

一种快速实时语音个人特征改变方法

毛启容 詹永照 杜守富

(江苏大学计算机科学与通信工程学院 镇江 212013)

摘要: 为保护实时语音通信中的个人特征, 该文提出了一种新的实时语音个人特征改变方法, 该方法采用 PLAR(Pseudo Log Area Ratio)系数曲线变换方法和基于线性预测的基音同步叠加(LP-PSOLA)算法分别对语音信号的谱参数和韵律参数进行修改, 从而实现语音信号个人特征的改变; 此外, 针对目前时长规整大多采用的同步叠加(SOLA)算法计算量大、不适合实时语音处理的缺点, 采用课题组提出的一种新的基于同步叠加方法的时长规整算法——自适应同步叠加(ASOLA)算法, 对个人特征改变后的语音信号进行时间上的弥补, 保证语音处理的实时性。最后, 利用该方法实现了实时语音的隐私保护。实验结果表明, 该方法合成的语音质量高、实时性好。

关键词: 语音改变; 隐私保护; 音高规整; 时长规整; 实时

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2007)02-0434-05

A Fast Modification Method for Personal Characteristics of Real-Time Speech

Mao Qi-rong Zhan Yong-zhao Du Shou-fu

(Department of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: To protect the privacy during the real-time speech communication, in this paper, a novel fast modification method for real-time speech is presented, in which the transformation method of Pseudo Log Area Ratio (PLAR) curve and Pitch-Synchronous-OverLap-and-Add (LP-PSOLA) algorithm based on Linear Prediction are respectively adopted to modify the spectrum parameters and the prosodic parameters of the speech, then the personal characteristics of the speech are modified. In addition, for Synchronous-OverLap-and-Add(SOLA) method popularly used for Time-Scale Modification (TSM) has a lot of computation load, it can not be used for processing the real-time speech. Therefore, in this paper, a novel TSM algorithm——Adaptive Synchronous OverLap and Add(ASOLA) presented by author's workgroup is adopted for TSM of the speech, the personal characteristics of which have been modified, to assure the real-time property of the speech. Finally, the fast modification method for real-time speech is used to protect the privacy during the real-time speech communication, and the test shows that, the speech synthesized by this method has high quality, and this approach can assure the real-time speech communication.

Key words: Speech modification; Privacy protection; Pitch-scale; Time-scale; Real-time

1 引言

随着网络的不断发展, 隐私保护问题逐渐成为人们关注的焦点之一, 而实时语音通信中的隐私保护作为隐私保护问题的一个重要方面, 正受到越来越多专家学者的关注。为保护实时语音通信中的个人隐私, 必须改变说话人的个人特征, 即对语音信号的特征参数(语谱参数和韵律参数)进行修改。由于改变后的语音在时间长度上与原来语音不同, 不能在实时通信中使用, 为保证实时环境中语音的连贯性, 必须对个人特征改变后的语音进行时间长度的弥补, 即语音的时长规整。

目前, 以声道面积函数的演化参数PLAR(Pseudo Log Area Ratio)作为被改变的语谱参数, 实时性好, 算法简单^[1]; 基于线性预测的基音同步叠加算法(LP-PSOLA)是广泛使用的语音处理的音高规整算法, 它有效地估计了语音产生的声道模型的系统函数^[2]; 而时长规整大多采用同步叠加(Synchronous OverLap and Add, SOLA)方法^[2,3], 但该方法计算量大, 搜索时间长, 不适合实时处理。Yim等人提出的GLS-TSM算法^[4], Wong等人提出的包络匹配算法^[5], 采用了诸如过零率匹配等方法。这两种算法的优点是, 减少了互相关系数计算带来的计算量, 提高了算法的执行速度。但是, 由于他们采用的是波形相似的判断方法, 其准确度不如采用互相关系数的判断方法, 因此, 合成的语音质量不高, 有时还会出现杂音。本文提出一种新的快速实时语音改变方法,

