 10.7652/xjtuxb201304003

高斯混合模型下的残留回波抑制算法

张琦¹, 王霞¹, 孙焘²

(1. 西安交通大学电子与信息工程学院, 710049, 西安; 2. 中兴通讯股份有限公司, 710065, 西安)

摘要: 为提高通话环境中的回波抑制效果,提出了基于高斯混合模型下的残留回波抑制方法。首先以线性预测系数为特征提取近端用户语音的特征向量,然后在高斯混合模型下对其进行训练,得到用户语音模板,通过对比用户语音模板和残留回波信号特征向量,判断非用户语音帧,最后将非用户语音帧予以消除。仿真实验结果表明,所提算法能正确区分用户语音帧与非用户语音帧,具有更好的回波抑制效果。在误判率约为 1%~2% 的情况下,可将自适应滤波器的回波消除量提高 7~10 dB。

关键词: 自适应回波抑制;线性预测系数;高斯混合模型;残留回波抑制

中图分类号: TN912.3 **文献标志码:** A **文章编号:** 0253-987X(2013)04-0011-06

An Algorithm for Cancelling Residual Echoes Based on Gaussian Mixture Model

ZHANG Qi¹, WANG Xia¹, SUN Tao²

(1. School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. ZTE Corporation, Xi'an 710065, China)

Abstract: An efficient way based on the Gaussian mixture model (GMM) is proposed to improve the performance of echo suppression in telephone. The speech feature vector of near-end users is extracted using linear prediction coefficients, and the user speech template is obtained by training user speech samples under the GMM. Then a judgment for a non-user speech frame can be made by comparing the residual echo signals with the user speech template, and the non-user speech frame is cancelled. Simulation results show that the proposed algorithm can distinguish user speech frame and non-user speech frame very well and has a better echo cancellation quantity. The echo cancellation quantity of the adaptive filter is improved by 7~10 dB, when the error rate is about 1%~2%.

Keywords: adaptive echo suppression; linear coefficient prediction; Gaussian mixture model; residual echo suppression

在移动通信中,尤其是在免提模式下,常常会出现回波现象,回波的存在降低了用户的通话质量,严重时甚至导致用户无法正常通信。回波抑制技术目前多采用自适应滤波器模拟回波路径,以得到模拟回声信号,并使含有回声的信号减去该模拟回声信

号,在回声传到用户之前消除其影响。回波抑制原理如图 1 所示,误差信号即为回波抑制后的输出信号,图中自适应滤波器常常由 FIR 线性滤波器实现^[1-2]。然而,在实际应用中会有很多因素,如双端通话问题、非线性回波问题等,影响了自适应滤波器

收稿日期: 2012-07-29。 作者简介: 张琦(1988—),女,硕士生;王霞(通信作者),女,副教授。 基金项目: 国家自然科学基金资助项目(60902045);中兴通信股份有限公司产学研项目。

