

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
G10L 21/02 (2006.01)  
A61F 2/72 (2006.01)



# [12] 发明专利说明书

专利号 ZL 200710017563.1

[45] 授权公告日 2009 年 12 月 30 日

[11] 授权公告号 CN 100576320C

[22] 申请日 2007.3.27

[21] 申请号 200710017563.1

[73] 专利权人 西安交通大学

地址 710049 陕西省西安市咸宁路 28 号

[72] 发明人 万明习 赵 钦 王素品 王卫波  
刘汉军

[56] 参考文献

US4669643A 1987.6.2

CN1776809A 2006.5.24

CN1070817A 1993.4.14

审查员 吴 娟

[74] 专利代理机构 西安通大专利代理有限责任公  
司

代理人 李郑建

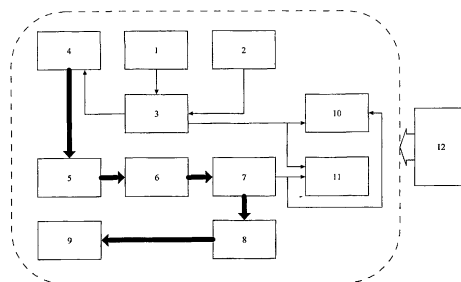
权利要求书 7 页 说明书 18 页 附图 4 页

[54] 发明名称

一种自动电子喉的电子喉语音增强系统与控制方法

[57] 摘要

本发明涉及一种自动电子喉的电子喉语音增强系统，基于双 DSP 的硬件系统平台和基于计算机的应用软件平台两种实现方式，系统包括应变传感采集模块、图像传感采集模块、电子喉振动模块、语音数据采集模块、A/D 转换模块、控制模块、数据处理模块、D/A 转换模块、滤波整形、功率放大和语音输出模块、电源模块、扩展程序存储模块、扩展数据存储模块；通过应变传感和图像运动检测传感的控制方式，对电子喉工作状态自动控制，从而使电子喉使用者无须手持操作电子喉，并对电子喉语音进行去噪与增强，改善发声质量，提高电子喉语音重建的便捷程度和语音质量。使用的信号处理算法经过软件仿真检验，有效的提高了电子喉语音的客观和主观评价水平。



1. 一种自动电子喉的电子喉语音增强系统，其特征在于，该系统基于双 DSP 的硬件系统平台和基于计算机的应用软件平台两种实现方式，包括：应变传感采集模块、图像传感采集模块、电子喉振动模块、语音数据采集模块、A/D 转换模块、控制模块、数据处理模块、D/A 转换模块、滤波整形、功率放大和语音输出模块、电源模块、扩展程序存储模块、扩展数据存储模块；

应变传感采集模块、图像传感采集模块、电子喉振动模块与控制模块连接，电子喉振动模块与语音数据采集模块相连，语音数据采集模块通过 A/D 转换模块与数据处理模块连接；控制模块和数据处理模块分别与扩展程序存储模块、扩展数据存储模块连接，数据处理模块还通过 D/A 转换模块与滤波整形、功率放大和语音输出模块连接，电源模块负责整个系统的电源供应；

其中，在双 DSP 硬件系统平台中，控制模块、数据处理模块由 DSP 芯片实现，A/D 转换模块、D/A 转换模块、扩展程序存储模块、扩展数据存储模块由专门芯片实现；

在计算机应用软件平台中，控制模块、数据处理模块由计算机 CPU 和相应的软件实现，A/D 转换模块、D/A 转换模块由计算机声卡实现，扩展程序存储模块、扩展数据存储模块由计算机内存、硬盘实现；

应变传感方式、图像传感方式为自动电子喉的两种自动控制方式，当外部环境不适合采用前两种自动工作模式时，使用者采取手动控制实现电子喉的语音增强；

所述应变传感方式为：应变传感采集模块以 2Hz 的频率不断循环采集面部两侧应变信号，并将应变信号转换为 8 位数字量传送到控制模块，得出面部两侧应变的平均值  $s_M$ ，并与设定判断电子喉开关的阈值  $S_0$  比较，当两侧采集到的应变信号都高于阈值  $S_0$ ，向电子喉振动模块发送控制信号以启动电

子喉，同时向数据处理模块发送信号，开始进行语音数据采集和增强处理；电子喉开始工作后，面部应变信号依然以原频率进行采集，当连续两帧的两侧应变信号都低于阈值 $S_0$ 时，认为使用者停止发声，关闭电子喉同时向数据处理模块发送信号，停止语音数据采集和处理；

所述图像传感方式为：图像传感采集模块以 2Hz 的频率循环拍摄  $80 \times 60$  的 8 位灰度口部图像，将每帧图像传送到控制模块，划分为 12 个  $20 \times 20$  的子区域；系统开机后自动通过摄像头记录当前静止状态下的口部 8 位灰度图像作为参考；在每一个子区域与参考图像的相应子区域进行进行 SAD 计算，并将结果与设定的阈值进行比较；如果一个图像子区域的 SAD 运算结果超过了阈值，则认为该子区域发生了显著运动，当达到半数的子区域发生显著运动时，则可判定使用者开始发声，由控制模块向电子喉振动模块发送控制信号以启动电子喉，同时向数据处理模块发送信号，开始进行语音数据采集和增强处理；在电子喉工作过程中，仍然以原有频率进行口部图像采样，在每采集一帧图像时，保留当前帧和上一帧的图像信息，如果当前帧与上一帧比较，发生显著运动的子区域不到半数，则再将当前帧与参考图像进行比较，如果发生显著运动的子区域仍然不到半数，则认为使用者停止发声，关闭电子喉，同时向数据处理模块发送信号，停止语音数据采集和处理；

数据处理模块在接受控制模块发来的信号之后执行中断程序，开始控制语音数据采集模块对语音进行 8kHz, 16bit 的采样，每取得 256 个采样点，即经 A/D 转换模块送入数据处理模块进行 FFT 运算，记录所得频谱的相角，再对实部和虚部取平方和，得到此 256 点语音的功率谱，之后对该 256 点语音进行 LPC 运算，得到其 16 阶的 LPC 系数，根据该系数求出当前语音帧感知加权滤波器的频响，再进一步求出当前语音帧的谱减系数，通过谱减系数与之前储存的噪声功率谱得出纯净语音功率谱估值，再将纯净语音估计功率谱开方后再与前面记录下的带噪语音相角相乘，并作 IFFT 变换，得到纯净语音估值，最后的结果经过滤波整形之后，通过 D/A 输出至标准音频接

口。

2. 权利要求 1 所述的自动电子喉的电子喉语音增强系统的控制方法，其特征在于，包括如下步骤：

1) 系统启动后，采集当前静息状态下的应变值、口部图像以及电子喉辐射噪声；

2) 根据实际环境和使用者需要选择工作模式，所述的工作模式为应变传感方式、图像传感方式和手动控制三种，当外部环境不适合采用前两种自动工作模式时，则选择手动控制工作模式；

3) 如果采用自动工作模式，则循环采集控制量信号，即应变或图像变化信号，采样频率为 2Hz；

4) 将采集到的控制量信号与设定阈值比较，超过阈值则启动电子喉；

5) 电子喉开始工作后，通过麦克风分帧采集使用者发出的电子喉语音，并进行数字化处理；

6) 通过基于听觉掩蔽效应的电子喉语音增强算法，结合之前采集的辐射噪声，对每帧电子喉语音进行增强；

7) 将增强后的电子喉语音经标准音频接口输出，并通过压缩算法储存；

8) 当采集到的控制量信号低于阈值，则关闭电子喉，完成一个工作周期。

3. 如权利要求 2 所述的方法，其特征在于，所述的应变传感方式的电子喉自动控制实现步骤如下：

尽量保持面部平静松弛，将两个应变传感器分别置于面部两侧，开机后传感器将自动读取当前的应变值，并通过 8 位模数转换送入 DSP 芯片，求出面部两侧应变的平均值；

设面部两侧应变经传感器和模数转换送入 DSP 的值分别为  $s_L$  和  $s_R$ ，则求出两侧应变平均值  $s_M = (s_L + s_R)/2$ ；

由于模数转换为 8 位精度，故最大值为 255，判断电子喉开关的阈值  $s_0$  设定为：

$$S_0 = s_M + (255 - s_M) \times 0.1$$

根据上式求出阈值  $s_0$  并储存起来，并以 2Hz 的频率不断循环采集面部两侧应变信号；由于人发声时，面部两侧肌肉动作应该是一致的，因此当两侧采集到的应变信号都高于阈值后，判定为使用者开始发声，则启动电子喉；

电子喉开始工作后，面部应变信号依然以原频率进行采集，以判断电子喉停止工作的时刻；由于发声过程中的语音间隔可能也导致面部低应变状态，因此为了避免这种误差影响，当采集到低于阈值的应变值时，并不停止电子喉工作，而是当连续两帧的两侧应变信号都低于阈值  $S_0$  时，才认为使用者停止发声，关闭电子喉。