2005-07-26 收到, 2006-02-15 改回

国家自然科学基金(60273040), 江苏省青蓝工程(1191170004)和江苏大学青年科学基金(1241170005)资助课题

首先对语音的个人特征改变，包括 PLAR 特征参数的改变和明显影响听觉的音高进行改变，然后采用本课题组提出的快速的时长规整算法^[6]对个人特征改变后的语音信号进行时间长度的弥补，从而保证高质量和快速的语音个人特征的改变。

2 语音改变方法流程

为减少计算量和误差，对采集到的语音信号进行静音检测，把语音信号分解成有声语音帧，在进行线性预测分析前，利用过零率和短时能量对语音信号进行初步的清浊音判决，将得到的浊音帧进行线性预测分析，得到浊音激励信号和线性预测系数。然后利用 PLAR 曲线变换方法对浊音激励信号进行谱参数的修改，对谱参数修改后的语音信号进行第二次清浊音判决，对分解后周期性明显的脉冲激励信号，进行基音标注和音高规整。从而从两个方面改变语音信号的个人特征。然后，基于线性预测模型，对个人特征改变后的语音信号重建新的语音，得到时间长度变化了的语音信号，再利用时长规整对其进行时间长度调整，使其保持与原来语音具有相同的时间长度。实时语音改变方法总体流程如图 1 所示。

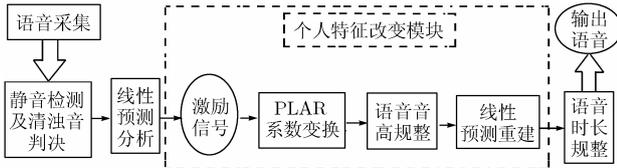


图 1 语音改变流程图

3 个人特征改变方法

3.1 PLAR 特征参数的改变方法

PLAR 系数定义为

$$h_0 = 0, \quad h_i = h_{i-1} + \lg \left(\frac{1 - k_i}{1 + k_i} \right) \quad (1)$$

其中 $i = 1, 2, \dots, p$ ， p 为预测阶数。 k_i 为第 k 阶反射系数，即部分相关系数(PARCOR)，可以从线性预测递推算法得到^[4]。

声道面积比对数函数为

$$g_i = \lg \left(\frac{A_{i+1}}{A_i} \right) = \lg \left(\frac{1 - k_i}{1 + k_i} \right) \quad (2)$$

其中 A_i 是声道的多节无损声管模型中第 i 节圆管的截面积，声道面积比函数具有变化平稳的性质。由式(2)得到：

$$k_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}} \quad (3)$$

由式(1)，式(3)可以递推得到

$$h_i = \lg \frac{A_{i+1}}{A_1} \quad (4)$$

从式(4)可以看出，PLAR 系数实际表示无损声管的截面积。图 2 给出了从成人男子发汉语的“人”的语音帧中提取的 PLAR 系数曲线图。其中， h_1 是声道前端， h_p 是声道后

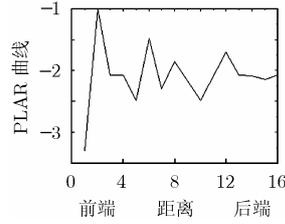


图 2 PLAR 曲线图

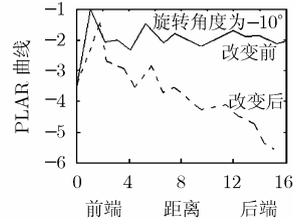


图 3 PLAR 曲线变换图

端。Corveleyn 等人在文献[1]中指出，只要保证 PLAR 曲线的波峰和波谷的包络不变，语音的内容就保持不变，因此，将 PLAR 曲线旋转一定的角度，就可以达到改变语音特征的目的。如图 3 所示(语音与图 2 相同)，虚线表示改变后的 PLAR 曲线。

设 $S(x)$ 表示曲线在 $x = 1, 2, \dots, p$ 时的 PLAR 系数曲线插值函数，可采用坐标旋转公式为

$$S'(x') = x \sin \theta + S(x) \cos \theta \quad (5)$$

$$x' = x \cos \theta - S(x) \sin \theta \quad (6)$$

其中 $S'(x)$ 为 $S(x)$ 旋转后的新的 PLAR 曲线， x' 为 x 旋转以后的横坐标值， θ 为旋转角度，当 $\theta > 0$ 为逆时针旋转曲线，当 $\theta < 0$ 是顺时针旋转曲线。

