

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号
WO 2012/003602 A1

(43) 国际公布日
2012年1月12日 (12.01.2012)

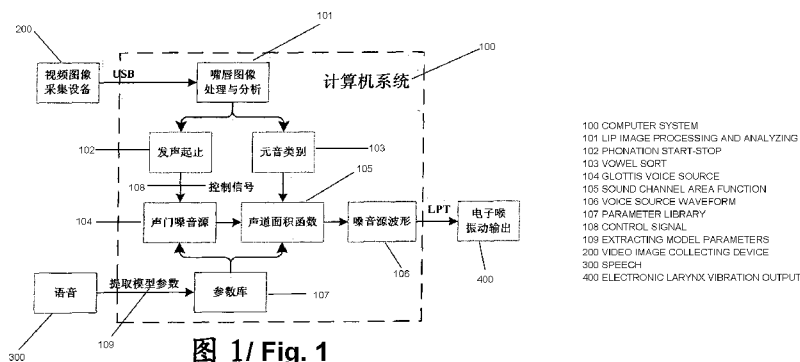
PCT

- (51) 国际分类号：
G10L 15/24 (2006.01) G06T 1/40 (2006.01)
G10L 13/02 (2006.01) A61F 2/70 (2006.01)
G06K 9/00 (2006.01)
- (21) 国际申请号：PCT/CN20 10/00 1022
- (22) 国际申请日：2010年7月9日 (09.07.2010)
- (25) 申 请 语 言：中文
- (26) 公 布 语 言：中文
- (71) 申请人 (对除美国外的所有指定国)：西安交通大学 (XI'AN JIAOTONG UNIVERSITY) [CN/CN]; 中国陕西省西安市咸宁路28号, Shaanxi (CN)。
- (72) 发明人及
(75) 发明人/申请人 (仅对美国)：万明习 (WAN, Mingxi) [CN/CN]; 中国陕西省西安市咸宁路28号, Shaanxi (CN)。吴亮 (WU, Liang) [CN/CN]; 中国陕西省西安市咸宁路28号, Shaanxi (CN)。王素品 (WANG, Supin) [CN/CN]; 中国陕西省西安市咸宁路28号, Shaanxi (CN)。牛志峰 (NIU, Zhifeng) [CN/CN]; 中国陕西省西安市咸宁路28号, Shaanxi (CN)。万聪颖 (WAN, Congying) [CN/CN]; 中国陕西省西安市咸宁路28号, Shaanxi (CN)。
- (74) 代理人：北京中原华和知识产权代理有限公司 (BEIJING ZHONGYUAN HUAHE INTELLEC TUAL PROPERTY AGENCY CO., LTD.); 中国北京市朝阳区北辰东路8号汇宾大厦A座909室, Beijing 100101 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。
- 本国际公布：
- 包括国际检索报告(条约第21条(3))。

[续页]

(54) Title: METHOD FOR RECONSTRUCTING ELECTRONIC LARYNX SPEECH AND SYSTEM THEREOF

(54) 发明名称：一种电子喉语音重建方法及其系统



(57) Abstract: A method for reconstructing electronic larynx speech and a system thereof are provided. The method includes the following steps: firstly, extracting model parameters from the collected speech as a parameter library; then collecting facial images of a sounder, and transmitting the facial images to an image analyzing and processing module to obtain phonation start-stop moments and phonation vowel sorts; then synthesizing voice source waveform by a voice source synthesis module; finally, outputting the voice source waveform by an electronic larynx vibration output module. The voice source synthesis module firstly sets glottis voice source model parameters, thereby synthesizing glottis voice source waveform, then simulates the sounds traveling in a sound channel by a waveguide model, and selects shape parameters of the sound channel according to the phonation vowel sorts, thereby synthesizing electronic larynx voice source waveform. The speech reconstructed by the method and the system thereof is much closer to the sound of the sounder himself.

(57) 摘要：

[续页]



WO 2012/003602 A1



提供了一种电子喉语音重建方法及其系统。该方法包括以下步骤：首先，从采集的语音中提取模型参数作为参数库；接着采集发声者的面部图像，传输面部图像给图像分析与处理模块，得到发声起止时刻与发声元音类别；接着通过噪音源合成模块合成噪音源波形；最后，通过电子喉振动输出模块将该噪音源波形输出。噪音源合成模块首先设置声门噪音源模型参数，从而合成声门噪音源波形，再用波导模型模拟声音在声道中传播，并根据发声元音类别选择声道的形状参数，从而合成电子喉噪音源波形。该方法及其系统重建的语音更接近发声者本身的声音。

一种电子喉语音重建方法及其系统

技术领域

本发明属于病变语音重建领域，特别涉及一种电子喉语音重建方法及其系统。

背景技术

语音或者语言是人类表达感情和相互交流的主要手段，然而，据统计每年全世界有成千上万的人因为各种喉部外科手术而暂时或永久丧失发声能力。鉴于此，各种嗓音康复技术应运而生，其中，以食管语音、气管食管语音，以及人工电子喉语音最为常见，而人工电子喉因为使用简单、适用范围广、可长时间发声而被广泛应用。

中国发明专利申请第 20091 0020897. 3 号公开了一种自动调节的咽腔电子喉语音通讯的方法，其去除了其他的噪音，从而提高了重建语音的质量。电子喉的工作原理是提供缺失的噪音振动源，并通过换能器将振动传递进入声道进行语音调制，最后通过唇端辐射产生语音。由此可见，提供缺失的噪音振动源是电子喉最根本的任务，然而，目前市面上所见的电子喉所提供的振动噪音源多为方波或脉冲信号，改进的线性换能器虽然能输出声门噪音源，但是这些都不符合实际使用过程中缺失的噪音振动源。无论是颈外式还是口腔式电子喉，振动传递进入声道的位置都不是声门，而且对于不同病人的不同手术情况，不仅是声带缺失，而且包括部分声道的缺失，这些都需要在电子喉振动源中得到补偿，因此从电子喉本质方面改进以提高电子喉语音质量是十分必要的。

鉴于以上问题，实有必要提供一种可以解决上述技术问题的电子喉语音重建方法及其系统

发明内容

本发明所要解决的技术问题是提供一种电子喉语音重建方法及其系统，通过本方法重建的语音不仅补偿了缺失声道的声学特性，而且保留了使用者的个性特点，更接近使用者本身的声音特点，语音质量更好。

为实现上述目的，本发明提供了一种电子喉语音重建方法，首先从采集的语音中提取模型参数作为参数库，接着采集发声者的面部图像，将该图像传输给图像分析与处理模块，图像分析与处理模块分析处理完之后，得到发声起止时刻与发声元音类别，再接着，以发声起止时刻和发声元音类别控制噪音源合成模块并合成噪音源波形，最后，通过电子喉振动输出模