网络出版时间: 2013-03-07

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20130307.0828.008.html>

的回波处理效果。

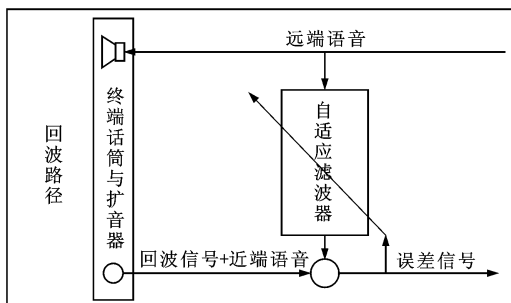


图1 自适应回波抑制原理图

目前,一些学者利用各种非线性处理方法来实现回波消除:文献[3]采用基于神经网络的自适应滤波器方法进行回波抑制,该方法较已有算法具有更好的收敛性,但是它主要用来处理多通道回波抑制问题;文献[4]将盲分离技术和自适应滤波技术结合起来消除回波,在自适应滤波器进行训练之前将远端语音和噪声信号分离,使得训练序列为纯净的远端语音,以此提升自适应滤波器的模拟效果,但是此方法仅在噪声较大的情况下才具有优势;文献[5]针对双端通话提出了基于高斯混合模型的频域检测算法,在高斯混合模型下将信号的相关序列作为特征序列进行训练,分别得到单端通话和双端通话的模板,对输入语音进行辨别,当出现双端通话时自适应滤波器将不再更新,以免滤波器性能恶化,然而这种方法只是在双端通话时对自适应滤波器的性能进行了保持,自适应滤波器只有在阶数和房间的单位冲激响应匹配时才能达到不错的效果,并且它也仅能消除线性回波。为进一步提升自适应滤波器法的性能,本文对自适应回波抑制后的残留回波抑制问题作了进一步的研究,提出了基于高斯混合模型的残留回波抑制算法(residual echo cancellation based on Gaussian mixture model, RECG)。RECG算法首先建立手机用户的语音模板,然后依据远端语音信号设置判决门限,通过比较自适应滤波器输出的误差信号和用户语音模板的差异进行用户语音帧判决,消除非用户语音帧以减少残留回波信号(包含滤波器阶数不匹配产生的残留线性回波和自适应滤波器不能处理的非线性回波)。本文算法提高了回波抑制器的性能,仿真实验证明了本文算法的有效性。

1 自适应滤波器法进行回波抑制

设 $x(n)$ 是滤波器输入参考信号(远端语音信号), $y(n)$ 是远端信号经过回波路径后产生的回波

信号, $v(n)$ 是近端用户语音信号,则期望信号 $d(n) = y(n) + v(n)$ 。 N 阶自适应滤波器对回波信号的估值为 $\hat{y}(n) = \sum_{i=0}^{N-1} \hat{w}(i)x(n-i)$, 此时误差信号可表示为 $e(n) = \mathbf{x}^T(n)(\mathbf{w} - \hat{\mathbf{w}}) + v(n)$, 其中 $\mathbf{w}$ 、 $\hat{\mathbf{w}}$ 分别是房间单位冲激响应系数向量和滤波器系数向量, $\mathbf{x}(n)$ 是远端语音信号向量。

本文的自适应回波抑制采用 NLMS 算法,该算法的滤波器系数更新公式为

$$\hat{\mathbf{w}}(n+1) = \hat{\mathbf{w}}(n) + \frac{\mu}{\mathbf{x}^T(n)\mathbf{x}(n) + \alpha} e(n)\mathbf{x}(n) \quad (1)$$

式中: μ 是一个正常数($\mu < 2$), 通常称为自适应步长因子; α 为任意小的正数。

理想状态下的自适应滤波器可以完全模拟回波信号,此时误差信号中将不再含有回波信号。然而,在利用自适应滤波器进行回波抑制时滤波器阶数需提前设定,当设定的阶数与真实房间单位冲激响应阶数不匹配时,自适应滤波器的输出将出现残留回波信号。针对这一问题,本文提出了 RECG 算法来进一步提高回波抑制性能。

2 RECG 算法模块

人与人之间在发音器官上存在着差异,同时讲话时发音习惯也有不同,这使得每个人的语音具备独一无二的点。本文提出的 RECG 算法利用声源的不同特点,将需要的语音信息保留,对自适应滤波器输出的误差信号作进一步的处理,消除残留回波,从而解决自适应滤波器参数设置不恰当以及存在非线性回波所带来的残留回波问题。

RECG 算法中包含有用户语音模型的建立、实时语音的处理以及语音的判决输出模块。

2.1 用户语音模型

在处理残留回波之前,首先建立用户语音模型,并将此模型存储在终端中作为用户语音帧的模板。该模型对录制的近端用户语音依次进行端点检测和特征提取,最后建立高斯混合模型。

由于用户提前录制的语音序列 $f(n)$ 不完全是语音帧,还存在说话的停顿间隔,所以需要对其进行端点检测和重组处理。将 $f(n)$ 分成 K 帧 $f_i (i=1, 2, \dots, K)$, 帧长为 L , 则帧能量为 $E_i = \sum_L f_i^2$ 。定义能量判决参数为

$$E_{\text{level1}} = \max(E_i) + T_N \quad (2)$$

$$E_{\text{level2}} = \max(E_i) + T_L \quad (3)$$

式中: $\max()$ 代表取 K 帧中能量最大的值; T_N 、 T_L 分别为语音能量门限 ($T_N < T_L < 0$)。

首先找到 $E_i > E_{\text{level2}}$ 的帧, 将帧号设为 M_1 , 从该帧向后判断, 设当判断至 $i = M_2$ 时出现 $E_{i+1} > E_i$, 则帧编号为 M_1 、 M_2 之间的语音帧为一段语音段, $M_2 + 1$ 即为下一语音段的开始帧 M_3 , 以此可依次找出 $M_4, \dots, M_j$ 。为了更加精确地找到语音的起止点, 需要作进一步的判断。当 $E_{M_j} < E_{\text{level1}}$ 时, 若 j 为奇数, 则 $M_j = M_j - 1$; 若 j 为偶数, 则 $M_j = M_j + 1$ 。最后, 可得出只含有用户语音的序列 $f'(n) = [f_{M_1}, \dots, f_{M_2}, \dots, f_{M_3}, \dots, f_{M_j}]$ 。

