时序不受限制，可以先执行前 者，后执行后者，或者同时执行前者和后者。

本实施例中的匹配结果包括：匹配成功与否，或者匹配分值；其中根据匹 配分值是否在预设范围内来判定是否匹配成功。

步骤105：根据所述匹配结果判断是否需要唤醒所述终端，若是，执行步骤 106，若否，则执行步骤107。

具体地，当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结 果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端；否则，判断不需要唤醒所述终端。

本步骤需要判断语音识别结果是否匹配成功和声纹特征是否匹配，这两个 判断过程的时序可以任意，具体包括以下情况：

1、先判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果 是否匹配成功，若是，则再判断所述语音信号的声纹特征与所述存储的语音唤 醒词声纹特征是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情 况下均判定不需要唤醒终端。

2、先判断所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否 匹配成功，若是，则再判断所述语音信号的语音识别结果与所述存储的语音唤 醒词识别结果是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情 况下均判定不需要唤醒终端。

3、同时判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结 果是否匹配成功、以及所述语音信号的声纹特征与所述存储的语音唤醒词声纹 特征是否匹配成功，若均成功，则判定需要唤醒所述终端；若有一个匹配不成 功则判定不需要唤醒终端。

本步骤中匹配成功可以包括：匹配分值在预设范围内，或者相似度在预设 范围之内。

步骤106：唤醒所述终端。

本步骤可以通过发送中断事件唤醒终端；当语音唤醒步骤在低功耗音频芯 片上执行时，低功耗音频芯片判断需要唤醒终端后，会上传中断事件给终端的主处理器唤醒处于休眠状态下的主处理器。

步骤107：不进行任何处理。

error rate to wake enhance the user experience of voice wake-up method, shown in Figure 1, the speech wake method of this embodiment includes the following steps:

Step 101: end-user acquisition wake-word speech signals emitted by obtaining the wake wakeup word recognition results voiceprint characteristic word speech signals and stored.

This embodiment of the method may be employed voice collecting device, such as a microphone to collect end-user voice signal sent. The method of this embodiment, the word recognition result obtained wake voiceprint characteristic timing is not limited, you can get the recognition result, after obtaining voiceprint may be acquired at the same time.

In this embodiment wake word recognition result is a wake-word speech recognition voice signal obtained wake-word text or commands. This embodiment voiceprint example is the speech signal may represent the user sound mode parameters, such as frequency of sound, loudness and so on.

In this step, a voice wake preparatory steps before occurred in a user-defined voice wake-up word in the process, this step may be performed by AP-side terminal, that is the main chip terminals, such as mobile phone main CPU ARM chip carry out.

In order to improve the accuracy of speech wake, the present embodiment method in the acquisition phase of sample characteristics, you can record multiple wake-word speech signals issued by the user, and then get the best wake-up word for word the speech signal wake-up times recorded results voiceprint recognition feature. As shown in Figure 2 for the terminal to obtain user-defined voice wake-word sample characteristics of the process, completed by the AP-side terminal:

In the terminal starts after a user-defined set wake-up word, the first recorded user of said wake-word speech signal, and then generate a wake wake-up signal word word speech recognition result of a voiceprint features 1, 1 and the recognition result voiceprint 1 storage to a custom wake-word memory area;

The second terminal of said user recorded wake-word speech signal, and then generate a wake wake-up word word speech recognition result

本实施例方法中当判断不需要唤醒终端时，可以忽略当前输入的语音信号，执行唤醒步骤的芯片或器件可以进入休眠状态，以节省终端功耗。

本实施例提供的语音唤醒方法，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征(即识别结果和声纹特征)作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同(同一用户的发声方式相同)，避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本实施例方法加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

考虑到噪声环境对用户输入的语音信号的影响，进一步提高用户唤醒率降低误唤醒率，本实施例方法在上述步骤102之后，步骤103之前还可以包括：

判断所述语音信号是否满足预设信号要求，若是，则执行步骤103。

具体地，可以判断输入的语音信号的信噪比是否在预设范围内，若是，则满足语音信号要求，或者语音信号的其他质量参数是否在预设范围内，若是，则满足语音信号要求，或者将输入的语音信号的声纹特征用用较简单的波形匹配，若成功，则满足信号要求。该步骤可以在低功耗音频芯片上完成。

本实施例方法在判断语音信号不满足预设信号要求时，可以输入的语音信号丢弃；或者当该判断步骤是在低功耗音频芯片上置信时，该芯片进入休眠状态，以节省电能。

优先地，如图3所示，上述判断所述语音信号是否满足预设信号要求的过程可以包括：

步骤301：对所述语音信号进行处理得到用户语音数据和背景噪声数据。

由于外界输入的语音信号包括用户自己发出的语音信号和当前用户环境产生的噪声信号，所以在进行信号判断时，可以将语音信号分离成用户语音数据和背景噪声数据。

signal 2 voiceprint characteristic 2, when the wake-up word recognition results 2 and 1 wake-up word recognition results are the same, optimize the use of voiceprint 2 voiceprint 1, when the wake-up word recognition results 2 and 1 wake word recognition results are not the same, re-recorded; recorded or re-enter the third record from the first start;

The third terminal of said user recorded wake-word speech signal, and then generate a wake wake-up word recognition results 3 voiceprint characteristic word speech signal 3, when the wake-up word recognition results 3 and 1 wake-up word recognition results are the same, optimize the use of voiceprint 3 above have been optimized voiceprint 1;

And so on, until the n-th recording is completed. Figure 2, n is 4, for example, the method of the present embodiment n values can be set according to actual demand.

Step 102: the terminal is in a dormant state, the voice signal detection and external input.

This step can be detected by the microphone terminal outside of the speech signal, when the terminal sleep state, the microphone can remain active.

In this step, the terminal is in hibernation from a hardware point of view may include: a main processor of the terminal is in a dormant state, from a software point of view may include: terminal operating system is in sleep or locked.

Step 103: obtaining a speech recognition result of the speech signal voiceprint features.

This step is obtaining a speech recognition result voiceprint feature no timing constraints, you can get the speech recognition result, after obtaining voiceprint, or you can get voiceprint, after obtaining the speech recognition result, you can also get voice results voiceprint recognition feature.

Speech recognition result

In this step, the voice signal is voice information obtained by speech recognition or text commands. This embodiment voiceprint example is the speech signal may represent a user utterance mode

Petitioner Exhibit 1004, Page 951 of 1518

步骤302：将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声 模型进行匹配。

本步骤中用户语音数据和背景噪声数据匹配时序不受限制，可以先后，也 可以同时。

本实施例方法可以预置一些较典型的生活场景下的稳态或非稳态噪声声学 统计模型，例如，汽车中，酒吧里，餐馆，家庭，办公室，地铁上，会议中等。

具体地，本步骤可以包括如下步骤：

将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分 值；

将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分 值。

步骤303：根据匹配结果判断所述语音信号是否满足预设信号要求。

本步骤中的匹配结果可以包括上述两个匹配过程得到的匹配分值，此时具 体地判断过程可以包括：

采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到 所述语音信号的匹配分值；

判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设 信号要求，若否，则判定不满足预设信号要求，若否，则不满足预设信号求。

例如对两个匹配分值进行加权处理得到一个匹配分值，然后判断该匹配分 值是否超出置信阀值，若是，则判定满足预设信号要求。

进一步考虑语音唤醒时终端的功耗问题，本实施例可以由一个音频处理器 来完成语音唤醒的主要过程，事先将声音样本特征(唤醒词识别结果和唤醒词语音信号的声纹特征)下载到该音频处理器的存储区；在终端处于休眠状态时 该音频处理器处于半sleep待机状态，由mic事件控制其工作唤醒；当mic检 测到声音输入时唤醒多核音频处理器，执行语音唤醒；这样可以使得终端的主 处理在休眠状态时，通过低功耗的音频处理器响应和处理外界输入的声音信息， 节省终端的功耗，提高终端的待机时间。

优先地，上述音频处理器可以采取分步式核心处理，Core1用于处理接收到 mic接收到外部声音信号的唤醒事件声音初步分析，即判断语音信号是否满足预 设要求，Core2用户后续的语音识别以及鉴权认证，即判断是否唤醒。当mic检 测到语音输入

parameters, such as the sound frequency, loudness and so on.

The main steps 103-106 to process step voice wake-up, which may be in a low-power audio chip performs, such as low-power DSP chip. Before voice wake, wake up in advance of the above word recognition results and the wake-up word speech signals voiceprint get downloaded to the low-power audio chip memory area, in preparation for the wake-up voice.

Step 104: Get wakeup word recognition result of the speech recognition result and storage, as well as matches the voiceprint and voiceprint stored.

In this step, get wake-up word recognition result of the speech recognition result matches were stored methods include: speech recognition result stored wake-up word recognition results were matched to obtain the voiceprint and voiceprint stored in way matching results include: the voiceprint features will be acquired and stored voiceprint match; this step get wake-up word recognition result matches the result with a stored voice, and access the stored voiceprint and voiceprint results matching process timing is not limited, you can execute the former, the latter after the execution, or both the execution of the former and the latter.

This embodiment results matching comprising: a successful match or not, or match scores; according to which the matching score is within a predetermined range to determine whether the match succeeds.

Step 105: according to the matching results to determine whether the need to wake up the terminal, and if so, step 106, and if not, proceed to step 107.

In particular, when the voice recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the characteristics of the voiceprint stored speech voice signal wake-up word voiceprint matching success when it is determined that the terminal needs to wake up; otherwise, the determination is no wakeup terminal.

This step needs to determine whether the results match the success of the speech recognition feature voiceprint matches, the timing of these two

时唤醒Core1，Core1执行判断语音信号是否满足预设要求的步骤； 当Core1判断为是时，唤醒Core2，并将语音信号传输给Core2(优先地，Core1 将语音信号存储在Core1、2共享的存储区中，由Core2提取)；Core2唤醒后执 行上述步骤103-106，完成唤醒鉴别，若通过，则由Core2向终端主处理器(例 如终端ARM芯片)传送硬件中断事件，终端主处理器接收到中断事件后被唤醒 由sleep模式切换到Normal模式正常运行。

应当理解的是：本实施例还可将语音识别结果匹配或者声纹匹配的过程放 在Core1中执行，只有当匹配成功时才唤醒Core2，然后Core2唤醒之后执行声 纹匹配或者语音结果匹配。

本实施例通过多核音频处理器的分步式核心处理设计，可以实现在节省终 端功耗或电能的效果，提升了用户体验。

下面通过如图4所示的语音唤醒系统来具体介绍本实施例方法；如图4所 示，该系统包括：DSP芯片(多核DSP芯片，可以包括至少两个核)、终端主CPU 中的ARM芯片和麦克风；其中DSP芯片包括两个核单元以及一个唤醒存储单元 (例如Flash)，第一核单元Core1和第二核单元Core2；ARM芯片包括唤醒词处 理单元和唤醒词存储单元；图4所示的系统进行语音唤醒的具体过程包括：

步骤501：ARM芯片中唤醒词处理单元录制用户所说的唤醒词语音信 号，并 生成唤醒词语音识别结果和声纹特征，将唤醒词语音识别结果和声纹特征存储 在唤醒词存储单元中。

具体地可以生成方式可以参考上述图2的描述。

步骤502：ARM芯片将唤醒词语音识别结果、唤醒词语音信号的声纹特征和 预置的背景噪声模型下载到DSP芯片，并存储在唤醒存储单元中。

步骤503：在ARM芯片处于休眠状态下，麦克风动态监测外界的语音输入， 当检测到语音信号时发送唤醒事件唤醒Core1。

步骤504：Core1对检测到的语音信号进行处理得到用户语音数据和背景语 音数据。

步骤505：Core1将所述用户语音数据与所述预置的背景噪声模型进行匹配 得到第一匹配分值，将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

步骤506：Core1采用预设分值算法对所述第一匹配分值和所述第二匹配分 值进行计算得到所述语音信

processes can be any judgment, including the following:

1, first determine the speech recognition result and the stored speech of the speech signal wake-up word recognition result if the match is successful, and if so, then determine the voiceprint of the speech signal with the stored wakeup word speech voiceprint match is successful, if it is judged need to wake up the terminal; judgment is NO circumstances determination no wakeup terminal.

2, first determine the voiceprint of the speech signal with the stored speech wake term voiceprint match is successful, and if so, then judge the speech recognition result with the storage of the speech signal of voice wake word recognition result if the match is successful, if it is judged need to wake up the terminal; judgment is NO circumstances determination no wakeup terminal.

3, while the determination of the speech recognition result of the speech signal stored speech recognition result if the wake-up word matching is successful, and the voiceprint of the speech signal and the wake-up word stored speech characteristics whether the voiceprint successful match, if all is successful, it is determined that the terminal needs to wake up; if it is determined that a match is unsuccessful no wakeup terminal.

In this step, a match may include: match scores within a preset range, or similarity within a preset range.

Step 106: wake up the terminal.

In this step, you can wake up the terminal by sending an interrupt event; when the voice wake-up steps to be performed on a low-power audio chip, low-power audio chip Analyzing need to wake up after the terminal, the terminal will upload interrupt events to wake up the main processor main processor sleep state.

Step 107: no processing.

This embodiment of the method when the judge is not required to wake the terminal, you can ignore the current input speech signal, the wake-up procedure can chip or device into hibernation to save power terminal.

号的匹配分值。

步骤507：Core1判断所述语音信号的匹配分值是否大于预设阈值，若是， 则执行步骤508，若否，则执行步骤512。

步骤508：Core1将语音信号传输给Core2同时唤醒Core2。

步骤509：Core2唤醒后，对所述语音信号进行语音识别获得语音识别结果， 并与存储的唤醒词识别结果进行匹配，若匹配成功，则执行步骤510，若失败， 则执行步骤513。

步骤510：Core2提取所述语音信号的声纹特征，并与存储的声纹特征进行 匹配，若成功，执行步骤511，若否，执行步骤513。

步骤511：Core2发送唤醒中断事件给ARM芯片唤醒ARM芯片。

步骤512：Core1重新休眠，DSP芯片休眠。

步骤513：Core2重新休眠，语音唤醒流程结束。

DSP芯片中Core1和Core2的语音算处理能力不同功耗有差异，当DSP芯片 只有Core1工作时因为其运算能力和主频和处理能力有限，其运转时是DSP处 在相对低功耗工作状态(Normal)。如果DSP芯片Core2工作时其运算和处理能 力较高，功耗相对较高，Core2被唤醒时DSP芯片处于Work态。但唤醒DSP芯 片全速运行work状态的功耗也是远小于终端主CPU芯片Normal态的运行功耗。

本实施例可以将DSP芯片的运行状态分为低耗能Core1运行状态和相对高 耗能运算的Core2运行态。从上图5可以看出，即使在复杂强噪音环境下，只 要没有清晰人声特征输入，Core2将不被唤醒，会使DSP芯片处于稳定低功耗状 态。即使Core2被唤醒如果人声输入不是自定义的唤醒指令，或者唤醒指令不 是录制唤醒特征的用户输入，手机主CPU芯片也不会被唤醒。通过硬件低功耗DSP芯片的阶梯层次的唤醒机制，个体化唯一化的鉴权机制保证了手机语音唤醒 的功耗控制，和误醒率的降低。

应用本实施例方法，带有声纹特征的自定义唤醒词可以在降低了智能机中 语音检测功耗控制优化难题的基础上，更好的适应特殊人群或者个体的个性化需求，在便捷操作的同时，让用户的语音控制更贴近于人们生活实际使用场景。

实施例二：

During voice wake, since the sound;

Example embodiments provide a voice to wake this method to end-users own voice as a sound sample, end-users get the sound characteristics of individuals (ie, the recognition result voiceprint characteristic) as an audio sample characteristics Sample features voiceprint and voiceprint wakeup extraction (same same user sound pattern), to avoid the user individual sound pattern difference in the sound sample user sound pattern caused wake-up failed to provide a wake-up rate; in addition, Example embodiments of the present joined the speech recognition result voiceprint matching characteristics, can effectively avoid the noisy environment caused by waking up, so that the noise environment can not wake up the terminal, reducing the error rate wake; the method of the present invention may also supply custom wake-term, to meet the needs of individual users, to enhance the user experience.

Further, since different users different sound pattern is different, that voiceprint, the application of the present embodiment method only to wake up my end user terminal, compared with the prior art, but also to improve the information security of the terminal.

Taking into account the environmental impact of noise on the voice signal input by the user, to further enhance the user reduces the error rate to wake wakeup rate, the method of this embodiment, after step 102 and before step 103 may also include:

Analyzing the voice signal meets the requirements of a preset signal, and if so, step 103 is performed.

In particular, based on the input speech signal to noise ratio is within a preset range, and if so, to meet the requirements of the speech signal, or other quality parameters of the speech signal is within a preset range, and if so, to meet the voice signal requirements, or to enter the voiceprint of the speech signal with the use of a relatively simple waveform matching, if successful, will meet the signaling requirements. This step can be done on a low-power audio chip.

