3.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法的系统，其特征在于，所述系统包括用于用户个性化信 息注册、唤醒词识别和声纹匹配的语音控制系统(1)，用于接收并存 储从语音控制系统(1)中传递来的已训练好的用户声纹模型和已识 别出的电视频道信息的信息存储单元(2)和接收从语音控制系统(1) 和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操 作的电视机主控制器(3)组成，以电信号方式连接，其中：

所述语音控制系统(1)由语音采集单元(11)，语音识别单元(12) 和声纹识别单元(13)组成，其中：

所述语音采集单元(11)用于完成用户语音信号的采集和处理；

所述语音识别单元(12)用于识别用户所说的电视频道信息和唤 醒词；

所述声纹识别单元(13)用于从用户语音中提取出代表其身份信 息的声纹信息进行声纹建模和声纹匹配。

4.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤A个性化信息注册，具体实现 步骤为：

A1.开启语音控制系统(1)，使其处于"用户注册"状态；

A2.若用户为首次进行个性化信息注册，执行该步骤，否则转向 步骤A6开始执行，用户对着电视机的语音采集单元(11)说话，说 话内容为任意文本，持续时间至少30秒；

A3.语音采集单元(11)采集输入的说话人语音信号，并对其进 行A/D转换，滤波处理；

A4.声纹识别单元(13)对经步骤A3处理后的语音信号作预处理， 包括加窗分帧、预加重、端点检测，提取声纹特征参数，然后调用信 息存储单元(2)中已有的声纹模型进行声纹匹配，若匹配成功，转 向步骤A7；反之，进入下一步操作；

A5.调用声纹识别单元(13)对从步骤A4中提取的特征参数进行 声纹模型的训练，将训练好的声纹模型存储到信息存储单元(2)；

A6.用户对着电视机的语音采集单元(11)说出自己最喜欢的电 视频道信息，如频道代码"3"代表湖南卫视，则说出数字"3"，转 向步骤A3重新开始；

operation is complete, just after the operation step C, and to complete the wake of its personalized voice TV;

When the user changes the TV channel preferences, by re-executing step A, user personalization information registration information storage unit to update the information television channel, this time without the need for a voiceprint model, only conduct voiceprint matching and speech recognition.

3. Voiceprint as claimed in using voice recognition and speech personalized TV system wake-up method according to claim 1, characterized in that the system comprises a user personalization information Sign up, wake up word recognition voiceprint matching voice control system (1) for receiving and storing transmitted from the voice control system (1) to have trained user voiceprint model and have identified television channel information information storage unit (2) and receiving control information transmitted from the voice control system (1) and the information storage unit, responsible for the implementation of personalized television voice Wake TV host controller operations (3), with an electrical signal connected, where:

The voice control system (1) by the voice acquisition unit (11), the voice recognition unit (12) voiceprint identification unit (13), wherein:

The voice acquisition unit (11) for the completion of the acquisition and processing user speech signal;

The voice recognition unit (12) for identifying a user of said television information and wake-up word; the voiceprint identification unit (13) is used to extract from a speech on behalf of its user identity information voiceprint information voiceprint model voiceprint match.

4. Claim and speech recognition using the voiceprint of claim 1, wherein the personalized television voice awakened, characterized in that said step A personalized information registration, concrete implementation steps to:

A1 open the voice control system (1), it is in "User Registration" state;.

A2. If the user is registered for the first time personalized information, perform this step,

A7.语音识别单元(12)对语音采集单元(11)采集到的语音进 行识别，将识别出的电视频道代码存储到信息存储单元(2)中并与 声纹匹配成功的声纹模型进行关联；

A8.退出个性化信息注册。

5.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤C语音唤醒，具体实现步骤为：

C1.当语音控制系统(1)处于"语音唤醒"开启状态时，用户对 着电视机语音采集单元(11)说出唤醒词；

C2.语音采集单元(11)采集输入的说话人语音信号，并对其进 行A/D转换，滤波处理；

C3.语音识别单元(12)对上述处理过的语音进行识别，将识别 出的内容与预先设定的唤醒词进行匹配，以判断采集到的输入语音信 号是否为电视机的语音唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器(3)，否则退出唤醒系统；

C4.声纹识别单元(13)将处理过的语音与信息存储单元(2)中 的声纹模型进行声纹匹配，若匹配成功，则将与用户声纹模型关联的 电视频道信息和唤醒操作指令输出给电视机主控制器(3)，否则退出 唤醒系统；