对处理后的用户语音序列 $f'(n)$ 进行特征提取处理。特征提取就是对语音信号进行分析处理, 获得语音信号中表征人的基本特征的信息。目前, 常用的 2 种语音特征参数为美尔 (Mel) 频率倒谱系数 (Mel-frequency cepstrum coefficient, MFCC) 和线性预测系数 (linear prediction coefficient, LPC)。由于 MFCC 利用了人耳的听觉原理和倒谱解的相关特性, 对噪声的敏感程度较差, 并且计算复杂度较高^[6], 所以本文采用 LPC 作为用户的语音特征参数。本文通过使实际语音采样值和线性预测采样值之间的均方误差最小, 得到一组唯一的线性预测系数。将 $f'(n)$ 分成 K_1 帧 $f'_i, i = 1, 2, \dots, K_1$, 帧长为 L_1 , 设线性预测系数的阶数为 P , 利用 Levinson-Durbin 算法^[7]求解第 i 帧的线性预测系数 $a_{p,i} (p = 1, \dots, P)$ 。

设提取的近端用户的语音特征向量为 Φ_{user} , 则

$$\Phi_{\text{user}} = [\Phi_{\text{user},1}, \Phi_{\text{user},2}, \dots, \Phi_{\text{user},K}] \quad (4)$$

式中: Φ_{user} 为 $P \times K_1$ 阶矩阵; $\Phi_{\text{user},i}$ 为 $P \times 1$ 列向量

$$\Phi_{\text{user},i} = [a_{1,i}, a_{2,i}, \dots, a_{P,i}]^T \quad (5)$$

将线性预测系数输入模型中, 建立用户语音模板向量。设 M 阶用户语音的模板向量为

$$q(\Phi_{\text{user},i} | \lambda_{\text{user}}) = [a_1 p_1(\Phi_{\text{user},i}), \dots, a_M p_M(\Phi_{\text{user},i})]^T \quad (6)$$

式中: $q_j(\Phi_{\text{user},i}) = \frac{1}{(2\pi)^{P/2} |\Sigma_j|^{1/2}} \exp \left\{ -\frac{1}{2} (\Phi_{\text{user},i} - m_j)^T (\Sigma_j)^{-1} (\Phi_{\text{user},i} - m_j) \right\}$, P 是向量 $\Phi_{\text{user},i}$ 的维数;

$\lambda_{\text{user}} = \{a_j, m_j, \Sigma_j\}, j = 1, \dots, M$, λ_{user} 是近端用户高斯混合量的特征集合, m_j 、 Σ_j 分别代表对应的均值和方差向量; $\sum_{j=1}^M a_j = 1$, a_j 是第 j 阶高斯混合量的权重。

设模板型的初始协方差矩阵 Σ_{j0} 为单位矩阵, 初始权重 $a_{j0} = 1/M$, 均值向量 μ_{j0} 的数值为随机数, 则

λ_{user} 可由 EM 迭代算法^[8]求得, 即

$$m_j = \frac{1}{a_j P} \sum_{k=1}^{K_1} q(j | \Phi_{\text{user},k}) \Phi_{\text{user},k} \quad (7)$$

$$\Sigma_j = \frac{1}{a_j P^2} \sum_{k=1}^{K_1} q(j | \Phi_{\text{user},k}) \Phi_{\text{user},k} (\Phi_{\text{user},k})^T \quad (8)$$

$$a_j = \frac{1}{P} \sum_{k=1}^{K_1} q(j | \Phi_{\text{user},k}) \quad (9)$$

$$q(j | \Phi_{\text{user},i}) = \frac{a_j q_j(\Phi_{\text{user},i})}{\sum_{j=1}^M a_j q_j(\Phi_{\text{user},i})} \quad (10)$$