Example of the present embodiment in determining the speech signal does not meet the requirements of the preset signal, the speech signal can be input to discard; or when the judging step is confidence in the low-power audio chip, the

为进一步提升用户体验，在终端唤醒之后，还行可以进行多样性的快速和 自动响应，例如打开应用程序等，本实施例在上述语音唤醒方法的基础上，提供了一种终端处理方法，如图6所示，包括如下步骤：

步骤601：利用实施例一所述的语音唤醒方法唤醒所述终端。

具体唤醒过程参考实施例一的具体描述。

步骤602：所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对 应的预设操作。

上述步骤602可以由终端中的主处理器完成。

本实施例方法可以预先设置自动定义唤醒词对应的响应操作，不同的唤醒 词可以对应不同的响应操作；在用户通过语音唤醒词唤醒终端之后，终端根据 用户输入语音信号的识别结果执行对应的响应操作，例如自动打电话或者自动 打开聊天软件或者开启驾驶模式等。这里的唤醒词是文字内容，识别结果也是 用户所说的文字内容。

优先地，当通过唤醒中断事件来唤醒终端情况下，可以设置不同的唤醒词 对应不同的唤醒中断事件，如果一个用户预先录制多个唤醒词，不同唤醒词对应不同的唤醒中断事件，终端对不同的唤醒中断事件，执行不同的预设操作或 者架构不同的应用场景。这样达成即使在手机黑屏休眠时用户也可以通过不同 的用户唤醒指令，进行多样性的快速和自动响应。例如用户自定义唤醒词为"打 电话"时，在终端处于休眠状态下，在用户说出"打电话"时终端发出对应的 唤醒中断事件来唤醒自身，并且终端在唤醒之后会根据预设的电话号码拨打电 话。

因此，上述步骤601语音唤醒方法中唤醒所述终端的过程包括：根据所述 语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

上述步骤602具体包括：根据所述唤醒中断事件的类型执行对应的预设操 作。

下面列举应用本实施例方法的具体场景：

场景一：如果用户录制"急救SOS，救命，妈妈"此类和用户情急时呼救口 语化相关的指令做为唤醒词时，给用户一个默认的紧急唤醒的设置方法。允许用户设置的应急号码，即使在强噪音数据下，此指令的敏感度分值在用户无法 手动呼叫的时候第一时间协助用户拨打出急救电话，或者发送出急救信息给接 收者，同时后台开启录音或摄像功能存放在紧急文件目录中，并在wifi环境下 上传网络云盘(此目

chip enters sleep mode to conserve electricity. preferentially, shown in Figure 3, the above-described voice signal meets a preset determination required signal process may include:

Step 301: the voice of the voice signals are processed to obtain user data and background noise data.

Since the external input voice signal includes a speech signal emitted and the current user's own user environment generated noise signal, during signal determining can be a voice signal into voice data and user data background noise.

Step 302: the voice data and the user data of the background noise and the background noise are preset to match the model.

In this step, the user voice data and background noise unrestricted data matching the timing may have to be at the same time.

Example of the present embodiment can be preset number of stationary and non-stationary noise acoustic statistical model than the typical life scenarios, such as a car, a bar, a restaurant, home, office, subway, meeting moderate.

In particular, this step may comprise the steps of: the user voice data and background noise model of the pre-match points to get the first matching value;

The data of the background noise and background noise model matching the preset value to obtain a second division match.

Step 303: according to the matching results to determine whether the speech signal meets a preset signal requirements.

This step may include the above two matching results obtained by the matching process to match the score, this time to determine specific process may include:

With the default value of the first matching algorithm score and the second match scores were calculated match score of the speech signal; match scores to judge the speech signal is greater than a predetermined threshold, if it is determined to meet the requirements of the preset signal, if not, it is determined that the signal does not meet

录只能通过密码，用户手动删除，不被系统格式化删除)， 紧急文件生成时会以短信息的形式通知用户设置的紧急联系人。

场景二：当监听到用户说"OK，LetGo！"，手机根据外界噪音环境是否类 似汽车载内环境，根据陀螺仪和加速度传感器检测手机的动能位移增量变化，根据光感器检测是否手机屏幕置于非遮挡状态，为用户直接开启中兴驾驶助手。 用户拿起手机下车时，根据噪音环境和传感器动量检测的变化自动关闭驾驶助 手。

场景三：幼儿英文发音练习：通过自定义唤醒词提前录制小朋友正学习中 的英文语句，其成功识别后对应设置儿童喜欢的特色应用或登陆模式。一个很喜欢玩手机游戏的小朋友，家长通过自定义唤醒词关联手机儿童模式，为了引 导小朋友在游戏的同时巩固练习英文发音，可以将正在学习中的小朋友发音和 语句预置为唤醒词，只有小朋友记忆正确并发音标准后才可以进入儿童模式进 行游戏。不经意间规范小朋友的行为，提高学习热情。

其他场景：也可以通过手机在线服务或者内置不同的歇后语，对联，或时 尚俏皮话供用户选择录制，当用户说了对应的时尚俏皮话的内容时，允许用户。

本实施例方法提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒 率，实现趣味个性化的语音应用场景实现。

实施例三：

如图7所示，本实施例提供了一种语音唤醒装置，其特征在于，包括：采 样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块； 所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音 信号的唤醒词识别结果和声纹特征并存储；

所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、 以及所述声纹特征与存储的声纹特征的匹配结果；

所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

the preset requirements, and if not, do not seek to meet preset signal.

For example two match scores were weighted to obtain a matching score, the match points and then determine whether the value exceeded the threshold of confidence, if it is judged to meet the requirements of the preset signal.

Consider further voice terminal power consumption when the wake-up, the present embodiment may be composed of a main audio processor to complete voice wake-up process, prior sound sample characteristics (wake-up word recognition results and wake voiceprint word speech signals) to download the audio processor storage area; when the terminal is in a dormant state in a semi-sleep of the audio processor standby mode, the mic event controlled wake of its work; when detecting wake mic audio multicore processors sound input to the implementation of voice wake-up; this can make the main terminal processing in the sleep state, the low-power audio processor in response to external input and processing of sound information, save power terminals, improve end-of standby time. priority, said audio processor core can be taken step-wise process, Core1 mic for processing a received signal received by the external sound a wake-up event sounds preliminary analysis, which determines the voice signal meets the preset requirement, Core2 users to follow speech recognition and authentication authentication, that is, whether to wake up. Wake When mic input Core1 speech is detected, the preset step required Core1 execute judgment speech signal meets; when Core1 determined to be, wake Core2, and voice signals to the Core2 (preferentially, Core1 voice signal stored in Core1,2 shared storage area extracted by the Core2); Core2 wake after performing the above steps 103-106, to wake up to identify, if adopted by the Core2 processor to the main terminal (for example terminal ARM chips) and transfer the hardware interrupt event, terminal host processor receives an interrupt wake-up event after the switch from sleep mode to normal mode normal operation.

It should be understood: the speech recognition result can also be matched to the present embodiment or voiceprint matching process on Core1 performed only when the match is successful only wake Core2, and then Core2 wake

Petitioner Exhibit 1004, Page 956 of 1518

所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

本实施例提供的语音唤醒装置，以终端用户自己声音作为声音样本，获取 终端用户个体的声音特征(即识别结果和声纹特征)作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同(同 一用户的发声方式相同)，避免了用户个体的发声方式差异于声音样本用户的发 声方式导致唤醒失败，提供了唤醒率；另外，本实施例装置中加入了语音识别 结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环 境无法唤醒终端，降低了误唤醒率；本实施例的装置还可以供用户自定义唤醒 词，满足用户个性化的需求，提升了用户的体验。

进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本 实施例装置可实现只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

优先地，如图8所示，在上述装置的基础上，本实施例语音唤醒装置还包 括：第二判断模块；

所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第 一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信 号的语音识别结果和声纹特征。

优先地，所述第二判断模块用于：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行 匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

优先地，如图9所示，在上述装置的基础上，本实施例唤醒装置中所述第 一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音 频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第 一核单元中；

所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是 否满足预设信号要求，若是，唤醒所述第二核单元；

up after a voiceprint match result of the execution or voice match.

In this embodiment, a multi-core audio processor core sub-step process designed to achieve savings in power terminals or electrical effects, enhance the user experience.

The following voice wake-up system shown in FIG. 4 will be specifically described embodiment the method of the present embodiment; FIG. 4, the system comprising: DSP chip (multicore DSP chip may include at least two nuclei), the main CPU terminal the ARM chip and microphone; wherein the DSP chip includes two nuclear units, and a wake-up storage unit (such as Flash), the first and second nuclear unit Core1 nuclear unit Core2; ARM chip includes a processing unit and a wake-up wake-up word word storage unit; specific process system shown in Figure 4 comprises a voice wakeup:

Step 501: ARM chip in the wake of word processing unit recording the user said wake-word speech signal, and generates a wake-word speech recognition result voiceprint feature will wake word speech recognition result is stored in the voiceprint feature wake term storage unit. can specifically be described with reference to the above-described generation method of FIG.

Step 502: ARM chip will wake word speech recognition result, wake-word speech signals voiceprint and a preset background noise model downloaded to the DSP chip, and stored in the wake storage unit.

Step 503: In the ARM chip is in sleep mode, the dynamic monitoring external microphone voice input, sends wake-up event when the wake-up Core1 voice signal is detected.

Step 504: Core1 of detected voice signals are processed to obtain user speech data and background speech data.

Step 505: Core1 voice data to the user and the background noise model to obtain a first preset matching match score, the background noise model and the background noise data to obtain a first pre-matching two match score;

Step 506: Core1 using preset algorithm to score the first and second match score of the match

所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的 语音识别结果和声纹特征。

优先地，本实施例中所述第一判断模块用于当所述语音信号的语音识别结 果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

本实施例装置提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒 率。

如图10所示，本实施例还提供了一种终端，包括处理模块和如上任一所 述的语音唤醒装置；

所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音 信号的语音识别结果执行对应的预设操作。

优先地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果 产生预设类型的唤醒中断事件唤醒所述终端；

所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

优先地，本实施例中处理模块集成在ARM芯片；所述语音唤醒装置集成在 低功耗音频芯片上。

本实施例提供的终端提高用户个体的语音唤醒率，降低噪音和他人声音的 误唤醒率，实现趣味个性化的语音应用场景。

以上内容是结合具体的实施方式对本发明所作的进一步详细说明，不能认 定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干简单推演或替换， 都应当视为属于本发明的保护范围。

points value, calculated match score of the speech signal.

Step 507: Core1 determination of the speech signal match score is greater than a predetermined threshold, If YES, step 508, and if not, step 512 is performed.

Step 508: Core1 voice signal to wake up at the same time Core2 Core2.

Step 509: After Core2 wake of the speech signal for speech recognition speech recognition result obtained, Wake word recognition results and stored to match, if the match is successful, step 510, if failed, step 513.

Step 510: Core2 extracted voiceprint of the speech signal, with stored voiceprint match, if successful, step 511, and if not, step 513.

Step 511: Core2 send a wakeup interrupt events to wake ARM ARM chip chip.

Step 512: Core1 back to sleep, DSP chip to sleep.

Step 513: Core2 back to sleep, wake voice flow ends.

There are differences DSP chip Core1 Core2 and voice processing capabilities of different calculation power only when the DSP chip Core1 work because of the limited computing power and clock speed and processing power when it is operating in a relatively low-power DSP job status (Normal). If the DSP chip Core2 work when the computational processing power and high power consumption is relatively high, when the Core2 wakes DSP chip in the Work state. But wake DSP chip running at full speed work state power is much smaller than the main CPU chip operating power terminal Normal state.

This embodiment of the operating state of the DSP chip can be divided into low-power operating state and Core1 relatively high energy-consuming operation Core2 run state. As can be seen from Figure 5, even in a complex environment of strong noise, not only to clear voice input feature, Core2 will not wake up, make DSP chip in a stable low-power state. Even if the person is awakened Core2 voice input is not custom instructions to wake, or wake-up instruction is not recorded wake-up feature user input, mobile phone main CPU chip will not be awakened. Low-power DSP

chip hardware ladder level wake-up mechanism, the only of individual authentication mechanisms to ensure the cell phone voice wake of power control, and reduce the error rate in the wake.

Example

embodiment of the present application, the custom with voiceprint wake term basis can reduce the power of intelligent speech detection machine control optimization problems, the better adapt to the individual needs of special groups or individuals, in at the same time easy to operate, allowing users to voice control closer to people's real life usage scenarios.

Second Embodiment:

To further enhance the user experience, after the wake-up terminal, the line can also be fast and automatic response to diversity, such as opening applications, etc., on the basis of the present embodiment, the above-described voice wake-up method, there is provided a terminal processing The method, as shown in Figure 6, comprising the steps of:

Step 601: Use a voice wake-up method described in Example wake the terminal.

Detailed description of a process of Reference Example

particular wake.

Step 602: After waking up in the terminal, according to a speech recognition result of the speech signal should perform the preset action. above step 602 can be done at the end of the main processor.

Example of the present embodiment may be preset to automatically defines the wake-word corresponding to the response action, different wake-up word may correspond to different response actions; after the user wakeup word wake terminals via voice, the terminal performs according to the identification result of user input speech signal corresponding response operations, such as automatic call or chat software to automatically open or turn driving mode. Here is the text of the wake-up word, but also the recognition result of said user text. priority, when the wake-up through an interrupt event to wake the terminal cases, you can set different wake-up

word corresponds to a different wake-up interrupt events, if a plurality of user pre-recorded wake-up word, different word corresponding to different wake wakeup interrupt events, terminal to different wake-up interrupt events, perform different preset operation or architecture different scenarios. Such reached even when the phone user can also awakened dormant black instruction by different users, for quick and automatic response diversity. For example, after a user-defined wake-up word is ''call'' in the dormant terminal, the terminal issues when the user says ''Call'' corresponds to the wake-up interrupt events to wake up to itself, and in the wake of the terminal according to a preset call the phone number.

Therefore, the above method in step 601 voice wake wake of the terminal process comprising: generating a preset type of wake-up interrupt events to wake the terminal according to the recognition result of the speech signal;

602 steps above include: wake-up according to the type of interrupt event executes the corresponding preset operation.

The following are applications of the present embodiment method specific scenarios:

Scene One: If the user record ''emergency SOS, help, Mom'' when such a desperate cry for help and user instructions related to oral language of the word as a wake-up, giving the user a default emergency wake-setting method. Allows the user to set the emergency number, even with strong noise data, the sensitivity value of this instruction the first time to assist users to make an emergency call when the user can not manually call, or send an emergency message to the recipient at the same time open the background recording or camera functions stored in the file directory, emergency, and in wifi network environment uploaded cloud disk (this directory can only be deleted by password, user manual, are not deleted formatted system), emergency file is generated in the form of short messages Notify the user to set emergency contact.

Scene Two: When listening to the user says ''OK, LetGo!'', A mobile phone according to the external environment is similar to the noise inside the car contained environment, according to the gyroscope and acceleration sensor to detect the phone's kinetic displacement incremental change, according to the light sensor to detect Are mobile

phone screen into a non-blocking state for the user to directly open ZTE driving assistant. When the user picks up the phone off, according to changes in the environment and noise detection sensor automatically turns off the momentum driving assistants.

Scene Three: Children practice English pronunciation: wake-word recording by customizing the children are learning English in the statement in advance, after the success of its children like identifying the corresponding set of characteristic application or landing mode. A very like to play mobile games of children, parents, custom wake-up word association Kid Mode phone, in order to guide the children in the game, while the consolidation exercise English pronunciation can be children are learning the pronunciation and statements preset wake-up word, only children memory and correct pronunciation standards before it can enter the game mode for children. Inadvertently regulate children's behavior, increase enthusiasm for learning.

Other scenarios: You can also phone or online service built different twisters, couplets, or when the recording is still one-liners for users to choose, when the user says the fashion pun content corresponding time, allowing the user.

This embodiment of a method to improve the individual user's voice wake-up rates, lower error rates wake-noise sound and others, to achieve interesting personalized voice Scenario.

Third Embodiment:

7, the present embodiment provides a wake-speech apparatus, characterized by comprising: a sampling module, a detection module, a first acquisition module, a second acquisition module, and a first judgment module wakeup module; the sampling module is used to wake-word speech signal acquisition end-user issue, get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

The detection module for the terminal is in a dormant state, the detection of the external input speech signal;

The first acquisition module for acquiring speech recognition result of the speech signal voiceprint feature;

The second acquisition module for acquiring the speech recognition result wakeup word recognition result and storage, as well as matches the voiceprint and voiceprint stored;

The first module is used for judging according to the matching results to determine whether to wake the terminal; in the wake of the first judging module for judging module is YES, the wake-up terminal.

During voice wake, since the sound;

Example provide voice wake present embodiment means for end-users own voice as a sound sample, end-users get the sound characteristics of individuals (ie, the recognition result voiceprint characteristic) as an audio sample characteristics Sample features voiceprint and voiceprint wakeup extraction (same same user sound pattern), to avoid the user individual sound pattern difference in the sound sample user sound pattern caused wake-up failed to provide a wake-up rate; in addition, this embodiment of the device is added to the speech recognition result voiceprint matching characteristics, can effectively avoid the noisy environment caused by waking up, so that the noise environment can not wake up the terminal, reducing the error rate wake; apparatus of the present embodiment can also provide the user custom wake term, to meet the needs of individual users, to enhance the user experience.

Further, since different users different sound pattern is different, that voiceprint, the application of the present embodiment of the device can be realized only to wake up my end user terminal, compared with the prior art, but also to improve the information security of the terminal.