块将上述噪音源波形输出，电子喉振动输出模块包括前置电路和电子喉振动器，所述噪音源合成模块的合成步骤如下：

1) 合成声门噪音源波形，即根据使用者发声的个性特征在参数库内选择声门噪音源模型参数，其中，发声起止时刻控制噪音源合成的开始和结束，所述声门噪音源合成采用LF模型，具体数学表示如下：

$$\begin{cases} u'_g(t) = E_0 e^{\alpha t} \sin(\omega_g t) & (0 \leq t \leq t_e) \\ u'_g(t) = -\left(\frac{E_e}{\epsilon t_a}\right) [e^{-\epsilon(t-t_e)} - e^{-\epsilon(t_c-t_e)}] & (t_e \leq t \leq t_c) \end{cases}$$

上式中， E_e 为幅度参数， t_p 、 t_e 、 t_a 、 t_c 均为时间参数，分别代表气流最大峰值时刻、最大负峰值时刻、指数回复段时间常数和基频周期，其余参数可由以上五个参数按照以下公式联合求得：

$$\begin{cases} \epsilon t_a = 1 - e^{-\alpha t_e} \\ \omega_g = \frac{\pi}{t_p} \\ U_e = 50 [e^{\alpha t_e} \{ \alpha \sin \omega_g t_e - \omega_g \cos \omega_g t_e \} + \omega_g] / (\alpha^2 + \omega_g^2) \\ E_e = -E_0 e^{\alpha t_e} \sin \omega_g t_e \\ U_e = \frac{E_e t_a}{2} K_\alpha \\ K_\alpha = \begin{cases} 2.0 & R_\alpha < 0.1 \\ 2 - 2.34 R_\alpha + 1.34 R_\alpha^2 & 0.1 \leq R_\alpha \leq 0.5 \\ 2A6 - 13.2 R_\alpha + 0.64 (R_\alpha - 0.5)^2 & R_\alpha > 0.5 \end{cases} \\ R_\alpha = \frac{t_a}{t_c - t_e}; \end{cases}$$

2) 根据发声元音类别选择声道的形状参数，利用波导模型模拟声音在声道中传播，按照以下公式计算出噪音源波形

$$\begin{cases} u_{i+1}^+ = (1 - r_i) u_i^+ - r_i u_M^- = u_i^+ - r_i (u_i^+ + u_{i+1}^-) & r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}} \\ u_i^- = (1 + r_i) u_{i+1}^- + r_i u_i^+ = u_{i+1}^- + r_i (u_i^+ + u_{i+1}^-) \\ \text{glottis : } u_1^+ = \frac{1 - r_g}{2} u_g - r_g u_1^- = \frac{1}{2} u_g - r_g \left(\frac{\Lambda}{2} u_g + u_1^- \right) & r_g \approx -1 \\ \text{lips : } u_{out} = (1 - r_N) u_N^+ = u_N^+ - u_N^- & r_N \approx -1 \end{cases}$$

声道由多个均匀截面积的声管级联表示，上式中， A_i 和 A_{i+1} 为第 i 个和第 $i+1$ 个声管的面积函数， u_i^+ 和 u_i^- 分别为第 i 个声管中的正向声压和反向声压， r_i 是第 i 个和第 $i+1$ 个声管相邻界面的反射系数。

作为本发明的优选实施例，所述图像分析与处理模块包括如下步骤：

步骤一：初始化参数，即预设分析矩形框范围、面积阈值和神经网络权系数，然后采集一帧视频图像，其中面积阈值为分析矩形框面积的百分之

步骤二：利用基于肤色的检测方法对嘴唇区域进行检测，即在 YUV 色彩空间按照下述公式计算矩形框范围的唇色特征值，并归一化为 0-255 灰度级：

$$Z = 0.493 \Lambda - 0.589G + 0.0265$$

步骤三：利用改进的最大类间方差法计算唇色特征值灰度图像的最佳分割阈值，然后，以此阈值对图像进行二值化分割，从而，得到嘴唇的初步分割图像；

步骤四：采用面积阈值的方法，将初步分割图像中面积小于阈值的区域作为噪声消去，得到最终的嘴唇分割图像；

步骤五：对嘴唇区域进行外轮廓和中心点提取：设定椭圆长轴与 X 轴成零度角，利用椭圆模型对嘴唇外轮廓进行匹配，通过一维哈夫变换检测得到椭圆长轴的大小；

步骤六：以归一化半长轴、归一化半短轴、长短轴之比和嘴唇归一化面积值作为一组参数，计算发声起止时刻和发声元音类别，其中，所述归一化半长轴、归一化半短轴，以及嘴唇归一化面积均是指以不发声时静态半长轴、半短轴、嘴唇面积为标准的归一化值。

作为本发明的另一优选实施例，所述图像分析与处理模块的步骤六中，采用人工神经网络算法计算发声起止时刻和发声元音类别。

作为本发明的另一优选实施例，所述人工神经网络算法为三层网络，包括输入层、隐含层，以及输出层，其中，输入层包含四个输入，即归一化半长轴、归一化半短轴、长短轴之比和嘴唇归一化面积值，输出层包括六个输出，即不发声、/ə/、/e/、/i/、/o/、/u/ 五个元音。

作为本发明的另一优选实施例，所述噪音源合成过程中，以声道咽腔下部声压波形作为颈部施加的噪音源波形。

作为本发明的另一优选实施例，所述噪音源合成过程中，以口腔位置声压波形作为口腔内施加的噪音源波形。

为了实现上述目的，本发明还提供了一种电子喉语音系统，包括 CMOS 图像传感器、与 CMOS 图像传感器的输出端相连的 FPGA 芯片、与 FPGA 芯片的输出端相连的语音芯片，以及与语音芯片的输出端相连的电子喉振动器。

本发明电子喉语音重建方法及其系统至少具有以下优点：首先，在噪音源合成模块的声门噪音源 LF 模型中，声门波形由幅度参数 Ee 以及 t_p 、 t_e 、 t_a 、 t_c 四个时间参数共同表征，而这五个参数可以从语音中提取出来，因此对

于不同的使用者来说，可以从其失声前保留的语音中提取出来作为合成参数，故重建语音具有使用者的个性特点；另外，在噪音源合成模块的声道波导模型中，根据视频信号判断的发声元音类别选择声道形状参数，依据使用者咽喉部手术切除情况，选择合适的振动器施加位置，因此，针对施加部位合成对应声道空间位置的声压波形作为电子喉噪音源波形，如此，不仅符合使用者的实际情况，而且极大的保留了使用者的个性特征，使重建语音更接近使用者本人的原始语音，改善重建语音质量。