式中: $j = 1, 2, \dots, M; i = 1, 2, \dots, K_1$ 。

不断迭代直至迭代前后的参数集 ($|a_j - a'_j|^2 + |m_j - m'_j|^2 + |\Sigma_j - \Sigma'_j|^2$)^{1/2} $< \epsilon$ 停止迭代, 一般 $\epsilon = 10^{-5}$ 。

通过以上步骤可以得到用户语音的模板向量 $q(\Phi_{\text{user},i} | \lambda_{\text{user}})$, 而用户语音的每一帧对应一个 M 阶的模板向量, 则 K_1 帧信号得到的模板向量组成 $M \times K_1$ 阶矩阵。对此矩阵的行向量值取平均得到高斯混合阶数相同时的模板向量的平均值 $q(\Phi_{\text{user},a} | \lambda_{\text{user}})$, 该值即为需要存储的模板向量。

2.2 实时处理残留回波

实时处理残留回波的过程与 2.1 节中所述的处理方法相同: 设利用归一化最小均方 (normalized least mean square, NLMS) 自适应滤波器进行回波抑制后的输出误差信号为 $e(n)$, 将 $e(n)$ 通过与上一节相同的方法设置信号能量门限 T_1 、 T_2 排除噪声帧, 然后对处理后信号分帧, 并进行相同的特征提取, 设帧长为 L_2 , 第 i 帧得到特征向量 Φ_i , 将特征向量代入式 (7)~式 (9) 进行 EM 法迭代, 可得到测量向量 $q(\Phi_i | \lambda_{\text{user}})$ 。

2.3 判别非近端语音帧

本文利用残留回波的测量向量与存储的模板向量之间的距离 Λ_i 来判别非近端语音帧。定义 Λ_i 为

$$\Lambda_i = \frac{\|q(\Phi_i | \lambda_{\text{user}}) - q(\Phi_{\text{user},a} | \lambda_{\text{user}})\|}{\|q(\Phi_{\text{user},a} | \lambda_{\text{user}})\|} \quad (11)$$

式中: $\| \cdot \|$ 表示计算向量的 2 范数。

随机取第 K 帧远端信号 $x(n)$ 代入高斯混合模型中得到远端语音的向量 $q(\Phi_{\text{far},k} | \lambda_{\text{user}})$, 定义非近端语音帧判决的门限为

$$T = G \frac{\|q(\Phi_i | \lambda_{\text{user}}) - q(\Phi_{\text{far},k} | \lambda_{\text{user}})\|}{\|q(\Phi_{\text{far},k} | \lambda_{\text{user}})\|} \quad (12)$$

式中: G 为门限增益, 一般取 3~5。

如果 $\Lambda_i > T$, 就判断该帧为非近端语音帧, 反之则为近端语音帧。

判为非近端语音帧时,将其消除以减少残留回波。将经 RECG 算法模块处理后的信号 $e'(n)$ 传向远端用户,此时残留回波已经得到有效的抑制。

3 仿真实验

本文随机选取两男两女,利用计算机声卡录制语音,将 4 段语音组合成不同的实验组,在近端用户手机终端的上行链路提取回波残留信号,进行 4 次实验验证提出方法的有效性。实验用语音如图 2 所示,采样频率均为 8 kHz。

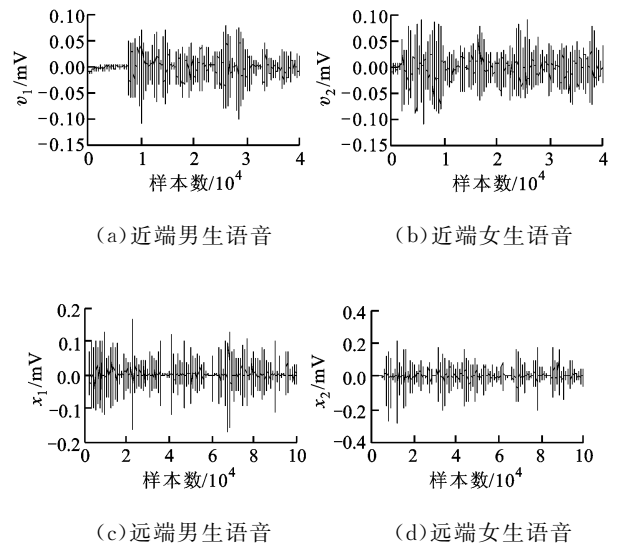


图 2 RECG 算法的近端、远端语音

回波信号 $y_1(n)$ 和 $y_2(n)$ 分别为远端语音 $x_1(n)$ 和 $x_2(n)$ 通过房间回波路径后的回波信号。实验房间大小为 5 m×3 m×3 m,近端用户手机在房间中心位置。表 1 为 4 次实验语音的设置。

