(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 108564956 A
(43)申请公布日 2018.09.21

(21)申请号 201810253151.6

(22)申请日 2018.03.26

(71)申请人 京北方信息技术股份有限公司
地址 100089 北京市海淀区西三环北路25号青政大厦7层

(72)发明人 冉承祥 高昊江 杨飞

(74)专利代理机构 北京品源专利代理有限公司 11332
代理人 孟金喆

(51)Int.Cl.
G10L 17/02(2013.01)
G10L 25/24(2013.01)

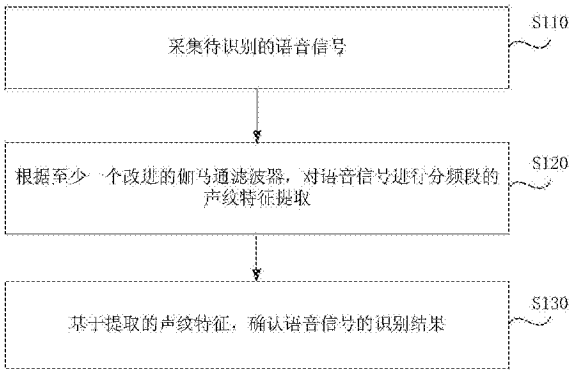
权利要求书2页 说明书12页 附图5页

(54)发明名称

一种声纹识别方法和装置、服务器、存储介质

(57)摘要

本发明实施例公开了一种声纹识别方法和装置、服务器、存储介质，其中，该方法包括：采集待识别的语音信号；根据至少一个改进的伽马通滤波器，对语音信号进行分频段的声纹特征提取；基于提取的声纹特征，确认语音信号的识别结果。本发明实施例可以解决现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题，可以提高滤波器对语音高频部分的分辨率，提高声纹特征提取的准确性，进而提高对语音的高频部分的识别效果，并且可以降低声纹识别涉及的运算复杂度以及响应时间。



CN 108564956 A

1.一种声纹识别方法，其特征在于，包括：

采集待识别的语音信号；

根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取；

基于提取的声纹特征，确认所述语音信号的识别结果。

2.根据权利要求1所述的方法，其特征在于，所述根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取，包括：

根据所述至少一个改进的伽马通滤波器，利用以下公式得到的滤波器的频率响应对所述语音信号的能量谱进行滤波，得到所述语音信号的滤波能量谱，

$$\hat{G}_i(k)=\begin{cases} G_i(k),\ 0<k\le \mathrm{NFFT}/4 \\ G_i(\mathrm{NFFT}/2+1-k),\ \mathrm{NFFT}/4<k\le \mathrm{NFFT}/2 \\ G_i(k-\mathrm{NFFT}/2),\ \mathrm{NFFT}/2<k\le 3\mathrm{NFFT}/4 \\ G_i(k),\ 3\mathrm{NFFT}/4<k\le \mathrm{NFFT} \end{cases},$$

其中，$\hat{G}_i(k)$表示第i个改进的伽马通滤波器的频率响应，$G_i(k)$表示标准的伽马通滤波器的频率响应，k表示所述语音信号的周期频谱上的第k个采样点，NFFT表示采样点数；

根据所述滤波能量谱得到所述语音信号的声纹特征。

3.根据权利要求2所述的方法，其特征在于，所述根据所述滤波能量谱得到所述语音信号的声纹特征包括：

对所述滤波能量谱取对数，得到对数频谱；

对所述对数频谱做离散余弦变换，得到混合耳蜗频率倒谱系数；

利用所述语音信号的平均短时对数能量替换所述混合耳蜗频率倒谱系数中的零阶系数，得到所述语音信号的声纹特征。

4.根据权利要求1所述的方法，其特征在于，在所述根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取之前，所述方法还包括：

对所述语音信号进行快速傅里叶变换，并进行归一化处理；

对所述归一化处理之后得到的频谱取平方，得到所述语音信号的能量谱，以便进行所述声纹特征提取。

5.根据权利要求1所述的方法，其特征在于，在所述采集待识别的语音信号之后，所述方法还包括：

采用改进的自扰动最小二乘法对所述语音信号进行自适应地语音加强。

6.根据权利要求1所述的方法，其特征在于，所述基于提取的声纹特征，确认所述语音信号的识别结果，包括：

基于所述提取的声纹特征，利用预先训练好的隐马尔可夫模型，得到所述语音信号的识别结果，其中，所述隐马尔可夫模型是基于训练语音的声纹特征训练得到，所述训练语音的声纹特征是根据所述至少一个改进的伽马通滤波器，进行分频段的声纹特征提取后得到。

7.根据权利要求6所述的方法，其特征在于，所述识别结果包括所述语音信号的说话人信息。

8.一种声纹识别装置，其特征在于，包括：

声纹采集模块，用于采集待识别的语音信号；

声纹特征提取模块，用于根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取；

声纹识别模块，用于基于提取的声纹特征，确认所述语音信号的识别结果。

9.一种服务器，其特征在于，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1～7中任一所述的声纹识别方法。

10.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求1～7中任一所述的声纹识别方法。

一种声纹识别方法和装置、服务器、存储介质

技术领域

[0001] 本发明实施例涉及声音识别技术领域，尤其涉及一种声纹识别方法和装置、服务器、存储介质。

背景技术

[0002] 随着用户的安全防范意识不断增加，越来越多的识别方式考虑以用户生理特征作为身份识别特征。声纹识别是一项重要且方便的识别方式。声纹识别涉及的语音特征的常用表征包括线性预测倒谱系数(Linear prediction cestrum coefficient，LPCC)、梅尔频率倒谱系数(Mel frequency cestrum coefficient，MFCC)和耳蜗频率倒谱系数(Gammatone frequency cestrum coefficient，GFCC)。

[0003] LPCC特征提取主要基于线性预测原理，认为语音采样点可由过去一段时间的语音采样线性组合来预测，可以表征一定的语音相关信息。考虑到人耳听到的声音与频率并不是线性的关系，基于LPCC特征提取的声纹识别效果往往较差。MFCC特征提取主要基于梅尔刻度，其设计模仿人的听觉，利用了人的听觉对频率的非线性感应，人耳听觉系统比任何自动识别系统更具有可靠性与便捷性，因此，MFCC特征提取是目前主流的声纹特征提取方法，该方法具有一定噪声鲁棒性。MFCC特征提取主要描述声道特征，其中蕴含的语义信息往往要强于说话人信息，在无噪声或低噪声下能作为较好的特征表达，其性能优于LPCC特征提取。然而，对于背景噪声较大，信道失真严重的语音，基于MFCC特征提取的声纹识别的抗噪能力还不够好，影响识别效果。GFCC特征提取模拟了人耳耳蜗听觉模型，利用了人耳对噪声的抗噪能力，在声纹识别方面具有较为不错的表现，鲁棒性更加优异。

[0004] 但是，通过观察伽马通(Gammatone)滤波器组的频率响应发现，滤波器组在高频部分频率分辨率较低，即基于GFCC特征提取的声纹识别没有完全利用语音的高频部分所含的语音信息，导致语音识别效果较差。

发明内容

[0005] 本发明实施例提供一种声纹识别方法和装置、服务器、存储介质，以解决现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题。

