(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103824564 A
(43)申请公布日 2014.05.28

(21)申请号 201410099003.5

(22)申请日 2014.03.17

(71)申请人 上海申磬产业有限公司
地址 201203 上海市浦东新区张江高科技园区松涛路489号C座2楼
申请人 金会庆
宋扬
金来

(72)发明人 金会庆 宋扬 金来 章志达
魏晶 张毅 刘想德 徐晓东

(74)专利代理机构 上海智力专利商标事务所 31105
代理人 瞿承达

(51)Int.Cl.
G10L 21/0232(2013.01)
G10L 15/20(2006.01)

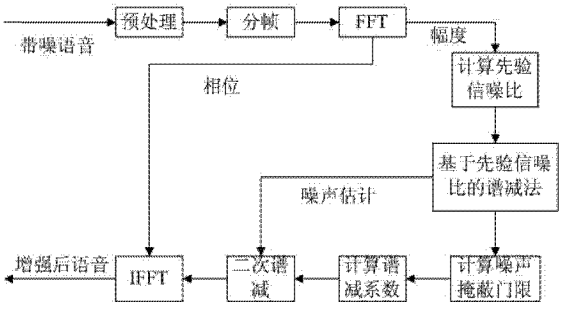
权利要求书4页 说明书9页 附图1页

(54)发明名称

一种电动轮椅语音识别过程中的语音增强方法

(57)摘要

本发明涉及电动轮椅语音识别过程中的语音增强方法，该增强方法在信号空间，运用听觉掩蔽效应和先验信噪比相结合的语音增强算法进行增强，克服了在智能轮椅鲁棒语音识别在噪声环境中识别率偏低的问题。本发明的方法对于特定人语音识别系统具有较好的通用性，其用于电动轮椅语音控制系统中的语音识别，能够提高系统在噪声环境下的识别率，实现了对电动轮椅的精确控制，达到用户与电动轮椅之间语音交互的目的。



CN 103824564 A

1. 一种电动轮椅语音识别过程中的语音增强方法，所述电动轮椅语音识别过程包括有输入语音信号预处理、特征参数提取、训练阶段中处理特征参数为每个命令词得到一个模块保存为模板库、识别阶段中先得到语音参数生成测试模板，该测试模板与模板库中的参考模板匹配，将匹配相似度最好的参考模块作为识别结果，其特征在于，所述的语音增强方法包括如下步骤：

第一步，对带噪声的输入语音信号进行预处理、分帧以及快速傅里叶变换，得到带噪音语音信号的频谱；

第二步，计算每一帧带噪声语音信号的先验信噪比，通过第一次谱减法，得到较为纯净的语音信号频谱；

第三步，通过听觉掩蔽效应计算噪声掩蔽门限，得到噪声掩蔽门限后计算自适应的谱减系数；

第四步，进行第二次谱减法，得到纯净的语音信号频谱；

第五步，通过逆快速傅里叶变换，得到增强后的语音信号。

2. 根据权利要求 1 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，所述第二步中的计算先验信噪比时，根据谱减法用线性时变滤波器形式表示，即$|Y_k|$乘以增益函数G_k，如式（1）所示：

$$|\hat{S}_k|=G_k|Y_k| \quad 0\le G_k\le 1 \tag{1}$$

增益函数可以通过先验信噪比表示，如式（2）所示：

$$G(k,i)=\sqrt{\frac{SNR_{pri}(k,i)}{SNR_{post}(k,i)}} \tag{2}$$

式中的先验信噪比可以通过“直接判决”法进行估计得到，如式（3）所示：

$$SNR_{pri}(k,i)=\eta\frac{|\hat{S}(k,i-1)|^2}{|\hat{D}(k,i-1)|^2}+(1-\eta)\max[SNR_{post}(k,i)-1,0] \tag{3}$$

式中，$SNR_{post}(k,i)$为：

$$SNR_{post}(k,i)=1+SNR_{pri}(k,i) \tag{4}$$

式中i为当前帧，$i-1$为前一帧，$\hat{S}(k,i-1)$为前一帧的估计结果，η为调节系数，0.8 $<\eta<$ 1。

3. 根据权利要求 1 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，所述第三步中计算基于人耳掩蔽效应的听觉掩蔽阈值时需先划定临界带宽，该临界带宽是当纯音信号的功率等于该噪声信号的功率时，纯音信号刚好被人耳听到的临界状态的频率宽度，该临界宽度为实验测得的常数。

4. 根据权利要求 3 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，所述听觉掩蔽阈值的计算过程如下：先分别计算临界带宽的功率谱、扩展临界带宽的功率谱和噪声掩蔽扩展门限，再将计算得到的噪声掩蔽扩展门限与人耳听觉的绝对门限比较，取两者最大的一个作为听觉掩蔽阈值。