C5.电视机主控制器(3)在成功收到语音控制系统(1)传递来 的电视频道信息和唤醒操作指令后，执行电视开机操作并自动切换到 对应用户喜爱的电视频道上。

6.如权利要求4所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤A5中的声纹模型的训练方法 为总变化因子建模技术。

7.如权利要求6所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述总变化因子建模技术，给定一段用 户的语音，与说话人及信道相关的高斯混合模型GMM的均值超向量S 表示如下：

$$S=m+Tv+\varepsilon$$

其中，m为通用背景模型的均值超向量，T矩阵为总变化空间， ε为残差，v矢量是与当前目标说话人(用户)及信道相关的总变化因 子，它代表着当前说话人的身份信息，即用户的声纹模型。

otherwise to step A6 started, users facing the TV voice acquisition unit (11) to speak, speak the contents of any text, the duration of at least 30 sec;

. Speaker voice signal

A3 voice acquisition unit (11) gathers the input, and its A / D conversion, filter processing;

A4. Voiceprint identification unit (13) after step A3 by pretreatment of the speech signal processing, including windowing framing, pre-emphasis, endpoint detection, feature extraction voiceprint parameters, and then call the information storage unit (2) already voiceprint model voiceprint matches, if the match is successful, go to step A7; on the contrary, enter the next step;

.

A5 call voiceprint identification unit (13) performs voiceprint model from step A4 feature parameters extracted training, the trained model voiceprint information stored in the storage unit (2);

A6. Users facing the TV voice acquisition unit (11) Name your favorite TV channel information such as channel codes "3" represents Hunan Satellite TV, they say the number "3" to step A3 resume;

A7. Voice recognition unit (12) for voice acquisition unit (11) collected speech recognition, it will recognize the TV channel code is stored in the information storage unit (2) and with a voiceprint matching the success of voiceprint Model associate;

A8. Exit personalized information registration.

.

5 according to claim 1 using a voiceprint and voice recognition personalized TV voice awakened, characterized in that said step C wake voice, specific implementation steps:

C1 when the voice control system (1) in the "Voice wake up" turned on, the user facing the TV voice acquisition unit (11) Name wake word;.

. Speaker voice signal

C2 voice acquisition unit (11) gathers the input, and its A / D conversion, filter processing;

Petitioner Exhibit 1004, Page 902 of 1518

8.如权利要求7所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述ε和v均服从高斯分布：

ε～N(0,I)

v～N(0,Σ)

其中，Σ为对角协方差矩阵，为得到v，需要事先采用与电视机 语音唤醒场景近似的语音数据集合，利用期望最大化算法训练得到通 用背景模型，得出总变化空间和残差向量；对于任意电视用户的语音，就可利用已训练好的T，ε和m以及由用户语音构建的高斯混合模型 S，求出每个用户的总变化因子，即声纹模型。

9.如权利要求4，5所述的采用声纹和语音识别进行个性化电视 语音唤醒的方法，其特征在于，所述声纹匹配，使用的是基于总变化 因子的声纹识别系统，采用两个总变化因子v_1和v_2之间的余弦距离 来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

d (v 1, v 2) = < v 1, v 2 > | | v 1 | | · | | v 2 | |

如果d值大于预先设定的门限值则认为这两段语音来自同一个 说话人，声纹匹配成功，否则匹配失败。

10.如权利要求3所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法的系统，其特征在于，所述语音识别单元(12)声学模 型训练时，隐马尔可夫模型的状态数取5，每个状态使用一个含8个 混合高斯成分的高斯混合模型GMM表示。

11.如权利要求3所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法的系统，其特征在于，所述语音识别单元(12)的语音 解码，采用将声学、语言模型和发音字典进行有效的整合，对输入的 语音信号PLP特征参数在词网络中进行搜索和匹配，从而找到统计意 义上最匹配的词串作为最终识别结果的基于加权有限状态转换的 Viterbi解码。

C3. Voice recognition unit (12) on said processed speech recognition, the recognized content with a preset wake-up word matching, to determine whether the input speech signal collected wake of TV voice instruction, if the match is successful then the wake-up command is passed to the TV host controller (3), otherwise exit wake;

C4. Voiceprint identification unit (13) and the processed voice information storage unit (2) voiceprint model voiceprint matches, if the match is successful, then the TV channel information associated with the user's voiceprint model and wake-up operation command output to a television host controller (3), otherwise exit the wake;

C5. TV host controller (3) After the successful receipt of the voice control system (1) is passed to the TV channel information and wake operating instructions, perform TV start operation and automatically switch to the corresponding user's favorite TV channels.