实际中,利用 NLMS 进行回波抑制时,由于真实回波路径阶数与设定值不匹配等原因会导致自适应滤波器的失效,因此本文从手机回波抑制模块输出端提取失效的自适应滤波器输出信号 $e(n)$ 进行下一步处理。

表 1 RECG 算法实验的语音设置

实验组别	近端语音	回波信号
第 1 组	v_1	y_1
第 2 组	v_1	y_2
第 3 组	v_2	y_1
第 4 组	v_2	y_2

3.1 RECG 算法模块进行残留回波抑制

本文利用提出的方法对残留回波作进一步消

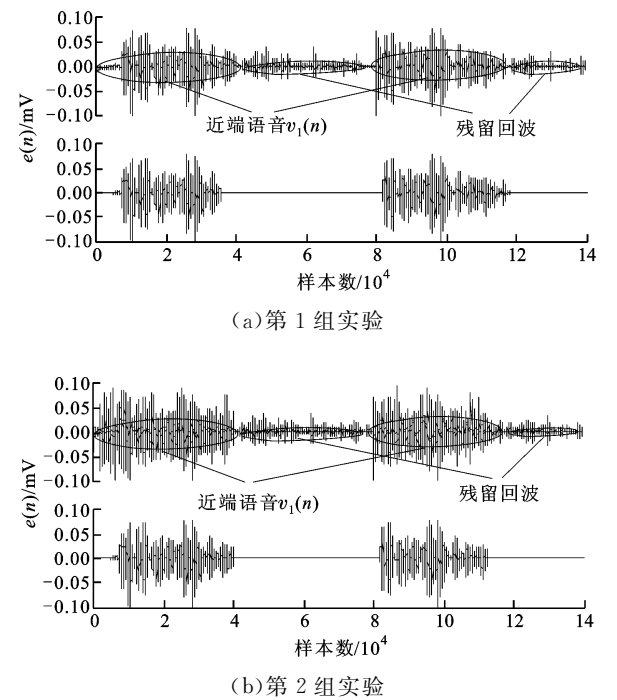
除。实验中近端用户的训练语音需要近端用户预先录制,其说话内容不同于测试语音。将训练语音依次进行端点检测、特征提取,从而建立高斯模型。用本文算法进行残留回波抑制时,每组实验中混合近端语音信号和残留回波信号 的参数设置如表 2 所示。

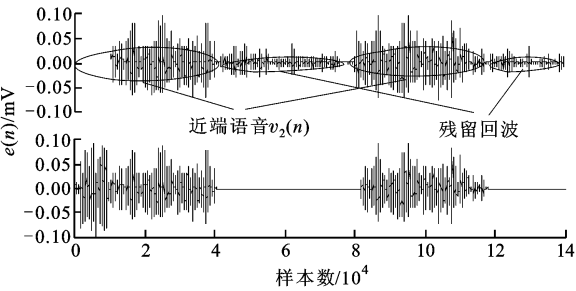
表 2 RECG 算法的参数设置

训练参数及取值	测试参数及取值	参数含义
$T_N, -30$	$T_1, -13$	语音能量门限
$T_L, -20$	$T_2, -10$	语音能量门限
$P, 19$	$P, 19$	LPC 线性预测阶数
$L, L_1, 64$	$L_2, 512$	语音信号帧长
$M, 10$	$M, 10$	高斯混合阶数
	$G, 3.4$	判决门限增益

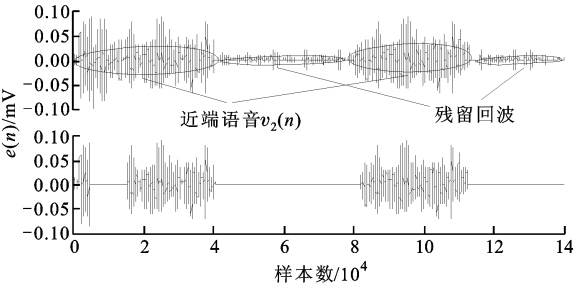
按照本文提出方法得到 4 组实验结果,输出信号 $e'(n)$ 如图 3 所示。