[0006] 第一方面，本发明实施例提供了一种声纹识别方法，该方法包括：

[0007] 采集待识别的语音信号；

[0008] 根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取；

[0009] 基于提取的声纹特征，确认所述语音信号的识别结果。

[0010] 第二方面，本发明实施例还提供了一种声纹识别装置，该装置包括：

[0011] 声纹采集模块，用于采集待识别的语音信号；

[0012] 声纹特征提取模块，用于根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取；

[0013] 声纹识别模块，用于基于提取的声纹特征，确认所述语音信号的识别结果。

[0014] 第三方面，本发明实施例还提供了一种服务器，包括：

[0015] 一个或多个处理器；

[0016] 存储装置，用于存储一个或多个程序，

[0017] 当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如本发明任一实施例所述的声纹识别方法。

[0018] 第四方面，本发明实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本发明任一实施例所述的声纹识别方法。

[0019] 本发明实施例通过根据至少一个改进的伽马通滤波器，对采集的待识别语音信号进行分频段的声纹特征提取，基于提取的声纹特征，确认语音信号的识别结果，解决了现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题，提高了滤波器对语音高频部分的分辨率，提高了声纹特征提取的准确性，进而提高了对语音的高频部分的识别准确性，并且降低了声纹识别涉及的运算复杂度以及响应时间。

附图说明

[0020] 图1是本发明实施例一提供的声纹识别方法的流程图；

[0021] 图2是本发明实施例二提供的声纹识别方法的流程图；

[0022] 图3是本发明实施例三提供的声纹识别方法的流程图；

[0023] 图4是本发明实施例四提供的声纹识别装置的结构示意图；

[0024] 图5是本发明实施例五提供的一种服务器的结构示意图。

具体实施方式

[0025] 下面结合附图和实施例对本发明作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释本发明，而非对本发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本发明相关的部分而非全部结构。

[0026] 实施例一

[0027] 图1是本发明实施例一提供的声纹识别方法的流程图，本实施例可适用于进行声纹识别的情况，该方法可以由声纹识别装置来执行，该装置可以采用软件和/或硬件的方式实现，并可集成在服务器上。如图1所示，该方法具体包括：

[0028] S110、采集待识别的语音信号。

[0029] 可以利用麦克风等具有声音采集功能的装置采集待识别的语音信号，并将其存储在任意的具有存储功能的装置中，以等待识别，也可以将采集的语音信号直接输出到进行声纹识别的装置中进行下一步处理。语音信号在采集过程中会进行相应的采样和量化处理，例如，用麦克风采集之后的语音信号便是已经过采样和量化处理的信号。

[0030] S120、根据至少一个改进的伽马通滤波器，对语音信号进行分频段的声纹特征提取。

[0031] 每一段语音信号均对应一定的能量谱，其高频部分和低频部分的频谱响应具有不同特点。伽马通滤波器的原理基于对人耳耳蜗的模拟，而人耳对于高频部分并不敏感，因此，标准的伽马通滤波器对于语音高频部分的分辨率并不高，滤波效果不够明显，导致语音

的高频信息丢失。利用改进的伽马通滤波器对语音信号分别进行高频和低频的声纹特征提取，相当于对语音高频部分和低频部分进行差异化滤波，这相比于标准的伽马通滤波器的滤波效果，可以避免语音高频信息的丢失，保证声纹特征提取结果的准确性，为准确的声纹识别奠定基础。其中，涉及的改进伽马通滤波器的具体数量，可以根据声纹特征提取需求进行设置。

[0032] 可选的，根据至少一个改进的伽马通滤波器，对语音信号进行分频段的声纹特征提取，包括：

[0033] 根据至少一个改进的伽马通滤波器，利用以下公式得到的滤波器的频率响应对语音信号的能量谱进行滤波，得到语音信号的滤波能量谱，

[0034] $$\hat{G}_i(k)=\begin{cases} G_i(k),\ 0<k\le \mathrm{NFFT}/4 \\ G_i(\mathrm{NFFT}/2+1-k),\ \mathrm{NFFT}/4<k\le \mathrm{NFFT}/2 \\ G_i(k-\mathrm{NFFT}/2),\ \mathrm{NFFT}/2<k\le 3\mathrm{NFFT}/4 \\ G_i(k),\ 3\mathrm{NFFT}/4<k\le \mathrm{NFFT} \end{cases},$$

[0035] 其中，$\hat{G}_i(k)$表示第i个改进的伽马通滤波器的频率响应，$G_i(k)$表示标准的伽马通滤波器的频率响应，k表示语音信号的周期频谱上的第k个采样点，NFFT表示采样点数；

[0036] 根据滤波能量谱得到语音信号的声纹特征。

[0037] 其中，k的取值范围0～NFFT，采样点数NFFT的取值与采样信号的采样频率f_s取值存在对应关系，将采样点数NFFT进行四等分即将采样频谱f_s进行四等分。示例性的，采样点区间(0,NFFT/4)对应频率区间(0,f_s/4)。由于能量频谱的轴对称特性，频率区间(0,f_s/4)与(3f_s/4,f_s)对应低频的语音信号的频率响应，频率区间(f_s/4,f_s/2)与(f_s/2,3f_s/4)对应高频的语音信号的频率响应。

[0038] 由以上公式可以看出，改进后的伽马通滤波器相当于混合滤波器，由标准的伽马通滤波器关于奈奎斯特频率的一半翻转得到。低频语音信号的滤波能量谱依然是标准的伽马通滤波器的频率响应，即充分利用了标准的伽马通滤波器对低频部分的高分辨率优势；而高频语音信号的滤波能量谱是在标准的伽马通滤波器的基础上进行变换得到，改进后得到的滤波频谱更加密集，即改进后的伽马通滤波器对应的高频滤波效果具有明显的提高，对语音高频部分的敏感度以及分辨率更高，声纹特征提取的准确性更高。示例性的，对于采样频率为16k的一段语音进行识别，滤波器通道数设置为8，改进后的伽马通滤波器的频率响应在4000～8000Hz的高频段的频谱明显要密集很多，而标准的伽马通滤波器在该频段的频率则相对稀疏，所以改进后的伽马通滤波器对语音高频部分的频率分辨率得到有效提高。

[0039] 换言之，对于达到相同的滤波效果，改进后的伽马通滤波器组所需要的通道数要少于标准的伽马通滤波器组所需的通道数，降低了声纹识别涉及的运算复杂度以及响应时间。示例性的，对一段语音进行识别，要达到较好的识别效果，标准的伽马通滤波器组所需的通道数为16，而对于改进后的伽马通滤波器组所需要的通道数设置为8即可。

[0040] S130、基于提取的声纹特征，确认语音信号的识别结果。

[0041] 基于得到的声纹特征，通过在云数据库或者本地语音数据库中进行特征的识别与匹配，便可以确认识别结果。

[0042] 在上述技术方案的基础上，可选的，根据滤波能量谱得到语音信号的声纹特征包

括：

[0043] 对滤波能量谱取对数，得到对数频谱；

[0044] 对得到的对数频谱做离散余弦变换(Discrete Cosine Transform,DCT)，得到混合耳蜗频率倒谱系数(Mix Gammatone frequency cestrum coefficient,Mix GFCC)；

[0045] 利用语音信号的平均短时对数能量替换混合耳蜗频率倒谱系数中的零阶系数，得到语音信号的声纹特征。

[0046] 其中，经过离散余弦变换处理，可以去除各个滤波器之间输出的相关性，去除一些不相关的量，保留关键的频谱特征信息。混合耳蜗频率倒谱系数中的零阶系数代表语音信号中的直流分量，对应的是语音信号中声音大与小的能量，直流分量的存在会影响声纹识别的准确性，因此需要将其舍去，同时考虑到语音信号的短时能量带有一定的语音信息，用其代替语音信号的直流分量，构成新的特征，避免了语音信息的丢失。零阶系数替换之后得到的混合耳蜗频率倒谱系数即可用来准确地表征待识别的语音信号的声纹特征。

