

VoIP 中一种基于 WebRTC 的回声消除改进算法

姚力 刘强

(浙江警察学院大数据重点实验室 杭州 310053)

摘要 回声现象是语音通信系统中的常见问题,一直困扰和影响着通信质量。提出一种基于 WebRTC(Web Real-Time Communication)开源框架的回声消除算法。该算法针对定点运算框架的局限性,采取浮点算法实现,提高了算法的效率和精度,同时还满足实时性要求。在移动设备上的仿真结果表明,所提算法在复杂度相当的前提下比原有算法有更高的回声返回损耗增益值,具备更好的回声消除性能。

关键词 VoIP, WebRTC, 回声消除

中图分类号 TN912.3 文献标识码 A

VoIP Acoustic Echo Cancellation Algorithm Based on WebRTC

YAO Li LIU Qiang

(Big Data Key Laboratory, Zhejiang Police College, Hangzhou 310053, China)

Abstract Echo phenomenon is a common problem in voice communication system, which affects the communication quality. This paper presented an echo cancellation algorithm based on WebRTC (Web Real-Time Communication). As the fixed-point realization is limited, the algorithm uses floating-point computation to improve algorithm efficiency and accuracy, while maintains algorithm speed. The experimental results based on mobile equipments show that the proposed algorithm outperforms the original one on echo return loss enhancement with comparable complexity.

Keywords VoIP, WebRTC, Echo cancellation

1 引言

近年来,随着计算机网络的不断发展,VoIP(Voice over Internet Protocol)技术及相关应用也迅速发展。与传统电话相比,IP 电话广泛利用互联网和全球 IP 互连环境的资源,具有通话成本低廉、带宽利用率高、能提供各种多媒体增值服务等优点。然而,由于其以数据封包(Data Packet)的形式在 IP 网络(IP Network)上做实时传递,存在数据丢包、延时和抖动等问题,使其服务质量 QoS(Quality of Service)得不到保证^[1]。特别地,当网络延时超过 45ms 时,远端通话者的声音被麦克风拾取后通过网络传回到远端,远端通话者能听到自己的回声,从而严重影响了通话质量和用户体验度。因此,回声消除技术在 VoIP 中起着至关重要的作用,其中最为普遍的做法是设计声学回声消除器 AEC(Acoustic Echo Cancellation)来实现回声的抑制^[2-3]。

由于软件具有实时、廉价、可移植性好等优点,本文采用的声学回声消除算法基于 WebRTC 实现。WebRTC 即网页实时通信(Web Real-Time Communication),是 Google 公司于 2010 年收购挪威语音及视频通信软件开发商 GIPS(Global IP Solutions)公司而获得的一项基于网页浏览器进行实时语音对话或视频对话的技术,并于 2011 年开放了其源代码。WebRTC 提供了视频会议的核心技术,其本质是将实时通信应用所需要的音视频捕获及处理模块、网络传输及会话控制等协议集成到网页浏览器^[4]。图 1 显示了 WebRTC 的总体框架。浏览器厂商利用 WebRTC 开源项目提供的相关功能

的 C++ API(如音视频的采集、处理引擎、会话和网络传输协议等)能有效地降低项目开发的成本和风险^[5]。

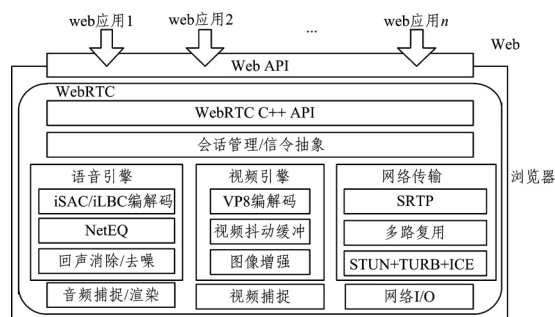


图 1 WebRTC 总体架构图

音频处理引擎是 WebRTC 中一个重要的、极具技术价值的组成模块,包含了 iSAC/iLBC 编解码器、语音 NetEQ 技术、回声消除/噪声抑制等三大部分。针对现今迅速发展的智能手机、平板电脑等移动终端,WebRTC 设计了专门的 AECM(AEC for Mobile)模块。该模块将语音信号变换到频域,采用回声抑制(Echo Suppression)的处理方法消除回声,自适应滤波器核采用的是 NLMS 算法^[6]。整个过程用定点运算实现,运算复杂度低,能得到较好的回声消除结果。