4. 如权利要求 2 所述的方法，其特征在于，所述的图像传感方式的电子喉自动控制实现步骤如下：

尽量保持面部、口部松弛平静，系统开机后自动通过摄像头记录当前静止状态下的口部 8 位灰度图像作为参考，图像分辨率设定为  $80 \times 60$ ；之后以 2Hz 的频率对口部图像进行采样，并将采集到的图像划分为 12 个  $20 \times 20$  的子区域，分别编号 1~12；在每一个子区域应用下式与原参考图像的相应子区域进行比较：

$$\varepsilon_k = \sum_{i=1}^{20} \sum_{j=1}^{20} |X_{i,j} - Y_{i,j}|$$

其中， $k$  为子区域编号， $1 \leq k \leq 12$ ， $(i, j)$  为子区域中任意一点的坐标， $X_{i,j}$  为参考图像中相应点的灰度值， $Y_{i,j}$  则为当前图像中相应点的灰度值；

最后求出第  $k$  个子区域中的灰度绝对值差分  $\varepsilon_k$ ；

在口部图像运动检测中，将超过 3% 的平均灰度变化视为显著运动，则

设定的阈值  $\varepsilon_0$  为:

$$\varepsilon_0 = 255 \times 0.03 \times 20 \times 20 = 3060$$

将每个  $\varepsilon_k$  ( $1 \leq k \leq 12$ ) 与阈值  $\varepsilon_0$  进行比较, 当  $\varepsilon_k \geq \varepsilon_0$  时, 则可认为第  $k$  个子区域图像发生显著运动; 设发生显著运动的子区域共计  $n$  个, 则当  $n \geq 6$  时, 即发生显著运动的子区域达到或超过半数, 则认为整个口部发生显著运动, 即可判定为使用者开始发声, 电子喉开始工作;

启动电子喉之后, 仍然以 2Hz 的频率进行图像采样, 将采集到的每一帧图像按照同样的求绝对值差分和算法, 与上一帧图像进行比较:

如果当前图像与上一帧图像相比, 显著运动的子区域达到或超过半数, 则认为运动没有停止, 电子喉保持工作状态;

如果当前图像与上一帧图像相比, 显著运动的子区域不到半数, 则采用相同的求绝对值差分和算法, 将当前图像再与最初储存的静止状态参考图像比较;

如果当前图像与参考图像相比, 显著运动的子区域达到或超过半数, 则认为使用者没有恢复口部静止状态, 电子喉保持工作状态;

如果当前图像与参考图像相比, 显著运动的子区域不到半数, 则认为使用者已恢复口部静止状态, 此时应关闭电子喉。

5. 如权利要求 2 所述的方法, 其特征在于, 所述的听觉掩蔽效应的电子喉语音增强算法及步骤如下:

设  $y(t) = s(t) + n(t)$ , 其中  $y(t)$  为带噪语音,  $s(t)$  为纯净语音,  $n(t) = n_1(t) + n_2(t)$ , 其中  $n_1(t)$  为周期性辐射噪声,  $n_2(t)$  为随机环境噪声, 这是建立在语音与噪声不相关, 因而具有加性的假设前提下的;

则求出  $y(t)$  的频谱为:

$$Y(\omega) = Y_R(\omega) + iY_I(\omega) = FFT[y(t)]$$

其中 $Y(\omega)$ 为 $y(t)$ 的频谱， $Y_R(\omega)$ 和 $Y_I(\omega)$ 分别为 $Y(\omega)$ 的实部和虚部；

相应的求出周期性背景噪声和随机噪声频谱估值为：

$$N(\omega) = N_R(\omega) + iN_I(\omega) = FFT[n(t)]$$

其中， $n_1(t)$ 保持在整个电子喉发声阶段基本稳定不变，而 $n_2(t)$ 则可能随时随环境发生变化；因此 $N(\omega)$ 也可能是随时间变化的，为了抵消这种变化带来的影响，实现谱减系数的动态调整，对带噪语音计算其感知加权滤波器的频响：

$$T(z) = \frac{A\left(\frac{z}{\sigma_1}\right)}{A\left(\frac{z}{\sigma_2}\right)} = \frac{1 - \sum_{k=1}^p a_k \sigma_1^k z^{-k}}{1 - \sum_{k=1}^p a_k \sigma_2^k z^{-k}}$$

其中 $T(z)$ 即为感知加权滤波器的 $z$ 域频响， $a_k$ 为当前语音帧的 $k$ 阶LPC系数， $p$ 为LPC系数的最高阶数， $\sigma_1$ 、 $\sigma_2$ 为控制共振峰能量偏移的系数， $0 \leq \sigma_2 \leq \sigma_1 \leq 1$ ；

由感知加权滤波器频响求出谱减系数：

$$\alpha = \alpha_{\max} \left( \frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}} \right) + \alpha_{\min} \left( \frac{T(\omega) - T(\omega)_{\max}}{T(\omega)_{\max} - T(\omega)_{\min}} \right)$$

$$\beta = \beta_{\max} \left( \frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}} \right) + \beta_{\min} \left( \frac{T(\omega) - T(\omega)_{\max}}{T(\omega)_{\max} - T(\omega)_{\min}} \right)$$

其中， $\alpha$ 、 $\beta$ 即为所求的谱减系数， $T(\omega)$ 为当前语音帧的感知加权滤波器频响， $T(\omega)_{\max}$ 、 $T(\omega)_{\min}$ 分别为当前语音帧中 $T(\omega)$ 的最大值和最小值， $\alpha_{\max}$ 和 $\beta_{\max}$ 分别为 $\alpha$ 和 $\beta$ 的最大值， $\alpha_{\min}$ 和 $\beta_{\min}$ 则分别为 $\alpha$ 和 $\beta$ 的最小值， $\alpha_{\max}$ 、 $\beta_{\max}$ 、 $\alpha_{\min}$ 和 $\beta_{\min}$ 根据语音中噪声的类型和程度确定；

根据上式求出的谱减系数，代入下式：

$$|\hat{S}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha |N(\omega)|^2 & , \text{当} \left[ \frac{|N(\omega)|}{|Y(\omega)|} \right]^2 < \frac{1}{\alpha + \beta} \text{时} \\ \beta |N(\omega)|^2 & , \text{其他} \end{cases}$$

求出纯净语音频谱估值为：

$$\hat{S}(\omega) = |\hat{S}(\omega)|^2 \times \left( \frac{Y_R(\omega)}{\sqrt{Y_R^2(\omega) + Y_I^2(\omega)}} + \frac{iY_I(\omega)}{\sqrt{Y_R^2(\omega) + Y_I^2(\omega)}} \right)$$

则纯净语音估值为：

$$\hat{s}(t) = IFFT[\hat{S}(\omega)]$$

在系统中，上述算法的实现步骤如下：

将电子喉置于正常使用位置并保持固定，系统启动后电子喉自动振动约 0.5 秒，以采集辐射噪声和当前环境噪声，将这段噪声做 FFT 变换，得到其功率谱，作为噪声谱估值记录下来；电子喉开始工作后，以 8kHz 的频率对语音进行采样，每 256 个记录点作为一帧；对每帧语音进行 FFT 变换，得到其频谱，并进一步得到带噪语音的功率谱；之后利用带噪语音的功率谱，通过线性预测分析得到当前语音帧的感知加权滤波器频率响应；再根据感知加权滤波器频响  $T(z)$  求出当前帧中的谱减系数  $\alpha$ 、 $\beta$ ，即可通过谱减运算得到纯净语音功率谱估值；经过相位还原和 IFFT 变换，则得到一帧纯净语音的估值；

在系统实现过程中，算法中的一些参数取值如下：

$$\sigma_1 = 1, \quad \sigma_2 = 0.8;$$

$$\alpha_{\min} = 1, \quad \alpha_{\max} = 6;$$

$$\beta_{\min} = 0, \quad \beta_{\max} = 0.02, \quad \text{当带噪语音信噪比低于 } -5\text{dB 时, } \beta_{\max} = 0.002。$$



## 一种自动电子喉的电子喉语音增强系统与控制方法

### 技术领域

本发明属于病变语音重建及语音增强技术领域，特别涉及一种自动电子喉的电子喉语音增强系统与控制方法。

### 背景技术

电子喉是喉切除等喉部病变患者最常使用的辅助发声工具之一。其使用方法易于掌握，但使用时需用手动控制电子喉的位置与开关状态，给使用者造成了很大不便。此外，电子喉语音中所包含的各种辐射噪声、背景噪声也会严重影响电子喉语音的质量。

目前国际上对电子喉位置与开关控制等问题进行了一定程度的研究，但采用的方法主要集中在肌电信号提取与控制方面，这就造成了系统应用中，必须通过相应的传感器，提取并处理肌电信号，从而实现自动控制电子喉的目的。由于肌电信号的提取放大与处理所需的设备要求较高，因此这种方式会造成整个系统的成本提高。

### 发明内容

基于上述现有技术存在的局限或不足，本发明的目的在于，提供一种自动电子喉的电子喉语音增强系统及其控制方法，该系统基于双 DSP 的硬件系统平台和基于计算机的应用软件平台两种实现方式，实现对电子喉的自动控制，从而使电子喉无须手持工作，使用更为便捷。同时，对产生的电子喉语音进行增强去噪等处理，提高电子喉语音的质量。

本发明针对电子喉使用患者，在使用电子喉的过程中，实现自动控制及相应的语音增强功能，通过多种工作模式的选择，使电子喉的使用简便程度和发出的语音质量得到有效的提高。系统产生的电子喉语音还可实时记录，