Second judging module; : preferentially, 8, on the basis of the above means, in this case voice wake embodiment further includes means Before After

in the second determining module for external detection module input speech signal, the first acquisition module acquiring the speech recognition result of the speech signal and a voiceprint characteristic, judging whether the speech signal satisfies preset signal requirements; the first module for obtaining said second determination means determines YES, obtaining a speech recognition result of the speech signal and a voiceprint characteristic. priority, the second judging module for: of the speech signal is

processed to obtain user data and voice background noise data; the user data and voice data of the background noise and background noise models are preset to match; determine whether the speech signal according to the matching result satisfies preset signal requirements. preferentially, shown in Figure 9, in the above apparatus, the apparatus of the present embodiment in the wake of the first acquisition module, a second acquisition module, judging module, and the first wake-up module is integrated in the multi-core audio processor second nuclear unit; the second judgment module integrated in the multi-core audio processor's first core units;

The second determination module is used after the first wake-up core unit, it determines whether the speech signal is a signal complies with the preset requirement, and if so, waking the second core means;

The first acquisition module for, after said second core means wake, the speech signal to obtain a speech recognition result and a voiceprint characteristic. preferentially, in this case the judgment of the first voice recognition module for the speech signal when the result of the stored speech the voiceprint of the speech signal wake-up and implementation of word recognition result matches, and the stored voice wake-up word voiceprint match is successful, it is determined that the terminal needs to wake up.

Example apparatus of the present embodiment to improve the individual user's voice wake-up rates, lower error rate noise and wake others sound.

10, the present embodiment also provides a terminal includes a processing module, and took office as a wake-up means of said speech;

After the processing module for the voice wake wake the terminal apparatus, performs the corresponding operation according to the preset speech recognition result of the speech signal. priority, the voice wake-up means for generating a wake-up module preset type of wake-up interrupt events to wake the terminal according to the recognition result of the speech signal;

The processing module is used to perform the type of interrupt events corresponding operation according to the preset wake.

Example processing module

priority, the present embodiment is integrated in the ARM chip; the voice wake-up device is integrated in the low-power audio chip.

Terminal

This embodiment provides improved user individual voice wake-up rates, lower error rates wake-noise sound and others, to achieve interesting personalized voice scenarios.

The above is described in further detail with reference to specific embodiments of the present invention is made, the final version does not recognize the specific embodiment of the invention is only limited to these instructions. Those of ordinary skill in the art of the present invention, without departing from the inventive concept premise, you can also make a number of simple deduction or replacement, should be regarded as belonging to the scope of the present invention.

Cited by: CN105810196 A ; CN105898065 A ; CN106131292 A ; CN106297777 A ; CN106448663 A ; CN106502649 A ; WO18086033 A1 ;

(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 105575395 A
(43)申请公布日 2016.05.11

(21)申请号 201410541444.6

(22)申请日 2014.10.14

(71)申请人 中兴通讯股份有限公司
地址 518057 广东省深圳市南山区高新技术产业园科技南路中兴通讯大厦

(72)发明人 王婕

(74)专利代理机构 深圳鼎合诚知识产权代理有限公司 44281
代理人 薛祥辉 李发兵

(51)Int.Cl.
G10L 17/22(2013.01)
G10L 15/22(2006.01)

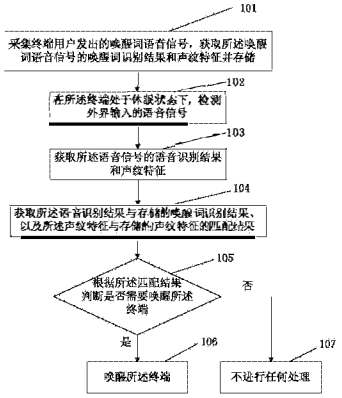
权利要求书3页 说明书12页 附图6页

(54)发明名称

语音唤醒方法及装置、终端及其处理方法

(57)摘要

本发明公开了一种语音唤醒方法及装置、终端及其处理方法。本发明的语音唤醒方法包括：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；在所述终端处于休眠状态下，检测外界输入的语音信号；获取所述语音信号的语音识别结果和声纹特征；获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端；本发明的语音唤醒方法能够解决现有语音唤醒技术存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题。



CN 105575395 A

Petitioner Exhibit 1004, Page 965 of 1518

1. 一种语音唤醒方法，其特征在于，包括如下步骤：

采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

在所述终端处于休眠状态下，检测外界输入的语音信号；

获取所述语音信号的语音识别结果和声纹特征；

获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端。

2. 如权利要求 1 所述的方法，其特征在于，在检测外界输入的语音信号之后，获取所述语音信号的语音识别结果和声纹特征之前，还包括：

判断所述语音信号是否满足预设信号要求，若是，获取所述语音信号的语音识别结果和声纹特征。

3. 如权利要求 2 所述的方法，其特征在于，所述判断所述语音信号是否满足预设信号要求的步骤包括：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

4. 如权利要求 3 所述的方法，其特征在于，所述将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配的步骤包括：

将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值；

将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

所述根据匹配结果判断所述语音信号是否满足预设信号要求的步骤包括：

采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值；

判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设信号要求，若否，则判定不满足预设信号要求。

5. 如权利要求 1-4 任一项所述的方法，其特征在于，所述根据所述匹配结果判断是否需要唤醒所述终端的步骤包括：

当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

6. 如权利要求 2-4 任一项所述的方法，其特征在于，所述判断所述语音信号是否满足预设信号要求的步骤包括：

多核音频处理器中的第一核单元唤醒后，所述第一核单元判断所述语音信号是否满足预设信号要求；

所述获取所述语音信号的语音识别结果和声纹特征的步骤包括：

当所述第一核单元判断为是时，所述第一核单元唤醒所述多核音频处理器中的第二核单元；

所述第二核单元获取所述语音信号的语音识别结果和声纹特征；

Petitioner Exhibit 1004, Page 966 of 1518

所述获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果的步骤包括：

所述第二核单元获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

所述根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端的步骤包括：

所述第二核单元根据所述匹配结果判断是否需要唤醒所述终端，若是，则所述第二核单元唤醒所述终端。

7. 如权利要求 6 所述的方法，其特征在于，还包括：

当所述第一核单元判断为否时，所述第一核单元进入休眠状态；

和 / 或

当所述第二核单元判断为否时，所述第二核单元进入休眠状态。

8. 如权利要求 6 所述的方法，其特征在于，所述第二核单元唤醒所述终端的步骤包括：

所述第二核单元发送中断事件唤醒所述终端中处于休眠状态的主处理器。

9. 一种终端处理方法，其特征在于，包括如下步骤：

利用权利要求 1-5 任一项所述的语音唤醒方法唤醒所述终端；

所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设操作。

10. 如权利要求 9 所述的方法，其特征在于，所述唤醒所述终端的步骤包括：

根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

所述语音信号的语音识别结果执行预设的操作的步骤包括：

根据所述唤醒中断事件的类型执行对应的预设操作。

11. 一种语音唤醒装置，其特征在于，包括：采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块；

所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

12. 如权利要求 11 所述的装置，其特征在于，还包括：第二判断模块；

所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信号的语音识别结果和声纹特征。

13. 如权利要求 12 所述的装置，其特征在于，所述第二判断模块用于：

对所述语音信号进行处理得到用户语音数据和背景噪声数据；

Petitioner Exhibit 1004, Page 967 of 1518

将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

根据匹配结果判断所述语音信号是否满足预设信号要求。

14. 如权利要求 12 或 13 所述的装置，其特征在于，所述第一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是否满足预设信号要求，若是，唤醒所述第二核单元；

所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的语音识别结果和声纹特征。

15. 如权利要求 11-13 任一项所述的装置，其特征在于，所述第一判断模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

16. 一种终端，其特征在于，包括：处理模块和如权利要求 11-13 任一项所述的语音唤醒装置；

所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音信号的语音识别结果执行对应的预设操作。

17. 如权利要求 16 所述终端，其特征在于，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

Petitioner Exhibit 1004, Page 968 of 1518

语音唤醒方法及装置、终端及其处理方法

技术领域

[0001] 本发明涉及语音识别技术领域，尤其涉及一种语音唤醒方法及装置、终端及其处理方法。

背景技术

[0002] 语音在手机终端中已不单单存在于单独的程序应用中，而开始慢慢的作为一种独立于手指触摸输入的新增输入方式，不需触碰手机，遥控和更智能化友好的服务于手机用户。

[0003] 而语音唤醒技术为用户解放双手完全的语音指令遥控手机提供了前提，唤醒主流技术是技术软件唤醒，但软件运行在手机系统启动为前提，为保证能随时随地接收到用户的语音指令，语音引擎需要一直后台运行和监听，手机系统将无法进入休眠待机的节电状态，此类语音唤醒技术的应用对电量消耗异常强烈。目前语音技术在市场技术应用中在不单单考虑到功能友好性，还要关注手机电量消耗，同时要求对用户使用中的不被语音过多打扰，所以目前各厂商对于语音识别系统的唤醒方案多数采用休眠状态下硬件按键唤醒或者手机系统运行时软件按钮点击唤醒。

[0004] 随着技术人员的不断改进，目前已经出现语音低功耗唤醒技术，在手机休眠状态下应用低功耗的芯片实现直接用语音指令唤醒手机操作系统，无需用户手动进行硬件和软件操作。

[0005] 但对于现有技术的语音低功耗唤醒功能中，只能使用固定的唤醒词，通过语音模型生成此预置固定唤醒词的声学模型。具体地现有语音低功耗唤醒功能的实现方法，包括如下：

[0006] 录制一些大批量的“选定唤醒词”声音样本进行处理和人群声学特征培训和提取。

[0007] 将统计估算得出的“固定的唤醒词”声音特征预置在手机系统中。

[0008] 根据用户个体输入的声音波形与统计生成的声音特征进行比对。

[0009] 因低功耗性能音频芯片中运算处理能力有限，群体样本统计生成的声学模型在单独用户个体使用时会有以下问题：

[0010] 1、用户个体的发声方式差异性远远低于或超出采样统计模型特征提取的置信区间，造成用户唤醒率低。如果样本量太小或者或人群特征太单一，例如，全是普通话，全是成年人，会造成有些特定人群的声音例如儿童，老人的声音对唤醒失效。

[0011] 2、在噪音环境下复杂噪音环境生成了采样统计模型中类似的似然曲线，造成误唤醒率高。不该唤醒时乱唤醒，造成用户困扰。

[0012] 3、由于只能通过厂家事先录制的“固定的唤醒词”来唤醒终端，用户不能

[0013] 够自定义唤醒词，不能满足用户个性化的需求，降低了用户体验。

[0014] 因此，现有技术虽然实现了一定功耗控制下的语音唤醒，但是存在用户唤醒率低、误唤醒率高以及用户体验差的技术等问题。

Petitioner Exhibit 1004, Page 969 of 1518

发明内容

[0015] 本发明要解决的主要技术问题是，提供一种语音唤醒方法及装置、终端及其处理方法，能够解决现有语音唤醒技术存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题。

[0016] 为解决上述技术问题，本发明提供一种语音唤醒方法，包括如下步骤：

[0017] 采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

[0018] 在所述终端处于休眠状态下，检测外界输入的语音信号；

[0019] 获取所述语音信号的语音识别结果和声纹特征；

[0020] 获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0021] 根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端。

[0022] 进一步地，所述方法在检测外界输入的语音信号之后，获取所述语音信号的语音识别结果和声纹特征之前，还包括：判断所述语音信号是否满足预设信号要求，若是，获取所述语音信号的语音识别结果和声纹特征。

[0023] 进一步地，所述判断所述语音信号是否满足预设信号要求的步骤包括：

[0024] 对所述语音信号进行处理得到用户语音数据和背景噪声数据；

[0025] 将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

[0026] 根据匹配结果判断所述语音信号是否满足预设信号要求。

[0027] 进一步地，所述将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配的步骤包括：

[0028] 将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值；

[0029] 将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

[0030] 所述根据匹配结果判断所述语音信号是否满足预设信号要求的步骤包括：

[0031] 采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值；

[0032] 判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设信号要求，若否，则判定不满足预设信号要求。

[0033] 进一步地，所述根据所述匹配结果判断是否需要唤醒所述终端的步骤包括：

[0034] 当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

[0035] 进一步地，所述判断所述语音信号是否满足预设信号要求的步骤包括：

[0036] 多核音频处理器中的第一核单元唤醒后，所述第一核单元判断所述语音信号是否满足预设信号要求；

[0037] 所述获取所述语音信号的语音识别结果和声纹特征的步骤包括：

[0038] 当所述第一核单元判断为是时，所述第一核单元唤醒所述多核音频处理器中的第二核单元；

Petitioner Exhibit 1004, Page 970 of 1518

[0039] 所述第二核单元获取所述语音信号的语音识别结果和声纹特征；

[0040] 所述获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果的步骤包括：

[0041] 所述第二核单元获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0042] 所述根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端的步骤包括：

[0043] 所述第二核单元根据所述匹配结果判断是否需要唤醒所述终端，若是，则所述第二核单元唤醒所述终端。

[0044] 进一步地，所述方法还包括：

[0045] 当所述第一核单元判断为否时，所述第一核单元进入休眠状态；

[0046] 和/或

[0047] 当所述第二核单元判断为否时，所述第二核单元进入休眠状态。

[0048] 进一步地，所述第二核单元唤醒所述终端的步骤包括：

[0049] 所述第二核单元发送中断事件唤醒所述终端中处于休眠状态的主处理器。

[0050] 同样为了解决上述的技术问题，本发明还提供了一种终端处理方法，包括如下步骤：

[0051] 利用上述前五项中任一项所述的语音唤醒方法唤醒所述终端；

[0052] 所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0053] 进一步地，所述唤醒所述终端的步骤包括：

[0054] 根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0055] 所述语音信号的语音识别结果执行预设的操作的步骤包括：

[0056] 根据所述唤醒中断事件的类型执行对应的预设操作。

[0057] 同样为了解决上述的技术问题，本发明还提供了一种语音唤醒装置，包括：采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块；

[0058] 所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

[0059] 所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

[0060] 所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

[0061] 所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0062] 所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

[0063] 所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

[0064] 进一步地，所述装置还包括：第二判断模块；

[0065] 所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

[0066] 所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信号的语音识别结果和声纹特征。

Petitioner Exhibit 1004, Page 971 of 1518

[0067] 进一步地，所述第二判断模块用于：

[0068] 对所述语音信号进行处理得到用户语音数据和背景噪声数据；

[0069] 将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

[0070] 根据匹配结果判断所述语音信号是否满足预设信号要求。

[0071] 进一步地，所述第一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

[0072] 所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是否满足预设信号要求，若是，唤醒所述第二核单元；

[0073] 所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的语音识别结果和声纹特征。

[0074] 进一步地，所述第一判断模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

[0075] 同样为了解决上述的技术问题，本发明还提供了一种终端，包括处理模块和上述所述的语音唤醒装置；

[0076] 所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0077] 进一步地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0078] 所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

[0079] 本发明的有益效果是：

[0080] 本发明提供了一种语音唤醒方法及装置、终端及其处理方法，能够提高用户唤醒率、降低误唤醒率和提升用户体验。具体地，本发明的语音唤醒方法包括：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；在所述终端处于休眠状态下，检测外界输入的语音信号；获取所述语音信号的语音识别结果和声纹特征；获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；根据所述匹配结果判断是否需要唤醒所述终端，若是，则唤醒所述终端；本发明的语音唤醒方法，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征（即识别结果和声纹特征）作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同（同一用户的发声方式相同），避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本发明的方法加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

[0081] 另外，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

Petitioner Exhibit 1004, Page 972 of 1518

附图说明

[0082] 图 1 为本发明实施例一提供的一种语音唤醒方法的流程示意图；

[0083] 图 2 为本发明实施例一提供的一种获取用户自定义语音唤醒词样本特征的流程示意图；

[0084] 图 3 为本发明实施例一提供的一种判断语音信号是否满足预设信号要求的流程示意图；

[0085] 图 4 为本发明实施例一提供的一种语音唤醒系统的结构示意图；

[0086] 图 5 为本发明实施例一提供的一种语音唤醒系统进行语音唤醒的流程示意图；

[0087] 图 6 为本发明实施例二提供的一种终端处理方法的流程示意图；

[0088] 图 7 为本发明实施例三提供的第一种语音唤醒装置的结构示意图；

[0089] 图 8 为本发明实施例三提供的第二种语音唤醒装置的结构示意图；

[0090] 图 9 为本发明实施例三提供的第三种语音唤醒装置的结构示意图；

[0091] 图 10 为本发明实施例三提供的一种终端的结构示意图。

具体实施方式

[0092] 下面通过具体实施方式结合附图对本发明作进一步详细说明。

[0093] 实施例一：

[0094] 考虑到现有固定唤醒词语音唤醒方法存在的用户唤醒率低、误唤醒率高以及用户体验差的技术问题；本实施例提供了一种可以提高终端用户唤醒率、降低误唤醒率、提升用户体验的语音唤醒方法，如图 1 所示，本实施例的语音唤醒方法包括如下步骤：

[0095] 步骤 101：采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储。

[0096] 本实施例方法可以采用语音采集装置，例如麦克风来采集终端用户发出的语音信号。本实施例方法中获取唤醒词识别结果和声纹特征的时序不限，可以先获取识别结果，后获取声纹特征，也可以同时获取。