从图 3 可以看出,本文提出的算法可正确地鉴别近端语音信号和残留回波信号,最终残留回波信号被抑制。共进行了 4 组实验,每组实验判决 27 帧,即本实验共判决 108 帧,其中第 1 组实验和第 4 组实验中近端语音分别出现 1 帧误判。该方法对不同的信号具有一定的误判率,多次实验共统计 432 帧输出信号的误判帧数,算得帧误判率大约为 1%~2%。





(c)第 3 组实验



(d)第 4 组实验

图 3 4 组 RECG 算法处理前、后残留回波抑制效果对比

3.2 性能对比

为了通过对比来进一步说明本文方法对原自适应滤波器算法性能的提升,引入回声衰减增益 ER-LE(Echo Return Loss Enhancement)^[9]。当只有远端语音时,自适应滤波器消除回波的效果可以通过回声衰减增益衡量,定义衰减增益为

$$G = 10\lg(\sum d^2 / \sum e^2) \quad (13)$$

G 越大说明回波抑制的效果越好。在本文 3.1 节中,混合信号在 40 000~140 000 点处为远端产生的回波信号,此时间段为只有远端语音阶段,因此在此阶段可以观察本文提出算法与自适应滤波器法的回波消除量的比较。当远端信号分别为 $x_1(n)$ 和 $x_2(n)$ 时的性能比较如图 4、图 5 所示。可以看出,本文方

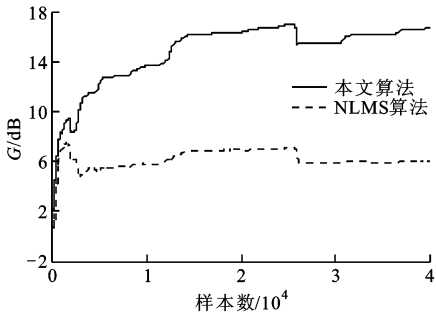


图 4 远端信号为 $x_1(n)$ 时 2 种算法的处理效果对比

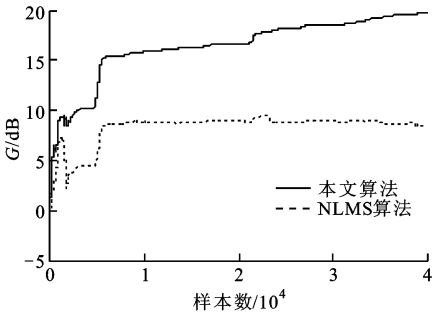


图 5 远端信号为 $x_2(n)$ 时 2 种算法的处理效果对比

法能在自适应滤波器阶数不匹配的情况下,将其回波消除量提高 7~10 dB。

4 结 论

本文提出了一种 RECG 算法模块,在自适应回波抑制处理后对其残留回波信号进行处理。其基本思想是以线性预测系数为特征,在高斯混合模型下对手机用户语音样本进行训练得到用户语音模板,通过对比误差信号和模板进行用户语音帧判决,将非用户语音帧置零以消除残留回波。本方法能在帧误判率为 1%~2% 的情况下,将回波消除量提高 7~10 dB。

参考文献:

[1] RAO H I K, FARHANG-BOROUJENY B. Fast LMS/Newton algorithms for stereophonic acoustic echo cancellation [J]. IEEE Transactions on Signal Processing, 2009, 57(8): 2919-2930.

[2] EHTIATI N, CHAMPAGNE B. Constrained adaptive echo cancellation for discrete multitone systems [J]. IEEE Transactions on Signal Processing, 2009, 57(1): 302-312.

[3] BEKRANI M, KHONG A W H, LOTFIZAD M. Neural network based adaptive echo cancellation for stereophonic teleconferencing application [C] // Proceedings of International Conference on Multimedia and Expo. Piscataway, NJ, USA: IEEE Press, 2010: 1172-1177.

[4] KIM K J, JANG C S, JEONG J M, et al. Acoustic echo cancellation using the DUET algorithm based blind separation in a noisy environment [C] // TENCON, IEEE Region10 Conference. Piscataway, NJ, USA: IEEE Press, 2006: 1-4.

[5] LEE K H, CHANG J H, NAM K S, et al. Frequency-domain double-talk detection based on the Gaussian mixture model [J]. Signal Processing Letters, IEEE,

2010, 17(5): 453-456.