为了提高运算精度,在计算复杂度相当的前提下进一步改善回声消除性能,本文在 WebRTC 的 AECM 基础上提出了一种改进算法。该算法使用浮点运算实现,同时优化了滤波器控制逻辑。实验结果表明,该算法能在复杂度相当的情况下产生更好的回声消除效果。

姚力(1985—),男,博士,讲师,主要研究方向为信号处理、多媒体技术, E-mail: zjyaoli@163.com。

2 声学回声消除算法

2.1 定点与浮点运算的转换规则

WebRTC中,AECM通过Q域值表示浮点数值,而所有变量、整个接口函数、运算操作都是定点数值。Q域值表示了浮点数小数点的所在位置。例如,定点操作得到某频域和值A为14054,Q域值为10,该值用C语言中的16位二进制数可以表示为0b110110,11100110。我们将小数点标在第10位上,那么A即为0b1101,1011100110(约为10进制数13.724)。在使用浮点实现的代码中,我们得到该数值为13.757,这表明除了精度不同,定点和浮点的数据应该可以相互转换和验证。通常有以下公式:

$$A_{float} = A_{fix} / (1 \ll Q) \quad (1)$$

同时,由定点数实现的开方、求对数等数学函数会相对复杂,当使用浮点实现时,这些数学函数都能从相应的C语言库函数中直接得到。

2.2 算法流程

图2描述了整个AECM算法的流程。首先,算法以块为单位,每次采样128个时域波形上的点,进行128点的快速傅立叶变换(FFT),得到复数的浮点频域值。根据远端信号和近端信号频谱的相似程度(相关性大小),算法估计出整个系统的延时,并将远端信号延时后与前段信号对齐。回声消除的核心模块是自适应滤波器^[7],如图3所示,远端信号 $x(k)$ 通过环境中的回声信道 h 产生回声 $y(k)$,近端信号 $d(k)$ 由 $y(k)$ 混合近端语音 $v(k)$ 共同组成。通过构造 M 抽头自适应滤波器 \hat{h} 来模拟回声信道 h ,可以使得 $\hat{y}^{\wedge}(k)$ 逼近回声信号 $y(k)$,使 $e(k)$ 逼近近端语音 $v(k)$,从而达到消除回声的目的。

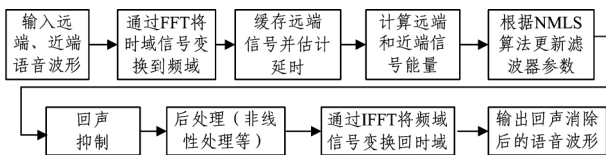


图2 AECM算法流程图

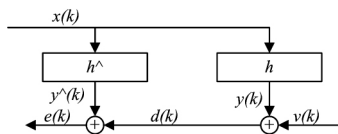


图3 自适应滤波器原理图

首先,AECM滤波器采用归一化最小均方差(NMLS)算法,得到信道滤波器系数;然后通过在频域上回声抑制的方法,将近端信号中的远端回声部分有效地去除;最后,通过非线性处理^[8]、添加舒适噪声^[9]、快速傅立叶反变换(IFTT)等模块输出回声消除后的声音信号。

2.3 延时估计

远端通话者的声音经过环境回声信道的反射后,通常产生了一定的延时^[10]。只有准确地估计延时,才能将近端信号与远端信号在时间轴上对齐。最常用的延时估计方法是计算两信号的互相关函数,通过遍历候选延时的互相关系数,选择取得最大值的候选时间为实际延时。但此方法在应用时计算复杂度高,尤其是应用在实时要求高的系统中。AECM算法简化了流程,在算法复杂度和性能上做了较好的权衡。

如表1所列,在快速傅立叶变换之后,我们得到了远端和近端语音信号的频谱far_spectrum和near_spectrum,其中远端频谱将被缓存起来作为候选匹配项。选择频谱中最重要的