[0097] 本实施例中唤醒词识别结果是对唤醒词语音信号进行语音识别得到的唤醒词文字内容或者命令。本实施例中声纹特征是语音信号中可以代表用户发声方式的参数，例如声音频率、响度等等。

[0098] 本步骤为进行语音唤醒之前的准备步骤，发生在用户自定义语音唤醒词的过程中，本步骤可以由终端的 AP 侧执行，即终端的主芯片，例如手机主 CPU 中 ARM 芯片来完成。

[0099] 为了提高语音唤醒的精确性，本实施例方法在采集样本特征的阶段，可以多次录制用户发出的唤醒词语音信号，然后根据多次录制的唤醒词语音信号得到最优的唤醒词识别结果和声纹特征。如图 2 所示，为终端获取用户自定义语音唤醒词样本特征的过程，由终端的 AP 侧完成：

[0100] 在终端启动用户自定义唤醒词设置之后，第一次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的唤醒词识别结果 1 和声纹特征 1，将识别结果 1 和声纹特征 1 存储至自定义唤醒词存储区中；

[0101] 终端第二次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的唤醒词识别结果 2 和声纹特征 2，当唤醒词识别结果 2 与唤醒词识别结果 1 相同时，利用声纹特征

Petitioner Exhibit 1004, Page 973 of 1518

2 优化声纹特征 1，当唤醒词识别结果 2 与唤醒词识别结果 1 不相同时，重新录制；进入第三次录制或者从第一次开始重新录制；

[0102] 终端第三次录制用户所说的唤醒词语音信号，然后生成唤醒词语音信号的唤醒词识别结果 3 和声纹特征 3，当唤醒词识别结果 3 与唤醒词识别结果 1 相同时，利用声纹特征 3 优化上述已经优化过的声纹特征 1；

[0103] 依次类推，直到第 n 次录制完成。图 2 中 n 取 4，本实施例方法中 n 的取值可以根据实际需求进行设定。

[0104] 步骤 102：在所述终端处于休眠状态下，检测外界输入的语音信号。

[0105] 本步骤中可以由终端的麦克风来检测外界的语音信号，在终端休眠状态时，麦克风可以保持工作状态。

[0106] 本步骤中终端处于休眠状态下从硬件的角度可以包括：终端的主处理器处于休眠状态，从软件的角度可以包括：终端的操作系统处于休眠或锁定状态。

[0107] 步骤 103：获取所述语音信号的语音识别结果和声纹特征。

[0108] 本步骤中获取语音识别结果和声纹特征没有时序限制，可以先获取语音识别结果，后获取声纹特征，也可以先获取声纹特征，后获取语音识别结果，还可以同时获取语音识别结果和声纹特征。

[0109] 本步骤中语音信号的语音识别结果是对语音信息进行语音识别得到的文字内容或者命令。本实施例中声纹特征是语音信号中可以代表用户发声方式的参数，例如声音频率、响度等等。

[0110] 步骤 103-106 为语音唤醒过程中主要步骤，其可以在一个低功耗音频芯片执行，例如低功耗的 DSP 芯片。在进行语音唤醒之前，事先将上述获取的唤醒词识别结果和唤醒词语音信号的声纹特征下载到低功耗音频芯片的存储区，为语音唤醒做准备。

[0111] 步骤 104：获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果。

[0112] 本步骤中获取语音识别结果与存储的唤醒词识别结果的匹配结果方式包括：将语音识别结果与存储的唤醒词识别结果进行匹配，获取所述声纹特征与存储的声纹特征的匹配结果的方式包括：将获取的声纹特征与存储的声纹特征进行匹配；本步骤中获取所述语音结果与存储的唤醒词识别结果匹配结果，与获取声纹特征与存储的声纹特征的匹配结果过程时序不受限制，可以先执行前者，后执行后者，或者同时执行前者和后者。

[0113] 本实施例中的匹配结果包括：匹配成功与否，或者匹配分值；其中根据匹配分值是否在预设范围内来判定是否匹配成功。

[0114] 步骤 105：根据所述匹配结果判断是否需要唤醒所述终端，若是，执行步骤 106，若否，则执行步骤 107。

[0115] 具体地，当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端；否则，判断不需要唤醒所述终端。

[0116] 本步骤需要判断语音识别结果是否匹配成功和声纹特征是否匹配，这两个判断过程的时序可以任意，具体包括以下情况：

[0117] 1、先判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果是否

Petitioner Exhibit 1004, Page 974 of 1518

匹配成功，若是，则再判断所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情况下均判定不需要唤醒终端。

[0118] 2、先判断所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否匹配成功，若是，则再判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果是否匹配成功，若是，则判定需要唤醒所述终端；判断为否的情况下均判定不需要唤醒终端。

[0119] 3、同时判断所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果是否匹配成功、以及所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征是否匹配成功，若均成功，则判定需要唤醒所述终端；若有一个匹配不成功则判定不需要唤醒终端。

[0120] 本步骤中匹配成功可以包括：匹配分值在预设范围内，或者相似度在预设范围之内。

[0121] 步骤 106：唤醒所述终端。

[0122] 本步骤可以通过发送中断事件唤醒终端；当语音唤醒步骤在低功耗音频芯片上执行时，低功耗音频芯片判断需要唤醒终端后，会上传中断事件给终端的主处理器唤醒处于休眠状态下的主处理器。

[0123] 步骤 107：不进行任何处理。

[0124] 本实施例方法中当判断不需要唤醒终端时，可以忽略当前输入的语音信号，执行唤醒步骤的芯片或器件可以进入休眠状态，以节省终端功耗。

[0125] 本实施例提供的语音唤醒方法，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征（即识别结果和声纹特征）作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同（同一用户的发声方式相同），避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本实施例方法加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本发明的方法还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

[0126] 进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例方法只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

[0127] 考虑到噪声环境对用户输入的语音信号的影响，进一步提高用户唤醒率降低误唤醒率，本实施例方法在上述步骤 102 之后，步骤 103 之前还可以包括：

[0128] 判断所述语音信号是否满足预设信号要求，若是，则执行步骤 103。

[0129] 具体地，可以判断输入的语音信号的信噪比是否在预设范围内，若是，则满足语音信号要求，或者语音信号的其他质量参数是否在预设范围内，若是，则满足语音信号要求，或者将输入的语音信号的声纹特征用用较简单的波形匹配，若成功，则满足信号要求。该步骤可以在低功耗音频芯片上完成。

[0130] 本实施例方法在判断语音信号不满足预设信号要求时，可以输入的语音信号丢弃；或者当该判断步骤是在低功耗音频芯片上置信时，该芯片进入休眠状态，以节省电能。

[0131] 优先地，如图 3 所示，上述判断所述语音信号是否满足预设信号要求的过程可以包括：

[0132] 步骤 301：对所述语音信号进行处理得到用户语音数据和背景噪声数据。

[0133] 由于外界输入的语音信号包括用户自己发出的语音信号和当前用户环境产生的

Petitioner Exhibit 1004, Page 975 of 1518

噪声信号，所以在进行信号判断时，可以将语音信号分离成用户语音数据和背景噪声数据。

[0134] 步骤 302：将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配。

[0135] 本步骤中用户语音数据和背景噪声数据匹配时序不受限制，可以先后，也可以同时。

[0136] 本实施例方法可以预置一些较典型的生活场景下的稳态或非稳态噪声声学统计模型，例如，汽车中，酒吧里，餐馆，家庭，办公室，地铁上，会议中等。

[0137] 具体地，本步骤可以包括如下步骤：

[0138] 将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值；

[0139] 将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值。

[0140] 步骤 303：根据匹配结果判断所述语音信号是否满足预设信号要求。

[0141] 本步骤中的匹配结果可以包括上述两个匹配过程得到的匹配分值，此时具体地判断过程可以包括：

[0142] 采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值；

[0143] 判断所述语音信号的匹配分值是否大于预设阈值，若是，则判定满足预设信号要求，若否，则判定不满足预设信号要求，若否，则不满足预设信号求。

[0144] 例如对两个匹配分值进行加权处理得到一个匹配分值，然后判断该匹配分值是否超出置信阀值，若是，则判定满足预设信号要求。

[0145] 进一步考虑语音唤醒时终端的功耗问题，本实施例可以由一个音频处理器来完成语音唤醒的主要过程，事先将声音样本特征（唤醒词识别结果和唤醒词语音信号的声纹特征）下载到该音频处理器的存储区；在终端处于休眠状态时该音频处理器处于半 sleep 待机状态，由 mic 事件控制其工作唤醒；当 mic 检测到声音输入时唤醒多核音频处理器，执行语音唤醒；这样可以使得终端的主处理在休眠状态时，通过低功耗的音频处理器响应和处理外界输入的声音信息，节省终端的功耗，提高终端的待机时间。

[0146] 优先地，上述音频处理器可以采取分步式核心处理，Core1 用于处理接收到 mic 接收到外部声音信号的唤醒事件声音初步分析，即判断语音信号是否满足预设要求，Core2 用户后续的语音识别以及鉴权认证，即判断是否唤醒。当 mic 检测到语音输入时唤醒 Core1，Core1 执行判断语音信号是否满足预设要求的步骤；当 Core1 判断为是时，唤醒 Core2，并将语音信号传输给 Core2（优先地，Core1 将语音信号存储在 Core1、2 共享的存储区中，由 Core2 提取）；Core2 唤醒后执行上述步骤 103-106，完成唤醒鉴别，若通过，则由 Core2 向终端主处理器（例如终端 ARM 芯片）传送硬件中断事件，终端主处理器接收到中断事件后被唤醒由 sleep 模式切换到 Normal 模式正常运行。

[0147] 应当理解的是：本实施例还可将语音识别结果匹配或者声纹匹配的过程放在 Core1 中执行，只有当匹配成功时才唤醒 Core2，然后 Core2 唤醒之后执行声纹匹配或者语音结果匹配。

[0148] 本实施例通过多核音频处理器的分步式核心处理设计，可以实现在节省终端功耗或电能的效果，提升了用户体验。

[0149] 下面通过如图 4 所示的语音唤醒系统来具体介绍本实施例方法；如图 4 所示，该系

Petitioner Exhibit 1004, Page 976 of 1518

统包括：DSP 芯片（多核 DSP 芯片，可以包括至少两个核）、终端主 CPU 中的 ARM 芯片和麦克风；其中 DSP 芯片包括两个核单元以及一个唤醒存储单元（例如 Flash），第一核单元 Core1 和第二核单元 Core2；ARM 芯片包括唤醒词处理单元和唤醒词存储单元；图 4 所示的系统进行语音唤醒的具体过程包括：

[0150] 步骤 501：ARM 芯片中唤醒词处理单元录制用户所说的唤醒词语音信号，并生成唤醒词语音识别结果和声纹特征，将唤醒词语音识别结果和声纹特征存储在唤醒词存储单元中。

[0151] 具体地可以生成方式可以参考上述图 2 的描述。

[0152] 步骤 502：ARM 芯片将唤醒词语音识别结果、唤醒词语音信号的声纹特征和预置的背景噪声模型下载到 DSP 芯片，并存储在唤醒存储单元中。

[0153] 步骤 503：在 ARM 芯片处于休眠状态下，麦克风动态监测外界的语音输入，当检测到语音信号时发送唤醒事件唤醒 Core1。

[0154] 步骤 504：Core1 对检测到的语音信号进行处理得到用户语音数据和背景语音数据。

[0155] 步骤 505：Core1 将所述用户语音数据与所述预置的背景噪声模型进行匹配得到第一匹配分值，将所述背景噪声数据与所述预置的背景噪声模型进行匹配得到第二匹配分值；

[0156] 步骤 506：Core1 采用预设分值算法对所述第一匹配分值和所述第二匹配分值进行计算得到所述语音信号的匹配分值。

[0157] 步骤 507：Core1 判断所述语音信号的匹配分值是否大于预设阈值，若是，则执行步骤 508，若否，则执行步骤 512。

[0158] 步骤 508：Core1 将语音信号传输给 Core2 同时唤醒 Core2。

[0159] 步骤 509：Core2 唤醒后，对所述语音信号进行语音识别获得语音识别结果，并与存储的唤醒词识别结果进行匹配，若匹配成功，则执行步骤 510，若失败，则执行步骤 513。

[0160] 步骤 510：Core2 提取所述语音信号的声纹特征，并与存储的声纹特征进行匹配，若成功，执行步骤 511，若否，执行步骤 513。

[0161] 步骤 511：Core2 发送唤醒中断事件给 ARM 芯片唤醒 ARM 芯片。

[0162] 步骤 512：Core1 重新休眠，DSP 芯片休眠。

[0163] 步骤 513：Core2 重新休眠，语音唤醒流程结束。

[0164] DSP 芯片中 Core1 和 Core2 的语音算处理能力不同功耗有差异，当 DSP 芯片只有 Core1 工作时因为其运算能力和主频和处理能力有限，其运转时是 DSP 处在相对低功耗工作状态（Normal）。如果 DSP 芯片 Core2 工作时其运算和处理能力较高，功耗相对较高，Core2 被唤醒时 DSP 芯片处于 Work 态。但唤醒 DSP 芯片全速运行 work 状态的功耗也是远小于终端主 CPU 芯片 Normal 态的运行功耗。

[0165] 本实施例可以将 DSP 芯片的运行状态分为低耗能 Core1 运行状态和相对高耗能运算的 Core2 运行态。从上图 5 可以看出，即使在复杂强噪音环境下，只要没有清晰人声特征输入，Core2 将不被唤醒，会使 DSP 芯片处于稳定低功耗状态。即使 Core2 被唤醒如果人声输入不是自定义的唤醒指令，或者唤醒指令不是录制唤醒特征的用户输入，手机主 CPU 芯片也不会被唤醒。通过硬件低功耗 DSP 芯片的阶梯层次的唤醒机制，个体化唯一化的鉴权

Petitioner Exhibit 1004, Page 977 of 1518

机制保证了手机语音唤醒的功耗控制，和误醒率的降低。

[0166] 应用本实施例方法，带有声纹特征的自定义唤醒词可以在降低了智能机中语音检测功耗控制优化难题的基础上，更好的适应特殊人群或者个体的个性化需求，在便捷操作的同时，让用户的语音控制更贴近于人们生活实际使用场景。

[0167] 实施例二：

[0168] 为进一步提升用户体验，在终端唤醒之后，还行可以进行多样性的快速和自动响应，例如打开应用程序等，本实施例在上述语音唤醒方法的基础上，提供了一种终端处理方法，如图 6 所示，包括如下步骤：

[0169] 步骤 601：利用实施例一所述的语音唤醒方法唤醒所述终端。

[0170] 具体唤醒过程参考实施例一的具体描述。

[0171] 步骤 602：所述终端在唤醒之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0172] 上述步骤 602 可以由终端中的主处理器完成。

[0173] 本实施例方法可以预先设置自动定义唤醒词对应的响应操作，不同的唤醒词可以对应不同的响应操作；在用户通过语音唤醒词唤醒终端之后，终端根据用户输入语音信号的识别结果执行对应的响应操作，例如自动打电话或者自动打开聊天软件或者开启驾驶模式等。这里的唤醒词是文字内容，识别结果也是用户所说的文字内容。

[0174] 优先地，当通过唤醒中断事件来唤醒终端情况下，可以设置不同的唤醒词对应不同的唤醒中断事件，如果一个用户预先录制多个唤醒词，不同唤醒词对应不同的唤醒中断事件，终端对不同的唤醒中断事件，执行不同的预设操作或者架构不同的应用场景。这样达成即使在手机黑屏休眠时用户也可以通过不同的用户唤醒指令，进行多样性的快速和自动响应。例如用户自定义唤醒词为“打电话”时，在终端处于休眠状态下，在用户说出“打电话”时终端发出对应的唤醒中断事件来唤醒自身，并且终端在唤醒之后会根据预设的电话号码拨打电话。

[0175] 因此，上述步骤 601 语音唤醒方法中唤醒所述终端的过程包括：根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0176] 上述步骤 602 具体包括：根据所述唤醒中断事件的类型执行对应的预设操作。

[0177] 下面列举应用本实施例方法的具体场景：

[0178] 场景一：如果用户录制“急救 SOS，救命，妈妈”此类和用户情急时呼救口语化相关的指令做为唤醒词时，给用户一个默认的紧急唤醒的设置方法。允许用户设置的应急号码，即使在强噪音数据下，此指令的敏感度分值在用户无法手动呼叫的时候第一时间协助用户拨打出急救电话，或者发送出急救信息给接收者，同时后台开启录音或摄像功能存放在紧急文件目录中，并在 wifi 环境下上传网络云盘（此目录只能通过密码，用户手动删除，不被系统格式化删除），紧急文件生成时会以短信息的形式通知用户设置的紧急联系人。

[0179] 场景二：当监听到用户说“OK，Let Go！”，手机根据外界噪音环境是否类似汽车载内环境，根据陀螺仪和加速度传感器检测手机的动能位移增量变化，根据光感器检测是否手机屏幕置于非遮挡状态，为用户直接开启中兴驾驶助手。用户拿起手机下车时，根据噪音环境和传感器动量检测的变化自动关闭驾驶助手。