附图的简要说明

图 1 是本发明电子喉语音重建方法的流程示意图；

图 2 是本发明嘴唇运动图像处理和提取控制参数提取程序流程图；

图 3 是本发明噪音源合成流程图；

图 4 是本发明不同发声和使用情况下合成的电子喉噪音源波形图；

图 5 是本发明电子喉振动输出模块示意图；

图 6 是本发明电子喉语音系统的一个结构框图。

实现发明的最佳方式

下面结合附图对本发明电子喉语音重建方法及其系统进行详细描述：

本发明以计算机系统为平台，根据使用者失声的具体情况及个人发声特点调整噪音源波形的合成，同时利用视频信号对噪音源合成进行实时控制，最终通过并口连接的电子喉振动输出模块将上述噪音源波形输出。

本发明电子喉语音重建方法的系统包括图像采集设备、与图像采集设备的输出端相连的图像处理及分析模块、与图像处理及分析模块的输出端相连的噪音源合成模块，以及与噪音源合成模块的输出端相连的电子喉振动输出模块。

请参阅图 1 所述，当系统启动后，图像采集设备，即摄像头采集使用者发声过程中的面部图像，并将该面部图像传输给图像处理及分析模块，图像处理及分析模块接收到该数据后进行处理与分析，即通过嘴唇检测、分割、边缘提取和拟合，从而得到嘴唇边缘的椭圆模型形状参数，之后再通过人工神经网络算法计算判断发声的起止时刻和发声元音类别并作为噪音源合成的控制信号；噪音源合成模块采用发声合成法原理，根据不同使用者的情况，包括手术情况、发声个性特点，以及提取的发声起止和发声元音类别，合成具有使用者个性特征和符合实际发声需要的噪音源波形，最后通过电子喉振动输出模块将上述合成的噪音源波形输出。

由上述可知，本发明电子喉语音重建方法主要包括三个部分，一、图像采集及处理；二、电子喉噪音源的合成；三、电子喉的振动输出。下面

详细描述：

本发明的第一部分为图像采集及处理，主要利用图像处理的方法对嘴唇的运动实现实时分析，并利用人工神经网络判断发声起止时刻与发声元音类别，作为控制信号用以控制电子喉噪音源的动态合成。

下面结合图2所示详细介绍第一部分的具体实现步骤：

- 1) 初始化参数，即预设分析矩形框范围、面积阈值和神经网络权系数，然后采集一帧视频图像，其中面积阈值为分析矩形框面积的百分之一；
- 2) 利用基于肤色的检测方法对嘴唇区域进行检测，即在YUV色彩空间按照下述公式(一)计算矩形框范围的唇色特征值以增强嘴唇区域的区分度，并归一化为0-255灰度级，从而，得到唇色特征值灰度图像，公式(一)如下：

$$Z = 0.4937R - 0.589G + 0.0265B \quad \text{公式(一)}$$

在上述公式(一)中，R、G、B分别代表红色、绿色和蓝色分量。

- 3) 利用改进的最大类间方差(Otsu)法计算唇色特征值灰度图像的最佳分割阈值，然后，以此阈值对图像进行二值化分割，从而，得到嘴唇的初步分割图像；

4) 采用面积阈值的方法，将初步分割图像中面积小于阈值的区域作为噪声消去，得到最终的嘴唇分割图像；

5) 对嘴唇区域进行外轮廓和中心点提取：假设椭圆长轴与X轴成零度角，利用椭圆模型对嘴唇外轮廓进行拟合，通过一维哈夫(Hough)变换检测得到椭圆长短轴的大小；

6) 以归一化半长轴、归一化半短轴、长短轴之比和嘴唇归一化面积值四个值作为一组参数，经过人工神经网络的计算得到发声起止时刻和发声元音类别，用以指导噪音源合成控制。

需要说明：在本发明中，归一化半长轴、归一化半短轴，以及嘴唇归一化面积均是指以不发声时静态的半长轴、半短轴、嘴唇面积为标准的归一化值。

在本实施方式中，以长短轴的比值和归一化参数作为神经网络的输入，因为它们不但能够准确反映嘴型的变化情况，而且可以判断发声起止时刻与元音类别，具有很好的距离不变性，可以克服由于使用者与摄像头距离变化造成图像中嘴唇面积大小改变而产生的判断错误，因此，得到的判断信号与语音波形具有很好的吻合度，判断准确率较高。

另外，为了满足实时性的要求，本发明的图像处理在嘴唇分割和椭圆模型参数匹配中都采用了时空域联合的跟踪控制方法，即基于说话时面部变化缓慢连续的假设，通过前一帧图像分割的区域信息和椭圆匹配的参数信息指导本帧图像的分割的矩形范围和匹配的参数范围，很好的利用了帧

内和帧间信息，不仅提高了处理的速度，还提高了计算精度。

本发明中的人工神经网络为三层前向神经网络，包括输入层（即归一化半长轴、归一化半短轴、长短轴之比和嘴唇归一化面积值）、隐含层（三十个节点）、输出层（即不发声、/ə/、/ɔ/、/ɪ/、/i/、以及/ɛ/五个元音），其中，神经网络的节点权系数预先经过样本训练得到，训练采用误差反向传播（BP）算法，样本为不发声静止状态和发各个元音时的嘴唇形状参数。

请继续参阅图 3 所示，本发明的第二个部分为噪音源的合成，利用发声合成法原理，通过源-滤波器两步合成电子喉噪音源，具体步骤如下：

第一步：合成声门噪音源波形：根据使用者发声的个性特征在参数库中选择并设置声门噪音源模型参数，图像采集及处理模块内得到的发声起止时刻控制噪音源合成的开始和结束，按照 LF 模型合成声门噪音源；

所述声门噪音源合成采用 LF 模型，具体数学表示如下：

$$\begin{cases} u'_g(t) = E_0 e^{\alpha t} \sin(\omega_g t) & (0 \leq t \leq t_e) \\ u'_g(t) = -\left(\frac{E_e}{\epsilon t_a}\right) [e^{-\epsilon(t-t_e)} - e^{-\epsilon(t_c-t_e)}] & (t_e \leq t \leq t_c) \end{cases}$$