以做进一步处理之用。

为实现上述任务，本发明采用如下的技术方案：

一种自动电子喉的电子喉语音增强系统，其特征在于，该系统基于双 DSP 的硬件系统平台和基于计算机的应用软件平台两种实现方式，包括：应变传感采集模块、图像传感采集模块、电子喉振动模块、语音数据采集模块、A/D 转换模块、控制模块、数据处理模块、D/A 转换模块、滤波整形、功率放大和语音输出模块、电源模块、扩展程序存储模块、扩展数据存储模块；

应变传感采集模块、图像传感采集模块、电子喉振动模块与控制模块连接，电子喉振动模块与语音数据采集模块相连，语音数据采集模块通过 A/D 转换模块与数据处理模块连接；控制模块和数据处理模块分别与扩展程序存储模块、扩展数据存储模块连接，数据处理模块还通过 D/A 转换模块与滤波整形、功率放大和语音输出模块连接，电源模块负责整个系统的电源供应；

其中，在双 DSP 硬件系统平台中，控制模块、数据处理模块由 DSP 芯片实现，A/D 转换模块、D/A 转换模块、扩展程序存储模块、扩展数据存储模块由专门芯片实现；

在计算机应用软件平台中，控制模块、数据处理模块由计算机 CPU 和相应的软件实现，A/D 转换模块、D/A 转换模块由计算机声卡实现，扩展程序存储模块、扩展数据存储模块由计算机内存、硬盘实现。

上述自动电子喉的电子喉语音增强系统的控制方法，其特征在于，包括如下步骤：

1) 系统启动后，采集当前静息状态下的应变值、口部图像以及电子喉辐射噪声；

2) 根据实际环境和使用者需要选择工作模式，所述的工作模式为应变传感方式、图像传感方式和手动控制三种，当外部环境不适合采用前两种自动工作模式时，则选择手动控制工作模式；

- 3) 如果采用自动工作模式, 则循环采集控制量信号, 即应变或图像变化信号, 采样频率为 2Hz;
- 4) 将采集到的控制量信号与设定阈值比较, 超过阈值则启动电子喉;
- 5) 电子喉开始工作后, 通过麦克风分帧采集使用者发出的电子喉语音, 并进行数字化处理;
- 6) 通过基于听觉掩蔽效应的电子喉语音增强算法, 结合之前采集的辐射噪声, 对每帧电子喉语音进行增强;
- 7) 将增强后的电子喉语音经标准音频接口输出, 并通过压缩算法储存;
- 8) 当采集到的控制量信号低于阈值, 则关闭电子喉, 完成一个工作周期。

本发明通过应变传感器与摄像头实现对控制量的采集, 利用基于 SAD 算法的差动量判别技术, 对电子喉进行自动控制。患者在使用中可以根据自己的实际需要, 采用应变传感方式、图像运动检测传感方式与手动控制三种工作模式。其中, 应变传感和图像运动检测传感是两种自动工作模式, 当外部环境不适于自动工作模式时, 可选择使用传统的手动控制模式。此外, 产生的电子喉语音通过麦克风进行采集、记录, 并利用基于听觉掩蔽效应的增强算法来消除语音中带来的周期性背景噪声和随机噪声, 提高电子喉语音质量。

基于应变传感方式的电子喉自动控制实现步骤如下:

尽量保持面部平静松弛, 将两个应变传感器分别置于面部两侧, 开机后传感器将自动读取当前的应变值, 并通过 8 位模数转换送入 DSP 芯片, 求出面部两侧应变的平均值。设面部两侧应变经传感器和模数转换送入 DSP 的值分别为  $s_L$  和  $s_R$ , 则求出两侧应变平均值  $s_M = (s_L + s_R)/2$ 。

由于模数转换为 8 位精度, 故最大值为 255。判断电子喉开关的阈值设定为:

$$S_0 = s_M + (255 - s_M) \times 0.1$$

根据上式求出阈值 $s_0$ ，并储存起来，并以 2Hz 的频率不断循环采集面部两侧应变信号。由于人发声时，面部两侧肌肉动作应该是一致的，因此当两侧采集到的应变信号都高于阈值后，可判定为使用者开始发声，则启动电子喉。电子喉开始工作后，面部应变信号依然以原频率进行采集，以判断电子喉停止工作的时刻。由于发声过程中的语音间隔可能也导致面部低应变状态，因此为了避免这种误差影响，当采集到低于阈值的应变值时，并不停止电子喉工作，而是当连续两帧的两侧应变信号都低于阈值 $s_0$ 时，才认为使用者停止发声，关闭电子喉。

基于图像传感方式的电子喉自动控制实现步骤如下：

尽量保持面部、口部松弛平静，系统开机后自动通过摄像头记录当前静止状态下的口部 8 位灰度图像作为参考，图像分辨率设定为  $80 \times 60$ 。之后以 2Hz 的频率对口部图像进行采样，并将采集到的图像划分为 12 个  $20 \times 20$  的子区域，分别编号 1~12。在每一个子区域应用下式与原参考图像的相应子区域进行比较：

$$\varepsilon_k = \sum_{i=1}^{20} \sum_{j=1}^{20} |X_{i,j} - Y_{i,j}|$$

其中， $k$  为子区域编号， $1 \leq k \leq 12$ ， $(i, j)$  为子区域中任意一点的坐标， $X_{i,j}$  为参考图像中相应点的灰度值， $Y_{i,j}$  则为当前图像中相应点的灰度值。最后求出第  $k$  个子区域中的灰度绝对值差分  $\varepsilon_k$ 。

在口部图像运动检测中，将超过 3% 的平均灰度变化视为显著运动，则可设定阈值：

$$\varepsilon_0 = 255 \times 0.03 \times 20 \times 20 = 3060$$

将每个  $\varepsilon_k$  ( $1 \leq k \leq 12$ ) 与阈值  $\varepsilon_0$  进行比较，当  $\varepsilon_k \geq \varepsilon_0$  时，则可认为第  $k$  个子区域图像发生显著运动。设发生显著运动的子区域共计  $n$  个，则当  $n \geq 6$  时

(即发生显著运动的子区域达到或超过半数), 可认为整个口部发生显著运动, 即可判定为使用者开始发声, 电子喉开始工作。

启动电子喉之后, 仍然以 2Hz 的频率进行图像采样。将采集到的每一帧图像按照同样的求绝对值差分和算法, 与上一帧图像进行比较。如果当前图像与上一帧图像相比, 显著运动的子区域达到或超过半数, 则认为运动没有停止, 电子喉保持工作状态; 如果当前图像与上一帧图像相比, 显著运动的子区域不到半数, 则采用相同的求绝对值差分和算法, 将当前图像再与最初储存的静止状态参考图像比较。如果当前图像与参考图像相比, 显著运动的子区域达到或超过半数, 则认为使用者没有恢复口部静止状态, 电子喉保持工作状态; 如果当前图像与参考图像相比, 显著运动的子区域不到半数, 则认为使用者已恢复口部静止状态, 此时应关闭电子喉。

基于听觉掩蔽效应的电子喉语音增强算法及实现步骤如下:

设  $y(t) = s(t) + n(t)$ , 其中  $y(t)$  为带噪语音,  $s(t)$  为纯净语音,  $n(t) = n_1(t) + n_2(t)$ , 其中  $n_1(t)$  为周期性辐射噪声,  $n_2(t)$  为随机环境噪声。这是建立在语音与噪声不相关, 因而具有加性的假设前提下的。

则求出  $y(t)$  的频谱为:

$$Y(\omega) = Y_R(\omega) + iY_I(\omega) = FFT[y(t)]$$

其中  $Y(\omega)$  为  $y(t)$  的频谱,  $Y_R(\omega)$  和  $Y_I(\omega)$  分别为  $Y(\omega)$  的实部和虚部。

相应的求出周期性背景噪声和随机噪声频谱估值为:

$$N(\omega) = N_R(\omega) + iN_I(\omega) = FFT[n(t)]$$

在一般情况下,  $n_1(t)$  可以保持在整个电子喉发声阶段基本稳定不变, 而  $n_2(t)$  则可能随时随环境发生变化。因此  $N(\omega)$  也可能是随时间变化的。为了抵消这种变化带来的影响, 实现谱减系数的动态调整, 对带噪语音计算其感知加权滤波器的频响:

$$T(z) = \frac{A\left(\frac{z}{\sigma_1}\right)}{A\left(\frac{z}{\sigma_2}\right)} = \frac{1 - \sum_{k=1}^p a_k \sigma_1^k z^{-k}}{1 - \sum_{k=1}^p a_k \sigma_2^k z^{-k}}$$

其中  $T(z)$  即为感知加权滤波器的  $z$  域频响,  $a_k$  为当前语音帧的  $k$  阶 LPC 系数,  $p$  为 LPC 系数的最高阶数,  $\sigma_1, \sigma_2$  ( $0 \leq \sigma_2 \leq \sigma_1 \leq 1$ ) 为控制共振峰能量偏移的系数。