32个频段(12-43),算法估计了频谱的均值threshold_spectrum并设其为门限值。当某个频段值大于门限值时,将该位设为1,反之则设为0。这样便得到了远端和近端信号的二值化的频谱。通过求解两者的按位异或值,选择相似度最高的候选远端信号并计算对应的延时。

表1 延时估计模块的工作流程

模块名:延时估计
输入:float far_spectrum[65],float near_spectrum[65]
处理过程:for (i=12;i<=43;i++) { threshold_spectrum[i]=EstimateMeanvalue(); //估计频谱的均值作为门限 if (spectrum[i] > threshold_spectrum[i]) SetBit(1); else SetBit(0);}
输出:int far_binary_spectrum[32],int near_binary_spectrum[32]; int similarity=far_binary_spectrum^near_binary_spectrum

2.4 滤波器设计

上文介绍了自适应滤波器的工作原理,但更新滤波器系数只能在远端单端说话时应用,在双端同时说话时,近端语音和远端声音的回声会混合在一起,从而使两者的线性关系大大下降,这时更新滤波器系数会适得其反,会减缓滤波器的收敛程度。可见,滤波器控制逻辑对滤波器收敛及整个回声消除的结果起到了非常关键的作用。

本文优化了AECM的滤波器控制逻辑,使用前置、后置两个滤波器共同管理^[11]。其中自适应滤波器始终在更新,而实际被使用的存储滤波器则是有条件地更新。只有认为当前自适应滤波器的系数值比存储滤波器中的值更好并满足一些条件时,才将其值赋给存储滤波器。图4显示了本文回声消除算法的流程图,设远端语音信号 $x(k)$ 经过LEM(Loudspeaker-Enclosure-Microphone)系统产生了回声 $y(k)$,该回声与背景噪声 $n(k)$ 以及可能存在的近端语音信号 $v(k)$ 组成混合信号 $d(k)$ 一起被麦克风拾取到。即 $d(k) = y(k) + n(k) + v(k)$ 。设前置滤波器为 $w_f(k) = [w_{f,0}(k), \dots, w_{f,N-1}(k)]^T$,其中, N 为滤波器阶数。通过该滤波器模拟得到的回声信号为 $y_f(k)$ 。我们衡量两者的差值,表示为麦克风端得到的误差信号为 $e_f(k)$:

$$e_f(k) = d(k) - y_f(k) = d(k) - w_f(k)^T x(k) \quad (2)$$

其中, $x(k) = [x(k), \dots, x(k-n+1)]^T$ 。同样地,得到后置滤波器的误差信号为:

$$e_b(k) = d(k) - y_b(k) = d(k) - w_b(k)^T x(k) \quad (3)$$

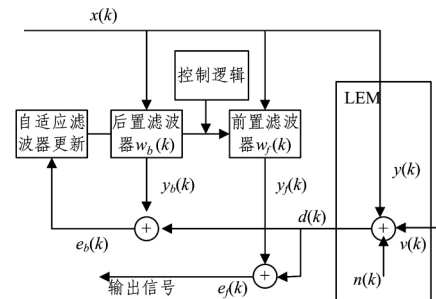


图4 AECM算法流程图

后置滤波器始终根据NMLS更新参数,其规则如下:

$$w_b(k+1) = w_b(k)^T + \frac{\mu e_b(k) x(k)}{\|x(k)\|^2 + \epsilon} \quad (4)$$

其中, $\|x(k)\|^2 = x(k)^T x(k)$, μ 为更新步长系数, ϵ 为归一化的正常数避免除数为0。由于之前已经计算了信号的频谱,因此不难得到它的能量:

$$P_x(k) = \frac{1}{M} \sum_{i=0}^{M-1} x^2(k-i) \quad (5)$$

其中, M 为常数。在 AECM 中, μ 的更新策略为:

$$\mu = 2^{-\left(\mu_{\max} - \mu_{\min}\right) \cdot \left(1 - \frac{P_{\text{far}} - p_{\min}}{P_{\text{max}} - p_{\min}}\right)} \quad (6)$$

其中, P_{far} 为远端信号能量, 剩下的为预设的参数。

滤波器的控制逻辑是 AEC 算法的核心, 它决定了何时将后置滤波器系数赋给前置滤波器。表 2 描述了控制器的工作流程, 式(7)表示当后置滤波器的收敛程度(MSE)明显差于前置滤波器时, 应重置后置滤波器。而式(8)表示满足该条件时, 考虑后置自适应滤波器得到较好的收敛值, 将其赋予前置存储滤波器。