为了保证旋转后得到的 PLAR 系数的稳定性，要保持旋转的角度在一定的范围并满足如下公式： $x - 1 < x' < x$ ；($\theta > 0$) 或者

$$x < x' < x + 1, \quad \theta < 0 \quad (7)$$

得到的新的曲线在原来坐标上长度会发生改变，也就是说声道的截面个数发生改变，即线性预测系数(LPC)分析的阶数发生变化。设新的阶数为 p' ，则 $p' = [T^{-1}(p)]$ ， T 为旋转变换，可重新计算在 $\hat{x} = 1, 2, \dots, p'$ 时的 PLAR 系数，从而得到改变后的 LPC 系数。

将式(6)代入式(7)得到

$$x < x \cos \theta - S(x) \sin \theta < x + 1, \quad \theta < 0 \quad (8)$$

解此不等式得到

$$0 < r \sin \frac{\theta}{2} \cos \left(\frac{\theta}{2} - \phi \right) < \frac{1}{2}, \quad \theta < 0 \quad (9)$$

其中 $r = \sqrt{x^2 + S^2(x)}$ ， $\phi = \arctg \frac{S(x)}{x}$ ； $x = 1, 2, 3, \dots, p$ 。

旋转以后得到新的 PLAR 系数， $s'(\hat{x})$ 在 $\hat{x} = 1, 2, \dots, p'$ 处的值可以插值得到，其公式为

$$s'(\hat{x}) = \frac{x' - (x - 1)}{x - (x - 1)} s'(x - 1) + \frac{x - x'}{x - (x - 1)} s'(x) \quad (10)$$

图 3 即为旋转变换的示意图，经验证 $\theta = 5^\circ \sim 10^\circ$ 都满足式(9)的条件。

3.2 音高规整

对 PLAR 曲线变换后的语音信号，进行基音同步叠加方法的音高规整，其方法是首先对浊音激励信号进行基音标注，然后对标注过的信号，进行音高规整^[7,8]。

(1) 基音标注 基音标注的实质是求取语音信号的基音周期，而求取基音周期最常用的方法是利用自相关函数^[9]。设 $s(n)$ 是一段加窗语音信号，它的非零区间为 $n = 0 \sim$

$(N-1)$ 。 $s_w(n)$ 的自相关函数称为语音信号 $s(n)$ 的短时自相关函数, 用 $R_w(l)$ 表示(其中 l 为语音信号的长度), 它的计算公式为

$$R_w(l) = \sum_{-\infty}^{\infty} s_w(n)s_w(n+l) = \sum_{n=0}^{N-l-1} s_w(n)s_w(N+l) \quad (11)$$

如果 $s(n)$ 是一个浊音性质的周期信号, 那么它的短时自相关函数也呈现出明显的周期性, 且 $R_w(l)$ 的周期即等于 $s(n)$ 的周期。相反, 清音信号近似于随机噪声, 它的短时自相关函数不具有周期性且随着 l 的增大而迅速减小。因此, 可利用这个特点来求出一个浊音信号的基音周期。在决定基音周期时, 利用了短时自相关函数在基音周期的各个整数倍点上具有较高的峰值这一特性。