- [6] 张学锋, 王芳, 夏萍. 融合 LPC 与 MFCC 的特征参数 [J]. 计算机工程, 2011, 37(4): 216-217.
ZHANG Xuefeng, WANG Fang, XIA Ping. Characteristic parameters based on LPC and MFCC [J]. Computer Engineering, 2011, 37(4): 216-217.
- [7] 曹华, 李伟, 谭艳梅. 线性预测及其 Matlab 实现 [J]. 现代电子技术, 2009(7): 133-135.
CAO Hua, LI Wei, TAN Yanmei. Linear prediction and its implementation by Matlab [J]. Modern Electronic Technique, 2009(7): 133-135.

- [8] REYNOLDS D A, ROSE R C. Robust text independent speaker identification using Gaussian mixture speaker models [J]. IEEE Transactions on Speech and Audio Processing, 1995, 3(1): 72-83.
- [9] 王永德, 王春霞. 一种改进的快速 LMS/Newton 算法及其在回声对消中的应用 [J]. 电子学报, 2000, 28(2): 97-100.
WANG Yongde, WANG Chunxia. An improved fast LMS/Newton algorithm and its application in echo cancellation [J]. Journal of Electronics, 2000, 28(2): 97-100.

(编辑 刘杨)

(上接第 5 页)

- [2] 徐学基, 诸定昌. 气体放电物理 [M]. 上海: 复旦大学出版社, 1996: 280-286.
- [3] HERLIN M A, BROWN S C. Electrical breakdown of a gas between coaxial cylinders at microwave frequencies [J]. Physical Review, 1948, 74(8): 910-913.
- [4] HERLIN M A, BROWN S C. Breakdown of a gas at microwave frequencies [J]. Physical Review, 1948, 74(3): 291-296.
- [5] ANDERSON D, JORDAN U, LISAK M, et al. Microwave breakdown in resonators and filters [J]. IEEE Transactions on Microwave Theory and Techniques, 1999, 47(12): 2547-2556.
- [6] TOMALA R, JORDAN U, ANDERSON D, et al. Microwave breakdown of the TE_{11} mode in a circular waveguide [J]. Journal of Physics: D Applied Physics, 2005, 38(14): 2378-2381.
- [7] RASCH J, ANDERSON D, LISAK M, et al. Microwave corona breakdown in a gas-filled rectangular resonator cavity [J]. Journal of Physics: D Applied Physics, 2009, 42(5): 1-6.
- [8] CHUNG T H, LIN M, HYUN J Y, et al. Two-dimensional fluid simulation of capacitively coupled RF electronegative plasmas [J]. Japanese Journal of Applied Physics, 1997, 36(5A): 2874-2882.
- [9] SOROLLA E, MATTES M. Corona discharge in microwave devices: a comparison of ionization rate model [J]. IEEE Microwave Review, 2010, 16(1): 41-46.
- [10] LISOVSKII V A. Criterion for microwave breakdown of gases [J]. Technical Physics, 1999, 44(11): 1282-1285.
- [11] JORDAN U, ANDERSON D, LAPIERRE L, et al. On the effective diffusion length for microwave breakdown [J]. IEEE Transactions on Plasma Science, 2006, 34(2): 421-430.
- [12] MACDONALD A D, GASKELL D U, GITTERMAN H N. Microwave breakdown in air, oxygen, and nitrogen [J]. Physical Review, 1963, 130(5): 1841-1850.
- [13] COBINE J D. Gaseous conductors [M]. New York, USA: Dover Publications Inc., 1958: 143-240.
- [14] 卡普佐夫 H A. 电子学: 上册 [M]. 吴全德, 译. 北京: 高等教育出版社, 1957: 130-171.
- [15] 杨津基. 气体放电 [M]. 北京: 科学出版社, 1983: 65-83.
- [16] MACDONALD A D. Microwave breakdown in gases [M]. New York, USA: John Wiley & Sons, Inc., 1966: 191-196.
- [17] BROWN S C, MACDONALD A D. Limits for the diffusion theory of high frequency gas discharge breakdown [J]. Physical Review, 1949, 76(11): 1629-1633.
- [18] 王欲知. 真空技术 [M]. 成都: 四川人民出版社, 1981: 38-41.

(编辑 刘杨)