5. 根据权利要求 4 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在

于，临界带宽个数的计算，

$$Z = 26.18f/(1960+f) - 0.53 \quad (5)$$

式中，Z为临界带宽个数，f为频率。

6. 根据权利要求 5 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，把每个临界带宽内语音信号的功率谱求和即可得到每个临界带宽的功率谱，设$P(k)$为信号快速傅里叶变换的功率谱，则每个临界带宽的功率谱为：

$$B_i = \sum_{k=l_i}^{h_i} P(k) \quad (6)$$

式中，l_i为临界带宽i的最小频率，h_i为临界带宽i的最大频率，i从 1 到i_{max}，语音信号的采样频率决定i_{max}。

7. 根据权利要求 5 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，将临界带宽功率谱转换为扩展临界带宽功率谱，用扩展函数SF_{ij}来估计不同临界带宽之间的掩蔽效应，如式(7)所示：

$$SF_{ij} = 15.81 + 7.5(i-j+0.747) - 17.5(1+(i-j+0.747)^2)^{\frac{1}{2}} \quad (7)$$

通过每个临界带宽的功率谱与扩展函数的卷积计算得到扩展临界带宽的功率谱，如式(8)所示：

$$C_i = SF_{ij} * B_j \quad (8)$$

其中，C_i为第i个扩展临界带宽的功率谱。

8. 根据权利要求 5 所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，确定带噪声语音信号是类似纯音还是类似噪声，其通过谱平坦测度(Spectral Flatness Measure, SFM)来确定：

$$SFM(i) = 10\log_{10}\frac{Gm_i}{Am_i} \quad (9)$$

其中，Gm_i是第i个扩展临界宽带功率的几何平均，Am_i是第i个扩展临界宽带功率的算术平均，

$$Gm_i = \left(\prod_{k=l_i}^{h_i} P(k)\right)^{\frac{1}{h_i-l_i+1}} \quad (10)$$

$$Am_i = \frac{1}{h_i-l_i+1}\left[\sum_{k=l_i}^{h_i} P(k)\right] \quad (11)$$

根据 SFM 来产生参数α，该参数表明功率谱为浊音的程度，即：

$$\alpha = \min\left(\frac{SFM(i)}{SFM_{max}}, 1\right) \quad (12)$$

其中$SFM_{max}=-60dB$，当$SFM(i)=0dB$时，$\alpha=0$表示完全是噪声；当$SFM(i)=-60dB$时，$\alpha=1$表示完全是纯音，实际的语音信号既不完全是噪声，也不完全是纯音，所以α介于0到1之间。

9. 根据权利要求8所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，

根据语音频谱的清音与浊音特性产生的门限偏移量表示式(13)所示：

$$O_i=\alpha_i\times(14.5+i)+(1-\alpha_i)\times5.5 \quad (13)$$

由此可得到扩展噪声掩蔽阈值为：

$$TSF_i=10^{\log_{10}(C_i)-(O_i/10)} \quad (14)$$

噪声的掩蔽阈值通过阈值归一化，并通过绝对听阈值进行比较得到，如式(15)所示：

$$T_i=\max\left\{T_{abs}(i),\frac{TSF_i}{\sum_{j=1}^{i}SF_j}\right\} \quad (15)$$

其中绝对听阈值$T_{abs}(i)$由非线性函数求得：

$$T_{abs}(i)=3.64\times(f/1000)^{0.8}-6.5e^{-0.6(f/1000-3.3)^2}+10^{-3}\times(f/1000)^4 \quad (16)。$$

10. 根据权利要求9所述的一种电动轮椅语音识别过程中的语音增强方法，其特征在于，所述第三步中自适应的谱减法系数计算，得到听觉掩蔽阈值后，通过下式计算得到自适应的谱减法系数：

$$\frac{T_{max}-T_i}{\alpha(k)-\alpha_{min}}=\frac{T_i-T_{min}}{\alpha_{max}-\alpha(k)} \quad (17)$$

$$\frac{T_{max}-T_i}{\beta(k)-\beta_{min}}=\frac{T_i-T_{min}}{\beta_{max}-\beta(k)} \quad (18)$$

其中T_i为Bark频率段的听觉掩蔽阈值，T_{max}和T_{min}是每一帧的听觉掩蔽阈值的最大值和最小值；

通过下式计算得到自适应的谱减系数：

$$\alpha(k)=\frac{(T_{max}-T_i)(\alpha_{max}-\alpha_{min})}{T_{max}-T_{min}}+\alpha_{min} \quad (19)$$

$$\beta(k)=\frac{(T_{max}-T_i)(\beta_{max}-\beta_{min})}{T_{max}-T_{min}}+\beta_{min} \quad (20)$$

其中$\alpha_{min}=1,\alpha_{max}=6,\beta_{min}=0$和$\beta_{max}=0.02$，其为通过大量实验得到的经验常数值；

式(1)中的增益函数如下算法得到：

$$G(k)=\begin{cases}(1-\alpha\left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma})^{1/\gamma} & \left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma}\le\frac{1}{\alpha+\beta}\\(\beta\left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma})^{1/\gamma}\end{cases} \quad (21)$$