每帧语音信号的基音周期估计值的标注方法为: 在每帧(帧长为 30ms)浊音激励信号中, 选取一段时间范围(比如 2.5~15ms)计算出 $R_w(l)$ 。首先, 要找到 $R_w(l)$ 的最大值点 l_{\max} , 它显然是基音周期的第一候选者, 假设 $R_w(l)$ 在这一点最大值为 $R_{\max} = R_w(l_{\max})$; 其次, 寻找 $R_w(l)$ 的若干局部最大点作为基音周期的其他候选者, 这些局部最大点必须满足以下 4 个条件: 第一, 其取值应在 $R_{\max} - R_{\text{TH}} \sim R_{\max}$ 的范围内(R_{TH} 是一个恰当选取的阈值); 第二, 各个局部最大点之间以及它们与全局最大点之间的间隔不得小于 l_{TH} (l_{TH} 是一适当选取的间隔值, $l_{\text{TH}} \geq N_l$, N_l 应根据试验确定)。如图 4 所示, l_1 是满足上述条件的一个局部最大点; 第三, 对各个局部最大点和全局最大点进行“清晰度”检查, 找出这些最大点中的清晰点(所谓清晰点是指作为谷值或峰值的局部最大点); 第四, 在所有清晰点最大点中找出最左边的点 l_c , 该点即为该帧语音的基音周期估计值。

清晰度检查即寻找所有的清晰点, 其方法如下: 若某局部最大点的前 $0.8 \cdot N_l$ 个局部最大点的函数值均大于该局部最大点的函数值, 且该局部最大点的后 $0.8 \cdot N_l$ 个局部最大点的函数值均小于该局部最大点的函数值, 则该局部最大点不是清晰点, 反之则是。找出满足要求的所有清晰点, 则其中函数值最小的清晰点即为本帧的基音周期。

同时, 在连续浊音帧边界处的周期, 或者对有检测到异常变化的周期数值, 应根据最近几帧的平均周期分别予以估计或舍去。

(2) 残差信号的基音同步叠加 基音同步叠加算法是建立分析语音的基音位置和重建语音的基音位置的映射, 设分析语音改变前的基音位置之间的间隔为 t_a , 重建语音的基音位置之间的间隔为 t_s , 则 $t_s = \beta t_a$ ($\beta < 1$ 音调升高; $\beta > 1$ 音调降低)。采用 PSOLA 法进行音高的规整时, 若要降低基

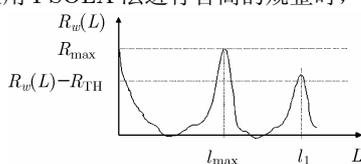


图 4 短时自相关法求基音周期示例

音频率, 则意味着 t_s 要比 t_a 大, 重建语音的周期变长; 若要升高基音频率, 则意味着 t_s 要比 t_a 小, 重建语音的周期变短; 改变激励信号的基音周期, 采用插值法和抽取法, 其公式如下:

$$y(i) = x(i) + (i \cdot \beta - i) \cdot [x[\text{INT}(i \cdot \beta) + 1] - x[\text{INT}(i \cdot \beta)]] \\ 1 \leq i \leq Pt_s \quad (12)$$

其中 $y(i)$ 为采样幅度值, $x(i)$ 为时间采样点, i 为采样点序号, Pt_s 为 t_s 间隔内的样点数。

4 自适应同步叠加时长规整算法

传统的时长规整常采用 SOLA 方法, 该方法采用归一化的互相关函数最大的准则^[3], 寻找同合成语音段中的叠加区域的信号波形最相似的波形进行叠加, 但该方法搜索时间长, 计算量大, 不适合实时环境。因此, 本文对 SOLA 进行了改进, 提出了自适应同步叠加(ASOLA)算法。

4.1 自适应最佳叠加点搜索算法思想

由于叠加方法的主要计算量来自于归一化的互相关函数值的计算, 特别是在高采样率的乐音信号的处理过程中, 为了减少计算量, 首先将互相关函数值的计算公式简化为

$$R_m[k] = \frac{\sum_{j=0}^{j=L_o-1} y(mL_s + k + j) \cdot x(mL_a + j)}{\sum_{j=0}^{j=L_o-1} |y(mL_s + k + j)| \sum_{j=0}^{j=L_o-1} |x(mL_a + j)|} \quad (13)$$