上式中， E_e 为幅度参数， t_p 、 t_e 、 t_a 、 t_c 均为时间参数，分别代表气流最大峰值时刻、最大负峰值时刻、指数回复段时间常数和基频周期，其余参数可由以上五个参数按照以下公式联合求得：

$$\left\{ \begin{array}{l} \epsilon t_a = 1 - e^{-\epsilon(t_c-t_e)} \\ \omega_g = \frac{\pi}{t_p} \\ U_e = E_0 [e^{\alpha t_e} (\alpha \sin \omega_g t_e - \omega_g \cos \omega_g t_e) + \omega_g] / (\alpha^2 + \omega_g^2) \\ E_e = -E_0 e^{\alpha t_c} \sin \omega_g t_e \\ U_e = \frac{E_e t_a}{2} K_\alpha \\ K_\alpha = \begin{cases} 2.0 & R_\alpha < 0.1 \\ 2 - 2.34R_\alpha^2 + 1.34R_\alpha^4 & 0.1 \leq R_\alpha \leq 0.5 \\ 2.16 - 1.32R_\alpha + 0.64(R_\alpha - 0.5)^2 & R_\alpha > 0.5 \end{cases} \\ R_\alpha = \frac{t_a}{t_c - t_e} \end{array} \right.$$

第二步：根据判断的发声元音类别选择声道的形状参数，利用波导模型模拟声音在声道中传播，根据以下公式计算出使用者实际使用时，振动传递进入声道处的声压波形，即为合成的电子喉噪音源：

所述模拟声音在声道中传播的波导模型的具体数学表示如下：

$$\begin{cases} u_{i+1}^+ = (1 - r_i) u_{i+1}^- - r_i u_i^- = u_i^+ - r_i (u_i^+ + u_{i+1}^-) \\ u_i^- = (1 + r_i) u_{i+1}^- + r_i u_i^+ = u_{i+1}^- + r_i (u_i^+ + u_{i+1}^-) \end{cases} \quad r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}$$

$$\begin{cases} \text{glottis} : u_1^+ = \frac{1 - r_g}{2} u_g - r_g u_1^- = \frac{1}{2} u_g - r_g \left(\frac{1}{2} u_g + u_1^- \right) & r_g \approx -1 \\ \text{lips} : u_{\text{out}} = (1 - r_N) u_N^+ = u_N^+ - u_N^- & r_N \ll -1 \end{cases}$$

其中，声道由多个均匀截面积的声管级联而成，表示为面积函数 A_i ， u_i^+ 和 u_i^- 分别为第 i 个声管中的正向声压和反向声压， r_i 是第 i 个和第 $i+1$ 个声管相邻界面的反射系数，由相邻声管的截面积 A_i 和 A_{i+1} 确定，波导模型通过迭代可以计算出声道任意位置的声压。

需要说明的是：第一、在上述噪音源合成模块的 LF 模型中，声门噪音源波形由幅度参数 E_e 及 t_p 、 t_e 、 t_a 、 t_c 四个时间参数共同确定，对于不同的人而言，由于其解剖结构和发声特点不同，因此，声门噪音源波形具有个性差异，这些都可以体现在 LF 模型的五个参数中，而这几个参数都是从语音中提取出来的。例如，女性发声时基频普遍高于男性，因此，女性的 t_c 要比男性小等等。在本发明中，为了充分保留使用者的声音特点，重建出与患者失声前相同的语音，需要从患者失声前采集的语音中提取上述五个参数，存在参数库内，使用电子喉时，只需要在参数库中提取上述参数，即可重建出具有使用者发声特点的语音，而对于没有采集到失声前语音的患者来说，其可以选择自己喜欢的语音特点的参数，从而重建出自己喜欢的语音。

第二、在上述噪音源合成模块的波导模型中，唯一的参数就是声道的面积函数 A_i ，不同的人发声或者同一人发声不同，其对应的声道形状都不相同，故，本发明中采用元音类别的控制方法，根据不同的发声元音，选取不同的声道面积函数用于合成；而对于不同的使用者，我们首先建立一个元音-声道面积函数对应模板库，合成时只需要根据判断元音类别查找相应的声道函数即可，这个模板库的建立是利用反求的方法从使用者录制的语音中获得声道响应函数，再从声道响应函数求取最佳匹配的声道面积函数，这样可以使得使用者的发声个性特征得到保留。

由以上可知，通过两步合成，可以计算出声道中任意位置的声压信号，然而选取声道中哪一个位置的声压信号作为电子喉噪音源，需要根据使

用者具体的手术情况和使用方式来决定。

下面请参阅图 4 所示，为不同情况下合成的噪音源的波形图，例如，由于喉癌而进行喉切除手术但声道保留较完整的使用者，可以采用颈部施加振动的方式从而充分利用保留的声道作用，因此，选取声道咽腔下部的声压波形作为电子喉噪音源波形，图 4(a) 和图 4(c) 即分别为该情况下发元音为 /o/ 和 /% / 时合成的噪音源波形；对于咽部癌症的患者，需要进行咽切除术，如此，病人不仅丧失了声带，而且很大部分的声道也被破坏，此时必须选取口腔处的声压波形作为噪音源波形，图 4(b) 和图 4(d) 即分别为该情况下发元音为 /o/ 和 /% / 时合成的噪音源波形。

如此，从图 4 可以看出本发明针对不同的手术情况、使用情况和发声类别，从而，合成不同的电子喉噪音源波形，不仅符合实际使用的需要，而且保留了使用者的个性特点，很大程度上改善了电子喉重建语音的质量。

请参阅图 5 所示，本发明的第三个模块为电子喉的振动输出模块，包括电子喉振动器以及电子喉振动器前置电路，计算机通过 LPT 并口将合成的电子喉噪音源波形信号输入前置电路，经过数模转换和功率放大后，由音频接口输出模拟电压信号，最后电子喉振动器振动，从而输出噪音源。

所述电子喉振动器为线性换能器，即将电压信号线性转换成机械振动，因此，其可以按照合成的噪音源输出振动，同时为了满足口腔内施加的需要，使用导音管将振动导入口腔内部。

请继续参阅图 5 所示，电子喉振动器前置电路由输入输出接口、D/A 数模转换、功率放大和电源控制组成。输入输出接口分别为 25 针数字输入并口和 3.5mm 模拟输出音频接口，其中数字输入并口与计算机并口输出端相连，传输速度为 44100Byte/s，模拟输出音频接口与电子喉振动器相连接；D/A 数模转换器采用 DAC0832，数据精度 8 位，可以直接与 LPT 并口的数据位相连；功率放大器使用 Ti 公司的 TPA701 音频功率放大器，+3.5~+5.5V 供电，输出功率可达 700mW；电源控制为 5V 电池，提供 +5V 直流电压给各芯片。