6. As the use of voiceprint and voice recognition voice personalized TV Wake method according to claim 4, wherein said step of training methods A5 voiceprint Model total change factor modeling techniques.

7. Claim using the voiceprint and voice recognition according to claim 6 personalized TV voice awakened, characterized in that the total change factor modeling techniques given period the user's voice, the mean and the speaker and channel associated Gaussian mixture model GMM ultra-vector S is expressed as follows:

S = m Tv e where, m is the mean supervectors universal background model, T matrix for the total change in space, e is the residual, v vector is the current target speaker (the user) and total change factors associated channel, it represents the current the speaker's identity information, that user voiceprint model.

.

8 claim using the voiceprint and voice recognition personalize Television Voice wake method according to claim 7, wherein said e and v are Gaussian:

e ~ N (0, I) v ~ N (0, S) where, S is a diagonal covariance matrix, to obtain v, require prior recourse to wake the TV voice scene approximate

voice data collection, use EM algorithm to train to get general background model, gave a total change of space and residual difference vector; for any television user's voice, you can use the trained T, e and S m and Gaussian mixture model built by the user's voice, determine the total change factor for each user, that voiceprint model.

9. The as claimed in claim 4, 5 using voiceprint and voice recognition personalize the TV voice awakened, characterized in that said voiceprint matching, is based on the use of voiceprint recognition system total change factors, the total change in the use of two factor v_1 and v_2 to measure between two speech contained speaker sound the degree of matching pattern information, the formula is as follows: d (v 1, v 2) = and lt; v 1, v 2 and gt; | | v 1 | | $\cdot$ | | v 2 | |

If d is greater than a predetermined threshold is considered two paragraphs from the same speaker's voice, voiceprint matches, otherwise the match fails.

10. Voiceprint as claimed in using voice recognition and speech personalized wake television system according to the method of (12) acoustic models, characterized in that, the voice recognition unit training, state the number of HMM take 5, each state using a Gaussian mixture model GMM Gaussian mixture containing eight component representation.

11. Voiceprint as claimed in using voice recognition and speech personalized TV system wake-up method according to claim 3, characterized in that said speech voice recognition unit (12) decoding, using acoustic, language models and pronunciation dictionaries for effective integration of the voice signal input characteristic parameters PLP search and match the word network, in order to find the statistical sense of the word that best matches the string as a final recognition result based on the weighted Finite state transitions Viterbi decoding.

Description:

采用声纹和语音识别进行个性化电视语音唤醒的方法

Using the voiceprint and voice recognition personalize Television Voice Wake methods

技术领域

Technical Field

本发明涉及声纹识别和语音识别技术，通过声纹识别进行电视机使用用户 的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒。

背景技术

近年来，随着智能语音处理技术、互联网和云计算技术的快速发展，目前 市面上出现的智能电视大部分都开始支持简单的语音操作，如进行音量调节，资源搜索等，同时电视语音助手等应用也开始出现。但由于中国各地的方言种 类多、同音字多、各地方言的语音语调的不同、周围环境噪声等因素造成语音 识别的准确率很低，电视无法执行用户语音命令的现象时常出现。因此，对消 费者来说，目前的语音电视在一定程度上并不能满足其各项需求。同时，这些 人机语音交互功能大多需要通过云端计算来完成，存在一定的安全隐患。

现在几乎每个家庭都拥有一台或几台电视机，但一台电视通常会被几位家庭 成员使用，特别是放置在客厅的电视，而每位家庭成员对电视节目频道的喜好 却不尽相同，这时在打开电视后通常需要用遥控器对电视节目频道进行遍历切 换以找到自己最喜欢的频道，非常不方便且繁琐。

发明内容

为解决现有技术的不足，本发明提供一种采用语音进行个性化电视唤醒的 实现方法，能够对用户语音采用声纹识别和语音识别，实现对用户的身份进行监测并使用非常简单的唤醒词就能唤醒电视打开其最喜欢的电视频道。

众所周知，声纹识别又称说话人识别，是生物识别技术的一种，声纹识别 和语音识别均属于智能语音处理技术的范畴。具体是将说话人语音信号中的声纹特征参数与训练好的声纹模型进行模式匹配从而自动进行说话人身份确认或 辨认的技术。语音识别是通过模式识别和语言理解将语音信号中包含的文字信 息提取出来的技术，主要包括特征提取、声学模型和语言模型训练、解码技术 等。

本发明具体实现步骤如下：

A.个性化信息注册

采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别出用户最 喜爱的电视频道信息。

B.信息存储

The present invention relates to a voiceprint recognition and speech recognition technology, by voiceprint identification were TV user's identification, and through identification of a user's speech and the speech recognition result to control the TV personalize voice wake.