其中 L_o 为叠加长度 ($L_o = \lambda_1 T_p$, λ_1 为常数且 $1 \leq \lambda_1 \leq 2$, T_p 为基音周期), L_a 为分析时间间隔, L_s 为合成时间间隔, k 是叠加点搜索下标, $k \in [0, k_{\max}]$, k_{\max} 为最大搜索长度, $k_{\max} = N/2$ (N 为语音截成长度 $N = \lambda_2 T_p$, λ_2 为常数且 $4 \leq \lambda_2 \leq 8$), 压扩系数 $\alpha = L_s / L_a$ ($\alpha > 1$ 延长说话时间, $\alpha < 1$ 缩短说话时间)。最佳叠加点的求取, 本质上是求取语音信号的基音周期, 设 $x(n)$ 是周期为 T 的信号, 即 $x[n] = \sum_{r=-\infty}^{r=\infty} y(n + rT)$, r 为整数, $y(n)$ 只在一个周期内有值。语音信号是短时平稳的, 在浊音段具有准周期性, 利用此性质可以间隔基音周期的整数倍来计算互相关系数, 因此, 如果要从分析语音信号中截取一段(周期的整数倍长度)信号, 并从该语音信号剩下的信号段中找出与该段信号最相似的信号段, 只需依次向前滑动周期的整数倍并对每个间隔的互相关系数进行比较即可, 无需逐点分析。最佳叠加点的搜索方法如下: 首先对截取后剩下的语音信号段中每隔基音周期间隔的点计算互相关系数, 找出互相关系数最大的点作为初始最佳叠加点; 然后将上一遍搜索间隔的一半作为本次搜索的搜索间隔, 在初始最佳叠加点附近的较小区域内(该区域长度为上一遍搜索长度的一半)用同样的方法进行进一步的搜索, 如此下去, 直到搜索长度为 1, 则该位置为最终的最佳叠加点位置。如图 5 所示, 第一遍搜索以每隔基音周期间隔计算

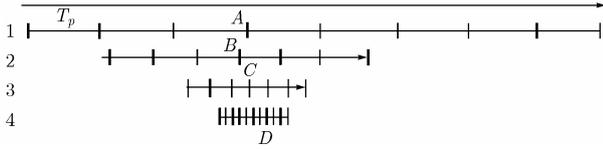


图 5 自适应搜索算法示意图

出每点的互相关系数，找出初始最佳叠加点，假设为 A 点，然后以 A 点为中心，在其两边区域以上一次搜索间隔的一半作为搜索间隔再次搜索，找到 B 点，如此下去直到找到 D 点为最佳叠加点。

在进行语音叠加时，因为语音在短时期内基音周期变化具有连续性，所以每次叠加的第一遍搜索的搜索间隔用上次叠加的第一遍搜索的搜索间隔来更新，自适应地找到最佳的叠加点。对于同一次叠加内的各遍搜索，后一遍搜索的搜索间隔为前一遍的一半，搜索间隔的计算参见文献[6]。

设信号的采样率为 R ，对于第 m 次叠加时，各参数的含义为： K_{\max} 表示最大搜索长度， $K_{\max} = N/2$ ； $K_m(n)$ 表示第 m 次叠加时，第 n 遍搜索的最大搜索长度，则 $K_m(1) = K_{\max}$ ； F_m 表示第 m 次叠加时语音信号的基音频率； P_m 表示第 m 次叠加时语音信号的基音周期的样点数； $I_m(n)$ 表示第 m 次叠加时，第 n 遍搜索时需要的搜索间隔，则 $I_m(1) = P_m$ ； $T_m(n)$ 表示第 m 次叠加时，第 n 遍搜索时需要的搜索次数。