由感知加权滤波器频响求出谱减系数:

$$\alpha = \alpha_{\max} \left( \frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}} \right) + \alpha_{\min} \left( \frac{T(\omega) - T(\omega)_{\max}}{T(\omega)_{\max} - T(\omega)_{\min}} \right)$$

$$\beta = \beta_{\max} \left( \frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}} \right) + \beta_{\min} \left( \frac{T(\omega) - T(\omega)_{\max}}{T(\omega)_{\max} - T(\omega)_{\min}} \right)$$

其中,  $\alpha, \beta$  即为所求的谱减系数,  $T(\omega)$  为当前语音帧的感知加权滤波器频响,  $T(\omega)_{\max}, T(\omega)_{\min}$  分别为当前语音帧中  $T(\omega)$  的最大值和最小值,  $\alpha_{\max}$  和  $\beta_{\max}$  分别为  $\alpha$  和  $\beta$  的最大值,  $\alpha_{\min}$  和  $\beta_{\min}$  则分别为  $\alpha$  和  $\beta$  的最小值。 $\alpha_{\max}, \beta_{\max}, \alpha_{\min}$  和  $\beta_{\min}$  一般根据语音中噪声的类型和程度确定。

根据上式求出的谱减系数, 代入下式:

$$|\hat{S}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha |N(\omega)|^2 & , \text{当 } \left[ \frac{|N(\omega)|}{|Y(\omega)|} \right]^2 < \frac{1}{\alpha + \beta} \text{ 时} \\ \beta |N(\omega)|^2 & , \text{其他} \end{cases}$$

求出纯净语音频谱估值为:

$$\hat{S}(\omega) = |\hat{S}(\omega)|^2 \times \left( \frac{Y_R(\omega)}{\sqrt{Y_R^2(\omega) + Y_I^2(\omega)}} + \frac{iY_I(\omega)}{\sqrt{Y_R^2(\omega) + Y_I^2(\omega)}} \right)$$

则纯净语音估值为:

$$\hat{s}(t) = \text{IFFT}[\hat{S}(\omega)]$$

在本系统中, 上述算法的实现步骤如下:

将电子喉置于正常使用位置并保持固定, 系统启动后电子喉自动振动约 0.5 秒, 以采集辐射噪声和当前环境噪声。将这段噪声做 FFT 变换, 得到其

功率谱，作为噪声谱估值记录下来。电子喉开始工作后，以 8kHz 的频率对语音进行采样，每 256 个记录点作为一帧。对每帧语音进行 FFT 变换，得到其频谱，并进一步得到带噪语音的功率谱。之后利用带噪语音的功率谱，通过线性预测分析得到当前语音帧的感知加权滤波器频率响应。再根据感知加权滤波器频响  $T(z)$  求出当前帧中的谱减系数  $\alpha$ 、 $\beta$ ，即可通过谱减运算得到纯净语音功率谱估值。经过相位还原和 IFFT 变换，则得到一帧纯净语音的估值。

在系统实现过程中，算法中的一些参数取值如下：

- 1)  $\sigma_1 = 1$ ， $\sigma_2 = 0.8$ ；
- 2)  $\alpha_{\min} = 1$ ， $\alpha_{\max} = 6$ ；
- 3)  $\beta_{\min} = 0$ ， $\beta_{\max} = 0.02$ （当带噪语音信噪比低于 -5dB 时， $\beta_{\max} = 0.002$ ）。

本发明的方法所使用的自动控制算法与信号处理算法经过仿真检验，可以有效的实现电子喉的自动控制，并提高电子喉语音的质量。在主观与客观评测中，增强后的电子喉语音都有了明显的改善。通过对电子喉的自动控制和对电子喉语音的增强处理，可以使喉部病变患者的替代发声更为简便易行，并取得良好的效果。

### 附图说明

图 1 为本发明的自动电子喉及电子喉语音增强系统的结构图：其中的标号分别表示：1、应变传感采集模块，2、图像传感采集模块，3、控制模块，4、电子喉振动模块，5、语音数据采集模块，6、A/D 转换模块，7、数据处理模块，8、D/A 转换模块，9、滤波整形、功率放大及语音输出模块，10、程序扩展存储模块，11、数据扩展存储模块，12、电源模块；

图 2 为本系统实现应变控制部分的算法框图。

图 3 为本系统实现图像控制部分的算法框图。

图 4 为基于感知加权滤波器的电子喉语音增强算法框图。

图 5 为基于感知加权滤波器的电子喉语音增强效果仿真示意图，其中 (a) 图为增强之前的电子喉语音信号波形，(b) 图为增强之后的电子喉语音信号波形。

以下结合附图对本发明作进一步的详细描述。

### 具体实施方式

参见附图，本发明分为基于应力应变传感与图像传感两种工作模式，从其他角度提出了电子喉自动控制的实现思路与实施方案，并将技术方案分为基于 DSP 硬件系统平台和基于计算机应用软件平台两种实施方式。

本发明在硬件系统平台中，采用基于双 DSP 核心的硬件系统实现；在计算机软件平台中，则采用专门开发的 Windows 应用软件实现。通过面部应变检测传感和口部图像运动检测传感的方式实现对电子喉的自动控制，并利用基于感知加权滤波器的谱相减方法，对电子喉语音进行去噪和增强，以消除语音中所带有的辐射噪声和随机噪声，提高语音质量。

面部应变检测传感方式的电子喉自动控制基于以下原则：人在发声时，面部两侧肌肉、皮肤会随着口形变化而产生应变。一般而言，面部两侧的应变应保持一致，且相比于静息状态时，始终保持正应变状态。因此，在使用者避免面部不必要动作的前提下，只要检测出符合条件的应变，即可作为发声判定的依据。实现应变自动控制的方法如下：佩戴好装置，尽量保持面部平静松弛，将两个应变传感器分别置于面部两侧，开机后传感器将自动读取当前的应变值，通过 8 位模数转换送入 DSP 芯片，求出面部两侧应变的平均值  $s_M$ ，并设定判断电子喉开关的阈值为：

$$S_0 = s_M + (255 - s_M) \times 0.1$$

根据上式求出阈值  $s_0$ ，并储存起来，并以 2Hz 的频率不断循环采集面部两侧应变信号。当两侧采集到的应变信号都高于阈值后，可判定为使用者开始发声，则启动电子喉。电子喉开始工作后，面部应变信号依然以原频率进



行采集，以判断电子喉停止工作的时刻。由于发声过程中的语音间隔可能也会导致面部低应变状态，因此为了避免这种误差影响，当采集到低于阈值的应变值时，并不停止电子喉工作，而是当连续两帧的两侧应变信号都低于阈值  $s_0$  时，才认为使用者停止发声，关闭电子喉。

口部图像运动检测传感方式的电子喉自动控制基于以下原则：在发声时，口部形状和图像必然会发生变化，因此只要检测出符合条件的口部图像变化，即可作为发声判定的依据。对口部图像运动的检测采用基于 SAD (Sum-Absolute-Difference) 的算法。SAD 算法可应用于图像运动检测、运动追踪等方面，在本系统中，不需要追踪图像运动的轨迹，而只要确认口部图像开始运动，即可作为判定开始发声的控制信号。实现图像传感控制的方法如下：尽量保持面部、口部松弛平静，系统开机后自动通过摄像头记录当前静止状态下的口部 8 位灰度图像作为参考，图像分辨率设定为  $80 \times 60$ 。之后以 2Hz 的频率对口部图像进行采样，并将采集到的图像划分为 12 个  $20 \times 20$  的子区域，分别编号 1~12。在每一个子区域应用下式与原参考图像的相应子区域进行比较：

$$\varepsilon_k = \sum_{i=1}^{20} \sum_{j=1}^{20} |X_{i,j} - Y_{i,j}|$$

其中， $k$  为子区域编号， $1 \leq k \leq 12$ ， $(i, j)$  为子区域中任意一点的坐标， $X_{i,j}$  为参考图像中相应点的灰度值， $Y_{i,j}$  则为当前图像中相应点的灰度值。最后求出第  $k$  个子区域中的灰度绝对值差分  $\varepsilon_k$ ，并设定阈值为：

$$\varepsilon_0 = 255 \times 0.03 \times 20 \times 20 = 3060$$

将每个  $\varepsilon_k$  ( $1 \leq k \leq 12$ ) 与阈值  $\varepsilon_0$  进行比较，当  $\varepsilon_k \geq \varepsilon_0$  时，则可认为第  $k$  个子区域图像发生显著运动。设发生显著运动的子区域共计  $n$  个，则当  $n \geq 6$  时（即发生显著运动的子区域达到或超过半数），可认为整个口部发生显著运动，即可判定为使用者开始发声，电子喉开始工作。

启动电子喉之后，仍然以 2Hz 的频率进行图像采样。将采集到的每一帧图像按照同样的求绝对值差分算法，与上一帧图像进行比较。如果当前图像与上一帧图像相比，显著运动的子区域达到或超过半数，则认为运动没有停止，电子喉保持工作状态；如果当前图像与上一帧图像相比，显著运动的子区域不到半数，则采用相同的求绝对值差分算法，将当前图像再与最初储存的静止状态参考图像比较。如果当前图像与参考图像相比，显著运动的子区域达到或超过半数，则认为使用者没有恢复口部静止状态，电子喉保持工作状态；如果当前图像与参考图像相比，显著运动的子区域不到半数，则认为使用者已恢复口部静止状态，此时应关闭电子喉。