在以上实施方式中，该电子喉的语音系统是以视频采集设备、计算机和电子喉振动输出模块为基础实现的，然而，为了便于实现，还可以采用另外一种实施方式，如图 6 所示，在该实施方式中，电子喉语音系统包括用于采集图像的 CMOS 图像传感器、与 CMOS 图像传感器的输出端相连并用于对采集到的图像进行分析处理和噪音源合成的 FPGA 芯片、与 FPGA 芯片的输出端相连并用于对合成的电子喉噪音源波形进行 D/A 转换和功率放大的语音芯片，以及与语音芯片的输出端相连的电子喉振动器。

所述 CMOS 图像传感器采用 MICRON 公司的 MT9M011，最大分辨率为 640 x 480，在该分辨率下的帧率为 60 帧/s，用于采集使用者发声过程中的面

部图像。

FPGA 芯片支持 SOPC 技术，实现以视频数据为输入，经过视频数据处理分析和电子喉噪音源合成，最终输出电子喉噪音源波形数据的功能；该 FPGA 芯片除了包含与 CMOS 图像传感器及语音芯片相连的接口外，还包括 LCD、FLASH，以及 SDRAM，其中，LCD 为液晶显示屏，用于显示相关数据，FLASH 为闪存，SDRAM 为同步动态随机存储器。

语音芯片采用 AIC23，包括 D/A 转换器和功率放大功能，经过 D/A 转换和功率放大后，由音频接口输出到电子喉振动器。

以上所述仅为本发明的一种实施方式，不是全部或唯一的实施方式，本领域普通技术人员通过阅读本发明说明书而对本发明技术方案采取的任何等效的变换，均为本发明的权利要求所涵盖。

权 利 要 求

1、一种电子喉语音重建方法，首先从采集的语音中提取模型参数作为参数库，接着采集发声者的面部图像，将该图像传输给图像分析与处理模块，图像分析与处理模块分析处理完之后，得到发声起止时刻与发声元音类别，再接着，以发声起止时刻和发声元音类别控制噪音源合成模块并合成噪音源波形，最后，通过电子喉振动输出模块将上述噪音源波形输出，电子喉振动输出模块包括前置电路和电子喉振动器，其特征在于：所述噪音源合成模块的合成步骤如下：

1) 合成声门噪音源波形 根据使用者发声的个性特征在参数库内选择声门噪音源模型参数 其中 发声起止时刻控制噪音源合成的开始和结束 所述声门噪音源合成采用 LF 模型，具体数学表示如下：

$$\begin{cases} u'_g(0) = E_0 e^{\alpha t} \sin(\omega_g t) & (0 \leq t \leq t_e) \\ u_g(t) = -\left(\frac{E_e}{\varepsilon t_a}\right) [e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}] & (t_e \leq t \leq t_c) \end{cases}$$

上式中， E_e 为幅度参数， t_p 、 t_e 、 t_a 、 t_c 均为时间参数，分别代表气流最大峰值时刻、最大负峰值时刻、指数回复段时间常数和基频周期，其余参数可由以上五个参数按照以下公式联合求得：

$$\begin{cases} \varepsilon t_a = 1 - e^{-\varepsilon(t_c-t_e)} \\ \omega_g = \frac{\pi}{t_p} \\ U_e = E_0 [e^{\alpha t_e} (\alpha \sin \omega_g t_e - \omega_g \cos \omega_g t_e) + \omega_g] / (\alpha^2 + \omega_g^2) \\ E_e = -E_0 e^{\alpha t_e} \sin \omega_g t_e \\ U_e = \frac{E_e t_a}{2} K_\alpha \\ K_\alpha = \begin{cases} 2.0 & R_\alpha < 0.1 \\ 2 - 2.34R_\alpha^2 + 1.34R_\alpha^4 & 0.1 \leq R_\alpha \leq 0.5 \\ 2.16 - 1.32R_\alpha + 0.64(R_\alpha - 0.5)^2 & R_\alpha > 0.5 \end{cases} \\ R_\alpha = \frac{t_a}{t_c - t_e}; \end{cases}$$

2) 根据发声元音类别选择声道的形状参数，利用波导模型模拟声音在声道中传播，按照以下公式计算出噪音源波形：

$$\begin{cases} u_{i+1}^+ = (1 - r_i)u_i^+ - r_i u_{i+1}^- = u_i^+ - r_i(u_i^+ + u_{i+1}^-) \\ u_i^- = (1 + r_i)u_{i+1}^- + r_i u_i^+ = u_{i+1}^- + r_i(u_i^+ + u_{i+1}^-) \end{cases} \quad r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}$$

$$\begin{cases} \text{glottis} & u_i^+ = \frac{1 - r_g}{2} u_g - r_g u_i^- = \frac{1}{2} u_g - r_g \left(\frac{1}{2} u_g + u_i^- \right) \quad r_g \approx -1 \\ \text{lips} & u_N^+ = (1 - r_N)u_N^- = u_N^- \quad r_N \approx -1 \end{cases}$$

声道由多个均匀截面积的声管级联表示，上式中， A_i 和 A_{i+1} 为第*i*个和第*i+1*个声管的面积函数， u_i^+ 和 u_i^- 分别为第*i*个声管中的正向声压和反向声压， r_i 是第*i*个和第*i+1*个声管相邻界面的反射系数。

2、如权利要求 1 所述的电子喉语音重建方法，其特征在于：所述图像分析与处理模块包括如下步骤：

步骤一：初始化参数：预设分析矩形框范围、面积阈值和神经网络权系数，然后采集一帧视频图像，其中面积阈值为分析矩形框面积的百分之一；

步骤二：利用基于肤色的检测方法对嘴唇区域进行检测，即在 YUV 色彩空间按照下述公式计算矩形框范围的唇色特征值，并归一化为 0-255 灰度级：

$$Z = 0A93R - 0.589G + 0.0265$$

步骤三：利用改进的最大类间方差法计算唇色特征值灰度图像的最佳分割阈值，然后，以此阈值对图像进行二值化分割，从而，得到嘴唇的初步分割图像；

步骤四：采用面积阈值的方法，将初步分割图像中面积小于阈值的区域作为噪声消去，得到最终的嘴唇分割图像；

步骤五：对嘴唇区域进行外轮廓和中心点提取：设定椭圆长轴与 X 轴成零度角，利用椭圆模型对嘴唇外轮廓进行匹配，通过一维哈夫变换检测得到椭圆长短轴的大小；

步骤六：以归一化半长轴、归一化半短轴、长短轴之比和嘴唇归一化面积值作为一组参数，计算发声起止时刻和发声元音类别，其中，所述归一化半长轴、归一化半短轴，以及嘴唇归一化面积均是指以不发声时静态半长轴、半短轴、嘴唇面积为标准的归一化值。