[0047] 本实施例的技术方案通过根据至少一个改进的伽马通滤波器，对采集的待识别语音信号进行分频段的差异化声纹特征提取，然后基于提取的声纹特征，确认语音信号的识别结果，解决了现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题，提高了滤波器对语音高频部分的分辨率，提高了声纹特征提取的准确性，进而提高了对语音的高频部分的识别准确性，并且降低了声纹识别涉及的运算复杂度以及响应时间。此外，利用语音信号的平均短时对数能量替换语音信号中的直流分量，避免了语音信息的丢失，保证了语音识别结果的准确。

[0048] 实施例二

[0049] 图2是本发明实施例二提供的声纹识别方法的流程图，本实施例是在上述实施例的基础上进一步进行优化。如图2所示，该方法包括：

[0050] S210、采集待识别的语音信号。

[0051] S220、采用改进的自扰动最小二乘法对语音信号进行自适应地语音加强。

[0052] 在采集待识别的语音信号之后，采用改进的自扰动最小二乘法(Improved Self-perturbing Recursive Least Squares,ISPRLS)对语音信号进行自适应地语音加强，可以同时达到语音加强和有效消除语音信号背景噪声的目的，提高语音信号的信噪比，为后续准确地提取声纹特征奠定基础。

[0053] S230、对语音加强处理后的语音信号进行快速傅里叶变换，并进行归一化处理。

[0054] 经过快速傅里叶变换，可以将语音信号由时域空间信号转换为频域空间信号，得到的频谱具有周期性和对称性的特点，这在信号的分析处理过程中是非常有用的。

[0055] S240、对归一化处理之后得到的频谱取平方，得到语音信号的能量谱，以便进行声纹特征提取。

[0056] S250、根据至少一个改进的伽马通滤波器，基于语音信号的能量谱，对语音信号进行分频段的声纹特征提取。

[0057] S260、基于提取的声纹特征，确认语音信号的识别结果。

[0058] 可选的，基于提取的声纹特征，确认语音信号的识别结果，包括：

[0059] 基于提取的声纹特征，利用预先训练好的隐马尔可夫模型(Hidden Markov Model,HMM)，得到语音信号的识别结果，其中，隐马尔可夫模型是基于训练语音的声纹特

征，利用BaumWelch算法训练得到，训练语音的声纹特征是根据至少一个改进的伽马通滤波器，进行分频段的声纹特征提取后得到。

[0060] 训练语音可以是从云数据库或者本地语音数据库中抽取的语音信号，也可以是定期采集的训练人员的语音信号。根据上述基于改进的伽马通滤波器的声纹提取方法，提取出训练语音的声纹特征，然后进行语音识别模型的训练。从识别模型的训练到语音信号的识别，所使用的语音特征，均是利用基于改进的伽马通滤波器提取的声纹特征，保证了整个声纹识别过程的方法一致性，减少了声纹识别的误差。此外，语音识别模型并不限于隐马尔可夫模型，隐马尔可夫模型是众多模型中比较简单且易实现的一种统计模型，有利于降低语音识别过程的复杂度，可以简化识别过程但并不影响识别结果。

[0061] 进一步的，语音信号的识别结果包括语音信号的说话人信息。

[0062] 确定出说话人信息即确定出该语音信号对应的说话人。具体的，利用预先训练好的隐马尔可夫模型进行语音识别，所涉及的算法是维特比(Viterbi)算法，在云数据库或者本地语音数据库中进行特征的识别与匹配，得到的概率最大的说话人对应于语音信号的说话人，即待识别的语音是该说话人发出的语音。需要说明的是，训练模型之前，需要首先采集训练语音，同时记录每一段训练语音的说话人信息，判别确定语音信号说话人的过程，也是在已记录的说话人中进行判别，即待识别语音的说话人可能是已采集的训练语音对应的说话人之一。如果根据维特比算法得出的待识别的语音属于训练模型的语音库的某个说话人的概率最大值，低于预设阈值，则表示待识别的语音不属于训练模型的语音库里的所有说话人，其中预设阈值可以根据识别需求进行适应性设置。此外，语音信号的具体语义信息可以通过基于改进的伽马通滤波器提取的训练语音中的声韵母、音素和音节等语音识别基元的语音特征预先训练好的隐马尔可夫模型对语音信号进行识别得到。

[0063] 示例性的，为了进一步验证改进的伽马通滤波器的有效性，将基于混合耳蜗频率倒谱系数(Mix GFCC)和传统的基于梅尔频率倒谱系数(MFCC)与标准耳蜗频率倒谱系数(GFCC)的声纹识别结果进行对比。选取由德州仪器(TI)、麻省理工学院(MIT)和坦福研究院(SRI)合作构建的声学-音素连续语音语料库(The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus,TIMIT)中10个人进行测试，其中每个人10段语音，每段语音4s～10s，语音格式为16kHz采样率，16比特(bit)的量化位数，wav格式，分别从音素紧凑句子(Phonetically compact sentences,SX)、方言句子(Dialect sentences,SA)和音素发散句子(Phonetically diverse sentences,SI)中选一段作为待识别的语音信号，另外的7段用于语音识别模型的预先训练，最后统计识别的准确率。识别模型采用为经典的高斯混合模型(Gaussian mixture model,GMM)-隐马尔可夫模型(HMM)，高斯混合数设置为64，隐马尔可夫模型状态数设置为6。结果表明，基于标准耳蜗频率倒谱系数的识别准确率为82.7%，要优于基于梅尔频率倒谱系数的识别准确率75.5%，基于混合耳蜗频率倒谱系数的识别准确率为85.8%，又高于标准耳蜗频率倒谱系数的识别准确率，这表明本实施例改进的伽马通滤波器有效地提高了声纹识别的识别准确率。

[0064] 本实施例的技术方案通过对采集的待识别语音信号进行自适应地语音加强，提高语音信号的信噪比，然后得到语音信号的能量谱，根据至少一个改进的伽马通滤波器，进行分频段的声纹特征提取，最后基于提取的声纹特征，确认语音信号的识别结果，解决了现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题，提高了滤波器对语音高

频部分的分辨率，提高了对语音的高频部分的识别准确率，并且降低了声纹识别涉及的运算复杂度。此外，利用基于改进的伽马通滤波器提取的声韵母、音速和音节等语音基元的语音特征预先训练好的隐马尔可夫模型进行语音识别，可以得到较好的语音识别效果。

[0065] 实施例三

[0066] 图3是本发明实施例三提供的声纹识别方法的流程图，本实施例是在上述实施例的基础上进一步进行优化。如图3所示，该方法包括：

[0067] S310、采集待识别的语音信号，并进行包括预加重、分帧、加窗、端点检测和自适应语音加强的预处理。

[0068] 采集的语音信号是原始语音信号，对原始语音信号预加重处理，提升其高频部分以使得整个频谱平坦，再进行语音分帧使得可以进行短时平稳处理，接着加汉明窗避免遗漏帧与帧之间的信息，然后进行端点检测，减少需要处理的数据量，最后通过自适应滤波进行背景噪声消除同时实现语音加强。

[0069] 其中，对采集的语音信号进行预处理的过程具体如下：

[0070] (1)使用一阶高通滤波器滤波对原始语音信号进行预加重处理，传递函数为：

[0071] $H(z)=1-\mu z^{-1}$，

[0072] 式中，μ为预加重系数，是介于0.95～0.97之间的常数，可选的，μ取值为0.97。

[0073] (2)对预加重处理过的信号进行分帧，得到语音帧。可选的，按帧长512点、帧移为256点(对应16k采样率的每帧32毫秒)进行分帧。

[0074] (3)对语音帧使用汉明窗加窗处理，窗函数为：

[0075] $$W(n)=\begin{cases}0.54-0.46\cos\left(\frac{2\pi n}{NH-1}\right), & 0<n<NH-1\\ 0, & others\end{cases},$$