首先计算初始搜索间隔，其计算方法参见文献[6]，自适应搜索算法的具体步骤可表示如下：

- (1) $m \leftarrow 1$ ，置首遍搜索间隔 $I_1(1) = P_1$ ；
- (2) 对 $0 \leq i' < \lceil \log_2 P_m + 1 \rceil$ ， $0 \leq j' \leq T_m(i')$ ，计算互相关函数 $R_m[(j' - T_m(i')/2) \cdot I_m(i')]$ ；
- (3) 根据互相关函数最大的准则寻找最佳叠加点，设找到的最佳叠加点位置为 k_m ，然后在该点上进行第 m 次叠加，其叠加公式参见文献[2]；
- (4) 准备进行第 $m+1$ 次叠加：设 $c_{m+1} \leftarrow k_m / I_m(1)$ ，
 - (a) 若 $c_{m+1} \in (0,1)$ ，则 $I_{m+1}(1) \leftarrow I_m(1)$ ， $K_{m+1}(1) \leftarrow K_{\max}$ ；
 - (b) 若 $c_{m+1} \in (1,2)$ ，则 $I_{m+1}(1) \leftarrow k_m$ ， $K_{m+1}(1) \leftarrow K_{\max}$ ；
 - (c) 否则，按初始搜索间隔进行搜索，跳到第(6)步；
- (5) $K_{m+1}(1) \leftarrow K_{\max}$ ， $T_m(1) \leftarrow K_{m+1}(1) / I_{m+1}(1)$ ；对

$1 \leq n \leq \lceil \log_2 I_{m+1}(1) + 1 \rceil$ ，做

$$I_{m+1}(n+1) \leftarrow \left\lfloor \frac{1}{2} I_{m+1}(n) \right\rfloor$$

$$K_{m+1}(n+1) \leftarrow \lceil K_{m+1}(n) / 2 \rceil$$

$$T_{m+1}(n+1) \leftarrow \lceil [K_{m+1}(n+1) / I_{m+1}(n+1)] \rceil$$

(6) $m \leftarrow m+1$ ，重复步骤(2)，(3)，(4)，(5)直到所有语音处理结束。

4.2 算法比较

SOLA 算法一次叠加须进行加法 $3L_o \cdot K_{\max}$ 次、乘法 $(L_o + 1) \cdot K_{\max}$ 次、比较 K_{\max} 次，因为叠加长度 $L_o = \lambda_1 T_p = (\lambda_1 / \lambda_2) N$ ，若取最大搜索样点个数为 $K_{\max} = N/2$ ，则算法的时间复杂度为 $O(N^2)$ 。而 ASOLA 算法一次叠加须执行

加法 $((K_{\max} / P_m) \cdot 3L_o \cdot \log_2 P_m)$ 次、乘法 $(L_o + 1) \cdot (K_{\max} / P_m) \cdot \log_2 P_m$ 次、比较 $(\log_2 P_m (K_{\max} / P_m))$ 次，取 $P_1 = N / \lambda_2$ ，而 $P_m > P_1 (m > 1)$ ，则算法的时间复杂度在 $O(N \log_2 N)$ 以内。故 ASOLA 的理论时间开销小于 SOLA。

5 实时语音改变效果分析

原语音经过音高规整后，得到的是音高和时间长度均改变的语音，然后运用 ASOLA 进行时长规整，此时，ASOLA 规整因子 α 和音高规整因子 β 之间的关系为 $\alpha\beta = 1$ ，即当音高升高时，合成的语音必须进行时间延伸规整；而音高降低时，合成的语音必须进行时间压缩规整。由于 ASOLA 的计算速度快，因此能实时地改变语音，并且保持了与原语音相同的时间长度。

另外，对女声改变时，音高规整因子 β ，在 $\beta < 1$ 的空间比较大，对男声改变时， $\beta > 1$ 的空间比较大。如果选用半音阶作为规整度量，设升高或降低 γ 个半音阶，则其与 β 的关系为 $\beta = 2^{\gamma/12}$ ； $-12 \leq \gamma \leq 12$ ；经过测试：男声语音在 $2 < \gamma < 10$ 时，能保证合成高质量地向女声转变；而女声语音在 $-8 < \gamma < -2$ 时，能保证合成高质量地向男声转变。

对一段男声进行了个人特征的改变，并采用 ASOLA 算法进行时长规整，图 6 给出了该段男声改变前后的波形比较图，经听觉测试表明：当 $\beta = 0.75$ 时，该段男声明显变成了女声，其波形如图 6(b)所示，当 $\beta = 1.25$ 时，该段男声明显变的低沉，如图 6(c)所示，并且语音改变能保持实时性。

邀请了 10 个人对两段男声和一段女声进行了该语音改变方法执行前后的听觉测试和清晰度比较，将采样率固定为 44.1kHz，测试该方法在不同语音长度 N 执行时的清晰度，结果见表 1，表明 N 越小清晰度越高。