3、如权利要求 2 所述的电子喉语音重建方法，其特征在于：所述图像分析与处理模块的步骤六中，采用人工神经网络算法计算发声起止时刻和发声元音类别。

4、如权利要求 3 所述的电子喉语音重建方法，其特征在于：所述人工神经网络算法为三层网络，包括输入层、隐含层，以及输出层，其中，输入层包含四个输入，即归一化半长轴、归一化半短轴、长短轴之比和嘴唇归一化面积值，输出层包含六个输出，即不发声、/ə/、/ɪ/、/ʊ/、/ɔ/、/ɑ/。

以及 /..l/ 五个元音。

5、如权利要求 1 或 4 所述的电子喉语音重建方法，其特征在于：所述噪音源合成过程中，以声道咽腔下部声压波形作为颈部施加的噪音源波形。

6、如权利要求 1 或 4 所述的电子喉语音重建方法，其特征在于：所述噪音源合成过程中，以口腔位置声压波形作为口腔内施加的噪音源波形。

7、一种应用权利要求 1 所述的方法的电子语音系统，其特征在于：包括 CMOS 图像传感器、与 CMOS 图像传感器的输出端相连的 FPGA 芯片、与 FPGA 芯片的输出端相连的语音芯片，以及与语音芯片的输出端相连的电子喉振动器。