BACKGROUND

In recent years, with the rapid development of intelligent voice processing technology, the Internet and cloud computing, smart TV currently on the market that appears most are beginning to support a simple voice operations, such as adjusting the volume, resource searches, etc., while TV voice assistant applications have begun to appear. However, due to the multiple dialects vary throughout China, homophone many different factors such as ambient noise around the dialect pronunciation and intonation of speech recognition accuracy caused by low TV does not execute user voice commands phenomenon often occurs. So, for consumers, the current television voice to some extent and can not meet its needs. At the same time, most of these machine voice interaction capabilities needed to accomplish through cloud computing, there are some security risks.

Now almost every family has one or several TV sets, but a TV is usually used by several family members, especially placed in the living room TV, and each family member on the television program channel preferences but not the same, then after the turn on the TV remote control is typically required to traverse switching television channels to find their favorite channel, very inconvenient and cumbersome.

Summary

To solve the problem of the prior art, the present invention provides a method of using speech realization of personalized TV Wake, it is possible for the user using the voiceprint voice recognition and speech recognition, to achieve the user's identity and to monitor the use of very simple Words can wake wakeup open their favorite TV channels.

As we all know, voiceprint identification, also known as speaker recognition is a biometric technology, voiceprint recognition and speech recognition are within the scope of intelligent voice processing technology. Specifically the speaker's

将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电视机信息存 储单元。

C.语音唤醒

用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可根据声纹 识别和语音识别系统对唤醒词的声纹和语音识别结果从电视机信息存储单元中调出用户的个性化信息，自动完成打开电视并自动切换到用户最喜爱的电视频 道。

以上所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特点 是，所述步骤A、步骤B，在用户作第一次操作时完成，以后只需操作步骤C， 即可完成其个性化语音电视的唤醒。

当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进行用户个性 化信息注册，对信息存储单元中的电视频道信息进行更新，此时无需进行声纹 建模，仅需进行声纹匹配和语音识别。

由于本发明使用了声纹识别技术，很容易对不同用户建立起其对应的声纹 密钥(声纹模型)，通过不同用户声纹密钥的唯一性实现对其喜爱的电视频道信息进行关联和保护，进而实现具有用户个性化的电视语音唤醒功能。

本发明与现有技术相比，具有以下特点：

第一、用户只需张嘴说简单的唤醒词，如"hello电视"就能自动打开电视 并立即切换到自己最喜欢的电视台节目频道，无需再手动开机遥控，给用户带来智能、时尚的应用体验。

第二、个性化注册时的声纹识别性能与所说的文本和使用的语言无关，方 便使用。

第三、用声纹作为密码，简单的电视频道代码和唤醒词语音作为交互方式， 因此语音识别系统仅需实现简单~~的关~~键词识别即可，识别~~速~~度快，且声纹识别 和语音识别仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便 且保证了家庭信息的安全性。

第四、可设置多位用户的个性化语音开机信息，除唤醒功能外，同时也可 用于个性化自动语音换台。

第五、本发明的算法实现模块具有很好的可移植性和可扩展性，可被移植 到有线电视的机顶盒、通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

附图说明

voiceprint speech signal characteristic parameters and trained voiceprint model for pattern matching to automatically perform speaker identification or identification technology. Speech recognition is through pattern recognition and language information contained in the understanding of the speech signal will be extracted from the technology, including feature extraction, acoustic model and language model training, decoding technology.

The present invention specific implementation steps are as follows:

A. Personalized Information Register using the voiceprint identification system to establish the user's voiceprint model, using voice recognition system identifies the user's favorite TV channel information.

B. Information storage

The user voiceprint model and favorite TV channels and stores information related to the TV information storage unit.

C. Voice Wake Up

Users only need to tell the television voice collecting unit pre-set wake-up word, you can wake up the words of the voiceprint and voice recognition results from the television tune information storage unit based on the voice recognition and speech recognition systems the user's personalized information, auto-complete turn on the TV and automatically switch to the user's favorite TV channels. voiceprint and voice recognition using personalized television speech wake method described above, characterized in that said step A, step B, at the time of completion of the user operation for the first time, just after the steps C to complete their personalized voice TVs wake.