其中 $\gamma=2$。

一种电动轮椅语音识别过程中的语音增强方法

技术领域

[0001] 本发明涉及语音控制领域，具体涉及到语音识别技术在电动轮椅中的应用。

背景技术

[0002] 随着社会的发展和人类文明程度的提高，对老年人、残疾人的服务需求会日益增加，他们需要运用现代高新技术来改善他们的生活质量和生活自由度。世界人口老龄化的进程正在加快，另外由于各种交通事故、天灾人祸和种种疾病，每年均有成千上万的人丧失一种或多种能力(如行走、动手能力等)，这种社会现实促进了无障碍技术在助老、助残方面的应用。因此，无障碍技术成为国内外科技人员研究的热点。

[0003] 无障碍技术是借助先进的科学技术为老年人和残障人士提供有效的辅助手段，使他们能够重新融入社会。人机交互技术是无障碍技术的重要研究内容之一。人机交互技术根据采用控制模式的不同可以分为两类：第一，通过硬件实施操作完成人机交互，如操作鼠标、键盘、操纵杆等。这种控制方式容易操作，但是并不适用于失去上肢或上肢存在缺陷的人群；第二，采用模式识别技术，利用人体自身的器官，如手、腕部、头部和脑电等完成人机交互。具体说来是通过语音识别、手势识别、头部运动、腕部运动、肌电信号和脑电信号(EEG)等来完成对电子设备的控制。这种人机交互方式具有非接触性，交互过程也比较直观，并且适用范围更广。因此，基于语音识别的无障碍技术具有潜在的研究价值和意义。

[0004] 语音识别研究是多门学科交叉的研究，语言交流是人与人之间最有效、最直接，也是最方便和快捷的交流方式，同时是人机交互的重要交流方式。语音识别作为人机交互的常用方法，具有巨大的优势。目前绝大多数的语音识别系统都只能在低噪声的环境下使用。虽然在安静环境下语音控制电动轮椅能达到比较好的效果，但是在噪声环境下不能达到预期的要求，因此鲁棒语音识别的研究是当前语音识别研究的前沿，具有重要的理论和实际应用价值，对于推动语音识别系统走向市场化应用具有重要的意义。

[0005] 电动轮椅作为一种代步工具，主要是为老年人和残障人士提供服务。它融合了多种技术，如自主导航、避障和人机交互等技术。传统意义上的电动轮椅是通过手动操纵杆来完成对运动的控制，但是并不适用于上肢不便的用户，因此应用的人群范围受到了限制。随着科技的迅猛发展，基于模式识别的新型控制技术已经在电动轮椅上得到了广泛应用，如手势、头部运动、肌电信号和基于脑电信号的BCI技术等。基于鲁棒语音识别技术的电动轮椅控制能给许许多多的残障人带来了方便，使他们能够参加正常的社交活动。电动轮椅作为人机交互的一个应用平台，融合了机器人研究领域的多种技术，具有十分广泛的应用前景。

发明内容

[0006] 本发明所要解决的技术问题针对现有语音识别在噪声环境下识别效果不理想的困境，提供一种基于语音增强的电动轮椅控制系统。本发明要用于电动轮椅语音控制系统中的语音识别，能够提高系统在噪声环境下的识别率，实现了对电动轮椅的精确控制，达到

用户与电动轮椅之间语音交互的目的。

[0007] 为了达到上述发明目的，本发明专利提供的技术方案如下：

一种电动轮椅语音识别过程中的语音增强方法，所述电动轮椅语音识别过程包括有输入语音信号预处理、特征参数提取、训练阶段中处理特征参数为每个命令词得到一个模块保存为模板库、识别阶段中先得到语音参数生成测试模板，该测试模板与模板库中的参考模板匹配，将匹配相似度最好的参考模块作为识别结果，其特征在于，所述的语音增强方法包括如下步骤：

第一步，对带噪声的输入语音信号进行预处理、分帧以及快速傅里叶变换算法，得到带噪声语音信号的频谱；

第二步，计算每一帧带噪声语音信号的先验信噪比，通过谱减法得到较为纯净的语音信号；

第三步，通过听觉掩蔽效应计算噪声掩蔽门限，得到噪声掩蔽门限后计算自适应的谱减系数；

第四步，进行第二次谱减，得到纯净语音信号的频谱；

第五步，通过逆快速傅里叶变换，得到增强后的语音信号。

[0008] 在计算先验信噪比时，根据谱减法可以用线性时变滤波器形式表示，即$|Y_k|$乘以增益函数G_k，如式（1）所示：

$$|\hat{S}_k|=G_k|Y_k| \quad 0\le G_k\le 1 \tag{1}$$

增益函数可以通过先验信噪比表示，如式（2）所示：

$$G(k,i)=\sqrt{\frac{SNR_{pri}(k,i)}{SNR_{post}(k,i)}} \tag{2}$$

式中的先验信噪比可以通过“直接判决”法进行估计得到，如式（3）所示：

$$SNR_{pri}(k,i)=\eta\frac{|\hat{S}(k,i-1)|^2}{|\hat{D}(k,i-1)|^2}+(1-\eta)\max[SNR_{post}(k,i)-1,0] \tag{3}$$