表 2 滤波器控制逻辑

模块名: 滤波器控制器	
输入:	float far_spectrum[65], float near_spectrum[65], float Wadapt[65], float WStored[65];
处理过程:	对于每一个滤波器采样时刻 n , for ($i=0; i \leq 64; i++$) { Energynear(n) += near_spectrum[i]; Energyadapt(n) += Wadapt[i] * far_spectrum[i]; Energystored(n) += WStored[i] * far_spectrum[i]; $MSE_adapt = \sum_{n=1}^N Energy_{adapt}(n) - Energy_{near}(n) $, N 为滤波器采样周期 $MSE_stored = \sum_{n=1}^N Energy_{stored}(n) - Energy_{near}(n) $ 设 MSE_adapt_old 和 MSE_stored_old 为上一次采样保存的值。 if ($MSE_adapt/MSE_store > A$) && ($MSE_adapt_old/MSE_stored_old > B$) // ResetAdaptFilter(); // 将前置滤波器值赋予后置滤波器 else if ($MSE_adapt/MSE_store < C$) && ($MSE_adapt < T$) && ($MSE_adapt_old < T$) // StoreAdaptFilter(); // 将后置滤波器值赋予前置滤波器 输出: float Wadapt[65], float WStored[65];

2.5 回声抑制算法

AECM 使用频域上回声抑制的方法消除回声^[12]。考虑到通过自适应滤波器模拟产生的预估回声信号和前端语音信号可能存在能量增益上的差异, 算法首先计算系统的能量增益 G 。估计的信道回声信号能量, 近端信号能量分别为 P_{estimate} 和 P_{near} 。应有 $P_{\text{estimate}} \cdot G \approx P_{\text{near}}$ 。设 $\alpha = P_{\text{estimate}} \cdot G / P_{\text{near}}$, 且 α 截断到 $0 \sim 1$ 。设近端语音信号的频谱为 F_{near} , 则回声抑制后的信号频谱 F_{out} 为:

$$F_{\text{out}} = F_{\text{near}} \cdot (1 - \alpha) \quad (9)$$

式(9)表明当 α 越接近 0 时, 预估回声信号越低, 远端接近静音, 此时应将近端信号尽可能传递。当 α 越接近 1 时, 预估回声信号接近近端信号, 通过 AECM 去除回声, 达到回声消除的目的。

3 实验系统及测量结果

为测试本文算法的性能, 在移动设备上搭建了通话系统。用于实验的语音信号采样率为 16kHz, 时长约 35s。图 5 显示了测试信号包括远端语音和近端播放的背景语音。该测试系统包含了远端静音、近端静音、双端通话和远端背景噪声等过程, 从而能对系统性能进行更全面的评判。

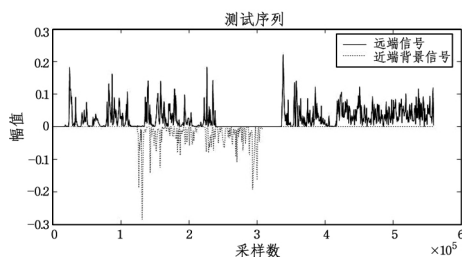


图 5 测试语音信号

我们在不同的移动设备和不同的工作模式下对比测试了本文算法和 WebRTC 算法。图 6 显示了在设备上采集到的经过 AECM 处理后的近端语音信号。可以看到, 本文提出的算法在 1~7s, 21~26s 的远端单端通话期间有更好的回声抑制结果; 而从实际主观听感上, 亦有如此结论。在 8~15s 的双端通话阶段, 其主客观结果难以比较。