When the user changes the TV channel preferences, by re-executing step A, user personalization information registration information storage unit to update the information television channel, this time without the need for a voiceprint model, only conduct voiceprint matching and speech recognition.

Since the present invention uses a voiceprint recognition technology, it is easy to establish the corresponding voiceprint key (voiceprint model) for

图1为本发明采用声纹和语音识别进行个性化电视语音唤醒的方法 系统组成框图；

图2为本发明实施例用户个性化信息注册系统流程图；

图3为本发明实施例个性化语音唤醒电视的流程图。

具体实施方式

以下结合附图和实施例对本发明作进一步的描述

本发明构建了一套采用声纹和语音识别进行个性化电视语音唤醒的系统 (如图1所示)，该系统包括用于用户个性化信息注册、唤醒词识别和声纹匹配 的语音控制系统1，用于接收并存储从语音控制系统1中传递来的已训练好的用 户声纹模型和已识别出的电视频道信息的信息存储单元2和接收从语音控制系 统1和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操作的 电视机主控制器3组成，以电信号方式连接。其中：

所述语音控制系统1由语音采集单元11，语音识别单元12和声纹识别单元 13组成，其中：

所述语音采集单元11用于完成用户语音信号的采集和处理；

所述语音识别单元12用于识别用户所说的电视频道信息和唤醒词；

所述声纹识别单元13用于从用户语音中提取出代表其身份信息的声纹信息 进行声纹建模和声纹匹配。

本发明的语音控制系统有二个状态按钮，即"用户注册"和"语音唤醒"。 当用户需要进行个性化信息注册时开启"用户注册"，平时均处于"语音唤醒" 状态，这样有利于节省电能。

下面结合附图分别介绍用户个性化信息注册过程和自动语音唤醒的具体实 施方式

用户首次进行个性化信息注册时(如图2所示)实施例，需对电视机说话 二次，第一次为任意文本的语音，用于声纹建模，第二次为用户自己最喜欢的 电视频道信息，用于将电视频道信息与用户声纹模型进行关联。具体步骤如下：

A1.开启语音控制系1，使其处于"用户注册"状态；

A2.若用户为首次进行个性化信息注册，执行该步骤，否则转向步骤A6开 始执行，用户对着电视机

different users of different users through a unique voiceprint key to their favorite TV channels correlate information and protection, so as to realize the TV voice with user personalization wake-up function.

Compared with the prior art, it has the following characteristics:

First, the user simply opened his mouth to say a simple wake-up words, such as "hello TV" will automatically turn on the TV and immediately switch to your favorite television program channels, without the need to manually turn the remote control, giving users a smart, fashion experience.

Second, voiceprint recognition performance and said the language used in the text and personalized registration has nothing to do, easy to use.

Third, using the voiceprint as a password, a simple television channel code and wake-word speech as an interactive way, so speech recognition system just to implement simple keyword recognition, recognition speed, and voiceprint recognition and speech Only trained and tested off-line identification can be completed without the need to send to the cloud server, easy to use and ensure the safety of family information.

Fourth, you can set personalized voice multiple users boot information, in addition to wake-up function, but also can be used to personalize automated voice for Taiwan.

Fifth, the algorithm of the present invention achieves module has good portability and scalability, can be ported to run on cable television set-top boxes, high-speed general-purpose processor or DSP chips, widely used in smart home related fields.

Brief Description

1 of the present invention using the voiceprint and voice recognition voice wake personalized TV system block diagram of the method;

2 is a schematic invention registered user personalization information systems implementation flow chart;

的语音采集单元11说话，说话内容为任意文本，持续 时间至少30秒；

A3.语音采集单元11采集输入的说话人语音信号，并对其进行A/D转换， 滤波处理；

A4.声纹识别单元13对经步骤A3处理后的语音信号作预处理，包括加窗分 帧、预加重、端点检测，提取声纹特征参数，然后调用信息存储单元2中已有的声纹模型进行声纹匹配，若匹配成功，转向步骤A7；反之，进入下一步操作；