[0076] 其中，NH为汉明窗长度。

[0077] (4)对加窗处理过后的语音帧采用双门限检测法进行端点检测。

[0078] (5)采用改进的自扰动最小二乘法对端点检测处理过的语音帧，进行自适应语音噪声消除，达到语音加强和降噪的目的，其自适应滤波器权值系数W(n)的更新公式为：

[0079] $W(n)=W(n-1)+k(n)[d(n)-u^{T}(n)W(n-1)]$，

[0080] 式中，粗体字母表示对应的矩阵，n为时间序列，d(n)为端点检测处理过后的语音帧信号d(n)的第n个点的值，是一个标量；

[0081] $u(n)=[u(1),u(2),u(3)\cdots,u(n)]$为参考噪声信号，在录音环境下会同步采集背景噪声，T表示矩阵转置；

[0082] k(n)为卡尔曼增益矢量，表达式为：$\boldsymbol{k}(n)=\frac{\boldsymbol{P}(n-1)\boldsymbol{u}(n)}{\lambda+\boldsymbol{u}^{T}(n)\boldsymbol{P}(n-1)\boldsymbol{u}(n)}$，

[0083] 式中，λ为遗忘因子，λ越大，则越早时间的语音信号数据对当前语音信号数据的影响越小，可选的，λ取值为0.95；

[0084] P(n)为参考信号的自相关矩阵的逆矩阵，其迭代式为：

[0085] $P(n)=[I-k(n)u^{T}(n)]P(n-1)+Q(n)$，

[0086] 式中，I为单位矩阵，Q(n)为自扰动项，其表达式如下：

[0087] $$Q(n)=\beta\cdot NINT\left\{\gamma\cdot\frac{|E[e^2(n)]-E[u^2(n)]|}{E[u^2(n)]}\right\}\cdot I,$$

[0088] 式中，NINT为取整函数，β与γ为常数，可选的，在仿真实验中，取β＝0.9，

[0089] γ＝0.8；e(n)＝d(n)-u^T(n)W(n)，E[e^2(n)]与E[u^2(n)]分别为：

[0090] E[e^2(n)]＝{(n-1)•E[e^2(n-1)]+e(n)}/n，

[0091] E[u^2(n)]＝{(n-1)•E[u^2(n-1)]+u(n)}/n，

[0092] 经过上述自适应滤波输出信号为e(n)，也就是预处理后得到的语音信号，将其记为x(n)，其中n为信号抽样时第n个采样点。

[0093] S320、对预处理后的语音信号进行快速傅里叶变换。

[0094] 预处理之后的语音信号记为x(n)，对其进行快速傅里叶变换后得到的频谱记为X(k)，其中k代表在周期频谱(0,f_s)上的第k个采样点，对应的频率为$k\cdot\frac{f_s}{NFFT}$，NFFT为快速傅里叶变换点数，k的取值范围是0～NFFT，f_s是采样信号的采样频率。

[0095] S330、对快速傅里叶变换后的频谱取平方，得到语音信号的能量谱，再利用改进的伽马通滤波器组进行能量谱滤波。

[0096] 伽马通滤波器组的时域脉冲响应为：

[0097] $$G_i(t)=At^{ng-1}e^{-2\pi b_i t}\cos(2\pi f_c(i)t+\phi_i)\,U_i(t),\ 1\le i\le p,$$

[0098] 式中，A为滤波器增益，可选的，A取值为1；p为滤波器个数，下标i表示第i个滤波器；ng为滤波器阶数，可选的，取ng＝4；Φ_i为滤波器的初始相位，因为人耳对相位不敏感，为了简化模型，可选的，取Φ_i＝0；f_c(i)为每个滤波器的中心频率；b_i为衰减因子，U_i(t)为阶跃函数。

[0099] 听觉临界频带用等效矩形带宽表示为：

[0100] EBR(f)＝24.7(4.37f/1000+1)，

[0101] 滤波器中心频率由以下公式给定：

[0102] $$f_c(i)=-QB_0+(f_u+QB_0)\exp\left\{\frac{i-1}{p}\cdot\ln\frac{f_1+QB_0}{f_u+QB_0}\right\},$$

[0103] 式中，Q为渐进因子，B_0为最小带宽，f_1、f_u分别表示滤波器组的最小中心频率与最大中心频率。

[0104] 对于中心频率f_c(i)，可得对应的衰减因子b_i：

[0105] b_i＝1.019EBR(f_c(i))，

[0106] 根据以上滤波器的特性，可以得到改进的伽马通滤波器的频率响应。

[0107] 具体步骤如下：

[0108] 可选的，给定渐进因子Q＝9.26449，最小带宽B_0＝24.7，滤波器组的最小中心频率f_1＝80，最大中心频率f_u＝f_s/2，f_s为采样信号的采样频率。需要说明的是，参数的具体取值，在此并不限定，在可以得到满足要求的滤波结果的情况下，参数值可根据需要进行适应性更改。

[0109] 对于第i个(i＝1,2……,p)滤波器：

[0110] 计算第i个滤波器的中心频率f_c(i)：

[0111] $$f_c(i) = -QB_0 + (f_u + QB_0)\exp\left\{\frac{i-1}{p}\cdot \ln\frac{f_1+QB_0}{f_u+QB_0}\right\};$$

[0112] 计算其等效矩形带宽EBR：

[0113] EBR(f_c(i))＝24.7(4.37f_c(i)/1000+1)；

[0114] 计算第i个衰减因子b_i：b_i＝1.019EBR(f_c(i))；

[0115] 进一步得到第i个4阶滤波器的时域脉冲响应：

[0116] $$g_i(\mathrm{t}) = t^3 e^{-2\pi b_i t}\cos(2\pi f_c(i)t)\,U(t),\ 1 \le i \le p,$$

[0117] 对伽马通滤波器的时域冲击响应在时间(0,wlen/f_s)上以f_s采样成离散冲激响应，并进行快速傅里叶变换，得到其频率响应G_i(k)，其中wlen表示帧长，可选的，wlen取值为512点。

[0118] 对每个滤波器幅频进行归一化处理：G_i(k)＝G_i(k)/max(G_i(k))，

[0119] 进一步的，对于每个滤波器，如果采样点k大于NFFT/4且k小于等于NFFT/2，则第k个采样点上的频率响应调整为：$\hat{G}_i(k) = G_i(NFFT/2+1-k)$，即将滤波器频率区间(0,$f_s$/4)内的频率响应关于$f_s$/4翻转并覆盖原有频谱；如果k大于NFFT/2且k小于等于3NFFT/4，则第k个采样点上的频率响应调整为：$\hat{G}_i(k) = G_i(k-NFFT/2)$，即将滤波器频率区间(3$f_s$/4,$f_s$)内的频率响应关于3$f_s$/4翻转并覆盖原有频谱。原有频谱是指根据标准伽马通滤波器进行滤波得到的频谱G_i(k)。其中，傅里叶变换后的频谱是轴对称的频谱，频率区间(0,f_s/4)与(3f_s/4,f_s)对应低频的语音信号的频率响应，频率区间(f_s/4,f_s/2)与(f_s/2,3f_s/4)对应高频的语音信号的频率响应。具体的，根据改进的伽马通滤波器进行分频段滤波的信号输出如下：