式中，$SNR_{post}(k,i)$为：

$$SNR_{post}(k,i)=1+SNR_{pri}(k,i) \tag{4}$$

式中i为当前帧，$i-1$为前一帧，$\hat{S}(k,i-1)$为前一帧的估计结果，η为调节系数，$0.8<\eta<1$。

[0009] 上述第三步计算基于人耳掩蔽效应的听觉掩蔽阈值时，需要先划定临界带宽，该临界带宽是当纯音信号的功率等于该噪声信号的功率时，纯音信号刚好被人耳听到的临界状态的频率宽度，该临界宽度为实验测得的常数。

[0010] 上述听觉掩蔽阈值的计算过程如下：先分别计算临界带宽的功率谱、扩展临界带宽的功率谱和噪声掩蔽扩展门限，再将计算得到的噪声掩蔽扩展门限与人耳听觉的绝对门限比较，取两者最大的一个作为听觉掩蔽阈值。

[0011] 在本发明电动轮椅语音识别过程中的语音增强方法中，临界带宽个数的计算，

$$Z=26.18f/(1960+f)-0.53 \tag{5}$$

式中，Z为临界带宽个数，f为频率。

[0012] 在本发明电动轮椅语音识别过程中的语音增强方法中，把每个临界带宽内语音

信号的功率谱求和即可得到每个临界带宽的功率谱，设$P(k)$为信号快速傅里叶变换的功率谱，则每个临界带宽的功率谱为：

$$B_i=\sum_{k=l_i}^{h_i}P(k) \tag{6}$$

式中，l_i为临界带宽i的最小频率，h_i为临界带宽i的最大频率，i从 1 到i_{max}，语音信号的采样频率决定i_{max}。

[0013] 由于各个临界带之间存在掩蔽效应，并且这种掩蔽效应随着临界带距离的增大而减弱。用一种扩展形式来表示临界带之间的相互掩蔽作用，所以将临界带宽功率谱转换为扩展临界带宽功率谱。将临界带宽功率谱转换为扩展临界带宽功率谱，用扩展函数SF_{ij}来估计不同临界带宽之间的掩蔽效应，如式(7)所示：

$$SF_{ij}=15.81+7.5(i-j+0.747)-17.5(1+(i-j+0.747)^2)^{\frac{1}{2}} \tag{7}$$

通过每个临界带宽的功率谱与扩展函数的卷积计算得到扩展临界带宽的功率谱，如式(8)所示：

$$C_i=SF_{ij}*B_j \tag{8}$$

其中，C_i为第i个扩展临界带宽的功率谱。

[0014] 研究表明存在两种噪声掩蔽阈值是纯音掩蔽噪声和噪声掩蔽纯音，在本发明电动轮椅语音识别过程中的语音增强方法中，确定带噪声语音信号是类似纯音还是类似噪声，其通过谱平坦测度(Spectral Flatness Measure, SFM)来确定：

$$SFM(i)=10\log_{10}\frac{Gm_i}{Am_i} \tag{9}$$

其中，Gm_i是第i个扩展临界宽带功率的几何平均，Am_i是第i个扩展临界宽带功率的算术平均，

$$Gm_i=\left(\prod_{k=l_i}^{h_i}P(k)\right)^{\frac{1}{h_i-l_i+1}} \tag{10}$$

$$Am_i=\frac{1}{h_i-l_i+1}\left[\sum_{k=l_i}^{h_i}P(k)\right] \tag{11}$$

根据 SFM 来产生参数α，该参数表明功率谱为浊音的程度，即：

$$\alpha=\min\left(\frac{SFM(i)}{SFM_{max}},1\right) \tag{12}$$

其中$SFM_{max}=-60dB$，当$SFM(i)=0dB$时，$\alpha=0$表示完全是噪声；当$SFM(i)=-60dB$时，$\alpha=1$表示完全是纯音，实际的语音信号既不完全是噪声，也不完全是纯音，所以α介于 0 到 1 之间。

[0015] 在本发明电动轮椅语音识别过程中的语音增强方法中，根据语音频谱的清音与浊音特性产生的门限偏移量表示式(13)所示：

$$O_i=\alpha_i\times(14.5+i)+(1-\alpha_i)\times5.5 \tag{13}$$

由此可得到扩展噪声掩蔽阈值为：

$$TSF_i=10^{\log_{10}(C_i)-(O_i/10)} \tag{14}$$

噪声的掩蔽阈值通过阈值归一化，并通过绝对听阈值进行比较得到，如式（15）所示：

$$T_i = \max\left\{T_{abs}(i), \frac{TSF_i}{\sum_{j=1}^{m} SF_j}\right\} \quad (15)$$

其中绝对听阈值$T_{abs}(i)$由非线性函数求得：

$$T_{abs}(i) = 3.64\times(f/1000)^{0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}\times(f/1000)^4 \quad (16)$$