A5.调用声纹识别单元13对从步骤A4中提取的特征参数进行声纹模型的训 练，将训练好的声纹模型存储到信息存储单2；

A6.用户对着电视机的语音采集单元11说出自己最喜欢的电视频道信息， 如频道代码"3"代表湖南卫视，则说出数字"3"，转向步骤A3重新开始；

A7.语音识别单元12对语音采集单元11采集到的语音进行识别，将识别出 的电视频道代码存储到信息存储单元2中并与声纹匹配成功的声纹模型进行关联；

A8.退出个性化信息注册。

当用户的电视频道喜好发生改变时，可通过重新进行个性化信息注册来更 新信息存储单元2中的电视频道信息，此时只需对着电视机的语音采集单元11说出电视频道信息即可，语音控制系统1会自动从上述的步骤A6开始执行，无 需对用户声纹重新建模。

自动进行个性化语音唤醒电视(如图3所示)实施例的具体步骤如下：

C1.当语音控制系统1处于"语音唤醒"开启状态时，用户对着电视机语音 采集单元11说出唤醒词；

C2.语音采集单元11采集输入的说话人语音信号，并对其进行A/D转换， 滤波处理；

C3.语音识别单元12对上述处理过的语音进行识别，将识别出的内容与预 先设定的唤醒词进行匹配，以判断采集到的输入语音信号是否为电视机的语音 唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器3，否则退出唤 醒系统；

C4.声纹识别单元13将处理过的语音与信息存储单元2中的声纹模型进行 声纹匹配，若匹配成功，则将与用户声纹模型关联的电视频道信息和唤醒操作

Figure 3 is a flowchart illustrating the invention personalized voice wake-TV implementation.

DETAILED DESCRIPTION

The following accompanying drawings and the embodiment of the present invention will be further described

The present invention is constructed using a voiceprint and voice recognition voice wake personalized TV system (Figure 1), which includes personalized information for users registered, Wake word recognition voiceprint matches Voice control system for receiving and storing passed from voice control system 1 to a good user voiceprint model and has identified information television channel information storage unit 2 has been training and receive voice control system 1 and the information store from control information transmission unit, responsible for the implementation of personalized television voice wake-up consisting of a TV host controller 3, connected to an electrical signal. Where:

The voice control system by the voice acquisition unit 11, a voice recognition unit 12 voiceprint identification unit 13, of which:

The voice collecting unit 11 for the completion of the acquisition and processing user speech signal;

The voice recognition unit 12 for identifying a user of said television information and wake-up word;

The voiceprint identifying unit 13 for extracting from a speech on behalf of its user identity information voiceprint information modeling voiceprint matches the voiceprint. voice control system of the present invention has two state button, the "Registration" and "Voice wake up." When the user needs to open the "User Registration", usually are in the "Voice wake-up" state, which will help save energy when personalized information registration. are described below in conjunction with the accompanying drawings of specific embodiments of the registration process the user personalized information and automated voice wake of

When users first time personalized information registered (Figure 2) Example, the need for television to speak twice, the first time any text of the speech, the voiceprint model for the second time for the user his favorite TV channel

指令输出给电视机主控制器3，否则退出唤醒系统；

C5.电视机主控制器3在成功收到语音控制系统1传递来的电视频道信息和 唤醒操作指令后，执行电视开机操作并自动切换到对应用户喜爱的电视频道上。

上述提到的预先设定的特定语音唤醒词可为任意设定的词组，如"打开电 视机"、"开电视"、"打开电视"、"hello电视"、"hello TV"等等。语音采集 单元11可为麦克风或其他语音采集设备，用于感知周边的语音信号。

本发明涉及到的声纹识别单元13所采用的识别算法为基于总变化因子(又 称为i-Vector)的说话人确认方法，其具体构建步骤如下：

一、语音信号的预处理

语音信号的预处理包括对语音信号进行预加重、加窗分帧和端点检测。对 于经过A/D转换后的数字语音信号，预加重使用如下的一阶数字滤波器来提升语音信号的高频分量：

$H(z)=1-0.97z^{-1}$

然后采用如下式的汉明窗窗函数对语音信号进行平滑，并在加窗过程中进行语 音信号的分帧操作，其中N为窗的长度。

$W(n)=0.54+0.46\cos[2pin/(N-1)]$

对于采样率为8000Hz的语音，分帧时使用的帧长为25毫秒，帧移为10毫 秒。本发明采用的是基于短时能量的端点检测法对语音信号进行端点检测，从 而判断语音信号的起始点和终止点，设x(·)为通过预加重后的语音信号，则表示 从语音信号的第n个点开始加窗时的短时能量E_n计算公式为：