[0120] $$\hat{G}_i(k) = \begin{cases} G_i(k),\ 0 < k \le \mathrm{NFFT}/4 \\ G_i(\mathrm{NFFT}/2+1-k),\ \mathrm{NFFT}/4 < k \le \mathrm{NFFT}/2 \\ G_i(k-\mathrm{NFFT}/2),\ \mathrm{NFFT}/2 < k \le 3\mathrm{NFFT}/4 \\ G_i(k),\ 3\mathrm{NFFT}/4 < k \le \mathrm{NFFT} \end{cases},$$

[0121] 采样点数NFFT的取值与采样信号的采样频率f_s取值存在对应关系，将采样点数NFFT进行四等分即将采样频谱f_s进行四等分。

[0122] 标准的伽马通滤波器组在语音高频部分的频率分辨率较低，是因为滤波器的等效矩形带宽正比于其中心频率，对于高频区域，滤波器的中心频率较高，导致滤波器间间隙较大，滤波分辨率便较低，因此，对语音的高频部分的识别效果较差。改进后的伽马通滤波器组对于高频区域的频谱分布比较密集，滤波器间间隙减小，恰好可以解决标准的伽马通滤波器组的低分辨问题。

[0123] S340、对滤波后的能量谱取对数，得到对数频谱。

[0124] 得到对数频谱表示为：$S_i = \ln[\sum_{k=0}^{NFFT-1}|X(k)|^2 G_i(k)]$，

[0125] 其中，X(k)是对预处理之后的语音信号进行快速傅里叶变换后得到的频谱形式。

[0126] S350、对得到的对数频谱做离散余弦变换，得到混合耳蜗频率倒谱系数。

[0127] 经过离散余弦变换处理，可以去除各个滤波器之间输出的相关性，得到的系数表

达式为：$GF_j = \sqrt{\frac{2}{p}} \sum_{i=1}^{p} S_i \cos(j\pi(i-0.5)/p)$，

[0128] 式中，j表示第j阶系数，j的取值0，1，2，……，p-1，p为滤波器个数或者滤波器通道数。在实际的语音识别系统中，并不是取全部阶数的系数，实验表明最前若干阶和最后若干阶的系数对语音的区分性能较大，因此，可选的，混合耳蜗频率倒谱系数取前26阶系数。其中，混合耳蜗频率倒谱系数的零阶系数GF_0，即表征语音帧的直流分量可表示为：

[0129] $GF_0 = \sqrt{\frac{2}{p}} \sum_{i=0}^{p-1} S_i = \sqrt{\frac{2}{p}} \sum_{i=0}^{p-1} \ln\left[\sum_{k=0}^{NFFT-1} |X(k)|^2 G_i(k)\right]$。

[0130] S360、利用语音帧的平均短时对数能量替换混合耳蜗频率倒谱系数中的零阶系数，便得到语音信号的声纹特征。

[0131] 计算每一帧语音信号的平均短时对数能量：

[0132] $\overline{E_0} = \ln\left(\sum_{n=0}^{wlen} x^2(n)\right)/wlen$，

[0133] 利用平均短时对数能量$\overline{E_0}$替换混合耳蜗频率倒谱系数的零阶系数GF_0，最终得到的系数便组成语音信号的声纹特征。

[0134] S370、利用预先训练好的隐马尔可夫模型对声纹特征进行识别，得到语音信号的识别结果。

[0135] 本实施例的技术方案通过对采集的语音信号进行预加重、分帧、加窗、端点检测和自适应语音加强的预处理，然后得到语音帧信号的能量谱，根据至少一个改进的伽马通滤波器，进行分频段的差异化声纹特征提取，并利用语音帧的平均短时对数能量替换语音帧的直流分量，得到准确的声纹特征，最后基于此声纹特征，确认语音信号的识别结果，解决了现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题，提高了滤波器对语音高频部分的分辨率，提高了声纹特征提取的准确性，进而提高了对语音的高频部分的识别准确性，保证了语音识别效果，并且降低了声纹识别涉及的运算复杂度以及响应时间。

[0136] 实施例四

[0137] 图4是本发明实施例四提供的声纹识别装置的结构示意图，本实施例可适用于进行声纹识别的情况。本发明实施例所提供的声纹识别装置可执行本发明任意实施例所提供的声纹识别方法，具备执行方法相应的功能模块和有益效果。如图4所示，该装置包括声纹采集模块410、声纹特征提取模块420和声纹识别模块430，其中：

[0138] 声纹采集模块410，用于采集待识别的语音信号。

[0139] 声纹特征提取模块420，用于根据至少一个改进的伽马通滤波器，对语音信号进行分频段的声纹特征提取。

[0140] 可选的，声纹特征提取模块420包括信号滤波单元和声纹特征确定单元，其中：

[0141] 信号滤波单元，用于根据至少一个改进的伽马通滤波器，利用以下公式得到的滤波器的频率响应对语音信号的能量谱进行滤波，得到语音信号的滤波能量谱，

[0142] $$\hat{G}_i(k)=\begin{cases} G_i(k),\ 0<k\le \mathrm{NFFT}/4 \\ G_i(\mathrm{NFFT}/2+1-k),\ \mathrm{NFFT}/4<k\le \mathrm{NFFT}/2 \\ G_i(k-\mathrm{NFFT}/2),\ \mathrm{NFFT}/2<k\le 3\mathrm{NFFT}/4 \\ G_i(k),\ 3\mathrm{NFFT}/4<k\le \mathrm{NFFT} \end{cases},$$

[0143] 其中，$\hat{G}_i(k)$表示第i个改进的伽马通滤波器的频率响应，$G_i(k)$表示标准的伽马通滤波器的频率响应，k表示语音信号的周期频谱上的第k个采样点，NFFT表示采样点数；

[0144] 声纹特征确定单元，用于根据滤波能量谱得到语音信号的声纹特征。

[0145] 可选的，声纹特征确定单元包括对数频谱确定子单元、倒谱系数确定子单元和直流替换子单元，其中：

[0146] 对数频谱确定子单元，用于对滤波能量谱取对数，得到对数频谱；

[0147] 倒谱系数确定子单元，用于对得到的对数频谱做离散余弦变换，得到混合耳蜗频率倒谱系数；

[0148] 直流替换子单元，用于利用语音信号的平均短时对数能量替换混合耳蜗频率倒谱系数中的零阶系数，得到语音信号的声纹特征。

[0149] 声纹识别模块430，用于基于提取的声纹特征，确认语音信号的识别结果。

[0150] 可选的，声纹识别模块430具体用于：基于提取的声纹特征，利用预先训练好的隐马尔可夫模型，得到语音信号的识别结果，其中，隐马尔可夫模型是基于训练语音的声纹特征训练得到，训练语音的声纹特征是根据至少一个改进的伽马通滤波器，进行分频段的声纹特征提取后得到。

[0151] 可选的，声纹识别模块430具体用于：基于提取的声纹特征，利用预先训练好的隐马尔可夫模型，得到语音信号的说话人信息。

[0152] 可选的，该装置还包括：傅里叶变换模块和能量谱确定模块，其中：

[0153] 傅里叶变换模块，用于对语音信号进行快速傅里叶变换，并进行归一化处理；

[0154] 能量谱确定模块，用于对归一化处理之后得到的频谱取平方，得到语音信号的能量谱，以便进行声纹特征提取。

[0155] 可选的，该装置还包括噪声消除模块，用于采用改进的自扰动最小二乘法对语音信号进行自适应地语音加强。

[0156] 本实施例的技术方案通过根据至少一个改进的伽马通滤波器，对采集的待识别语音信号进行分频段的差异化声纹特征提取，然后基于提取的声纹特征，确认语音信号的识别结果，解决了现有技术中由于语音的高频部分信息丢失导致的识别效果较差的问题，提高了滤波器对语音高频部分的分辨率，提高了声纹特征提取的准确性，进而提高了对语音的高频部分的识别准确性，并且降低了声纹识别涉及的运算复杂度以及响应时间。