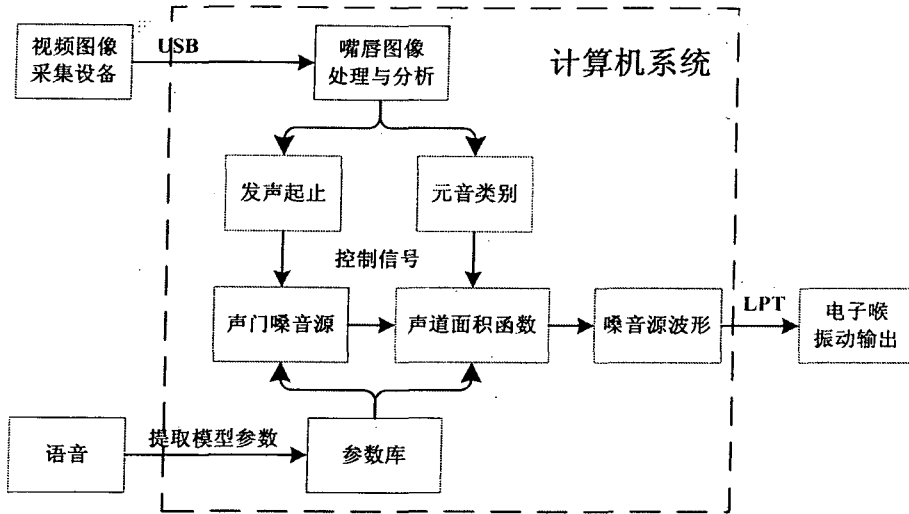


图 1

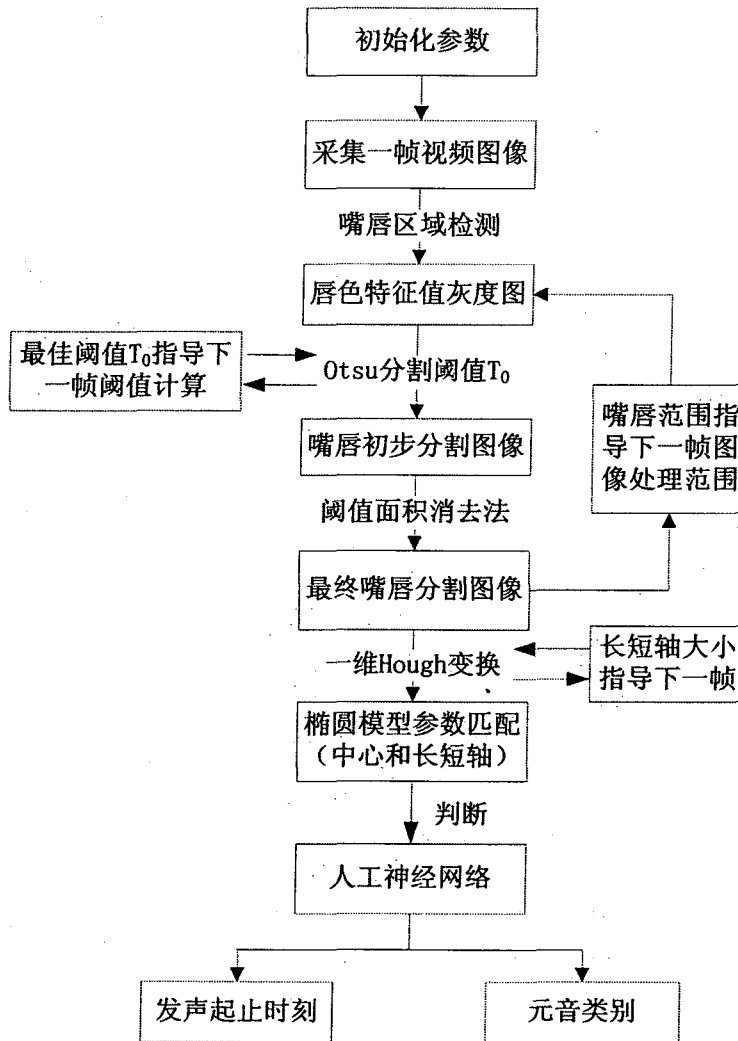


图 2

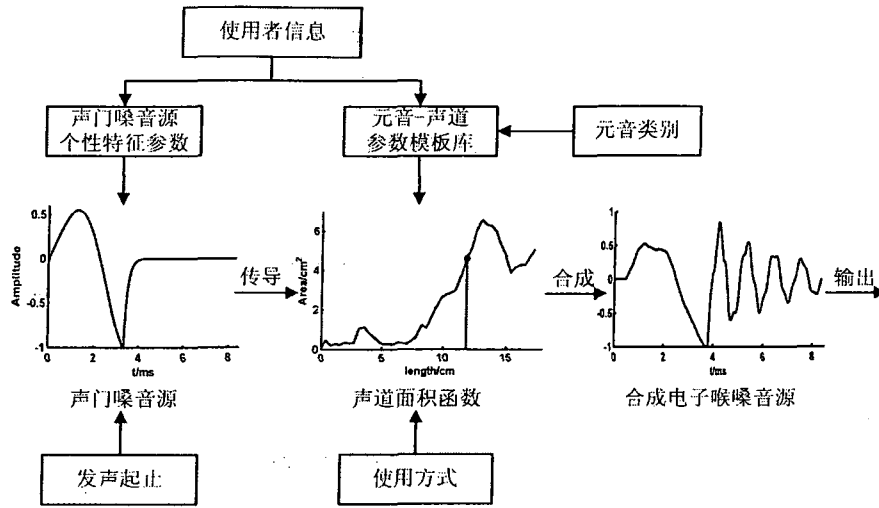


图 3

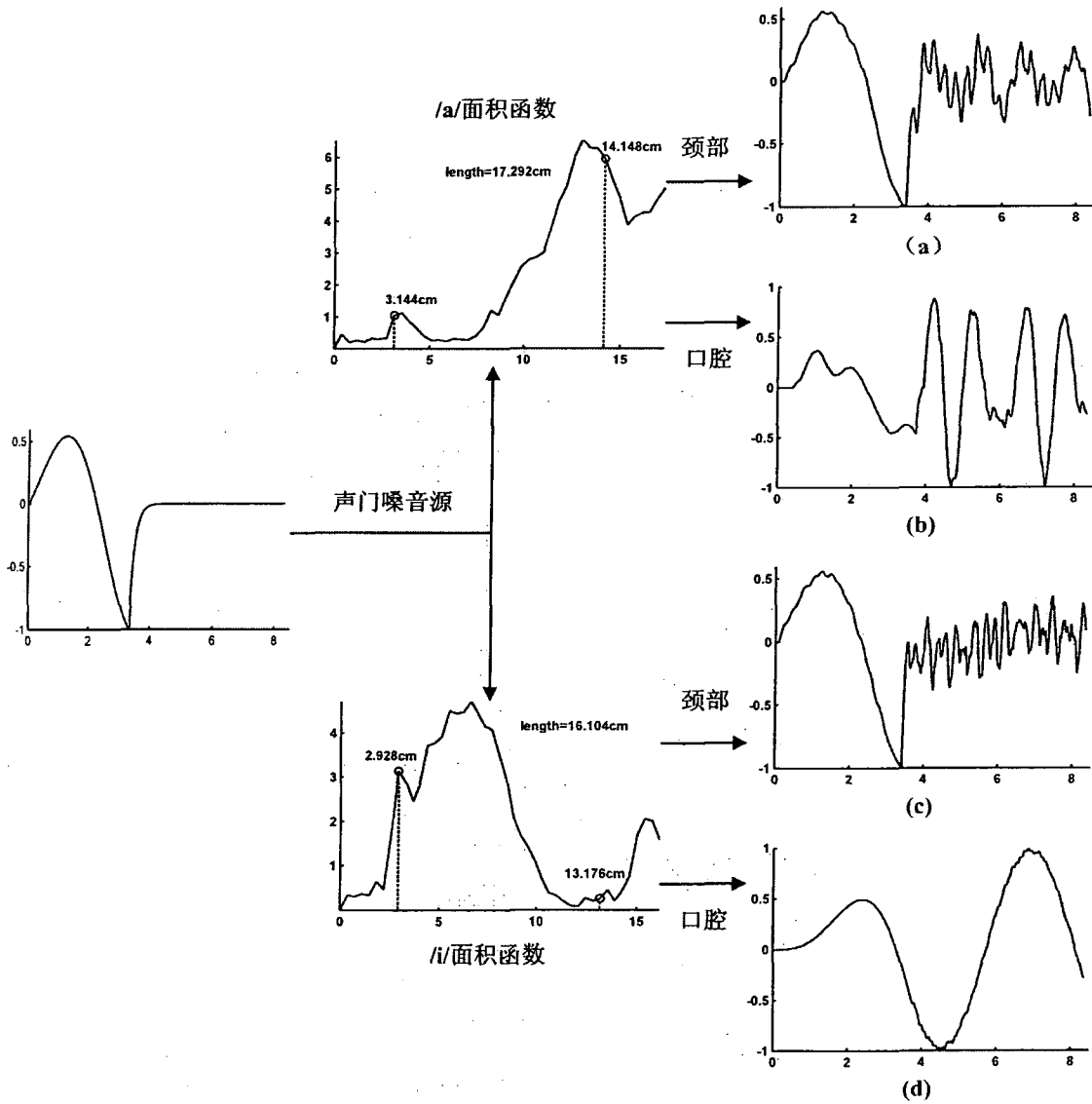


图 4

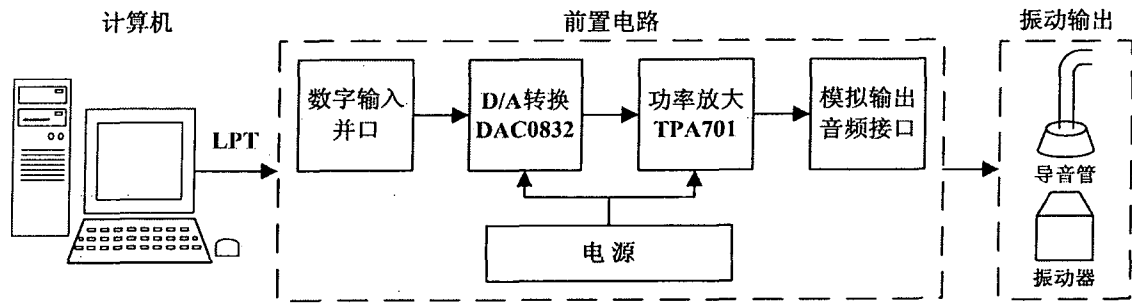


图 5

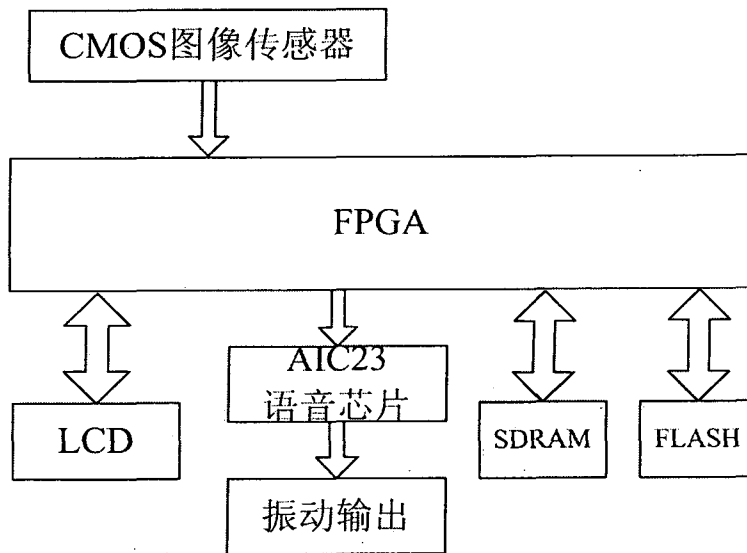


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1022

A. CLASSIFICATION OF SUBJECT MATTER

see the extra sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G10L; G06K; G06T; A61F; G06F17/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT,CNKI,WPI,EPODOC: electro, larynx, throat, guttural, palate, voice, sound, speech, phonation, synthesis, wave form, rebuild, reconstruct