在本发明电动轮椅语音识别过程中的语音增强方法中，所述第三步中自适应的谱减法系数计算，得到听觉掩蔽阈值后，通过下式计算得到自适应的谱减法系数：

$$\frac{T_{max}-T_i}{\alpha(k)-\alpha_{min}} = \frac{T_i-T_{min}}{\alpha_{max}-\alpha(k)} \quad (17)$$

$$\frac{T_{max}-T_i}{\beta(k)-\beta_{min}} = \frac{T_i-T_{min}}{\beta_{max}-\beta(k)} \quad (18)$$

其中T_i为 Bark 频率段的听觉掩蔽阈值，T_{max}和T_{min}是每一帧的听觉掩蔽阈值的最大值和最小值；

通过下式计算得到自适应的谱减系数：

$$\alpha(k) = \frac{(T_{max}-T_i)(\alpha_{max}-\alpha_{min})}{T_{max}-T_{min}} + \alpha_{min} \quad (19)$$

$$\beta(k) = \frac{(T_{max}-T_i)(\beta_{max}-\beta_{min})}{T_{max}-T_{min}} + \beta_{min} \quad (20)$$

其中$\alpha_{min}=1, \alpha_{max}=6, \beta_{min}=0$和$\beta_{max}=0.02$，其为通过大量实验得到的经验常数值；

式（1）中的增益函数如下算法得到：

$$G(k) = \begin{cases} (1-\alpha\left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma})^{1/\gamma} & \left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma} < \frac{1}{\alpha+\beta} \\ (\beta\left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma})^{1/\gamma} & \end{cases} \quad (21)$$

其中$\gamma=2$。

[0016] 基于上述技术方案，本发明专利在应用中取得了如下技术效果：

本发明电动轮椅语音识别过程中的语音增强方法在信号空间与特征空间提出相应的鲁棒语音识别算法，提高了系统在噪声环境下的识别率，用于电动轮椅语音控制系统中的语音识别，实现了对电动轮椅的精确控制，达到用户与电动轮椅之间语音交互的目的。

附图说明

[0017] 图 1 本发明电动轮椅语音识别过程中的语音增强方法中的语音识别原理框图。

[0018] 图 2 本发明电动轮椅语音识别过程中的语音增强方法中基于听觉掩蔽效应和先验信噪比的语音增强的原理图。

具体实施方式

[0019] 下面我们结合附图和具体的实施例来对本发明的基于语音增强的电动轮椅控制系统和控制方法做进一步的详细阐述，以求更为清楚明了地理解本发明的组成结构和工作原理，但不能以此来限制本发明的保护范围。

[0020] 本发明的目的是设计出一个电动轮椅语音识别控制系统，该系统包括：麦克风、基于微处理器操作平台、RS-232 串口通信模块、电动轮椅本体。麦克风类作为整个语音控制系统的输入；微处理器作为整个语音控制系统的上位机，用于处理语音信号；无线通信模块用于上位机和下位机之间的通信；电动轮椅是语音控制系统的下位机，用于完成相应的控制动作。

[0021] 电动轮椅语音识别的工作原理如图 1 所示，首先输入的语音信号要进行预处理，包括语音的预加重滤波、加窗分帧处理与双门限端点检测；其次，语音信号经过预处理后，接下来很重要的一环就是特征参数的提取；然后，在训练阶段，将特征参数进行一定的处理后，为每个命令词得到一个模板，保存为模板库；在识别阶段，语音信号经过相同的方法得到语音参数，生成测试模板，与模板库中的参考模板进行匹配，将匹配相似度最高的参考模板作为识别结果。

[0022] 如图 2 所示，在对语音进行识别时，首先对带噪声的语音信号进行预处理、分帧以及快速傅里叶变换（FFT）算法；然后，计算每一帧带噪声语音信号的先验信噪比通过谱减法得到较为纯净的语音信号；再次，通过听觉掩蔽效应计算噪声掩蔽门限，得到噪声掩蔽门限后计算自适应的谱减系数，然后进行第二次谱减得到纯净语音信号的频谱；最后通过逆快速傅里叶变换（IFFT）得到增强后的语音信号。

[0023] 在本发明的语音增强方法中，具体来说，上述语音增强过程的算法包括如下内容：

设定$s(t)$为纯净语音信号，$n(t)$为噪声信号，$y(t)$为带噪声的语音信号。那么，带噪声的语音信号可表示为：$y(t)=s(t)+n(t)$，经过快速傅里叶变换，结果为：$Y_k=S_k+N_k$，其功率谱密度为：$|Y_k|^2=|S_k|^2+|N_k|^2+S_kN_k^*+S_k^*N_k$，

由于语音信号与噪声是相互独立的，并且N_k为零均值的高斯分布，所以其可简化为：$E[|Y_k|^2]=E[|S_k|^2]+E[|N_k|^2]$，

每一帧的语音具有短时平稳性，其可表示为：$|Y_k|^2=|S_k|^2+\lambda_n(k)$，

$\lambda_n(k)$为无语音信号时$|N_k|^2$的统计平均值，由此可得到原始语音信号的估计：$|\hat{S}_k|=[|Y_k|^2-|N_k|^2]^{\frac{1}{2}}=[|Y_k|^2-\lambda_n(k)]^{\frac{1}{2}}$，将其写成增益形式为：