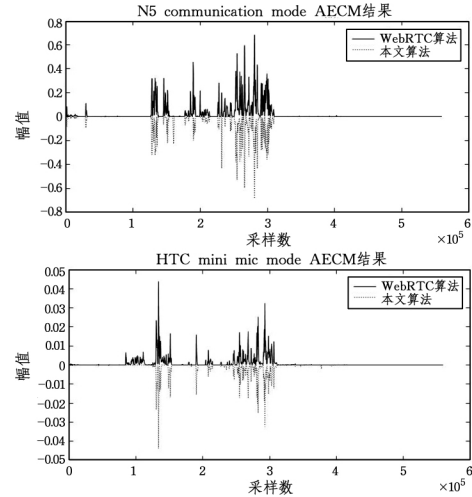


图 6 AECM 处理后的语音信号对比结果

为进一步比较算法的客观性, 我们绘制了回声返回损耗增益值 ERLE (Echo Return Loss Enhancement) 曲线^[13]。该曲线反映了原回声信号与剩余回声残差的比值, 即回声抵消量。ERLE 值越大表明回声消除性能越好。

图 7 示出了本文算法和 WebRTC 算法在不同设备上的 ERLE 曲线图。通过比较发现, 本文算法拥有更好的回声抑制效果。

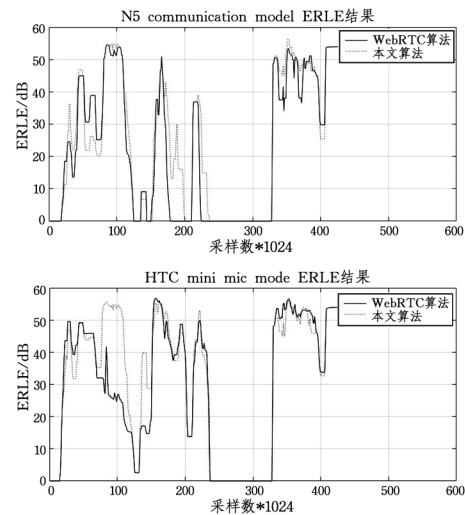


图 7 ERLE 对比结果

ERLE 的定义如下:

$$ERLE(n) = 10 \lg \left(\frac{\sum_{i=0}^{M-1} x^2(n-i)}{\sum_{i=0}^{M-1} e^2(n-i)} \right) \quad (10)$$

其中, M 取值 1024。

结束语 声学回声消除在 VoIP 通信系统中起着非常重要的作用。本文在 WebRTC 的 AECM 算法的基础上, 提出其浮点实现算法, 优化了滤波器控制逻辑, 旨在简化定点实现的程序可读性, 同时提高回声消除性能。实验结果表明, 用浮

(下转第 318 页)

由于 Predictive Tree 为每个对象建立与维护一棵树,对象与对象之间没有任何关联性,因此 Predictive Tree 的空间复杂度随着参数的增大而增加,甚至达到原始基础网络大小 ($\tau=30$)。AP-I 充分利用对象之间的关联性,不会存储重复的节点与边,具有更好的扩展性。同时,AP-I 能够更好地利用同向来减少空间复杂度。

最后通过实验验证 ρ 值和 t 值 (Pruning 操作运行间隔) 对索引大小以及更新操作时间的影响,如图 8 所示。

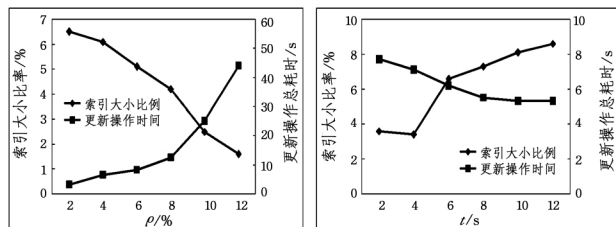


图 8 ρ 值和 t 值对索引大小和更新操作的影响

从实验结果中可以看出 ρ 值较大时 ($\rho \geq 0.08$), 虽然 AP-I 的大小可以更好地被控制, 但是对更新操作有较大的影响, 因此需要根据实际要求选择 ρ 值。本文的所有实验中设置 $\rho=0.04$ 。 t 值增加时, 运行 Pruning 操作的间隔加长, 导致 AP-I 中存在 AP 值过小的节点与边, 因此增加索引大小。然而, 增大的 t 值倾向于减少更新操作总耗时, 因为 Pruning 操作本身耗时减少。本文设定 $t=3\text{min}$ 。