[0157] 实施例五

[0158] 图5是本发明实施例五提供的一种服务器的结构示意图。图5示出了适于用来实现本发明实施方式的示例性服务器512的框图。图5显示的服务器512仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

[0159] 如图5所示，服务器512以通用服务器的形式表现。服务器512的组件可以包括但不限于：一个或者多个处理器516，存储装置528，连接不同系统组件(包括存储装置528和处理器516)的总线518。

[0160] 总线518表示几类总线结构中的一种或多种，包括存储装置总线或者存储装置控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构(Industry Subversive Alliance，ISA)总线，微通道体系结构(Micro Channel Architecture，MAC)总线，增强型ISA总线、视频电子标准协会(Video Electronics Standards Association，VESA)局域总线以及外围组件互连(Peripheral Component Interconnect，PCI)总线。

[0161] 服务器512典型地包括多种计算机系统可读介质。这些介质可以是任何能够被服务器512访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质。

[0162] 存储装置528可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器(Random Access Memory，RAM)530和/或高速缓存存储器532。服务器512可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储系统534可以用于读写不可移动的、非易失性磁介质(图5未显示，通常称为“硬盘驱动器”)。尽管图5中未示出，可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器，以及对可移动非易失性光盘，例如只读光盘(Compact Disc Read-Only Memory，CD-ROM)，数字视盘(Digital Video Disc-Read Only Memory，DVD-ROM)或者其它光介质)读写的光盘驱动器。在这些情况下，每个驱动器可以通过一个或者多个数据介质接口与总线518相连。存储装置528可以包括至少一个程序产品，该程序产品具有一组(例如至少一个)程序模块，这些程序模块被配置以执行本发明各实施例的功能。

[0163] 具有一组(至少一个)程序模块542的程序/实用工具540，可以存储在例如存储装置528中，这样的程序模块542包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块542通常执行本发明所描述的实施例中的功能和/或方法。

[0164] 服务器512也可以与一个或多个外部设备514(例如键盘、指向终端、显示器524等)通信，还可与一个或者多个使得用户能与该服务器512交互的终端通信，和/或与使得该服务器512能与一个或多个其它计算终端进行通信的任何终端(例如网卡，调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口522进行。并且，服务器512还可以通过网络适配器520与一个或者多个网络(例如局域网(Local Area Network，LAN)，广域网(Wide Area Network，WAN)和/或公共网络，例如因特网)通信。如图5所示，网络适配器520通过总线518与服务器512的其它模块通信。应当明白，尽管图中未示出，可以结合服务器512使用其它硬件和/或软件模块，包括但不限于：微代码、终端驱动器、冗余处理器、外部磁盘驱动阵列、磁盘阵列(Redundant Arrays of Independent Disks，RAID)系统、磁带驱动器以及数据备份存储系统等。

[0165] 处理器516通过运行存储在存储装置528中的程序，从而执行各种功能应用以及数据处理，例如实现本发明实施例所提供的声纹识别方法，该方法包括：

[0166] 采集待识别的语音信号；

[0167] 根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取；

[0168] 基于提取的声纹特征，确认所述语音信号的识别结果。

[0169] 实施例六

[0170] 本发明实施例六还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本发明实施例所提供的声纹识别方法，该方法包括：

[0171] 采集待识别的语音信号；

[0172] 根据至少一个改进的伽马通滤波器，对所述语音信号进行分频段的声纹特征提取；

[0173] 基于提取的声纹特征，确认所述语音信号的识别结果。

[0174] 本发明实施例的计算机存储介质，可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0175] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0176] 计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括——但不限于无线、电线、光缆、RF等等，或者上述的任意合适的组合。

[0177] 可以以一种或多种程序设计语言或其组合来编写用于执行本发明操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或终端上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。

[0178] 注意，上述仅为本发明的较佳实施例及所运用技术原理。本领域技术人员会理解，本发明不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本发明的保护范围。因此，虽然通过以上实施例对本发明进行了较为详细的说明，但是本发明不仅仅限于以上实施例，在不脱离本发明构思的情况下，还可以包括更多其他等效实施例，而本发明的范围由所附的权利要求范围决定。

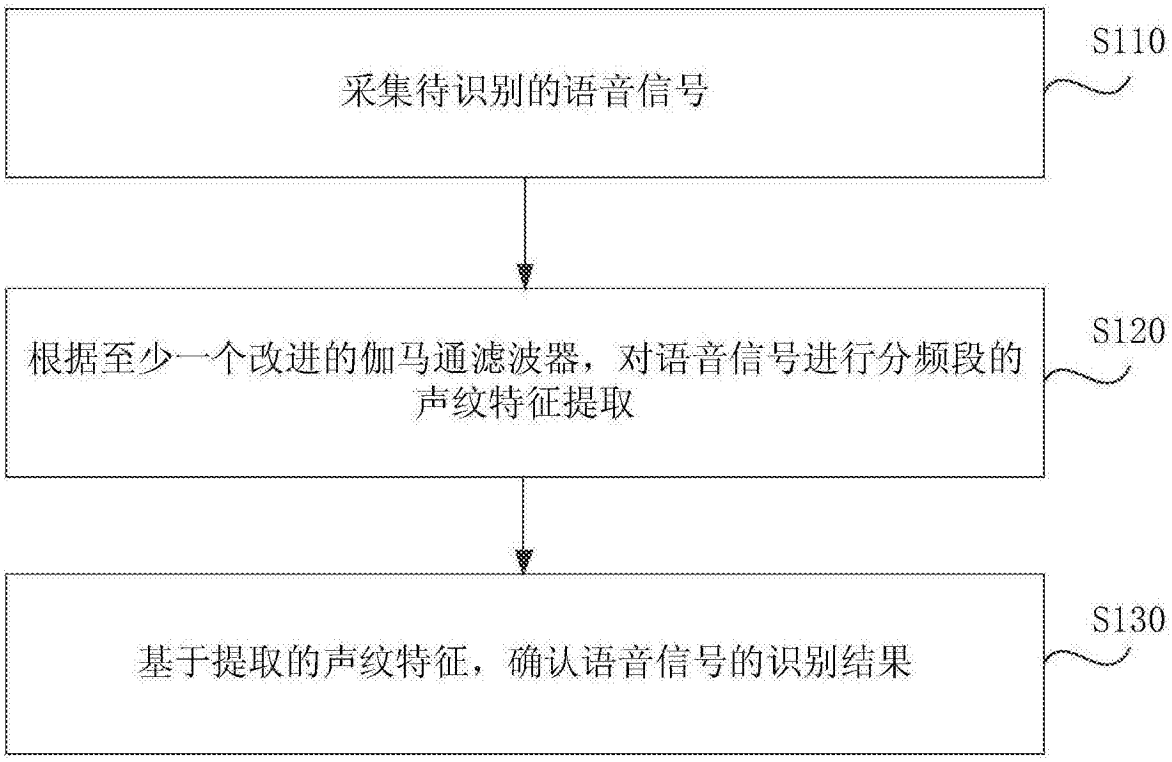

采集待识别的语音信号
S110
根据至少一个改进的伽马通滤波器，对语音信号进行分频段的声纹特征提取
S120
基于提取的声纹特征，确认语音信号的识别结果
S130