[0180] 场景三：幼儿英文发音练习：通过自定义唤醒词提前录制小朋友正学习中的英文

Petitioner Exhibit 1004, Page 978 of 1518

语句，其成功识别后对应设置儿童喜欢的特色应用或登陆模式。一个很喜欢玩手机游戏的小朋友，家长通过自定义唤醒词关联手机儿童模式，为了引导小朋友在游戏的同时巩固练习英文发音，可以将正在学习中的小朋友发音和语句预置为唤醒词，只有小朋友记忆正确并发音标准后才可以进入儿童模式进行游戏。不经意间规范小朋友的行为，提高学习热情。

[0181] 其他场景：也可以通过手机在线服务或者内置不同的歇后语，对联，或时尚俏皮话供用户选择录制，当用户说了对应的时尚俏皮话的内容时，允许用户。

[0182] 本实施例方法提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒率，实现趣味个性化的语音应用场景实现。

[0183] 实施例三：

[0184] 如图 7 所示，本实施例提供了一种语音唤醒装置，其特征在于，包括：采样模块、检测模块、第一获取模块、第二获取模块、第一判断模块和唤醒模块；所述采样模块用于采集终端用户发出的唤醒词语音信号，获取所述唤醒词语音信号的唤醒词识别结果和声纹特征并存储；

[0185] 所述检测模块用于在所述终端处于休眠状态下，检测外界输入的语音信号；

[0186] 所述第一获取模块用于获取所述语音信号的语音识别结果和声纹特征；

[0187] 所述第二获取模块用于获取所述语音识别结果与存储的唤醒词识别结果、以及所述声纹特征与存储的声纹特征的匹配结果；

[0188] 所述第一判断模块用于根据所述匹配结果判断是否需要唤醒所述终端；

[0189] 所述唤醒模块用于在所述第一判断模块判断为是时，唤醒所述终端。

[0190] 本实施例提供的语音唤醒装置，以终端用户自己声音作为声音样本，获取终端用户个体的声音特征（即识别结果和声纹特征）作为声音样本特征；在进行语音唤醒时，由于声音样本特征中声纹特征与唤醒时提取的声纹特征相同（同一用户的发声方式相同），避免了用户个体的发声方式差异于声音样本用户的发声方式导致唤醒失败，提供了唤醒率；另外，本实施例装置中加入了语音识别结果和声纹特征的匹配，可以有效地避免噪声环境导致的误唤醒，使得噪声环境无法唤醒终端，降低了误唤醒率；本实施例的装置还可以供用户自定义唤醒词，满足用户个性化的需求，提升了用户的体验。

[0191] 进一步，由于不同用户的发声方式是不同的即声纹特征不同，所以应用本实施例装置可实现只有终端用户本人才能唤醒终端，与现有技术相比，还提高了终端的信息安全。

[0192] 优先地，如图 8 所示，在上述装置的基础上，本实施例语音唤醒装置还包括：第二判断模块；

[0193] 所述第二判断模块用于在检测模块检测外界输入的语音信号之后，所述第一获取模块获取所述语音信号的语音识别结果和声纹特征之前，判断所述语音信号是否满足预设信号要求；

[0194] 所述第一获取模块用于在所述第二判断模块判断为是时，获取所述语音信号的语音识别结果和声纹特征。

[0195] 优先地，所述第二判断模块用于：

[0196] 对所述语音信号进行处理得到用户语音数据和背景噪声数据；

[0197] 将所述用户语音数据和所述背景噪声数据分别与预置的背景噪声模型进行匹配；

Petitioner Exhibit 1004, Page 979 of 1518

[0198] 根据匹配结果判断所述语音信号是否满足预设信号要求。

[0199] 优先地，如图 9 所示，在上述装置的基础上，本实施例唤醒装置中所述第一获取模块、第二获取模块、所述第一判断模块和所述唤醒模块集成在多核音频处理器的第二核单元中；所述第二判断模块集成在所述多核音频处理器的第一核单元中；

[0200] 所述第二判断模块用于在所述第一核单元唤醒之后，判断所述语音信号是否满足预设信号要求，若是，唤醒所述第二核单元；

[0201] 所述第一获取模块用于在所述第二核单元唤醒之后，获取所述语音信号的语音识别结果和声纹特征。

[0202] 优先地，本实施例中所述第一判断模块用于当所述语音信号的语音识别结果与所述存储的语音唤醒词识别结果匹配成功，且所述语音信号的声纹特征与所述存储的语音唤醒词声纹特征匹配成功时，则判定需要唤醒所述终端。

[0203] 本实施例装置提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒率。

[0204] 如图 10 所示，本实施例还提供了一种终端，包括处理模块和如上任一所述的语音唤醒装置；

[0205] 所述处理模块用于在所述语音唤醒装置唤醒所述终端之后，根据所述语音信号的语音识别结果执行对应的预设操作。

[0206] 优先地，所述语音唤醒装置中唤醒模块用于根据所述语音信号的识别结果产生预设类型的唤醒中断事件唤醒所述终端；

[0207] 所述处理模块用于根据所述唤醒中断事件的类型执行对应的预设操作。

[0208] 优先地，本实施例中处理模块集成在 ARM 芯片；所述语音唤醒装置集成在低功耗音频芯片上。

[0209] 本实施例提供的终端提高用户个体的语音唤醒率，降低噪音和他人声音的误唤醒率，实现趣味个性化的语音应用场景。

[0210] 以上内容是结合具体的实施方式对本发明所作的进一步详细说明，不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干简单推演或替换，都应当视为属于本发明的保护范围。