E n = Sigma m = - ∞ ∞ [x (m) W (n - m) 2 = x 2 (n) * W 2 (n)

二、声纹特征参数的提取

声纹特征参数的提取即从说话人的语音信号中提取出能反映其身份个性的 参数。本发明使用的是39维的感知线性预测系数(Perception Linear Predictive Coefficients，PLP)，具体提取过程为：首先对上述步骤一预处理 过后的语音做短时快速傅里叶变换，然后使用24阶子带的梅尔三角滤波器组进 行临界频率分析，接着对得到的语音能量谱进行等响度曲线预加重用以模拟人 耳的听觉感知系统对频率的敏感

information for the TV channel information and user voiceprint model association. Specific steps are as follows:

.

A1 open voice control system 1, it is in "User Registration" status;

.

A2 If the user is registered for the first time personalized information, perform this step, otherwise to step A6 started, users facing the TV voice collecting unit 11 to speak, speak any text content, duration of at least 30 seconds;

A3 11 gathering input voice Speaker voice signal acquisition unit, and its A / D conversion, filter processing;.

A4. Voiceprint identification unit after 13 pairs step A3 by pretreatment of the speech signal processing, including windowing framing, pre-emphasis, endpoint detection, feature extraction voiceprint parameters, and then call the information storage unit 2 has voiceprint model voiceprint matches, if the match is successful, go to step A7; on the contrary, enter the next step;

.

A5 call voiceprint identification unit 13 from step A4 extracted characteristic parameters training voiceprint model, the model will be trained voiceprint information stored in the storage unit 2;

.

A6 user facing the TV voice collecting unit 11 say their favorite TV channel information such as channel codes "3" represents Hunan Satellite TV, they say the number "3" to step A3 restart;

.

A7 voice recognition unit 12 speech acquisition unit 11 acquired speech recognition, will recognize the TV channel code is stored in the information storage unit 2 and the voiceprint matches the voiceprint model successful association;

A8. Exit personalized information registration.

When the user's preferences change TV channels, information storage unit can be updated by re-

度，即将能量谱用如下式中的等响度曲线进行 加权。

E (omega) = (omega 2 + 56.8 x 10 6) omega 4 (omega 2 + 6.3 x 10 6) 2 (omega 2 + 0.38 x 10 9) (omega 6 + 9.58 x 10 26)

然后使用指数为1/15的幂函数对能量进行压缩，接下来做离散傅里叶反变 换，用dubin算法计算M阶全极点模型，最后对得到的模型系数进一步转换为倒谱系数得到PLP特征参数。根据端点检测结果对采集到的信号中为语音信号 之外的信号，如噪声或静音帧对应的特征参数直接丢弃。

三、声纹模型训练

本发明中的用户声纹模型训练采用的是总变化因子建模技术，该技术的目 标在于对语音信号中含有的说话人身份信息和周围环境噪声、麦克风传输等信道信息变化空间进行同时建模。给定一段用户的语音，经过上述步骤一和二的 处理后，其与说话人及信道相关的高斯混合模型GMM的均值超向量S表示如下：

S＝m+Tv+ε

其中m为通用背景模型的均值超向量，T矩阵为总变化空间，ε表示残 差，v矢量是与当前目标说话人(用户)及信道相关的总变化因子，它是一个低 维矢量，代表着当前说话人的身份信息，即用户的声纹模型。ε和v均服从高 斯分布：

ε～N(0，I) v～N(0，∑)

其中Σ为对角协方差矩阵。对于本发明而言，为了得到v，需要事先采用 与电视机语音唤醒场景近似的语音数据集合，利用期望最大化算法(EM)训练得到通用背景模型，得出总变化空间和残差向量。对于任意电视用户的语音，就 可利用已训练好的T，ε和m以及由用户语音构建的高斯混合模型S，求出每 个用户的总变化因子，即声纹模型。

四、声纹匹配

本发明使用的基于总变化因子的声纹识别系统中，采用两个总变化因子v_1和v_2之间的余弦距离来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

d (v 1, v 2) = < v 1, v 2 > | | v 1 | | · | | v 2 | |

如果d值大于预先设定的门限值则认为这两段语音来自同一个说话人，声纹匹 配成功，否则匹配失

personalized information registered 2 television channel information, then just in front of the TV voice collecting unit 11 television channels say Information can be, the voice control system automatically A6 started from the above step, without the need for user voiceprint remodel. automatically personalized voice wake television specific procedure (Figure 3) to implement the following:

C1 when the voice control system is in the "Voice wake up" turned on, the user facing the TV voice collecting unit 11 wake-up word spoken;.