$|\hat{S}_k|=G_k|Y_k|$，$G_k=(1-1/\gamma_k)^{1/2}$，其中γ_k为后验信噪比，其如下式所示：$\gamma_k=|Y_k|^2/\lambda_n(k)$。

[0024] 1. 计算先验信噪比，根据谱减法可以用线性时变滤波器形式表示，即$|Y_k|$乘以增益函数G_k，如式（1）所示：

$$|\hat{S}_k|=G_k|Y_k| \quad 0\le G_k\le 1 \tag{1}$$

增益函数可以通过先验信噪比表示，如式（2）所示：

$$G(k,i)=\sqrt{\frac{SNR_{prio}(k,i)}{SNR_{post}(k,i)}} \tag{2}$$

式中的先验信噪比可以通过“Decision-Directed”法进行估计得到。如式（3）所示。

[0025] $$SNR_{prio}(k,i)=\eta\frac{|\hat{S}(k,i-1)|^2}{|\hat{D}(k,i-1)|^2}+(1-\eta)\max[SNR_{post}(k,i)-1,0] \quad (3)$$

式中，$SNR_{post}(k,i)$为：

$$SNR_{post}(k,i)=1+SNR_{prio}(k,i) \quad (4)$$

式中i为当前帧，$i-1$为前一帧，$\hat{S}(k,i-1)$为前一帧的估计结果，η为调节系数，一般在0.8~1之间。

[0026] 2. 人耳听觉掩蔽效应及掩蔽阈值计算，人耳听觉掩蔽效应是在一个很强的声音信号附近，弱的声音信号将不会被人耳听到，或者是声音B的出现会使声音A能被人耳听到的听觉阈值增大的现象。因此，A是被掩蔽声，B是掩蔽声。因为在耳蜗基底膜上，频率高的信号的传递距离小于频率低的信号传递的距离，所以，低频信号容易掩蔽高频信号。

[0027] 纯音信号被人耳感知受到噪声信号的影响，即噪声信号掩蔽纯音信号。达到更好地显现人耳听觉掩蔽效应，需要划定临界带宽。具有一定频带宽度以纯音信号的中心频率的连续噪声信号能掩蔽该纯音信号。临界带宽是当纯音信号的功率等于该噪声信号的功率时，纯音信号刚好被人耳听到的临界状态的频率宽度，其可以通过实验测得。在临界带宽的定义中，需要定义一个特殊的心理声学单位—Bark。1Bark 对应一个临界带宽的宽度。表 1 给出了通过实验得出的 24 个临界带宽的中心频率、编号和频率范围。

[0028] 听觉掩蔽阈值的计算需要分别计算临界带宽的功率谱、扩展临界带宽的功率谱和噪声掩蔽扩展门限。最后将计算得到的噪声掩蔽扩展门限与人耳听觉的绝对门限比较，取其中两者最大的一个作为听觉掩蔽阈值。

[0029] 根据式(5)可以计算语音信号的临界带宽个数。

[0030] $$Z=26.18f/(1960+f)-0.53 \quad (5)$$

式中，Z为临界带宽个数，f为频率。

[0031] 把每个临界带宽内语音信号的功率谱求和即可得到每个临界带宽的功率谱。设$P(k)$为信号快速傅里叶变换的功率谱，则每个临界带宽的功率谱为：

$$B_i=\sum_{k=1}^{A}P(k) \quad (6)$$

表 1 临界带宽划分表

编号（Bark）	中心频率（Hz）	频率范围（Hz）	带宽（Hz）	编号（Bark）	中心频率（Hz）	频率范围（Hz）	带宽（Hz）
1	50	20-100	80	13	1850	1720-2000	280
2	150	100-200	100	14	2150	2000-2320	320
3	250	200-300	100	15	2500	2320-2700	380
4	350	300-400	100	16	2900	2700-3150	450
5	450	400-510	110	17	3400	3150-3700	550
6	570	510-630	120	18	4000	3700-4400	700
7	700	630-770	140	19	4800	4400-5300	900
8	840	770-920	150	20	5800	5300-6400	1100
9	1000	920-1080	160	21	7000	6400-7700	1300
10	1170	1080-1270	190	22	8500	7700-9500	1800
11	1370	1270-1480	210	23	10500	9500-12000	2500
12	1600	1480-1720	240	24	13500	12000-15500	3500

上述公式（6）中，l_i为临界带宽i的最小频率，h_i为临界带宽i的最大频率，i从 1 到i_{max}，语音信号的采样频率决定i_{max}。

[0032] 由于各个临界带之间存在掩蔽效应，并且这种掩蔽效应随着临界带距离的增大而减弱。用一种扩展形式来表示临界带之间的相互掩蔽作用，所以将临界带宽功率谱转换为扩展临界带宽功率谱。