Petitioner Exhibit 1004, Page 980 of 1518

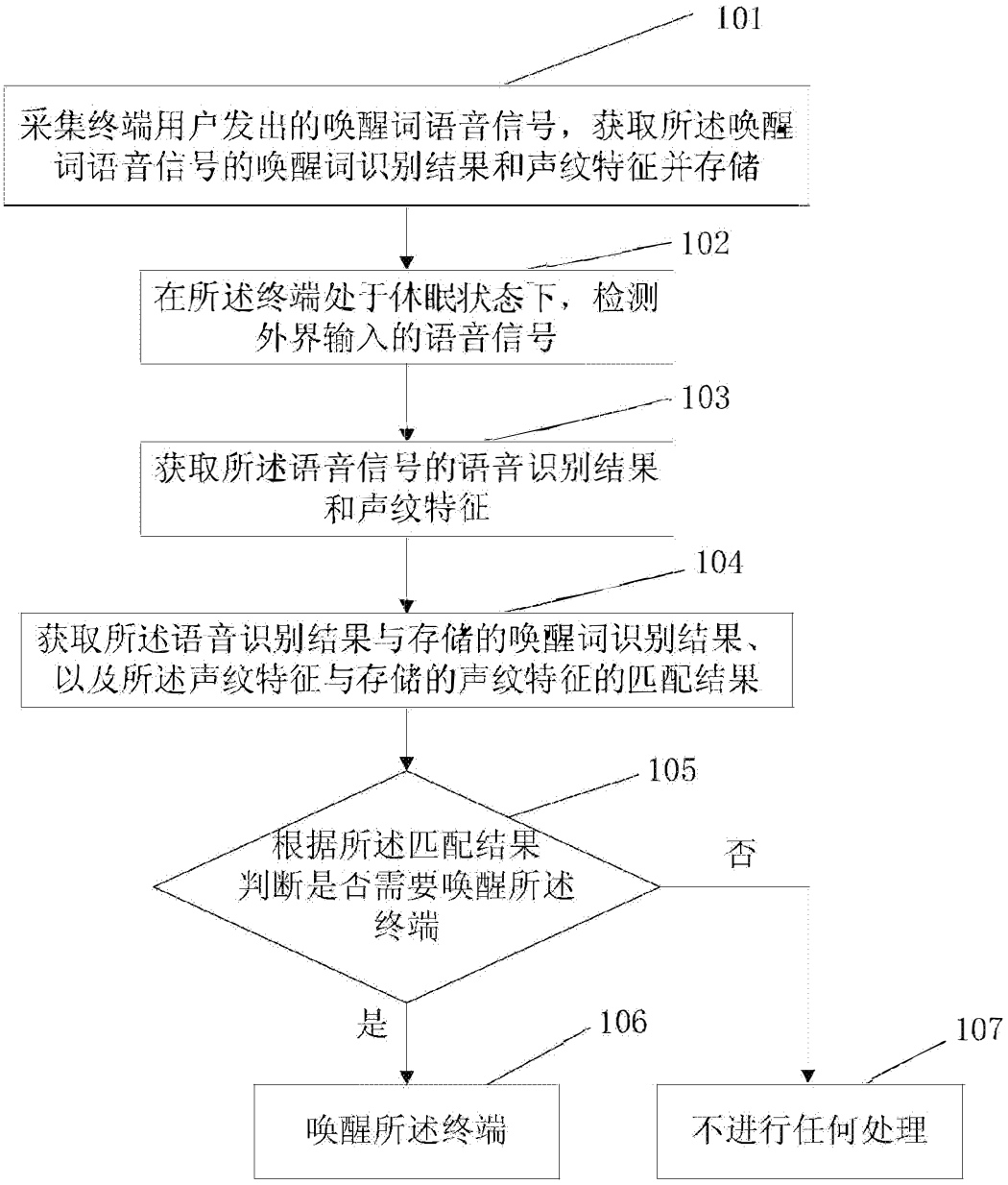


图 1

Petitioner Exhibit 1004, Page 981 of 1518

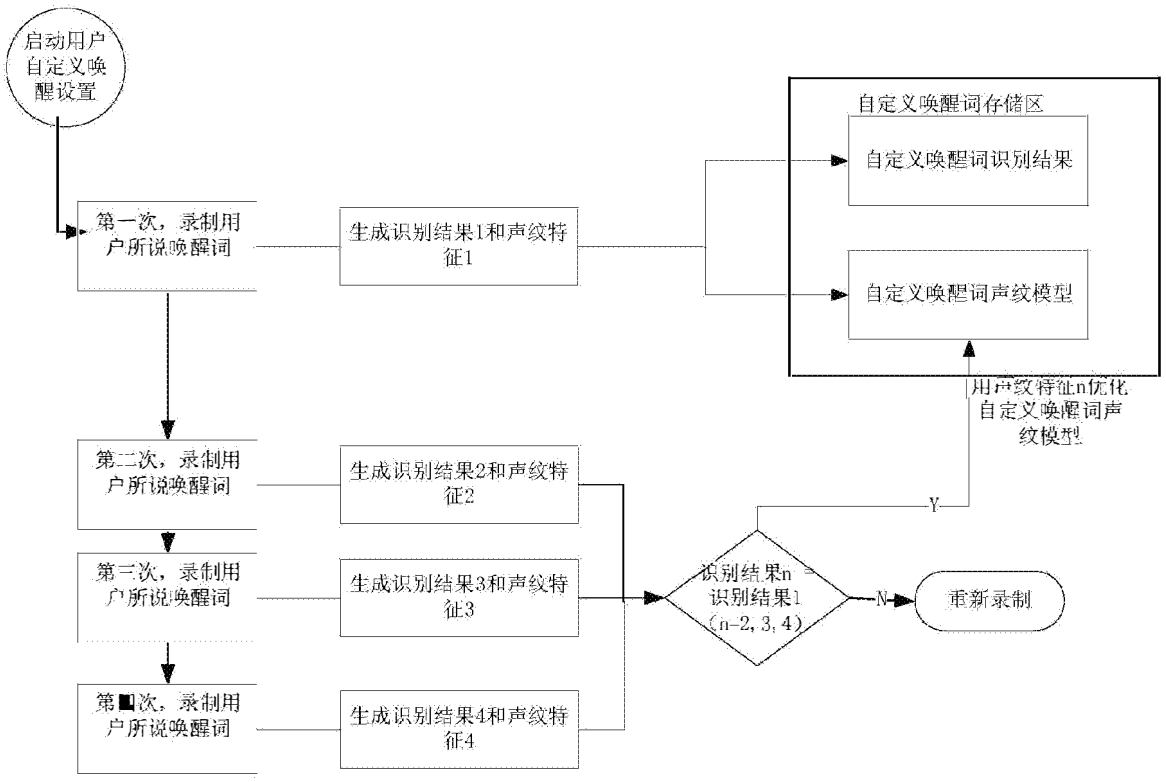


图 2

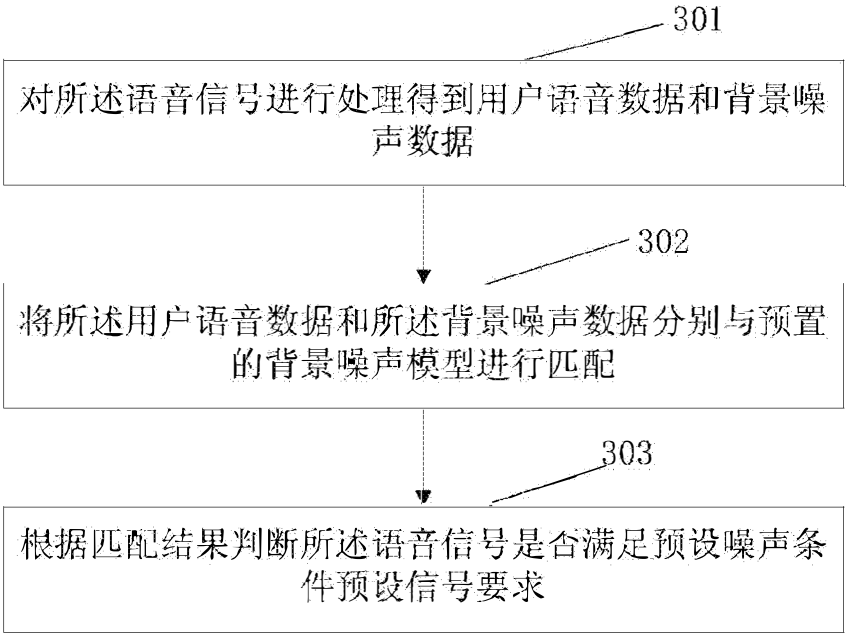


图 3

Petitioner Exhibit 1004, Page 982 of 1518

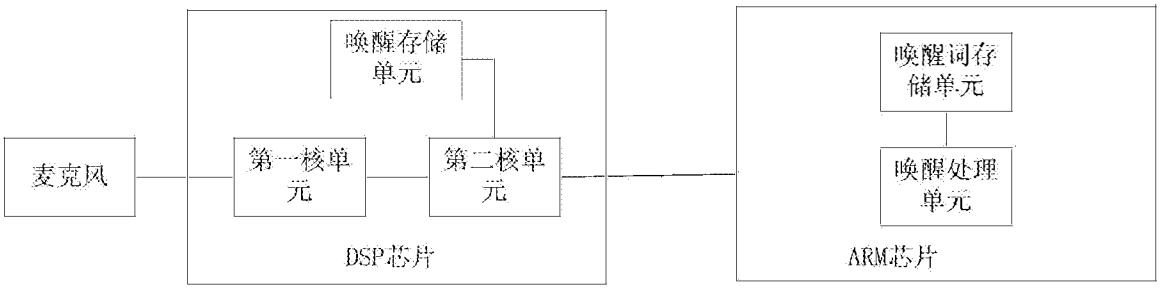


图 4

Petitioner Exhibit 1004, Page 983 of 1518

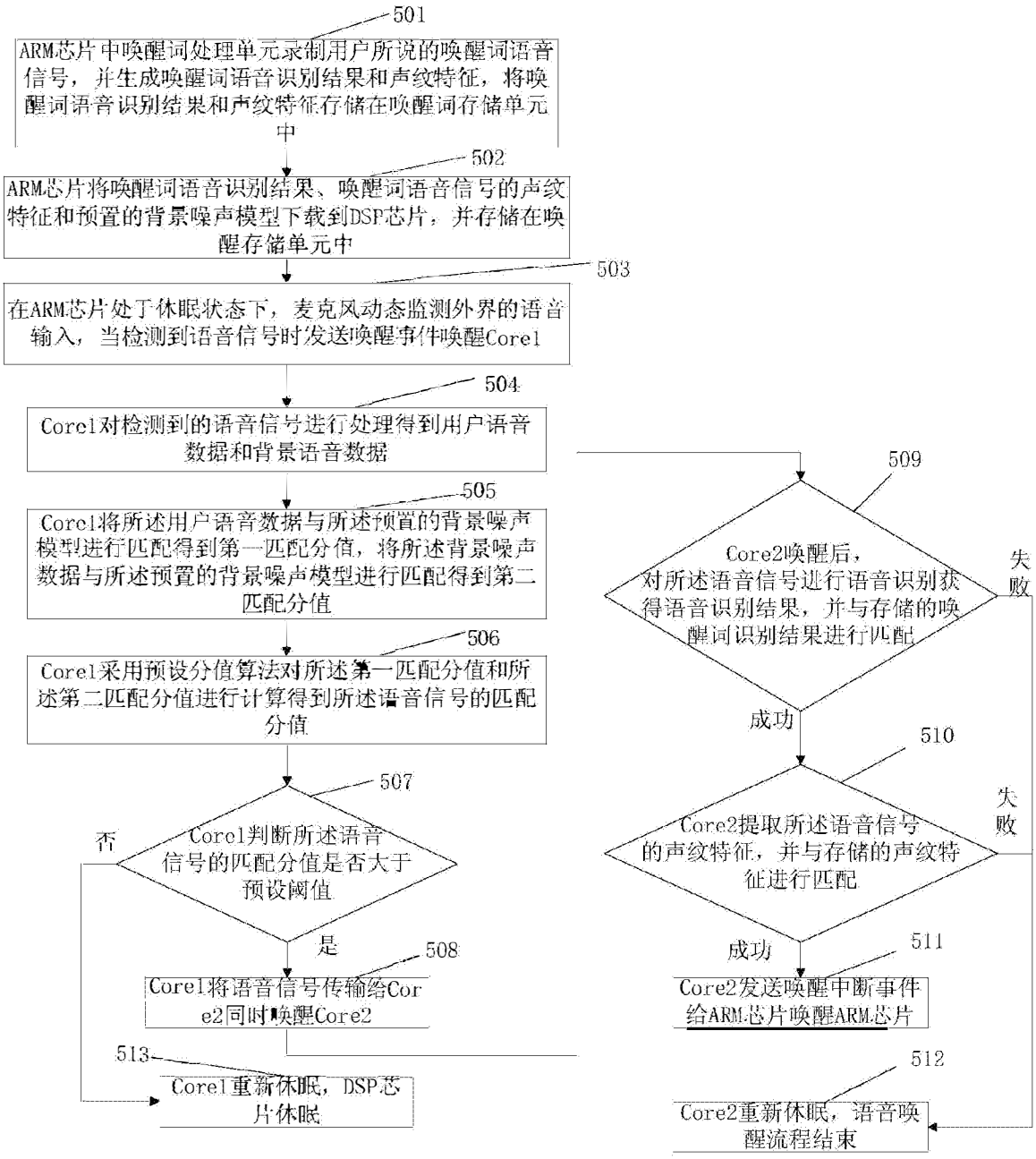


图 5

Petitioner Exhibit 1004, Page 984 of 1518

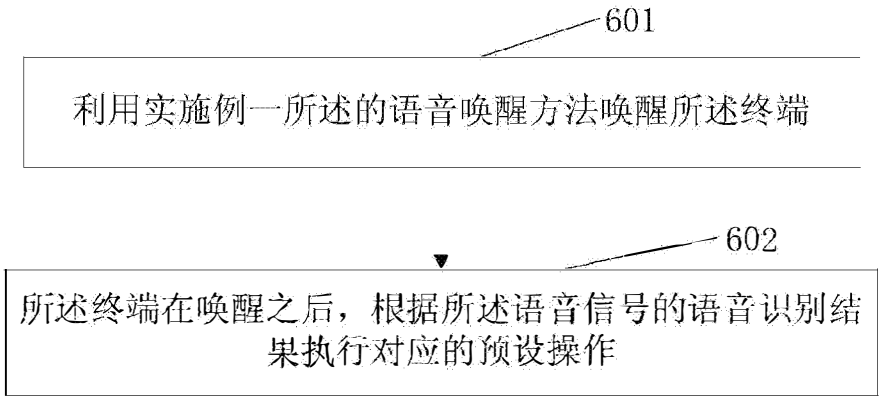


图 6

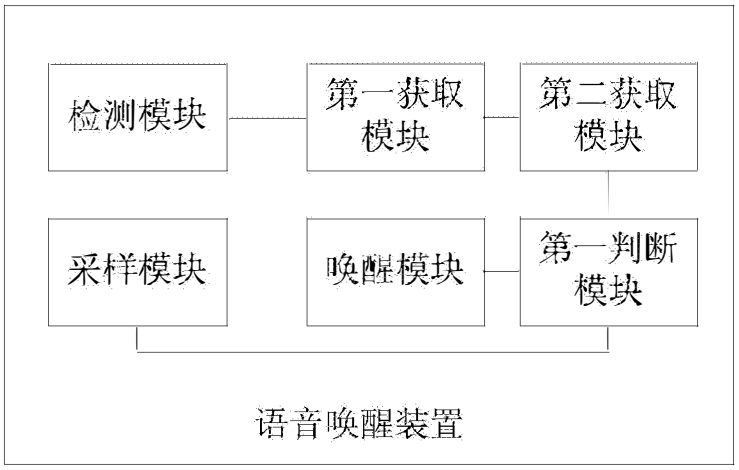


图 7

Petitioner Exhibit 1004, Page 985 of 1518

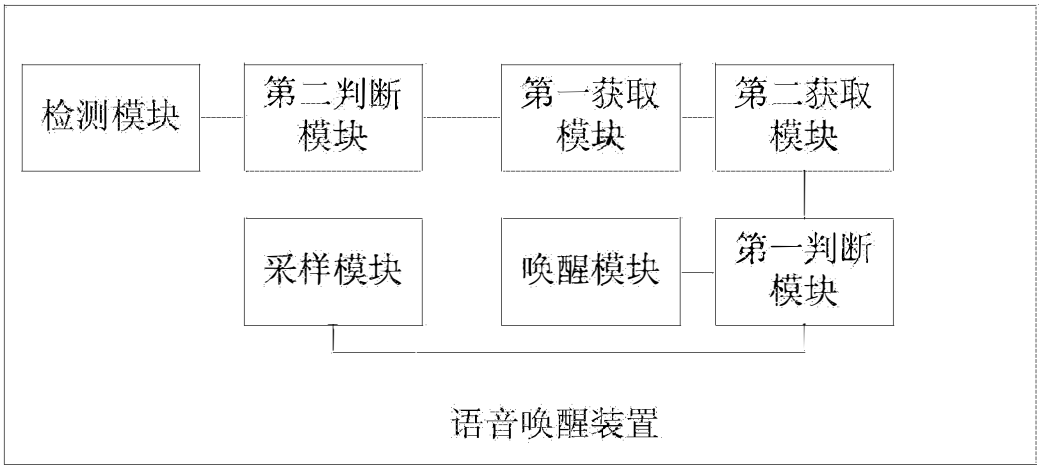


图 8

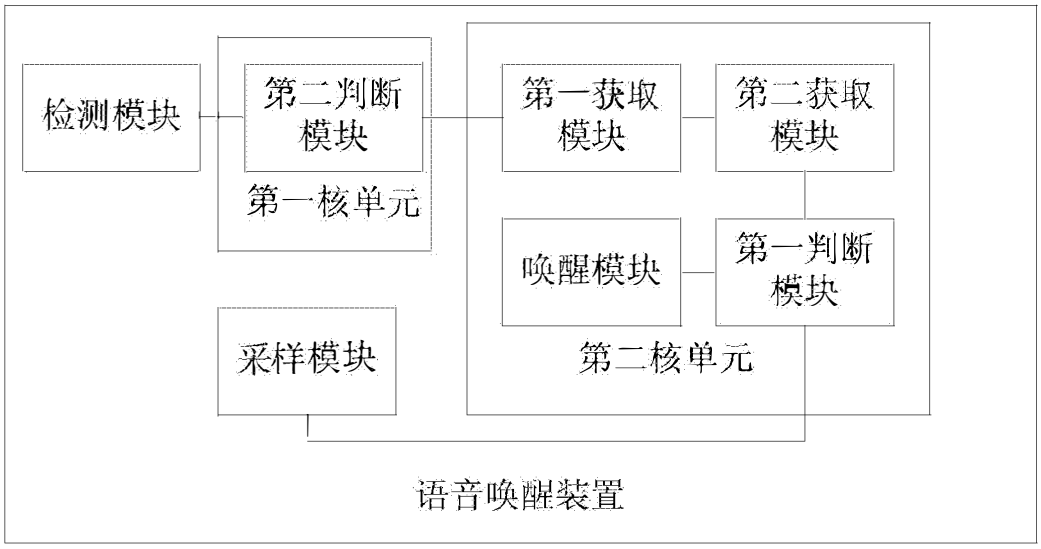


图 9

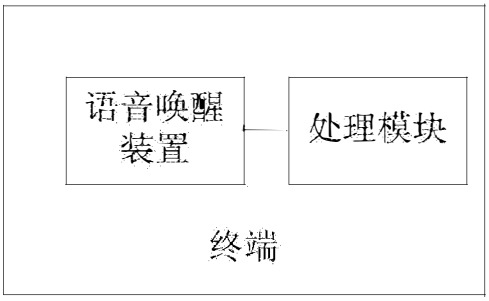


图 10

Petitioner Exhibit 1004, Page 986 of 1518

Publication number: CN101650943 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20100217
Application number: CN200810239892
Application date: 20081219
Priority: CN200810239892 20081219 CN20081239892;
Assignee: OF ACOUSTICS CHINESE ACADEMY OF SCIENCES INST ;
Assignee[std]: CHINESE ACAD INST ACOUSTICS ;
Inventor: YAN YONGHONG ; ZHANG QINGQING ; PAN JIELIN ;
Inventor[std]: QINGQING ZHANG ; JIELIN PAN ; YONGHONG YAN ;
International class[8]: G10L15/065; G10L17/06;
Family members: CN101650943 A

Title: NON-NATIVE SPEECH RECOGNITION SYSTEM AND METHOD THEREOF

Title: 一种非母语语音识别系统及方法

Abstract:

The invention relates to a non-native speech recognition system based on mixed model state correction and a method thereof. The non-native speech recognition system comprises a non-native speech interface, a native model module, a non-native model module, a native state decoding module, a non-native state forced alignment module, a native-non-native state similarity matrix computation module, a native-non-native state mapping table computation module and a non-native state correction model decoding module. In the system and the method thereof, a non-native acoustic model is corrected at statelevels based on a native acoustic model of a speaker and state mapping among different models, thus obtaining a model that better meets non-native pronunciation characteristics. The system and the method thereof have the advantages of obvious improvement of recognition performance compared with that of a recognition system which is not corrected by the method on the premise of only the native training data without increase of any non-native speech training data, no obvious speed fall of the recognition speech of the system and very high practicability.

Abstract:

本发明涉及一种基于混合模型状态修正的非母语语音识别系统及方法。该系统包括：非母语语音接口、母语模型模块、非母语模型模块、母语状态解码模块、非母语状态强制对齐模块、母语与非母语状态相似度矩阵计算模块、母语与非母语状态映射表计算模块及非母语状态修正模型解码模块，该系统及方法通过不同模型间的状态映射，利用说话人

The present invention relates to a non-native speech recognition system and method based on hybrid model state amended. The system comprises: a non-native speech interfaces, up model module, non-native model module, decoding module native state, non-native status forcing alignment module, native and non-native

母语的声学模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点的模型。该系统及方法有如下优点：在不增加任何非母语语音训练数据，仅依靠说话人母语训练数据的前提下，相对未采用该方法修正过的识别系统的识别性能有明显提高；同时系统识别语音的速度并未明显降低，具有很高的实用性。

state similarity matrix calculation module, native and non-native status mapping table calculation module and non-native state correction model decoding module, the system and method of mapping between the state of different models, the use of mother tongue speaker acoustic models on a state level for non-native acoustic model is modified to give more in line with the characteristics of non-native pronunciation model. The system and method has the following advantages: without adding any non-native speech training data, relying only native speaker training data are available, the relative performance of the method is not used to identify corrected identification system has significantly improved; the same system Speech Recognition It did not significantly reduce the speed, with high availability.

Claims:

1、一种非母语语音识别系统，其特征在于，所述系统包括: 一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非 母语状态强制对齐模块和母语状态解码模块; 一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提 供母语声学模型; 一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解 码模块提供非母语声学模型; 一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码， 得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分 段信息送入母语与非母语状态相似度矩阵计算模块; 一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行 强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非 母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块; 一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和 非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重 合时间大于预先设定的门限值时，认为这两个状态出现一次"同现"，统计出所有的 "同现"情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并 将该相似度矩阵信息送入母语与非母语状态映射表计算模块; 一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射 表; 及 一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射 表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母 语声学模型;最终利用该修正后的非母语声学模型完成非母语语音识别。

1, a non-native speech recognition system, wherein the system comprises: a non-native speech interface for data collection of non-native speech, and the non-native voice data into non-Native Status force alignment module and decoding module up state; a native model module to provide up to up status acoustic model decoding module and non-native state correction model decoding module; a non-native model module, used to force alignment to non-native state modules and non-native state correction model decoding modules provide non-native acoustic models; a native state decoding module for up acoustic model standard non-native voice data is decoded to give up the speech state-level segment information, namely the status of native speech decoding information and state-level native speech segment information is fed up with the non-native state similarity matrix calculation module; a non-native status forcing alignment module for non-native acoustic models will be compulsory for non-native speech data aligned to give non- native speech state-level segment information, namely non-native speech state reference information and non-native speech state level segment information is fed up with the non-native state similarity matrix calculation module; a native and non-native state similarity matrix calculation module for native speech state-level segment information and non-native speech state-level segment information on the time alignment, when the time is up speech state coincides with the non-native speech state is greater than a

2、根据权利要求1所述的非母语语音识别系统，其特征在于，所述非母语语音 状态对应于母语语音状态的相似度矩阵通过以下公式得到: lJ N ^count(t_ n=^ -其中，M，N分别为中文和英文的状态数量;A_(M,N)为相似度矩阵;A"为矩 阵的第严行尸列元素»tj为非母语语音状态，^i为母语语音状态;A;;。A，，^(M,N)J ^=1 M, ^=l N^ C_ (^ l_)是母语语音状态_和非母语语音状态^j之 间的"同现"出现次数。

书 求 要 3 禾 权 3、根据权利要求1所述的非母语语音识别系统，其特征在于，所述状态映射表 通过以下方式得到: 如果A。矩阵中n_行尸列元素是尸列的第n大元素，则状态叫是状态尸的第 n候选修正状态;每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正 状态，其中，n<Mc 4、根据权利要求1所述的非母语语音识别系统，其特征在于，所述非母语状态 修正模型解码模块在语音识别解码过程中，对于观察值0。，用n个候选状态修正以 后的非母语状态尸的观察概率巧"_·)变为: p_o_ = op^"^o^)to_o)Z^ory_1^(_)，。" 其中，0表示非母语状态的修正权重，勺是相应第叫个候选状态，n是候选数; 巧"^P·)"和p，"(ot)，""分别表示非母语状态尸和母语状态竹在观察值0·下的原始观 察概率。

5、一种非母语语音识别方法，包括如下步骤: (1)非母语语音接口采集一定数量的非母语语音数据，用于得到模型状态映射 表; (2)非母语状态强制对齐模块用非母语声学模型对非母语语音数据进行强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息; (3)母语状态解码模块用标准母语声学模型对非母语语音数据进行解码，得到 母语语音状态级分段信息，即母语语音状态解码信息; (4)母语与非母语状态相似度矩阵计算模块将得到的非母语语音状态级分段信 息和母语语音状态级分段信息在时间上对齐，当某两个状态的重合时间大于预先设 定的门限值时，这两个状态出现一次"同现"; (5)母语与非母语状态相似度矩阵计算模块统计出所有的"同现"情况并计算 得到非母语语音状态对应于母语语音状态的相似度矩阵; (6)母语与非母语状态映射表计算模块根据该相似度矩阵得到状态映射表; (7)非母语状态修正模型解码模块根据得到的状态映射表，在语音识别的解码 过程中，用映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得 到修正后的非母语声学模型。

6、根据权利要求5所述的非母语语音识别方法，其特征在于，所述非母语语音 状态对应于母语语音状态的相似度矩阵通过以下公式得到: 书 求 要 3 禾 权 A¨=_unt(t_SJ ·JN 散。u"·(、。_) n=^ 一 其中，M，N分别为中文和英文的状态数量;A^，^(M,N)为相似度矩阵;A"为矩 阵的第产行尸列元素;tj为非母语

predetermined threshold value, that the two states once "with the current", the statistics of all "with the current" situation, and calculate non-native speech state corresponds to the state of native speech similarity matrix, and the similarity matrix information into the native and non-native state map calculation module; a native and non-native status mapping table calculation module for obtaining the status of a mapping table according to the similarity matrix calculation; and a non-native state correction model decoding module for decoding the speech recognition process, with the state up acoustic model state map found amend non-native acoustic models corresponding state, to be non-native acoustic models after correction; the final use of non-native acoustic models after the completion of the correction of non-native speech recognition.

2, according to non-native speech recognition system according to claim 1, wherein said non-native speech state corresponds to the state of native speech similarity matrix obtained by the following formula: lJ N count (t_ n = - where, M, N, respectively the number of state Chinese and English; A_ (M, N) for the similarity matrix; A "is the first serious Xingshi matrix column elements tj non-native voice state, i-speaking voice status;. A;; A,, (M, N) J = 1 M, = l N C_ (l_) is native and non-native speech state _ voice status j between "co-occurrence" occurrences.

3 Wo books seeking the right to play according to claim non-native speech recognition system of claim 1, characterized in that said state mapping table in the following manner: If A. N_ Xingshi matrix column elements are the major elements of the corpse n columns, then the state is called the state of the corpse of the first amendment n candidate status; the status of each non-native language can be found n candidate correction native language in the matrix state, wherein, n and lt; Mc 4, according to non-native speech recognition system according to claim 1, characterized in that the non-native state correction model decoding module decoding the speech recognition process, the observed value of zero. With n candidate status to amend non-native state after the corpse of observation probabilities clever "_) becomes: p_o_ = op " o) to_o) Z ory_1 (_),. "Where 0 represents the right to amend non-native state of heavy, spoon the corresponding section called candidate status, n is the number of candidates; clever" P) "and p,"

语音状态，_为母语语音状态;A。A，，t(^,N)， r=l M, ^=1 N^ C^un^ (_>)是母语语音状态_和非母语语音状态_之 ^间的"同现"出现次数。

7、根据权利要求5所述的非母语语音识别方法，其特征在于，所述状态映射表 通过以下万式得到: 如果A"矩阵中叫行川列元素是尸列的第n大元素，则状态n"是状态尸的第 n候选修正状态;每个非母语语言的状态都能在矩阵中找到母语语言的n个候选修正 状态，其中，n<Mc 8、根据权利要求5所述的非母语语音识别万法，其特征在于，所述非母语状态 修正模型解码模块在语音识别解码过程中，对于观察值0。，用n个候选状态修正以 后的非母语状态尸的观察概率巧"^0。)变为: p_) = ^p ^" (o^)^a_O_o)Z^o4y_)，"" 其中，0表示非母语状态的修正权重，%是相应第叫个候选状态，n是候选数; 巴"(^·)"和p，"(^。)"分别表示非母语状态尸和母语状态叫在观察值0·下的原始观 察概率。

(ot), "" denote non-native status corpse and Bamboo native state under observation 0.5 original observation probabilities.