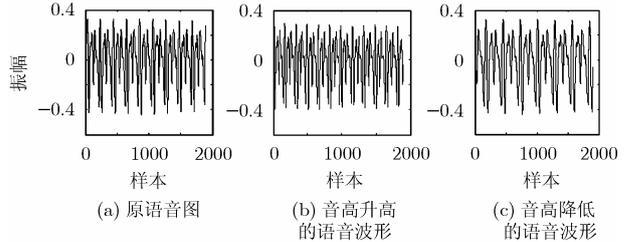


图 6 原语音与输出语音比较

表 1 不同区间长度对清晰度的影响

N	$(2 \sim 4) T_p$	$(4 \sim 8) T_p$	$(> 8) T_p$
清晰度	非常高	较高	高

6 结束语

本文提出了一种新的快速实时语音改变方法，该方法通过采用 PLAR 曲线变换方法和 PL-PSOLA 算法实现对语音信号个人特征的修改，并采用课题组提出的新的时长规整算法——ASOLA，对个人特征改变后的语音信号进行时间长度的调整。经过听觉测试和实时性实验，该方法能明显改变语音的个人特征，并能保证改变后的语音与原语音具有相同的节奏，合成语音质量高，满足实时性要求。同时，ASOLA

算法由于其速度快, 合成语音质量高, 还可以运用到实时音频压缩, 语音合成, 电子邮件语音应答, 乐音编辑等相关应用中。

参 考 文 献

- [1] Corveleyn S, Coose B, and Verhelst W. Voice modification and conversion using PLAR-parameters[C]. In Proc of IEEE Benelux Workshop on Model-Based Processing and Coding of Audio(MPCA-2002), Leuven, Belgium, 2002(11): 216–218.
- [2] Moulines E and Laroche J. Non-parametric technique for pitch-scale and time-scale modification of speech[J]. *Speech Communication*, 1995, 16(2): 175–205.
- [3] Laeol B and Fagam A D. A novel high quality efficient algorithm for time-scale modification of speech. 6th European Conference on Speech Communication and Technology[C]. Budapest, Hungary. 1999: 231–235.
- [4] Yim S and Pawate B I. Computationally efficient algorithm for time scale modification(GLS-TSM)[C]. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings. NJ, USA: IEEE, Piscataway, 1996, vol.2: 1009–1012.
- [5] Wong W and Au O C. Fast SOLA-based time scale modification using modified envelope matching. Proc. Of IEEE Int. Conf. on Acoustics, Speech and Signal Processing[C], Orlando, FL, Institute of Electrical and Electronics Engineers Inc. 2002, vol.3: 3188–3191.
- [6] 杜守富, 毛启容, 詹永照. 自适应同步叠加语音时长规整算法. *通信学报*, 2005, 2(26): 136–140.
- Du Shou-fu, Mao Qi-rong, and Zhan Yong-zhao. Adaptive synchronous overlap and add algorithm for time scale modification of speech. *Journal on Communications*, 2005, 2(26): 136–140.
- [7] Acero A. Source-filter models for time-scale pitch-scale modification of speech[C]. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings. NJ, USA: IEEE, Piscataway. 1998, vol.2: 881–884.
- [8] Jing Yinglong and Murphy P. Voice-source analysis for pitch modification of speech signals[C]. Proceedings of the COSTG-6 Conference on Digital Audio Effects(DAX-01), Limerick, Ireland, 2001: 125–128.
- [9] Laroche J. Autocorrelation method for high quality/pitch scaling. IEEE ASSP Workshop on Application of Signal Processing to Audio and Acoustics, Laroche, 1993: 200–204.
- 毛启容: 女, 1975 年生, 博士生, 讲师, 研究方向为语音处理, CSCW.
- 詹永照: 男, 1962 年生, 博士生导师, 教授, 主要研究方向为语音处理、CSCW、计算机图形学.
- 杜守富: 男, 1978 年生, 硕士, 研究方向为语音处理.