图1

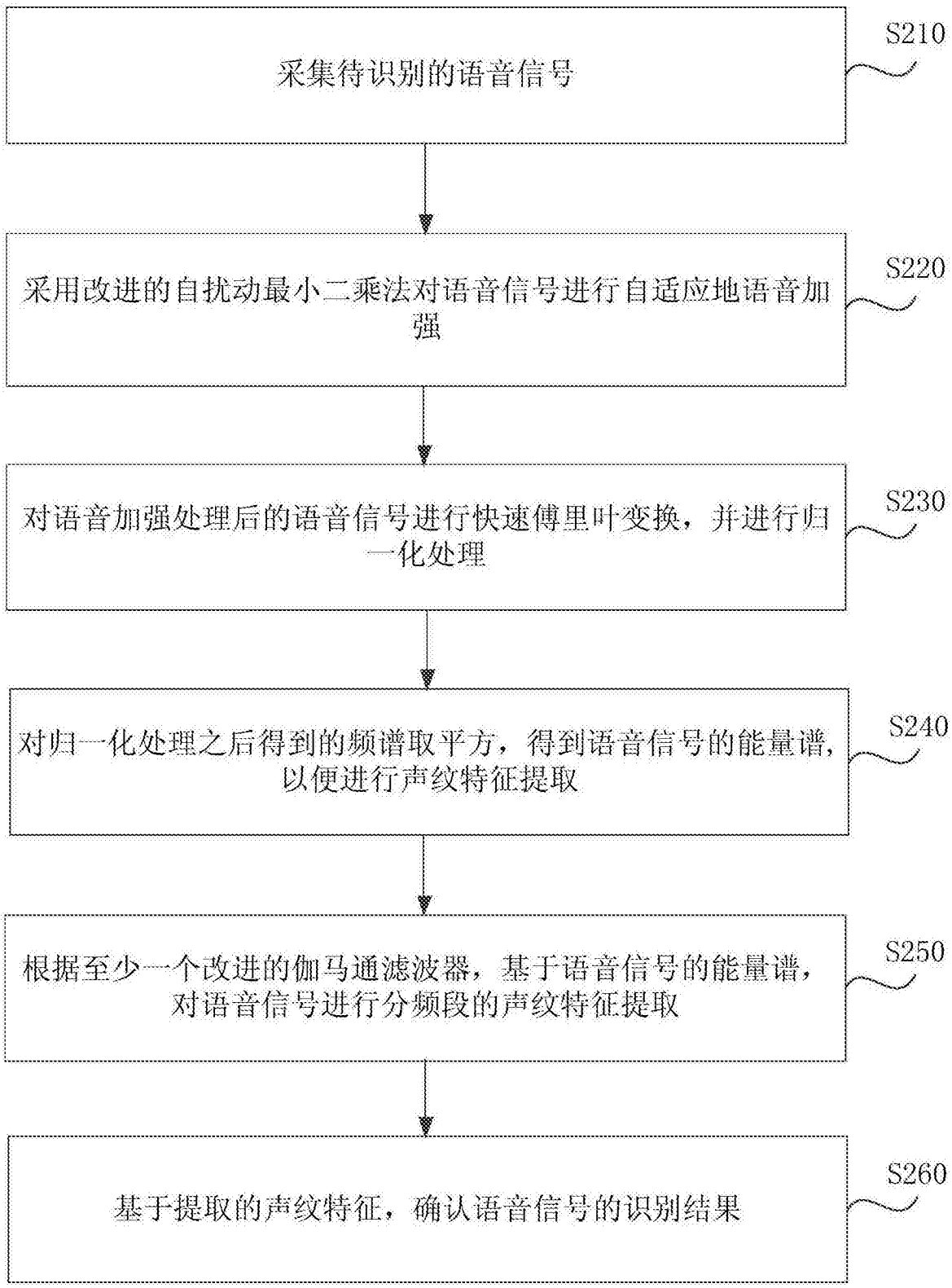

采集待识别的语音信号
S210
采用改进的自扰动最小二乘法对语音信号进行自适应地语音加强
S220
对语音加强处理后的语音信号进行快速傅里叶变换，并进行归一化处理
S230
对归一化处理之后得到的频谱取平方，得到语音信号的能量谱，以便进行声纹特征提取
S240
根据至少一个改进的伽马通滤波器，基于语音信号的能量谱，对语音信号进行分频段的声纹特征提取
S250
基于提取的声纹特征，确认语音信号的识别结果
S260