5, a non-native speech recognition method comprising the steps of: (1) non-native speech interface collecting a certain number of non-native voice data, for obtaining a model state map; (2) non-native status force alignment block with non-native acoustic model for non-native speech data alignment forced to give non-native speech state-level segment information, namely the status of non-native speech reference information; (3) the native state standard decoding module up Acoustics model of non-native speech data is decoded to give up the speech state-level segment information, namely the state of native speech decoding information; and (4) native and non-native state similarity matrix calculation module and the resulting non-native speech state-level segment information and native speech state-level segment information in time alignment, a time when the two states coincide greater than a preset threshold, these two states appear a "co-occurrence"; (5) Native and non-native status similar matrix calculation module statistics of all the "co-occurrence" the situation and calculate the non-native speech state corresponds to the state of native speech similarity matrix; (6) native and non-native state map state maps obtained computing module based on the similarity matrix Table; (7) non-native state correction model decoding module based on the obtained state map, in the decoding process of speech recognition, the acoustic model with native state mapping table to find the corresponding non-modified up acoustic model state, to be corrected non-native acoustic models.

6, according to non-native speech recognition method as claimed in claim 5, characterized in that the non-native speech state corresponds to the state of native speech similarity matrix obtained by the following equation: Book 3 Wo seek playing right A&#xfffd; = _unt (t_SJ &#xfffd; JN San.u "&#xfffd; (,._) n = one where, M, N, respectively Chinese and English, the number of state; A, (M, N) the similarity matrix; A "is the first production Xingshi column matrix element; tj state non-native speech, native speech _ to state; A.A,, t (, N), r = l M, = 1 "The same is now" is the native speech state and non-native speech state _ _ between the number of occurrences of

N C un (_ and; gt).

7, according to non-native speech recognition method as claimed in claim 5, characterized in that said state map obtained by the following square formula: if A "column of the matrix element is called row Nakagawa corpse column The n-th largest element, the state n "is the state of the corpse of the first amendment n candidate status; the status of each non-native language can be found n the status of candidate native language of the amendment in the matrix, where, n and lt; Mc 8 The non-native speech recognition methodology set forth in claim 5, wherein said non-native state correction model decoding module decodes the speech recognition process, the observed value of 0., With non-native state n candidate status after correcting corpse observation probabilities clever " 0) becomes:. P_) = p " (o) a_O_o) Z o4y_), ""' where, 0 Fixed right non-native state of weight, percent is the corresponding section called candidate status, n is the number of candidates; Pakistan "(*)" and p, "(.)" denote non-native status corpse and native state is called the observation The original observation probabilities under 0.5.

Description:

_1_ 5 tFL明书 一种非母语语音识别系统及方法 技术领域 本发明涉及语音识别系统及方法，特别涉及基于混合模型状态修正的非母语语 音识别系统及方法。

背景技术 随着现代社会信息的全球化，非母语语音识别成为语音识别技术领域研究热点 之一。对非母语的语音识别相对母语的语音识别而言，识别性能将会大幅度降低， 特别是对于有着严重口音的说话人。如何在保证语音识别系统在仅有少量非母语训练数据的前提下，对带有不同程度口音的非母语语音识别性能的提高是非母语语音 识别研究的重点。

文献(Bohn.O_S., F^ege,J·E., nThe production Of new and Similar Vowels by adult German learners Of English."」[MM] Stud. Second Lang. Acquis. 14, 131-)指出， 非母语说话人在进行目标语言发音时，可能会用说话人自身的母语发音进行替代，或者产生出有说话人母语发音特点和目标语言发音特点相结合的发音。这个结论启 发我们，说话人自身的母语语言训练数据，可能对说话人非母语语音识别有帮助， 特别是针对那些发音口音较重的语音数据。

_1_ 5 tFL Ming book a non-native speech recognition system and method TECHNICAL FIELD The present invention relates to a speech recognition system and method, and more particularly to non-native speech recognition system and method based on hybrid model state amended.

BACKGROUND With globalization, non-native speech recognition has become one of modern society information technology research focus in the field of speech recognition. Non-native speech recognition relative native speech recognition, the recognition performance will be significantly reduced, especially for people with serious accent speak. How to ensure that the voice recognition system in only a small number of non-native language training under the premise of the data, speech recognition performance for non-native accents with varying degrees of improvement are non-native speech recognition research.

Document (Bohn.O_S., F ege, J ◆ E., NThe production Of new and Similar Vowels by adult German learners Of English. "" [MM] Stud. Second

当前，说话人自适应技术(如MAP^ ML^R)已广泛的应用于处理非母语语音 识别中(Z. Wang,T. Schultz.A. Waibel, wComparison Of acoustic model adaptation techniques On non-native Speech" 「C^, Proc^ ICASSP 2003.)。这类方法主要是通过对少 量非母语语音数据进行自适应，使得母语发音声学模型能够在一定程度上逼近非母 语的发音特点。在这些万法中，自适应数据和测试数据的相似性是决定识别性能好 坏的关键因素。尽管自适应技术在提高非母语语音识别性能上有贡献，但通过自适 应以后的模型在非母语语音识别中的性能相比母语语音模型在母语语音识别的性 能，仍然较低。文献(J. Humlphries, P. Woodland, and D. Pearce. "Using accent-specific pronunciation modeling for robust Speech recognition." [C] In Proc. ICSLP '96, pages， 2324-2327, Philadelphia, PA, October 1996·)研究了自适应算法的局限性，指出导致非 母语语音识别率低的主要原因来自于说话人母语语系中无法涵盖的非母语语言发 音。如何便非母语语音声学模型较好的模拟出这类语言发音是非母语语音识别研究 工作的重点。

书 明 说 发明内容 为了克服上述现有技术的不足，本发明提供一种基于状态修正的非母语语音识 别系统及万法，该系统及万法通过不同模型间的状态映射，利用说话人母语的声学 模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点 的模型。

为了实现上述目的，本发明提出的一种非母语语音识别系统，包括: 一非母语语音接口，用于采集非母语语音数据，并将该非母语语音数据送入非 母语状态强制对齐模块和母语状态解码模块。

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提 供母语声学模型。

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解 码模块提供非母语声学模型。

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码， 得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分 段信息送入母语与非母语状态相似度矩阵计算模块。

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数据进行 强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非 母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

Lang. Acquis. 14, 131-) He notes that non-native speaker pronunciation language during target may be replaced with the speaker's own Pronounced, or have speakers produce Pronounced features and characteristics of the combination of the target language pronunciation pronunciation. This conclusion Kai made us, the speaker's own native language training data may be non-native speaker voice recognition helpful, especially for those heavier voice data pronunciation accent.

Currently, the speaker adaptation techniques (such as MAP ML R) has been widely used in the treatment of non-native speech recognition (Z. Wang, T. Schultz.A. Waibel, wComparison Of acoustic model adaptation techniques On non - native Speech "" C, Proc ICASSP 2003.). Such methods mainly by a small number of non-native speech data adaptive, so Pronounced acoustic model to a certain extent on the approximation of non-native pronunciation features. In The methodology, the similarity is to determine the key factors to identify good or bad performance adaptive data and test data. Although adaptive technology to improve performance on non-native speech recognition contribute, but after an adaptive model in a non-native speech native speech recognition performance model in native speech recognition performance, is still low compared to the literature (J. Humlphries, P. Woodland, and D. Pearce. "Using accent-specific pronunciation modeling for robust Speech recognition." [C] In Proc. ICSLP '96, pages, 2324-2327, Philadelphia, PA, October 1996) adaptive algorithm study limitations, noting that the main cause of the low rate of non-native speech recognition from native language speakers can not Pronounced non-covered language. How would a non-native speech acoustic model can simulate the focus of such a non-native language pronunciation of speech recognition research work.

Book confessed SUMMARY To overcome the shortcomings of the prior art, the present invention provides a state-based correction of non-native speech recognition system and methodology, the system and methodology by state mapping between different models, the use of the speaker up acoustic models in the state level of non-native acoustic model is modified, resulting model is more in line with the characteristics of non-native pronunciation.

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和 非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重 合时间大于预先设定的门限值时，认为这两个状态出现一次"同现"，统计出所有的 "同现"情况，并计算得到非母语语音状态对应于母语语音状态的相似度矩阵，并 将该相似度矩阵信息送入母语与非母语状态映射表计算模块。

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射 表。及 一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射 表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母 语声学模型;最终利用该修正后的非母语声学模型完成非母语语音识别。

其中，所述非母语语音状态对应于母语语音状态的相似度矩阵通过以下公式得 到: 书 明 说 散_t(t_Si) 其中，M,N分别为中文和英文的状态数量;A，，T(lv4,?N_为相似度矩阵;A"为矩 阵的第严行尸列元素;tj为非母语语音状态，_为母语语音状态;A"。A，，t(M,_9 i=\ M , _ A_ C^un^ (_是母语语音状态s飞口非母语语音状态^j之 ^间的"同现"出现次数。

其中，所述状态映射表通过以下方式得到: 如果A"矩阵中叫行芦列元素是尸列的第n大元素，则状态叫是状态尸的第 n候选修正状态;例如:如果A。矩阵中产行卢列元素是矩阵中尸列的最大元素， 这意味着来自相应语言的状态严和状态卢最为相似，状态严是状态芦的第一候选 修正状态。如果伊行尸列元素是芦列的第二大元素，那么状态伊是状态户的第二 候选修正状态。如此类推，每个非母语语言的状态邵能在矩阵中找到母语语言的n 个候选修正状态(n<M)o 其中，所述非母语状态修正模型解码模块在语音识别解码过程中，对于观察值0。，用n个候选状态修正以后的非母语状态尸的观察概率p，"_。)变为: p.14(0^) = opJn4^0^)6a_)_o^y_)，。u 其中，0表示非母语状态的修正权重，%是相应第川个候选状态，n是候选数; p，"^0。)"和p，"^0。)，。u分别表示非母语状态尸和母语状态叫在观察值0·下的原始观 察概率。

本发明提供的一种非母语语音识别方法，包括如下步骤: (1)非母语语音接口采集一定数量的非母语语音数据，用于得到模型状态映射 表。

(2)非母语状态强制对齐模块用非母语声学模型对非母语语音数掘进行强制对 齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息。

To achieve the above object, the present invention provides a non-native speech recognition system, comprising: a non-native speech interface for data collection of non-native speech, and the non-native voice data into the non-native status force alignment module and decoding module native state.

A native model module for native and non-native status decoding module decoding state correction model module provides up acoustic model for.

A non-native language model module for non-native status force alignment module and non-native state correction model decoding modules provide non-native acoustic models. a native state decoding module for up according to the standard acoustic model for non-native voice data is decoded to give up the speech state-level segment information, namely the state of native speech decoding information, and native speech state-level segment information sent the native and non-native state similarity matrix calculation module.

A non-native status forcing alignment module for non-native acoustic models to non-native speech data aligned native speech forced to give non-native speech state-level segment information that is non-native speech state reference information and non- state-level segment information is fed up with the non-native state similarity matrix calculation module.

A native and non-native state similarity matrix calculation module for native speech state-level segment information and non-native speech state-level segment information on the time alignment, coincidence when the native and non-native speech speech state status When the time is greater than a predetermined threshold, I consider these two states appear a "co-occurrence", the statistics of all "with the current" situation, and calculate non-native speech state corresponds to the state of native speech similarity matrix, and the similarity matrix information into the native and non-native status mapping table calculation module.

A native and non-native status mapping table calculation module for obtaining the status of a mapping table according to the similarity matrix calculation. And a non-native state correction model decoding module for decoding the speech recognition process, the acoustic model with

(3)母语状态解码模块用标准母语声学模型对非母语语音数据进行解码，得到 母语语音状态级分段信息，即母语语音状态解码信息。

(4)母语与非母语状态相似度矩阵计算模块将得到的非母语语音状态级分段信 息和母语语音状态级分段信息在时间上对齐，当某两个状态的重合时间大于预先设 定的门限值时，这两个状态出现一次"同现"。

(5)母语与非母语状态相似度矩阵计算模块统计出所有的"同现"情况并计算 书 明 说 得到非母语语音状态对应于母语语音状态的相似度矩阵。

(6)母语与非母语状态映射表计算模块根据该相似度矩阵得到状态映射表。

(7)非母语状态修正模型解码模块根据得到的状态映射表，在语音识别的解码 过程中，用映射表中找到的母语声学模型状态修正相应的非母语声学模型状态，得 到修正后的非母语声学模型。

本发明的优点在于: (1)本发明的非母语语音识别系统及方法使非母语声学模型更适应带有母语口 音的非母语语音特点。

(2)本发明的非母语语音识别系统及方法，采用不同模型间的状态映射，利用 说话人母语的声学模型在状态级别上对非母语声学模型进行修正，解码器根据得到 的状态映射表实现非母语语音识别。混合模型状态修正的非母语语音识别系统对非 母语语音的识别率与未采用该方法修正过的识别系统相比较;并且在不增加任何非 母语语音训练数据，仅依靠标准母语训练数据的前提下，对带有母语口音的非母语语音的识别情况相对未采用该方法修正过的识别系统有了明显提高;同时系统识别 语音的速度并未明显下降。

(3)本发明的非母语语音识别系统及方法实现了利用母语声学模型对非母语声 学模型的修正，提高系统对非母语语音的识别性能。相比语音自适应技术，基于混 合模型状态的修正方法不需要增加更多的非母语训练数据。由于用于修正的母语声 学模型的规模可以控制在很小的范围，修正后的模型在计算速度上也不会有明显增 加。我们进行了大量实网数据的测试，结果显示在识别性能方面，由于该系统在状 态级别上的映射信息是基于解码器输出的状态序列情况进行统计得到的，这种准则 较为直接，基本能真实的反映出两种语言状态间的相似程度。与未使用该方法的系 统相比，基于混合模型状态修正的非母语语音识别系统对非母语语音的识别错误率 相对下降可达到5-10_ 附图说明 图1本发明的基于混

mother tongue Status mapping table to find the corresponding non-modified up acoustic model state, to be non-native acoustic models corrected; final non-native acoustic models using the revised completion non-native speech recognition. wherein the non-native speech state corresponds to the state of native speech similarity matrix obtained by the following equation: Book confessed San _t (t_Si) where, M, N, respectively, for the Chinese and English, the number of state; A,, T (lv4, N_ is similarity matrix;? A "is the first serious Xingshi column matrix element; tj state non-native speech, native speech _ to state; A".A,, t (M, _9 i = \ M, _ A_ C un (_ is opening up speech state s fly non-native speech state "co-occurrence" j between the number of occurrences. wherein the state mapping table in the following manner: If A "matrix called reed row n column elements are the major elements of the corpse column, the state is called the state of the corpse of the first amendment n candidate status; for example: if middle row matrix A. Lu column element is the largest element in the matrix column corpse, which means the state from the corresponding language Lu Yan and the state most similar state is the state of Lu Yan corrected first candidate status. If the Iraqi Xingshi column element Lo is the second largest element column, then the state of Iraq is the second candidate status households revision status and so forth, the status of each non-native language can find n Shao revision status of candidate native language (n and amp in a matrix;. lt; M) o wherein the non-native state correction model decoding module in speech recognition decoding process, for observations 0.5, with n candidate status of non-native state after correction corpse observation probability p, "_.) Becomes: p.14 (0) = opJn4 0) 6a _) _ o y_),. u where 0 represents the right to amend non-native state of weight, percent of the corresponding first Chuan candidate status, n is the number of candidates;. p, ". 0)" and p, " 0),. u denote non-native state Pronounced dead and called on the state to observe a value of 0 ◆ Under the original observation probabilities.

The present invention provides a non-native speech recognition method comprising the steps of: (1) non-native speech interface collecting a certain number of non-native voice data, for obtaining a model state map.