基于听觉掩蔽效应的电子喉语音增强算法基于以下假设：噪声与语音不相关，且噪声和语音都保持短时平稳。将带噪语音作频谱变换，根据每一帧语音不同的谱减系数减去噪声谱，再利用人耳对相位不敏感的特性，用带噪语音频谱的相位替代纯净语音频谱的相位，恢复到时域即可得到纯净语音的估值。具体实现步骤可表述如下：

设  $y(t) = s(t) + n(t)$ ，其中  $y(t)$  为带噪语音， $s(t)$  为纯净语音， $n(t) = n_1(t) + n_2(t)$ ，其中  $n_1(t)$  为周期性辐射噪声， $n_2(t)$  为随机环境噪声。这是建立在语音与噪声不相关，因而具有加性的假设前提下的。

则求出  $y(t)$  的频谱为：

$$Y(\omega) = Y_R(\omega) + iY_I(\omega) = FFT[y(t)]$$

其中  $Y(\omega)$  为  $y(t)$  的频谱， $Y_R(\omega)$  和  $Y_I(\omega)$  分别为  $Y(\omega)$  的实部和虚部。

相应的求出周期性背景噪声和随机噪声频谱估值为：

$$N(\omega) = N_R(\omega) + iN_I(\omega) = FFT[n(t)]$$

在一般情况下， $n_1(t)$  可以保持在整个电子喉发声阶段基本稳定不变，而  $n_2(t)$  则可能随时随环境发生变化。因此  $N(\omega)$  也可能是随时间变化的。为了抵消这种变化带来的影响，实现谱减系数的动态调整，对带噪语音计算其感知

知加权滤波器的频响:

$$T(z) = \frac{A\left(\frac{z}{\sigma_1}\right)}{A\left(\frac{z}{\sigma_2}\right)} = \frac{1 - \sum_{k=1}^p a_k \sigma_1^k z^{-k}}{1 - \sum_{k=1}^p a_k \sigma_2^k z^{-k}}$$

其中  $T(z)$  即为感知加权滤波器的  $z$  域频响,  $a_k$  为当前语音帧的  $k$  阶 LPC 系数,  $p$  为 LPC 系数的最高阶数,  $\sigma_1, \sigma_2$  ( $0 \leq \sigma_2 \leq \sigma_1 \leq 1$ ) 为控制共振峰能量偏移的系数。

由感知加权滤波器频响求出谱减系数:

$$\alpha = \alpha_{\max} \left( \frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}} \right) + \alpha_{\min} \left( \frac{T(\omega) - T(\omega)_{\max}}{T(\omega)_{\max} - T(\omega)_{\min}} \right)$$

$$\beta = \beta_{\max} \left( \frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}} \right) + \beta_{\min} \left( \frac{T(\omega) - T(\omega)_{\max}}{T(\omega)_{\max} - T(\omega)_{\min}} \right)$$

其中,  $\alpha$ 、 $\beta$  即为所求的谱减系数,  $T(\omega)$  为当前语音帧的感知加权滤波器频响,  $T(\omega)_{\max}$ 、 $T(\omega)_{\min}$  分别为当前语音帧中  $T(\omega)$  的最大值和最小值,  $\alpha_{\max}$  和  $\beta_{\max}$  分别为  $\alpha$  和  $\beta$  的最大值,  $\alpha_{\min}$  和  $\beta_{\min}$  则分别为  $\alpha$  和  $\beta$  的最小值。  $\alpha_{\max}$ 、 $\beta_{\max}$ 、 $\alpha_{\min}$  和  $\beta_{\min}$  一般根据语音中噪声的类型和程度确定。

根据上式求出的谱减系数, 代入下式:

$$|\hat{S}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha |N(\omega)|^2 & , \text{当} \left[ \frac{|N(\omega)|}{|Y(\omega)|} \right]^2 < \frac{1}{\alpha + \beta} \text{时} \\ \beta |N(\omega)|^2 & , \text{其他} \end{cases}$$

求出纯净语音频谱估值为:

$$\hat{S}(\omega) = |\hat{S}(\omega)|^2 \times \left( \frac{Y_R(\omega)}{\sqrt{Y_R^2(\omega) + Y_I^2(\omega)}} + \frac{iY_I(\omega)}{\sqrt{Y_R^2(\omega) + Y_I^2(\omega)}} \right)$$

则纯净语音估值为:

$$\hat{s}(t) = IFFT[\hat{S}(\omega)]$$

在本系统中，上述算法的实现步骤如下：将电子喉置于正常使用位置并保持固定，系统开机后电子喉自动振动约 0.5 秒，以采集初始状态下的辐射噪声和当前环境噪声。将这段噪声做 FFT 变换，得到其功率谱，作为初始噪声谱估值记录下来。电子喉开始工作后，以 8kHz 的频率对语音进行采样，每 256 个记录点作为一帧。对每帧语音进行 FFT 变换，得到其频谱，并进一步得到带噪语音的功率谱。之后利用带噪语音的功率谱，通过线性预测分析得到当前语音帧的感知加权滤波器频率响应。再根据感知加权滤波器频响  $T(z)$  求出当前帧中的谱减系数  $\alpha$ 、 $\beta$ ，即可通过谱减运算得到纯净语音功率谱估值。经过相位还原和 IFFT 变换，则得到一帧纯净语音的估值。

在本系统工作过程中，电子喉的工作状态可能不是连续的，而是由使用者控制量变化决定的间断工作方式，即电子喉会随着使用者面部应变或口部图像的变化而随时进入开启或关闭状态。为了提高噪声估计的精度，对增强算法的实现方式进行了改进。具体改进方法如下，从使用者控制量恢复静息状态到电子喉停止工作之间存在一个时间延迟，该延迟为 0~1 秒，因此在一段电子喉语音停止之前，存在一个 0~1 秒的时间段，该时间段内使用者已经停止发声，但电子喉仍然保持工作状态，这段时间内采集到的语音信号应该仅包括电子喉的辐射噪声与环境噪声。将电子喉停止工作前采集到的最后一帧语音信号与初始噪声信号的功率谱作比较，当这两帧信号的能量相差不超过 10% 时，即可认为采集到的最后一帧语音信号为纯噪声，并将此帧信号作为更新后的噪声估值，替代初始噪声；当两帧信号的能量相差超过 10% 时，则采集到的最后一帧语音信号可能还包括语音成分，不能作为噪声估值，因此保持初始噪声不变。

在系统实现过程中，算法中的一些参数取值如下：

- 1)  $\sigma_1 = 1$ ， $\sigma_2 = 0.8$ ；
- 2)  $\alpha_{\min} = 1$ ， $\alpha_{\max} = 6$ ；

3)  $\beta_{\min} = 0$ ,  $\beta_{\max} = 0.02$  (当带噪语音信噪比低于-5dB 时,  $\beta_{\max} = 0.002$ )。

参看图 1, 图 1 是按照上述方法实现的一种自动电子喉及电子喉语音增强系统的原理图。整体系统包括: 应变传感采集模块 1、图像传感采集模块 2、电子喉振动模块 4、语音数据采集模块 5、A/D 转换模块 6、控制模块 3、数据处理模块 7、D/A 转换模块 8、滤波整形、功率放大和语音输出模块 9、电源模块 12、扩展程序存储模块 11、扩展数据存储模块 12;

应变传感采集模块 1、图像传感采集模块 2、电子喉振动模块 4 与控制模块 3 连接, 电子喉振动模块 4 与语音数据采集模块 5 相连, 语音数据采集模块 5 通过 A/D 转换模块 6 与数据处理模块 7 连接; 控制模块 3 和数据处理模块 7 分别与扩展程序存储模块 11、扩展数据存储模块 12 连接, 数据处理模块 7 还通过 D/A 转换模块 8 与滤波整形、功率放大和语音输出模块 9 连接, 电源模块 12 负责整个系统的电源供应;

应变传感采集模块 1、图像传感采集模块 2 与控制模块 3 连接, 控制模块 3 产生的控制信号送往电子喉振动模块 4, 产生的电子喉语音信号由语音数据采集模块 5 进行采集, 并通过 A/D 转换模块 6 送往数据处理模块 7 进行增强处理, 数据处理模块 7 通过 D/A 转换模块 8 与滤波整形、功率放大及输出模块 9 连接, 控制模块 3 和数据处理模块 7 还分别与程序扩展存储模块 10、数据扩展存储模块 11 连接, 整个电路系统由电源模块 12 负责供电。

上述原理图可分别实现于基于双 DSP 架构的硬件系统平台和基于计算机的 Windows 应用软件平台。在 DSP 硬件系统平台中, 控制模块 3、数据处理模块 7 由 DSP 芯片实现, A/D 转换模块 6、D/A 转换模块 8、扩展程序存储模块 11、扩展数据存储模块 12 由专门芯片实现; 在计算机应用软件平台中, 控制模块 3、数据处理模块 7 由计算机 CPU 和相应的软件实现, A/D 转换模块 6、D/A 转换模块 8 由计算机声卡实现, 扩展程序存储模块 11、扩展数据存储模块 12 由计算机内存、硬盘实现。