图2

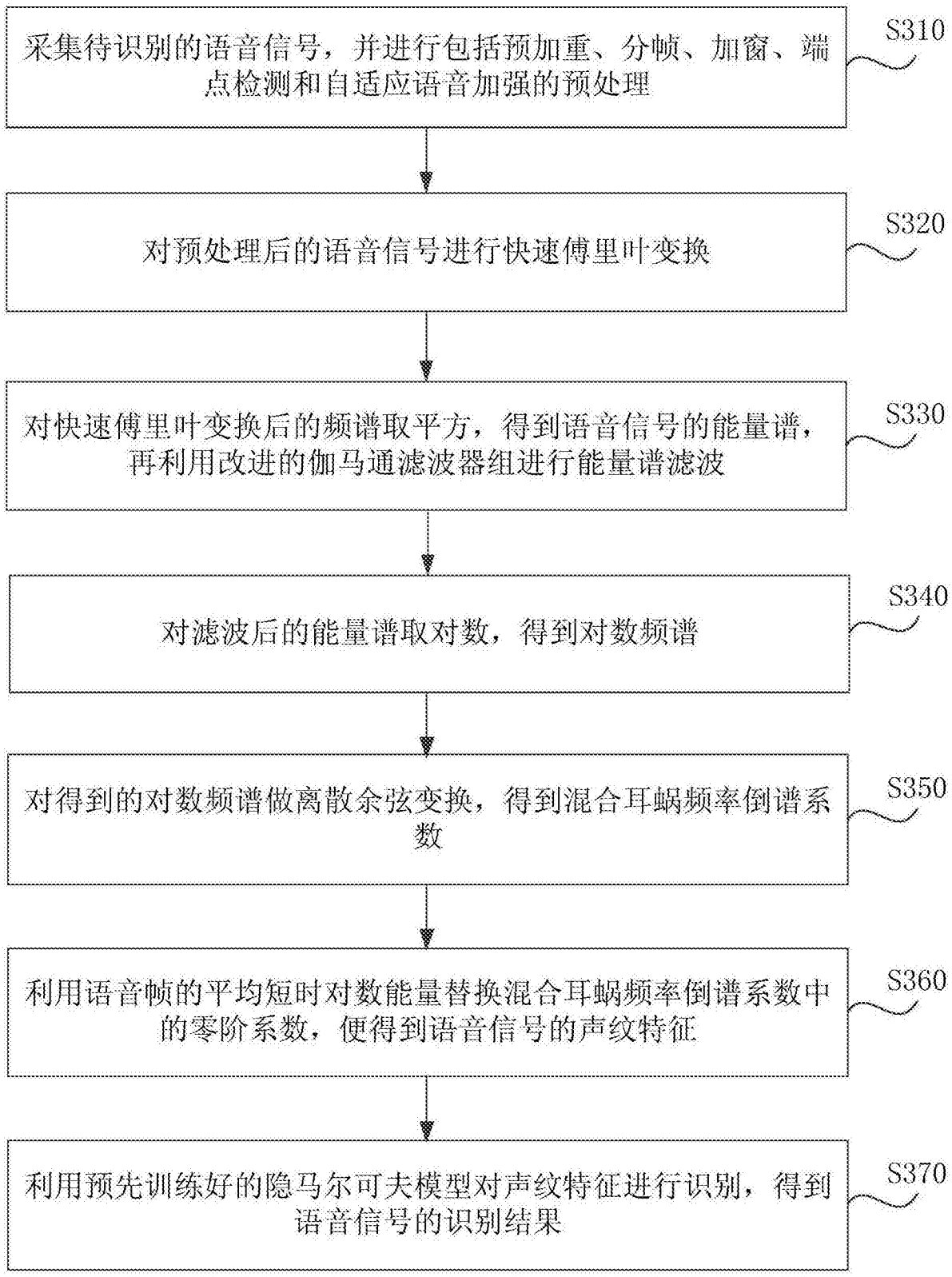

采集待识别的语音信号，并进行包括预加重、分帧、加窗、端点检测和自适应语音加强的预处理
S310
对预处理后的语音信号进行快速傅里叶变换
S320
对快速傅里叶变换后的频谱取平方，得到语音信号的能量谱，再利用改进的伽马通滤波器组进行能量谱滤波
S330
对滤波后的能量谱取对数，得到对数频谱
S340
对得到的对数频谱做离散余弦变换，得到混合耳蜗频率倒谱系数
S350
利用语音帧的平均短时对数能量替换混合耳蜗频率倒谱系数中的零阶系数，便得到语音信号的声纹特征
S360
利用预先训练好的隐马尔可夫模型对声纹特征进行识别，得到语音信号的识别结果
S370

图3

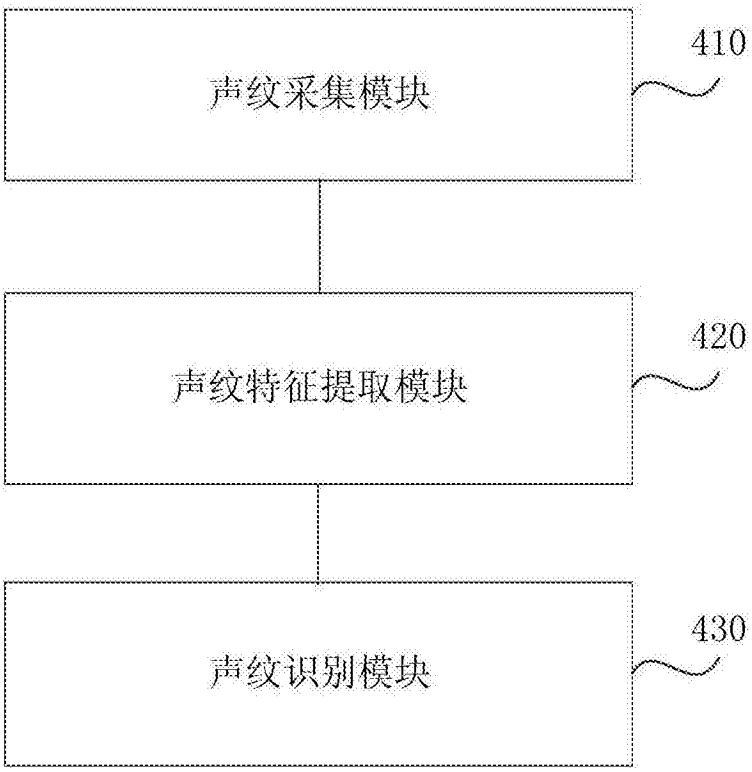


图4

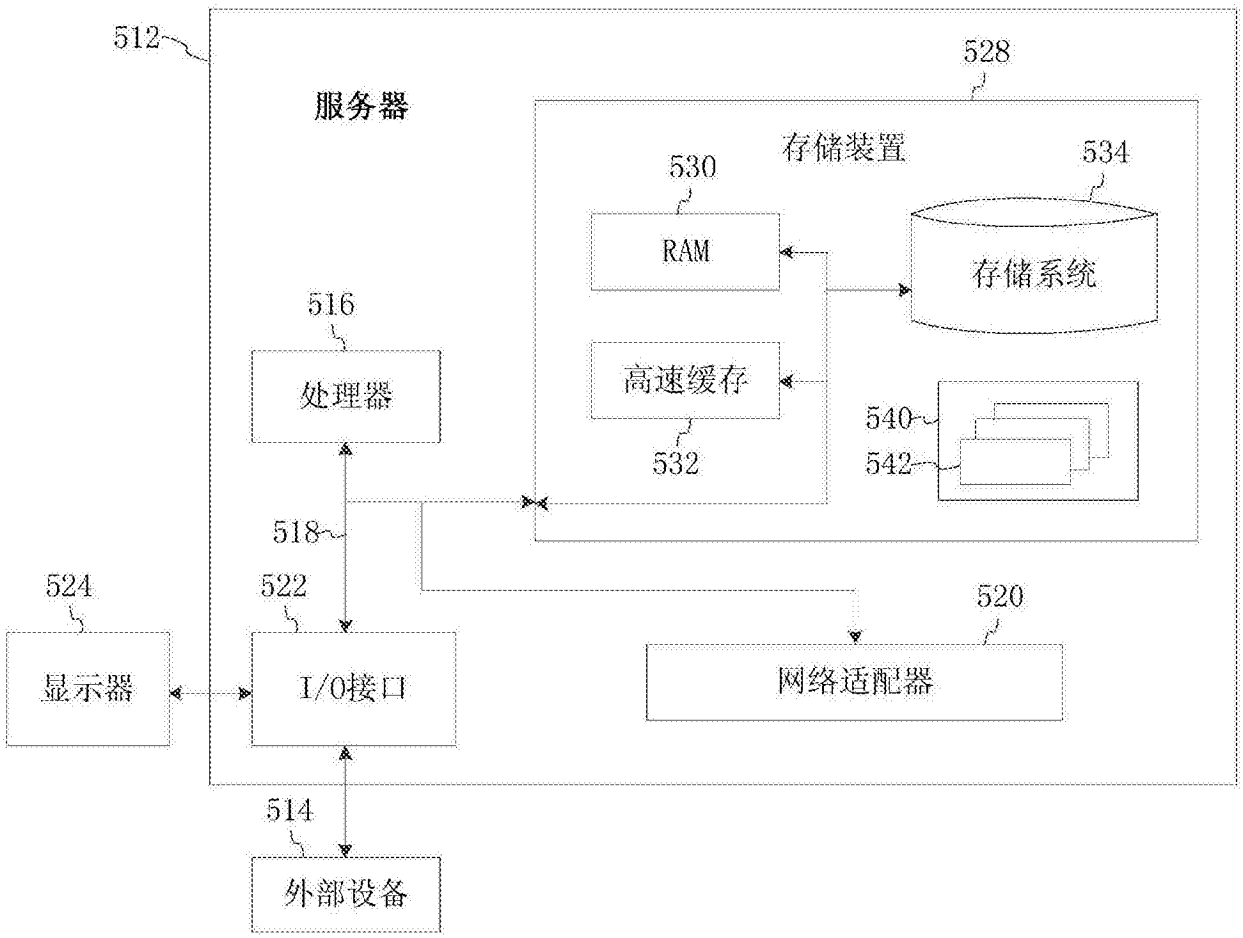

512
服务器
528
存储装置
530
RAM
534
存储系统
516
处理器
高速缓存
532
540
542
518
524
522
520
显示器
I/O接口
网络适配器
514
外部设备

图5