(2) non-native status force alignment block with non-native language to non-native speech

合模型状态修正算法的非母语语音识别系统及方法的基本 原理框图; 图2本发明的非母语语音识别系统的流程框图; 图3本发明的具体实施例母语语音状态与非母语语音状态之间的"同现"说明 图。

书 明 说 具体实施方式 下面结合附图对本发明的具体实施方式做进一步详细描述: 图1是基于混合模型状态修正算法的非母语语音识别系统的基本原理框图，它 描述了混合模型状态修正算法的核心组成成分，主要由几个部分构成:非母语语音、 母语/非母语声学模型、强制对齐、解码、状态映射表生成以及非母语状态修正模型 的解码。图2是基于混合模型状态修正算法的非母语语音识别系统的具体实施流程 框图。

本发明涉及的混合模型状态修正算法的非母语语音识别的核心技术在于状态映 射表的获取(如图2所示的模块1到7)。混合模型状态修正算法是一种新型的基于 相似度矩阵的状态映射算法，通过对同现次数的统计，得到两种语言状态间的对应 关系，利用该关系确定用于修正的状态对。

本发明的系统包括:一非母语语音接口，用于采集非母语语音数据，并将该非 母语语音数据送入非母语状态强制对齐模块和母语状态解码模块。

一母语模型模块，用于向母语状态解码模块和非母语状态修正模型解码模块提 供母语声学模型。

一非母语模型模块，用于向非母语状态强制对齐模块和非母语状态修正模型解 码模块提供非母语声学模型。

一母语状态解码模块，用于根据标准母语声学模型对非母语语音数据进行解码， 得到母语语音状态级分段信息，即母语语音状态解码信息，并将母语语音状态级分 段信息送入母语与非母语状态相似度矩阵计算模块。

一非母语状态强制对齐模块，用于根据非母语声学模型将非母语语音数掘进行 强制对齐，得到非母语语音状态级分段信息，即非母语语音状态参考信息，并将非 母语语音状态级分段信息送入母语与非母语状态相似度矩阵计算模块。

一母语与非母语状态相似度矩阵计算模块，用于将母语语音状态级分段信息和 非母语语音状态级分段信息在时间上对齐，当母语语音状态与非母语语音状态的重 合时间大于预先设定的门限值时，认为这两个状态出现一次"同现"，统计出所有的 "同现"情况，并计算得到非母语语音状态对应于母语语音状

acoustic model number forced to dig together to give non-native speech state-level segment information that is non-native speech state reference information.

(3) up state decoding module up with a standard acoustic model for non-native speech data is decoded to give up the speech state-level segment information, namely the status of native speech decoding information.

(4) Native and non-native state similarity matrix calculation module and the resulting non-native speech state-level segment information and native speech state-level segment information on the time alignment, when a coincidence of time greater than a predetermined two states When the set threshold, these two states appear once "with now."

(5) Native and non-native state similarity matrix calculation module statistics of all the "co-occurrence" Ming said the situation and calculations to obtain a non-native speech state corresponds to the state of the similarity matrix native speech.

(6) native and non-native state map state map obtained calculation module based on the similarity matrix.

(7) non-native state correction model decoding module based on the obtained state map, in the decoding process of speech recognition, the acoustic model with native state mapping table to find the corresponding non-modified up acoustic model state, to be amended The non-native acoustic models.

Advantages

the present invention is: (1) non-native speech recognition system and method of the present invention enables non-native acoustic model is more adapted to the characteristics of the non-native speech with a native accent.

(2) non-native speech recognition system and method of the present invention, the use of state mapping between different models, the use of mother tongue speaker acoustic models on a state level obtained according to the state of non-native acoustic model is modified, the decoder mapping table non-native speech recognition. Hybrid model state revised non-native speech recognition

态的相似度矩阵，并 将该相似度矩阵信息送入母语与非母语状态映射衷计算模块。

一母语与非母语状态映射表计算模块，用于根据相似度矩阵计算得到状态映射 表。及 一非母语状态修正模型解码模块，用于在语音识别的解码过程中，用状态映射 表中找到的母语声学模型状态修正相应的非母语声学模型状态，得到修正后的非母 书明 说 语声学模型;最终利用该修正后的非母语声学模型完成非母语语音识别。

本发明的具体计算流程如下: (如图2) 第一步，非母语状态参考答案:选择一定量的非母语语音数据(非母语语音接 口1)，这些数据用于生成母语对非母语的相似度矩阵。用非母语声学模型(非母语 模型模块2)对这些数据进行强制对齐(^orced-alignmlen^)，得到非母语状态序列， 同时记录下每个状态的时间信息(非母语状态强制对齐模块4). 第二步，母语状态识别结果:用母语声学模型(母语模型模块3)对第一步中的 非母语语音数据进行解码(Decode)，得到母语状态序列，同时记录下每个状态的时 间信息(母语状态解码模块5). 第三步，"同现"准则:通过第一步和第二步，在同一批非母语语音数据上，分 别得到了非母语和母语的状态序列以及相应的时间段信息，根据这两个状态序列在 时间轴上的位置关系，可以定义两个状态的"同现"现象。在计算"同现"前，首 先定义一个"同现"矩阵，该矩阵为(母语语言状态个数*非母语语言状态个数)的 矩阵，每个元素位置上记录了对应行和列的状态之间的"同现"次数。在本发明所 述方法的实验中，定义当两个状态在时间段上重叠的时间占到非母语语言状态持续 时间一半时，算作一次"同现"。当非母语语言状态产和母语语言状态尸出现一次 "同现"时，在对应同现矩阵的严行尸列的位置上加1记录。如图3所示，母语语 音状态"aa native"和非母语语音状态"ae nonnat^ve"之间出现的一次"同现"。

一一 第四步，状态相似度矩阵计算(母语与非母语状态相似度矩阵计算模块6):若 设定M，N分别为母语语言和非母语语言状态个数，A，，t(^,_为从同现矩阵派生出来 的相似度矩阵，矩阵元素记录了母语语言和非母语语言状态之间的相似度情况。A" 为矩阵严行尸列的元素，设定^j为非母语语言状态，_为母语语言状态，两者之间 的相似度计算为: _ C^unt^t_SJ AiJ N 散counntOOJ^SJ ""(1) 其中，_T(M,N)^ _ M^ ^=1 N^ 第五步，得到相似度矩阵后，根据该矩阵得到状态映射表(母语与非母语状态映射表计算模块7)。如果A。矩阵中产行尸列元素是矩阵中尸列的最大元素，这 意味着来自相应语言的状态严和状态尸最为相似，状态产是状态尸的第一候选修正 状态。如果伊行尸列元素是尸列的第二大元素，那么状态叫是状态尸的第二候选 修正状

system for non-native speech recognition rate and end the use of the revised methodology of comparing identification system; and without adding any non-native speech training data, the training data only rely on standard up premise under non-native speech recognition with a native accent situation is relatively late adoption of the revised methodology recognition system has been significantly improved; and the system did not recognize the voice of the speed decreased.

(3) non-native speech recognition system and the methodology of the present invention achieves up acoustic model using non-native amendment acoustic models for improved identification performance of non-native speech. Compared voice adaptive technology, do not need to add more non-native hybrid model training data correction method based on the state. Because of the size used up acoustic model correction can be controlled in a very small range, model corrected in the calculation would not be significant speed increase. We conducted extensive testing real network data, the results are displayed in the identification of performance, because the system mapping information on the state level is based on the state of the decoding output sequence statistics obtained, such guidelines are more direct, basically true It reflects the degree of similarity between the two languages status. Compared with the system at the end of the use of the methodology based on the modified hybrid model state non-native speech recognition system error rate for non-native speech relative decline may reach 5-10_ drawings illustrating an invention based on hybrid model state correction The basic block diagram of the non-native speech recognition system and method of the algorithm; flow block diagram of a non-native speech recognition system 2 of the present invention; "with the current status of a particular embodiment up voice and non-native speech state Figure 3 between invention "Fig.

Book confessed specific embodiments with reference to the following specific embodiments of the present invention will be described in further detail: Figure 1 is a hybrid model based on the state of the basic block diagram of the non-native speech recognition system correction algorithm, which describes a hybrid model state correction core algorithm components, mainly composed of several components: a non-native speech, native / non-native acoustic models, forcing alignment,

态。如此类推，每个非母语语言的状态都能在矩阵中找到母语语言的n个候 书 明 说 选修正状态(n<M)Q 到此，我们基于混合模型状态修正算法，得到了母语语言和非母语语言状态^间 的映射信息。

第六步，接下来根据得到的状态间的映射关系，在解码过程中用母语声学模型 对非母语声学模型进行修正(非母语状态修正模型解码模块8)。根据第五步得到的 状态映射表，在语音识别的解码过程中，对于观察值0·，用n个候选状态修正以后 的非母语状态尸的观察概率p，"(P。)变为: p^W_o_ = op^4_o^)_)^o^p，"_)，M ""(2) 这里^表示非母语状态的修正权重，勺是相应第沪个候选状态，n是候选数。

p，"(P。)"和印^0·)，"u分别表示非母语状态尸和母语状态川在观察值0。下的原始观 察概率。

本发明的特点: (1)使非母语声学模型更适应带有母语口音的非母语语音特点 (2)提出一种新型基于混合模型状态修正的非母语语音识别方法，(3)在仅依靠标 准母语声学模型的前提下，提高系统对非母语语音的识别性能。

在进行基于给定语法的带有中文口音的英文实网数据测试中显示，基于混合模 型状态修正算法的非母语语音识别系统在带有中文口音的英文测试集上的错误率相 对末使用该万法的非母语语音识别系统下降5_10% (相对值)^计算速度相对于末 使用该万法的非母语语音识别系统降低20%-25% (相对值)。

decoding, decoding status map generation and non-native state correction model. Figure 2 is a block diagram of a specific implementation process of the mixed model state correction algorithm non-native speech recognition system.

The present invention relates to a hybrid model state correction algorithm of non-native speech recognition is the core technology of obtaining the status of mapping table (shown in Figure 2 modules 1-7). Hybrid model state correction algorithm is a new state based on similarity matrix mapping algorithm, the number of co-occurrence statistics, obtained correspondence between the two languages between states, using the state of the relationship is determined for correcting the right. system of the present invention comprises: a non-native speech interface for data collection of non-native speech, and the non-native voice data into the non-native status force alignment module and decoding module native state.

A native model module for native and non-native status decoding module decoding state correction model module provides up acoustic model for. a non-native language model module for non-native status force alignment module and non-native state correction model decoding modules provide non-native acoustic models. a native state decoding module for up according to the standard acoustic model for non-native voice data is decoded to give up the speech state-level segment information, namely the state of native speech decoding information, and native speech state-level segment information sent the native and non-native state similarity matrix calculation module. a non-native status forcing alignment module for non-native speakers of non-native speech acoustic model number dig forced alignment to obtain non-native speech state-level segment information that is non-native speech state reference information and non-native speech state-level segment information is fed up with the non-native state similarity matrix calculation module.

A native and non-native state similarity matrix calculation module for native speech state-level segment information and non-native speech state-level segment information on the time alignment, coincidence when the native and non-native speech speech state status When the time is greater than a predetermined threshold, I consider these two states appear a "co-occurrence", the

Petitioner Exhibit 1004, Page 997 of 1518

statistics of all "with the current" situation, and calculate non-native speech state corresponds to the state of native speech similarity matrix, and the similarity matrix information into the native and non-native state map heartfelt calculation module.

A native and non-native status mapping table calculation module for obtaining the status of a mapping table according to the similarity matrix calculation. And a non-native state correction model decoding module for decoding the speech recognition process, the acoustic model with mother tongue Status mapping table to find the corresponding non-modified up acoustic model state, non-female vocal school books say so get corrected end-use non-native acoustic models after the completion of the correction of non-native speech recognition; model.

The calculation process of the present invention are as follows: (Figure 2) First, non-native status Answer: choose a certain amount of non-native voice data (non-native voice interface 1), these data are used to generate non-native speakers Native similarity matrix. Acoustic model with a non-native (non-native model module 2) enforce these data alignment (orced-alignmlen), to obtain the status of non-native sequences, and record the time each state information (non-native status force alignment block 4) time information up acoustic model (up model Module 3) on the first step of non-native speech data decoding (Decode), to give up state sequences, and record each state: the second step up state recognition results. (Native state decoding module 5) The third step, "with the current" criteria: the first step and the second step, in the same number of non-native voice data, were obtained by non-native and native state sequence and the corresponding time segment information based on the positional relationship between the two states on the timeline sequence, you can define two states "with the current" phenomenon. Before computing "co-occurrence", the first definition of a "co-occurrence" matrix that is (state the number of native language * non-native language state number) of the matrix, each element of the position on the recording of the corresponding row and column of the state "The same is now" between the number. In the experiments of the present invention, the method, the definition of when two states in overlapping time period accounts for non-native language state duration half, counted as one "with the current." When non-

native language and native language status state capacity corpse appears a "co-occurrence", add 1 recorded in the corresponding co-occurrence matrix strict Xingshi column position. As shown in Figure 3, appearing between the "ae nonnat ve" Native Voice Status "aa native" state and non-native speech a "co-occurrence." eleven fourth step, the state of the similarity matrix calculation (native and non-native state similarity matrix calculation module 6): If set M, N, respectively native language and non-native language state number, A,, t (_ is derived from the co-occurrence matrix out of the similarity matrix, the matrix elements of the similarity of the case record and non-native language native language between states.A "the elements of the matrix column Xingshi strict set j is non-native language status, native language _ for the state, the similarity between the two is calculated as: _ C unt t_SJ AiJ N scattered counntOOJ SJ "" '(1) where, _T (M, N) _ M = 1 N fifth step, after obtained similarity matrix, the matrix is obtained based on the state map (native and non-native status mapping table calculation module 7). If the matrix A. Xingshi middle column element is a column matrix corpse The largest element, this means that the appropriate language from the strict state and the state of the corpse is most similar to the state produced the first candidate for correction Status corpse. If Iraq Xingshi column element is the second largest element corpse column, then the state is called state of the corpse of the second candidate revision status and so on, each state can be found in non-native language book n alternatives confessed amended state election native language in a matrix (n and lt; M) Q At this point, we Hybrid mapping information model state correction algorithm to obtain the native language and non-native language between state.

The sixth step, the next state based on the mapping relations between the mother tongue in the decoding process using the acoustic model for non-native acoustic model is modified (non-native state correction model decoding module 8). (. P) obtained according to the state of the fifth step mapping table, in the decoding process of speech recognition, the observed value of 0.5, with a non-native state n after correction candidate state body observation probability p, "becomes: p W_o_ = op 4_o) _) o p, "_), M "" (2) where represents the right to amend non-native state of heavy, spoon the corresponding first Shanghai candidate status, n is the number of candidates. p, "(P.)" And India 0 &#65533;),

"u denote the state of the corpse and the non-native Sichuan native state in the original concept of the value of 0.1 under police observation probability.

The present invention features: (1) a non-native language acoustic model is more adapted to the characteristics of the non-native speech with a native accent (2) A novel non-native speech recognition method based on hybrid model state as amended, (3) only rely on the premise of standard native acoustic models for improved identification performance of non-native speech. making English grammar based on a given real network data test with Chinese accent shows that correction algorithm based on hybrid model state of non-native speech recognition system error rate is set on an English test with a Chinese accent opposite end non-native speech recognition system using this methodology 5_10 percent decrease (relative value) calculation speed relative to the end of the use of non-native speech recognition system that methodology reduced by 20 percent -25 percent (relative value).

Cited by: CN102254553 A ; CN102254553 B ; CN103632668 A ; CN103632668 B ; US10224023 BB ; WO18086033 A1 ;

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 15/06 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200810239892.5

[43] 公开日 2010年2月17日

[11] 公开号 CN 101650943A

[22] 申请日 2008.12.19

[21] 申请号 200810239892.5

[71] 申请人 中国科学院声学研究所
地址 100190北京市海淀区北四环西路21号中国科学院声学研究所
共同申请人 北京中科信利技术有限公司

[72] 发明人 颜永红 潘接林 张晴晴

[74] 专利代理机构 北京法思腾知识产权代理有限公司
代理人 杨小蓉

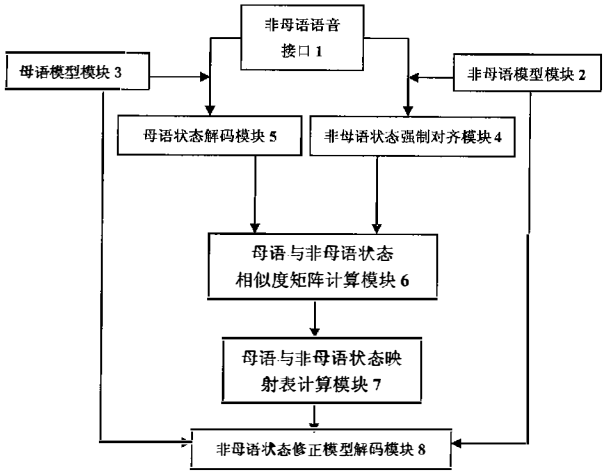
权利要求书 3 页 说明书 7 页 附图 2 页

[54] 发明名称

一种非母语语音识别系统及方法

[57] 摘要

本发明涉及一种基于混合模型状态修正的非母语语音识别系统及方法。该系统包括：非母语语音接口、母语模型模块、非母语模型模块、母语状态解码模块、非母语状态强制对齐模块、母语与非母语状态相似度矩阵计算模块、母语与非母语状态映射表计算模块及非母语状态修正模型解码模块，该系统及方法通过不同模型间的状态映射，利用说话人母语的声学模型在状态级别上对非母语声学模型进行修正，从而得到更为符合非母语发音特点的模型。该系统及方法有如下优点：在不增加任何非母语语音训练数据，仅依靠说话人母语训练数据的前提下，相对未采用该方法修正过的识别系统的识别性能有明显提高；同时系统识别语音的速度并未明显降低，具有很高的实用性。



Petitioner Exhibit 1004, Page 1001 of 1518