在硬件系统平台中，由于系统工作中需要采集的数据量和运算量较大，因此采用双 DSP 架构，整个系统的控制模块和数据处理模块分别由一片 DSP 芯片完成。双 DSP 芯片（控制模块和数据处理模块）、A/D 转换模块、D/A 转换模块、程序扩展存储模块、数据扩展存储模块、电源模块及相应的外部接口集成在一块电路板上。系统采用的 DSP 芯片为 TI 公司的 TMS320C5410A 芯片，其工作主频最高可达 120MHz，配有 64k-16bit 的内部程序 RAM。DSP 控制模块 3 和 DSP 数据处理模块 7 的外围电路包括：

1、程序扩展存储模块 10。采用 64k-16bit 的外部程序 RAM 芯片 CY7C1021；

2、数据扩展存储模块 11。采用 256k-16bit 的片外 Flash Memory 芯片 LH28F400BVE；

3、A/D 转换模块 6 和 D/A 转换模块 8。A/D、D/A 转换模块采用 TLC320AD50C，动态范围 88dB，信噪比 89dB，最大采样率 22.05kHz，采样精度 16bit，RCA 接口供模拟信号输入/输出；

4、电源模块 12。整个系统使用单一蓄电池（+5V）供电；

5、系统输出采用标准音频接口，便于与各种音频设备与通讯工具相连；

应变和图像传感采集的频率设定为 2Hz，这样可以保证在正常使用条件下，电子喉自动开启和关闭的时间延迟不会超过 1 秒，并且使 DSP 控制芯片有充足的时间来进行控制信号的采集、分析与提取。为了提取控制量以确定电子喉工作状态，在应变传感工作模式下，通过应变传感器采集到的模拟电信号经过 A/D 转换为数字量，并送入控制模块，每次采样仅需传送来自面部两侧的两个 8 位数据，故所需的数据量和运算量都很小，利用 DSP 片内的存储器即可完成工作。而在图像运动检测传感模式下，每次采集的图像数据量为  $80 \times 60 \times 8\text{bit}$ ，对每一个像素点需要进行一次减法、一次绝对值及一次求和运算，因此需要通过扩展数据存储器来进行数据的运算、暂存和结

果记录。

在电子喉语音增强部分，由于系统需要进行实时数据分析与处理，语音信号采集频率设定为 8kHz，每 256 采样点为一帧。此采样频率远低于 DSP 数据处理芯片的工作频率，可以保证帧间相邻两个采样点之间的时间间隔足以完成中断数据处理。在 8kHz 的采样率下，DSP 数据处理模块通过扩展数据存储区对 256 点的一帧语音进行 FFT、LPC、IFFT 等运算，并将结果通过 D/A 转换模块输出，可以保证实时的数据处理与输出，总延时不超过 0.5 秒。

系统上电之后，DSP 控制芯片开始正常工作，采集并提取控制信号，DSP 数据处理芯片则处于待机状态。当 DSP 控制芯片提取到满足条件的控制量以开启电子喉时，会同时向 DSP 数据处理芯片发送一个工作信号，触发中断，使 DSP 数据处理芯片开始采集语音信号并进行增强处理。当 DSP 控制芯片关闭电子喉时，也会向 DSP 数据处理芯片发送停止工作信号，使之恢复待机状态。

在计算机应用软件平台中，核心处理芯片的速度一般远超过 DSP 芯片，存储器容量也远大于硬件平台中的存储芯片，数据处理的实时性和数据存储量都是可以保证的，因此各种计算、处理和存储工作都可在计算机中执行，外部电路仅需要集成控制量传感采集模块即可。计算机与外部电路的接口包括 USB 数据接口和音频输入输出接口，通过配套开发的软件界面进行系统控制。

在使用软件平台工作时，需要将外部电路的 USB 数据线和音频输出线与计算机相连，并启动相应软件界面，通过专门开发的 Windows 应用软件实现系统控制和语音增强，增强后的语音可以通过计算机声卡的音频输出口进行输出。此外，系统工作过程中，使用者发出的电子喉语音还可以在计算机上记录为音频文件并保存，以供备份和后续研究使用。

系统中的电源模块除了要为电路芯片供电之外，还需要为电子喉振动模块供电，因此耗电量较大。系统中采用单一 5V 蓄电池供电，以保证足够的输出电流。

应变传感控制的数据流程如下所述：面部两侧的应变传感器以 2Hz 的频率循环检测应变，并将模拟电信号转换为 8 位数字量传送到控制模块，与之前记录的阈值进行比较。如当前两侧应变均大于阈值，则向电子喉振动模块发送控制信号以启动电子喉，同时向数据处理模块发送信号，开始进行语音数据采集和增强处理。在电子喉工作过程中，仍然以 2Hz 频率进行应变采样，当采集到的两侧应变都小于阈值时，则关闭电子喉，同时向数据处理模块发送信号，停止语音数据采集和处理，恢复待机状态。

图像运动检测传感控制的数据流程如下所述：固定于口部前方的摄像头以 2Hz 的频率循环拍摄  $80 \times 60$  的 8 位灰度口部图像，将每帧图像传送到控制模块，划分为 12 个  $20 \times 20$  的子区域，分别与参考图像的相应子区域进行 SAD 计算，并将结果与阈值进行比较。如果一个图像子区域的 SAD 运算结果超过了阈值，则认为该子区域发生了显著运动。当达到半数的子区域发生显著运动时，则可判定使用者开始发声，由控制模块向电子喉振动模块发送控制信号以启动电子喉，同时向数据处理模块发送信号，开始进行语音数据采集和增强处理。在电子喉工作过程中，仍然以 2Hz 频率进行口部图像采样，在每采集一帧图像时，保留当前帧和上一帧的图像信息。如果当前帧与上一帧比较，发生显著运动的子区域不到半数，则再将当前帧与参考图像进行比较，如果发生显著运动的子区域仍然不到半数，则认为使用者停止发声，关闭电子喉，同时向数据处理模块发送信号，停止语音数据采集和处理，恢复待机状态。

电子喉语音增强部分的数据流程如下所述：接受到控制模块发来的中断信号之后，数据处理模块执行中断程序，开始控制语音数据采集模块对语音进行 8kHz, 16bit 的采样，每取得 256 个采样点，即经 A/D 转换模块送入数



据处理模块进行 FFT 运算，记录所得频谱的相角，再对实部和虚部取平方和，得到此 256 点语音的功率谱。之后对该 256 点语音进行 LPC 运算，得到其 16 阶的 LPC 系数，根据该系数求出当前语音帧感知加权滤波器的频响，再进一步求出当前语音帧的谱减系数。将谱减系数与之前储存的噪声功率谱代入改进的谱减公式，求出纯净语音功率谱估值，再将纯净语音估计功率谱开方后再与前面记录下的带噪语音相角相乘，并作 IFFT 变换，得到纯净语音估值。最后的结果经过滤波整形之后，通过 D/A 输出至标准音频接口。此外，在基于语音网络传输的应用领域，还可以将处理之后的语音直接以数字方式进行传输，再配合摄像头采集的口部图像视频信号，可以有效的提高增强后语音的可懂度，达到良好的增强效果。

根据上述原理，本系统的工作过程可参考图 2 和图 3 所示的控制算法流程，如下所述：使用者应保证系统启动前，选择三种控制模式（应变传感模式、图像运动检测传感模式以及手动控制模式）之中的一种将应变传感器、摄像头及电子喉置于正常工作位置，但不进行发声动作，然后开启系统。在硬件系统平台中，系统上电后，256k-16bit 的 Flash ROM 映射为数据空间，用于启动时程序的自动装载（Boot Loading）；在应用软件平台中，启动软件界面，并点击开始工作。程序复位后，进入主程序，控制芯片记录当前静息状态下的面部两侧应变值和口部图像信息，然后启动电子喉工作约 0.5 秒，并向数据处理模块发送工作信号，使数据处理模块采集并计算保存辐射噪声的功率谱，之后电子喉关闭，数据处理模块进入待机程序，等待中断触发，系统进入正常工作状态。

在正常工作状态，当选择电子喉自动控制工作模式时，控制模块程序根据当前工作模式，按 2Hz 的采样频率循环执行控制量采集、分析和判断步骤（应变或图像），当检测到满足条件的控制信号时，则启动电子喉振动模块，同时向数据处理模块发送中断信号，使语音采集模块按 8kHz 的采样率进行电子喉语音采集，并将采集到的数据通过 A/D 模块转换为数字信号，

依次送入信号采集缓冲区。当采集到 256 个数据，信号采集缓冲区被填满之后，则进入数据处理程序，将信号采集缓冲区的一帧语音信号送入数据处理模块，计算带噪语音功率谱、感知加权滤波器频响以及谱减系数，并与之前记录的噪声功率谱估值一起代入增强算法，经处理后得到 256 点纯净电子喉语音估值信号。将此帧语音信号送入数据发送缓冲区，按照 8kHz 的频率送入 D/A 模块，转换为模拟语音信号输出。语音数据处理算法可参考图 4 所示的信号处理流程。数据处理模块循环进行电子喉语音采样、处理工作，直到控制模块发送停止工作信号，则停止语音采样，数据处理程序跳转回待机状态。