[0033] 用扩展函数SF_{ij}来估计不同临界带宽之间的掩蔽效应，如式（7）所示。

[0034] $$SF_{ij}=15.81+7.5(i-j+0.747)-17.5(1+(i-j+0.747)^2)^{\frac{1}{2}} \quad (7)$$

通过每个临界带宽的功率谱与扩展函数的卷积计算得到扩展临界带宽的功率谱，如式（8）所示。

[0035] $$C_i=SF_{ij}*B_j \quad (8)$$

其中，C_i为第i个扩展临界带宽的功率谱。

[0036] 研究表明存在两种噪声掩蔽阈值是纯音掩蔽噪声和噪声掩蔽纯音。为了确定带噪声语音信号是类似纯音还是类似噪声，其通过谱平坦测度（Spectral Flatness Measure，SFM）来确定。

[0037] $$SFM(i)=10\log_{10}\frac{Gm_i}{Am_i} \quad (9)$$

其中，Gm_i是第i个扩展临界宽带功率的几何平均，Am_i是第i个扩展临界宽带功率的算术平均。

[0038] $$Gm_i=\left(\prod_{k=l_i}^{h_i}P(k)\right)^{\frac{1}{h_i-l_i+1}} \quad (10)$$

$$Am_i=\frac{1}{h_i-l_i+1}\left[\sum_{k=l_i}^{h_i}P(k)\right] \quad (11)$$

根据 SFM 来产生参数α，该参数表明功率谱为浊音的程度，即：

$$\alpha=\min\left(\frac{SFM(i)}{SFM_{max}},1\right) \quad (12)$$

其中$SFM_{max}=-60dB$。当$SFM(i)=0dB$时，$\alpha=0$表示完全是噪声；当$SFM(i)=-60dB$时，$\alpha=1$表示完全是纯音。实际的语音信号既不完全是噪声，也不完全是纯音，所以α介于 0 到 1 之间。

[0039] 根据语音频谱的清音与浊音特性产生的门限偏移量表示式（13）所示：

$$O_i=\alpha_i\times(14.5+i)+(1-\alpha_i)\times5.5 \quad (13)$$

由此可得到扩展噪声掩蔽阈值为：

$$TSF_i=10^{\log_{10}(E_i)-(O_i/10)} \quad (14)$$

噪声的掩蔽阈值通过阈值归一化，并通过绝对听阈值进行比较得到，如式（15）所示。

[0040] $$T_i=\max\left\{T_{abs}(i),\frac{TSF_i}{\sum_{j=i_{low}}^{i_{high}}SF_{ij}}\right\} \quad (15)$$

其中绝对听阈值$T_{abs}(i)$由非线性函数求得：

$$T_{abs}(i)=3.64\times(f/1000)^{0.8}-6.5e^{-0.6(f/1000-3.3)^2}+10^{-3}\times(f/1000)^4 \quad (16),$$

噪声掩蔽扩展门限与人耳听觉的绝对门限比较，取其中两者最大的一个作为听觉掩蔽阈值。

[0041] 3. 自适应的谱减法系数计算，得到听觉掩蔽阈值后，通过下式计算得到自适应的谱减法系数。

[0042] $$\frac{T_{max}-T_i}{\alpha(k)-\alpha_{min}}=\frac{T_i-T_{min}}{\alpha_{max}-\alpha(k)} \quad (17)$$

$$\frac{T_{max}-T_i}{\beta(k)-\beta_{min}}=\frac{T_i-T_{min}}{\beta_{max}-\beta(k)} \quad (18)$$

其中T_i为 Bark 频率段的听觉掩蔽阈值，T_{max}和T_{min}是每一帧的听觉掩蔽阈值的最大值和最小值。

[0043] 通过计算得到自适应的谱减系数，如下式：

$$\alpha(k)=\frac{(T_{max}-T_i)(\alpha_{max}-\alpha_{min})}{T_{max}-T_{min}}+\alpha_{min} \quad (19)$$

$$\beta(k)=\frac{(T_{max}-T_i)(\beta_{max}-\beta_{min})}{T_{max}-T_{min}}+\beta_{min} \quad (20)$$

其中$\alpha_{min}=1,\alpha_{max}=6,\beta_{min}=0$和$\beta_{max}=0.02$，通过大量实验得到的经验值。利用上述得到的谱减系数求得增益函数，代入式（1）中，将上一次谱减之后得到的功率谱与增益相乘，

得到纯净的语音信号。

[0044] 最后，通过逆快速傅里叶变换（IFFT）得到增强后的语音信号。

[0045] 上述式（1）中的增益函数可以通过 Berouti 等人提出的算法得到，如下式所示。

[0046] $$G(k)=\begin{cases}\left(1-\alpha\left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma}\right)^{1/\gamma} & \left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma}<\frac{1}{\alpha+\beta}\\ \left(\beta\left[\frac{|D(k)|}{|Y(k)|}\right]^{\gamma}\right)^{1/\gamma}\end{cases} \quad (21)$$