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN101030384A, (UNIV XIAN JIAOTONG), 05 Sep. 2007(05.09.2007), the whole document	1-7
A	CN1776809A, (UNIV XIAN JIAOTONG), 24 May 2006(24.05.2006), the whole document	1-7
A	CN101474104A, (UNIV XIAN JIAOTONG), 08 Jul. 2009(08.07.2009), the whole document	1-7
A	WO2010/004397A1, (UNIV WITWATERSRAND JOHANNESBURG), 14 Jan. 2010(14.01.2010), the whole document	1-7

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
26 Mar. 2011(26.03.2011)Date of mailing of the international search report
07 Apr. 2011 (07.04.2011)Name and mailing address of the ISA/CN
The State Intellectual Property Office, the P.R.China
6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China
100088
Facsimile No. 86-10-62019451Authorized officer
XU,Wei
Telephone No. (86-10)62411668

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN20 10/00 1022

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN101030384A	05.09.2007	CN100576320C	30.12.2009
CN1776809A	24.05.2006	CN1003705 19C	20.02.2008
CN101474104A	08.07.2009	NONE	
WO2010/004397A1	14.01.2010	NONE	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN20 10/00 1022

Continuation of the second sheet

A. CLASSIFICATION OF SUBJECT MATTER:

G10L15/24 (2006.01) 1

G10L13/02 (2006.01) 1

G06K9/00 (2006.01) i

G06T1/40 (2006.01) i

A61F2/70 (2006.01) i

<p>A. 主题的分类</p> <p style="text-align: center;">参见附加页</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p style="padding-left: 20px;">IPC: G10L; G06K; G06T; A61F; G06F17/-</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称,和使用的检索词(如使用))CNPAT,CNKI,WPI,EPODOC: 电子喉, 语音, 声音, 噪音, 合成, 波形, 重建, 重构, 聘 ,electro, larynx, throat, guttural, palate, voice, sound, speech, phonation, synthesis, wave form, rebuild, reconstruct</p>																	
<p>C. 相关文件</p> <table border="1" style="width:100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%; padding: 5px;">类 型 *</th> <th style="width: 60%; padding: 5px;">引用文件, 必要时, 指明相关段落</th> <th style="width: 30%; padding: 5px;">相关的权利要求</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;">A</td> <td style="padding: 5px;">CN101030384A, (西安交通大学),05.9月2007(05.09.2007), 全文</td> <td style="padding: 5px;">1-7</td> </tr> <tr> <td style="padding: 5px;">A</td> <td style="padding: 5px;">CN1776809A, (西安交通大学),24.5月2006(24.05.2006), 全文</td> <td style="padding: 5px;">1-7</td> </tr> <tr> <td style="padding: 5px;">A</td> <td style="padding: 5px;">CN101474104A, (西安交通大学),08.7月2009(08.07.2009), 全文</td> <td style="padding: 5px;">1-7</td> </tr> <tr> <td style="padding: 5px;">A</td> <td style="padding: 5px;">WO2010/004397A1, (UNIV WITWATERSRAND JOHANNESBURG), 14.1月2010(14.01.2010), 全文</td> <td style="padding: 5px;">1-7</td> </tr> </tbody> </table>			类 型 *	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN101030384A, (西安交通大学),05.9月2007(05.09.2007), 全文	1-7	A	CN1776809A, (西安交通大学),24.5月2006(24.05.2006), 全文	1-7	A	CN101474104A, (西安交通大学),08.7月2009(08.07.2009), 全文	1-7	A	WO2010/004397A1, (UNIV WITWATERSRAND JOHANNESBURG), 14.1月2010(14.01.2010), 全文	1-7
类 型 *	引用文件, 必要时, 指明相关段落	相关的权利要求															
A	CN101030384A, (西安交通大学),05.9月2007(05.09.2007), 全文	1-7															
A	CN1776809A, (西安交通大学),24.5月2006(24.05.2006), 全文	1-7															
A	CN101474104A, (西安交通大学),08.7月2009(08.07.2009), 全文	1-7															
A	WO2010/004397A1, (UNIV WITWATERSRAND JOHANNESBURG), 14.1月2010(14.01.2010), 全文	1-7															
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p>																	
<table style="width:100%;"> <tr> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>* 引用文件的具体类型:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件</p> <p>"E" 在国际申请日的当天或之后公布的在先申请J%¼¼¼</p> <p>"L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)</p> <p>"O" 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>"P" 公布日先于国际申请日但迟于所要求的优先权日的文件</p> </td> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>"X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>"&" 同族专利的文件</p> </td> </tr> </table>			<p>* 引用文件的具体类型:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件</p> <p>"E" 在国际申请日的当天或之后公布的在先申请J%¼¼¼</p> <p>"L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)</p> <p>"O" 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>"P" 公布日先于国际申请日但迟于所要求的优先权日的文件</p>	<p>"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>"X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>"&" 同族专利的文件</p>													
<p>* 引用文件的具体类型:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件</p> <p>"E" 在国际申请日的当天或之后公布的在先申请J%¼¼¼</p> <p>"L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)</p> <p>"O" 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>"P" 公布日先于国际申请日但迟于所要求的优先权日的文件</p>	<p>"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>"X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>"&" 同族专利的文件</p>																
<p>国际检索实际完成的日期</p> <p style="text-align: center;">26.3月2011(26.03.2011)</p>	<p>国际检索报告邮寄日期</p> <p style="text-align: center;">07.4月2011(07.04.2011)</p>																
<p>ISA/CN 的名称和邮寄地址:</p> <p>中华人民共和国国家知识产权局</p> <p>中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号: (86-10)62019451</p>	<p>授权官员</p> <p style="text-align: center;">徐薇</p> <p>电话号码: (86-10) 62411668</p>																

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2010/001022

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN101030384A	05.09.2007	CN100576320C	30.12.2009
CN1776809A	24.05.2006	CN100370519C	20.02.2008
CN101474104A	08.07.2009	无	
WO20 10/004397 A 1	14.01.2010	无	

A. 主题的分类

续第 2 页

- G10L15/54 (2 006.01) ↓
- 610L13/02 (2 006.01) ↓
- 60639/00 (2 006.01) ↓
- O06N/110 (2 006.01) ↓
- A61F2/70 (2006.01) ↓