结束语 针对已有的 Predictive Tree 的局限性, 本文提

出了一种新颖的 AP-I (Adaptive Predication-Index) 索引, 通过有效挖掘移动对象之间的路径关联性以及 Pruning 操作, 避免大量的索引更新和重建操作, 提升索引效率。对不同数据进行实验, 结果表明, 在同等查询效率的基础上, AP-I 能够实现更好的准确度, 以及更高的索引空间效率与更新效率。

AP-I 主要利用自适应概率模型 (AP) 提升预测的准确性, 而如何提高 AP 的收敛速度 (即较快的收敛于最佳路径) 将是后续主要研究的内容; 同时, 如何结合以往历史数据从而提升准确度也是值得考虑的研究内容。

参考文献

- [1] ZHANG R, JAGADISH H V, DAI B T, et al. Optimized Algorithms from Predicative Range and KNN Queries on Moving Objects [J]. Information Systems, 2010, 35(8): 911-932.
- [2] AYHAN S, PESCE J, COMITZ P, et al. Predictive Analytics with Aviation Big Data [C]// Integrated Communications, Navigation and Surveillance Conference, 2013: 1-13.
- [3] JEUNG H, YIU M L, ZHOU X F, et al. Path Prediction and Predictive Range Querying in Road Network Databases [J]. The VLDB Journal, 2010, 19(4): 585-602.
- [4] HENDAWI A M, BAO J, MOKBEL M F, et al. Predictive Tree: An Efficient Index for Predictive Queries on Road Networks [C]// Proceedings of the International Conference on Data Engineering, 2015: 1215-1226.

(上接第 311 页)

点实现的 AECM 算法在性能上更加出色, 但保持了相当的算法复杂度, 适用于实时的 VoIP 回声消除系统。

参考文献

- [1] 李挥, 林茫茫, 胡海军, 等. VoIP 回声消除器设计及算法研究 [J]. 电子学报, 2007, 35(9): 1774-1778.
- [2] 王庆辉, 李永哲. VoIP 声学回声消除算法研究 [J]. 现代电子技术, 2009, 32(7): 157-159.
- [3] FUKUI K, SHIMAUCHI S, KOBAYASHI K, et al. Acoustic echo canceller software for VoIP hands-free application on smartphone and tablet devices [C]// 2014 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, 2014: 133-134.
- [4] TAHERI S. WebRTCbench: a benchmark for performance assessment of webRTC implementations [C]// 2015 13th IEEE Symposium on Embedded Systems For Real-time Multimedia (ESTIMedia), Amsterdam, 2015: 1-7.
- [5] 胡敏, 刘六程, 刘鹏. 基于 WebRTC 的视频会议系统的设计与实现 [J]. 电视技术, 2013, 37(1): 141-143, 150.
- [6] CIOCHINA S, PALEOLOGU C, BENESTY J, et al. An optimized NLMS algorithm for acoustic echo cancellation [C]// 2015 International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, 2015: 1-4.
- [7] BUCHNER H, KELLERMANN W. Improved Kalman gain com-

putation for multichannel frequency-domain adaptive filtering and application to acoustic echo cancellation [C]// 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, 2002: 1909-1912.

- [8] PADHI T, KAR A, CHANDRA M. Family of adaptive algorithms based on second order Volterra filters for nonlinear acoustic echo cancellation: A technical survey [C]// 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), New Delhi, 2014: 909-914.
- [9] SCHULD T, LINDSTROM F, CLAESSEON I. A Combined Implementation of Echo Suppression, Noise Reduction and Comfort Noise in a Speaker Phone Application [C]// 2007 Digest of Technical Papers International Conference on Consumer Electronics, Las Vegas, NV, 2007: 1-2.
- [10] LU Y H, FOWLER R, TIAN W, et al. Enhancing echo cancellation via estimation of delay [J]. IEEE Transactions on Signal Processing, 2005, 53(11): 4159-4168.
- [11] 汪萌, 戴礼荣, 宋彦, 等. 一种使用双滤波器的回声消除算法 [J]. 数据采集与处理, 2006, 21(1): 74-79.
- [12] NAKAYAMA K, OFUJI Y, HIRANO A. An adaptive control method for regression coefficient in frequency domain non-linear echo suppressor [C]// 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, 2012: 624-628.
- [13] 开跃春, 金涛, 贾宏志. IP 会议电话回声消除器的设计及仿真 [J]. 计算机仿真, 2010, 27(1): 322-325.