其中$\gamma=2$。

[0047] 上述语音信号的语音增强，谱减法的改进有很多种，如自适应和听觉掩蔽等，而本发明的语音增强将先验信噪比和听觉掩蔽结合的方法是之前没有运用的语音增强算法，由此通过本发明的语音增强算法对语音信号进行增强处理。基于改进感知非均匀谱压缩的特征提取算法是将每一帧语音信号的功率谱或者子带能量进行压缩，对其压缩因子进行改进，然后通过相应的算法提取语音信号的特征值。

[0048] 在电动轮椅鲁棒语音识别方法在操作中，先由麦克风获取语音信号，对语音信号进行预处理后，对其进行语音增强和特征提取，在信号空间，运用听觉掩蔽效应和先验信噪比相结合的语音增强算法进行增强；在特征空间，运用改进感知非均匀谱压缩的特征提取算法更加准确提取语音信号的特征值，用动态时间规整算法训练和识别语音信号，将语音识别结果转换成电动轮椅的底层驱动指令，调用相应的控制函数，驱动电动轮椅按照语音信号进行运动。

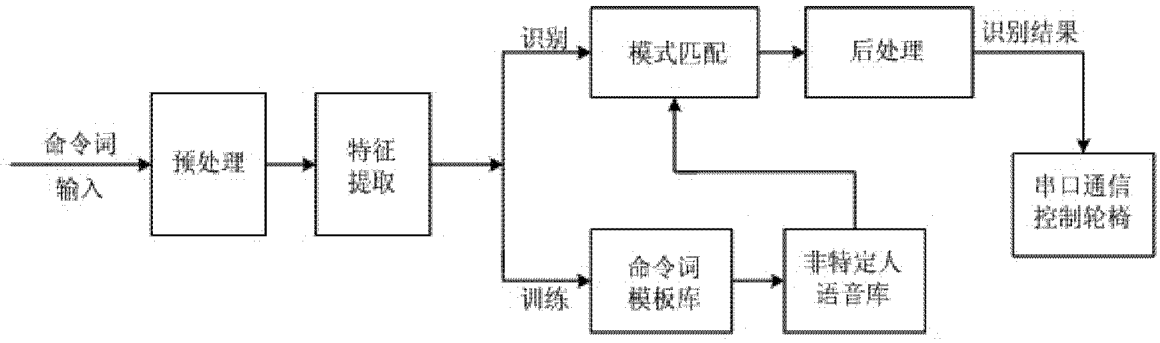

命令词
输入
预处理
特征
提取
识别
模式匹配
后处理
识别结果
串口通信
控制轮椅
训练
命令词
模板库
非特定人
语音库

图 1

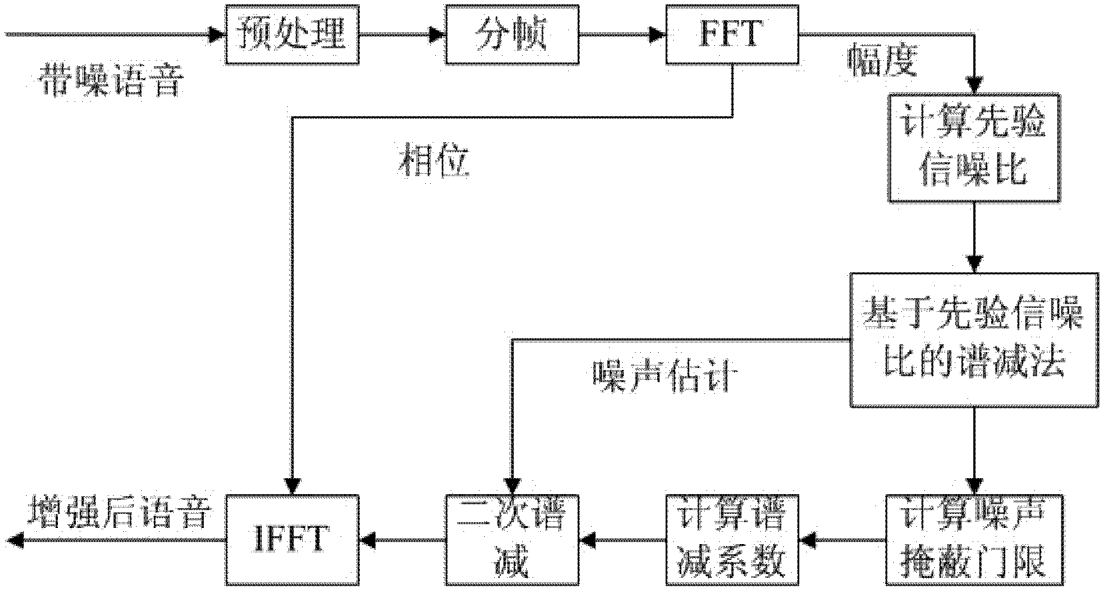

带噪语音
预处理
分帧
FFT
幅度
相位
计算先验
信噪比
基于先验信噪
比的谱减法
噪声估计
增强后语音
IFFT
二次谱
减
计算谱
减系数
计算噪声
掩蔽门限

图 2