当外部环境不适于使用自动控制模式时（如震动较剧烈，或光线明暗变化较大时），可选择传统的手动控制电子喉工作模式，此时的自动控制模块自动关闭，仅保留数据处理模块进行工作。

在任何一种控制模式下，都可以控制摄像头对口部图像进行采集、存储，并将每一帧口部图像信号实时输出，作为对语音信号的补充。在特定条件下，可以通过口部图像来辅助理解语音，并为发声矫正、语音康复评价等方面的工作积累相关材料。

电子喉的自动控制是一个相对新颖的课题，其核心问题在于控制量的选择、分析与提取。选择易于提取、误差较小且能精确反映电子喉工作状态的控制量，是整个系统的基础。谱相减算法在正常语音增强领域已经得到了广泛应用，但在病理语音特别是电子喉语音增强领域，需要通过谱减系数的设定和调整，对其进行相应的修改和完善，以达到更好的增强效果。将电子喉从工作控制到语音增强的整个工作流程进行集成化处理，并进行系统实现工作，能够给使用者带来更多便利和更好的语音质量，使喉部病变患者的语音重建工作达到更高的水平，并在语音通信、传输等领域发挥重要的作用。

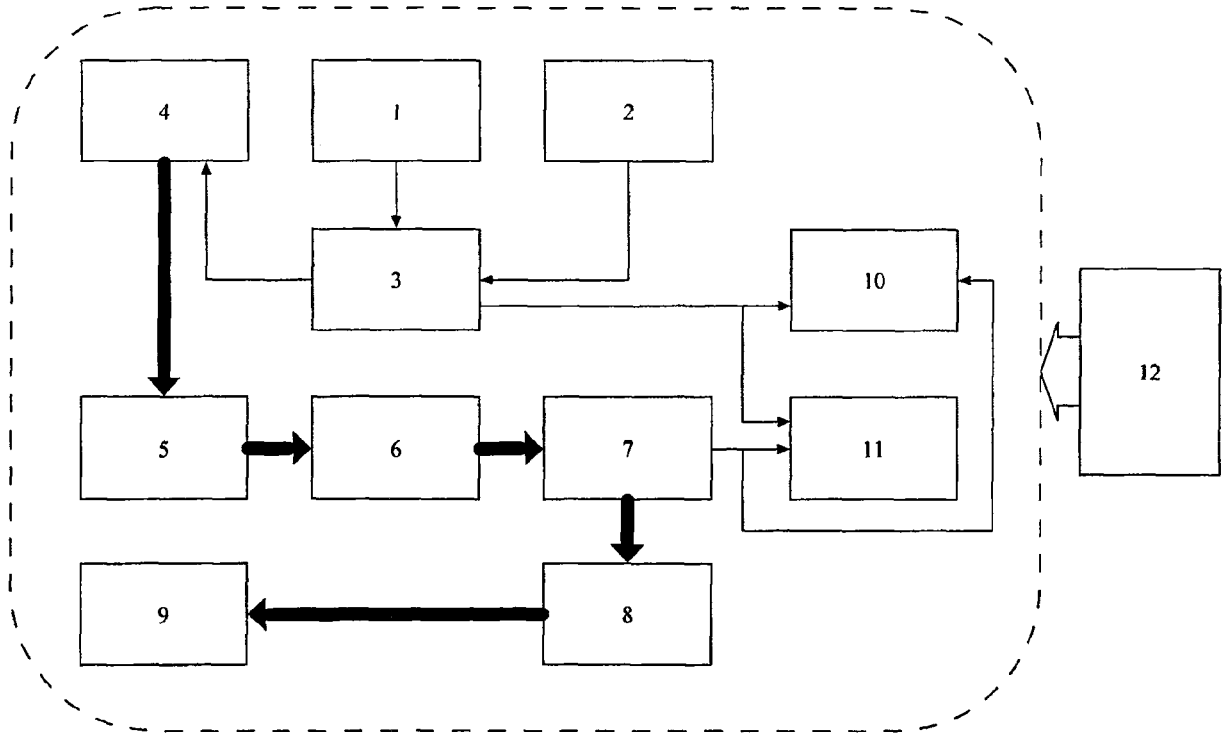


图 1

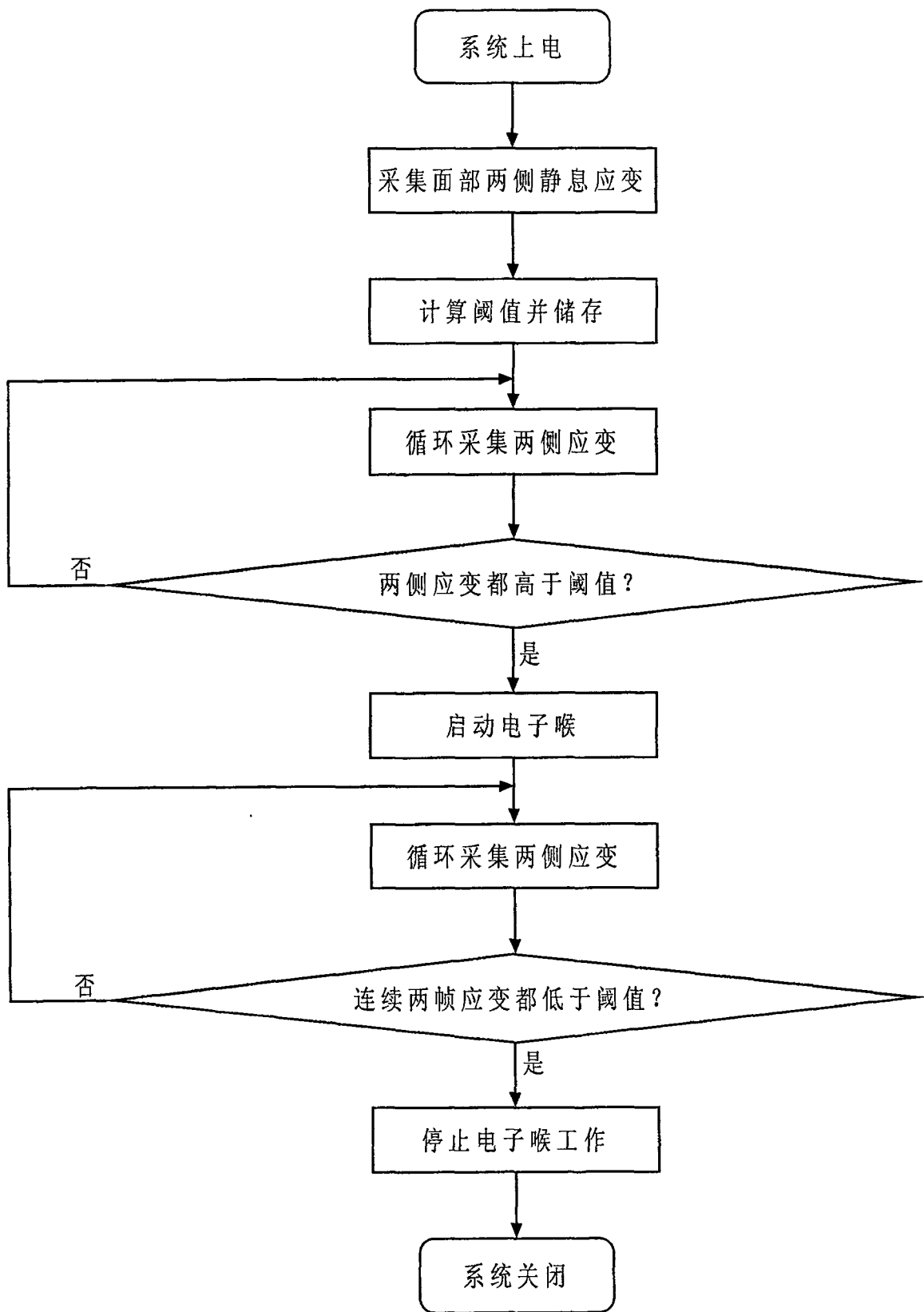


图 2

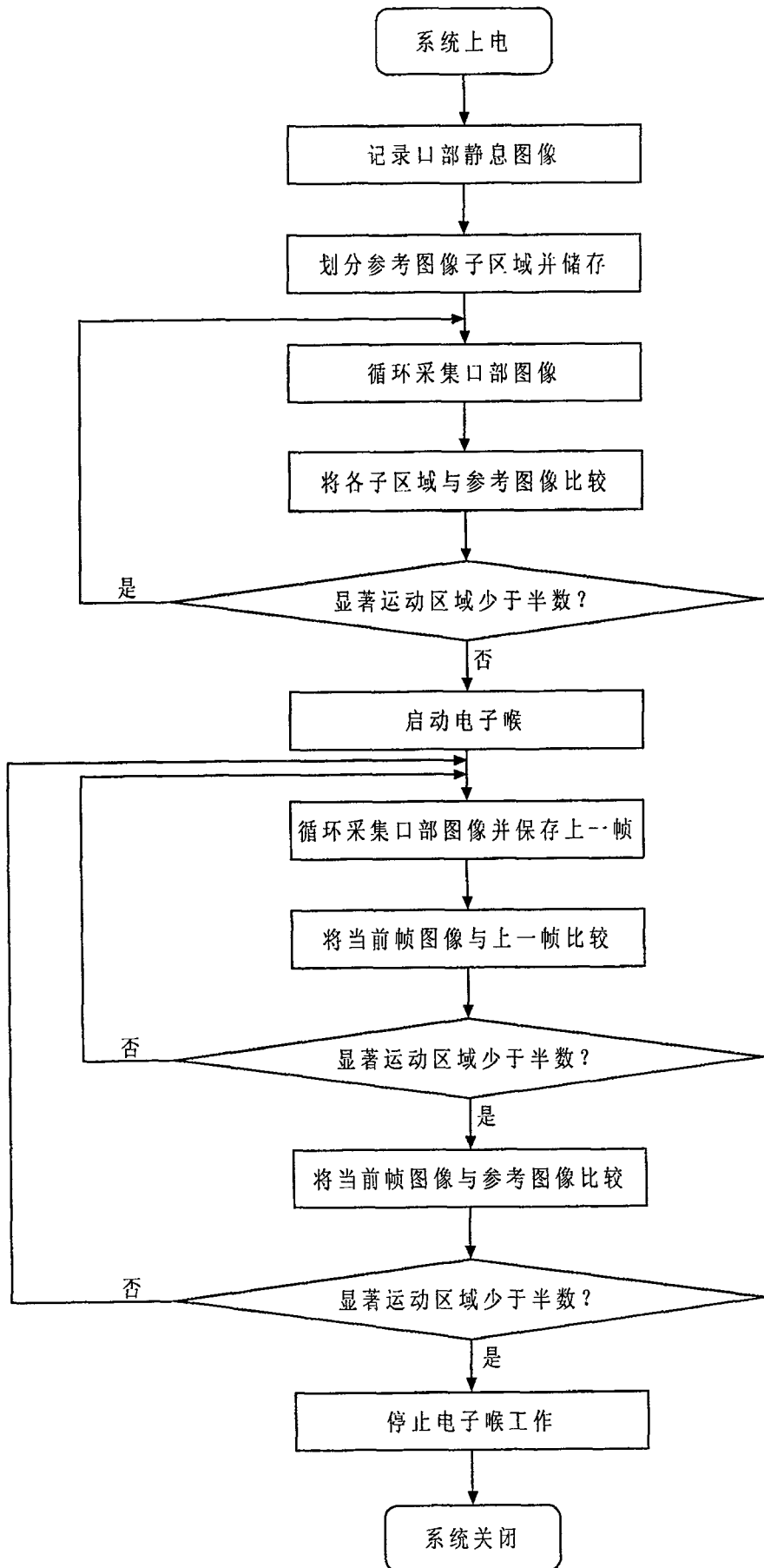


图 3

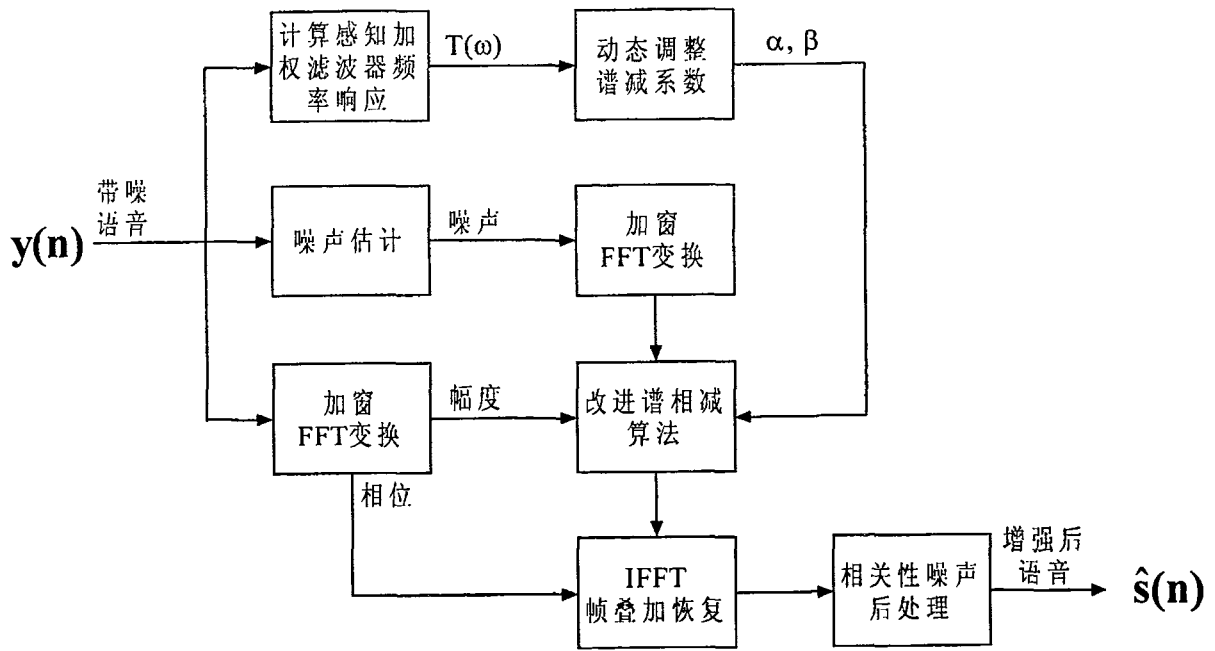


图 4

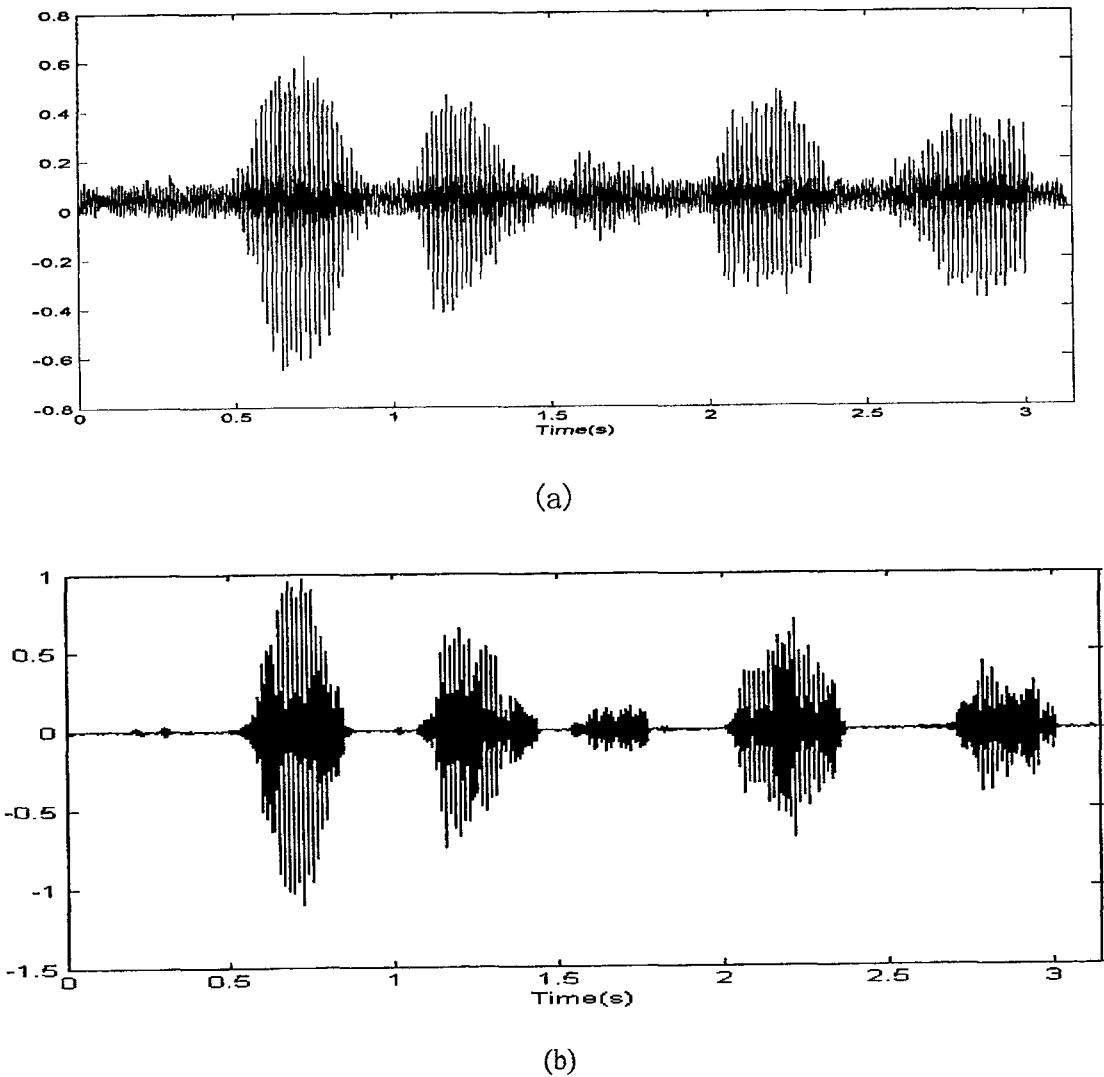


图 5