C2 11 gathering input voice Speaker voice signal acquisition unit, and its A / D conversion, filter processing;.

C3. Speech recognition unit 12 through the above-described processing speech recognition, the recognized content with a preset wake-word matches, collected to determine whether the input speech signal wakeup command is TV voice, If the match is passed to the TV main controller 3, or exit the wake success will wake operating instructions;

C4. Voiceprint identification unit 13 processed voice information storage unit 2 voiceprint model voiceprint matches, if the match is successful, then the model associated with the user's voiceprint information television channel and wake operating instructions output to a television host controller 3, or withdraw wake;

C5. TV host controller 3 successfully receive voice control system transfer to the TV channel information and operation instruction after wake-up, perform TV start operation and automatically switch to the corresponding user's favorite TV channels. above-mentioned predetermined specific voice can be set to wake word phrases, such as "Turn on the TV," "On TV," "turn on the TV", "hello television", "hello TV", etc. and so on. The voice collecting unit 11 may collect equipment for the microphone or other voice, the voice signal for sensing peripheral.

Recognition algorithm of the present invention relates to the voiceprint identification unit 13 uses for the speaker based on the total change factor (also known as i-Vector) confirmation method to build concrete steps are as follows:

First, the pre-processing of the speech signal

败。有研究结果显示，通过概率线性判别分析规整后的总 变化因子的鉴别能力更强，可进一步提升声纹识别的性能，但由于本发明主要 适用于家庭的电视唤醒，家庭成员数目并不多，一般为1到7之间，因此上述 基于总变化因子的声纹识别算法已足以保证本发明在声纹识别方面的准确率。

下面介绍本发明中涉及到的语音识别单元所采用的语音识别算法

由于本发明主要用于电视唤醒和频道切换，用户使用时涉及到的唤醒词和 用于表示电视频道信息的词语数量都非常少，因此语音识别单元中采用的语音识别算法只需一套简单的基于GMM-HMM(基于高斯混合模型的隐马尔可夫模型) 识别系统即可，且训练和识别过程均采用离线方法，这样能够大大提高语音识 别的速度。构建一套GMM-HMM语音识别系统主要包括声学特征参数的提取，声 学模型和语言模型训练，语音解码过程。具体实现步骤如下：

一、声学特征参数的提取

由于感知线性预测系数中同时含有说话人的身份信息和语义信息，因此上 述介绍的声纹识别算法中的PLP参数提取方法同样用于语音识别，但PLP参数的维数为52维，同时为了增强声学特征的区分能力，需要将52维的PLP参数 进行异方差线性判别分析(Heteroscedastic Linear Discriminant Analysis, HLDA)后再用于声学模型建模和解码。

二、声学模型训练

声学模型的训练采用最大似然估计准则，需要用到期望最大化算法EM和 Baum-Welch算法，前者用来处理隐马尔可夫模型中由状态序列隐变量带来的不 完全数据下的训练问题，而后者用于从训练数据中高效的累积统计量，从而有 效抽取模型参数更新所需要的信息。对于本发明的应用，在进行声学模型训练 时，隐马尔可夫模型的状态数取5，每个状态使用一个含8个混合高斯成分的 GMM模型表示，第j状态的高斯混合密度表达式如下：

b j (x) = Sigma m = 1 M c jm p jm (x)

其中 Sigma m = 1 M c jm = 1, c jm greater than or equal to 0

p jm (x) = 1 (2 pi) D / 2 | Sigma jm | 1 / 2 exp { - 1 2 (x - mu jm) T Sigma jm - 1 (x - mu jm) }

M＝8表示高斯混合数目，x为任意语音帧的特征参数，D为特征的维数，c_{jm}， micro $_{jm}$和Σ_{jm}为第j状

Pretreatment

speech signal comprises speech signal pre-emphasis, plus the window framing and endpoint detection. For after A / D converted digital voice signal pre-emphasis using the following first-order digital filter to enhance the high-frequency component of the speech signal:

H (z) = 1-0.97z $^{-1}$ and then using the following formula Hamming window window function smoothed speech signal, and framing operations in the speech signal windowing process, where N is the length of the window.

W (n) = 0.54 0.46cos [2pin / (N-1)]

For a sampling rate of 8000Hz voice, frame length framing the use of 25 milliseconds, a frame shift of 10 ms. The present invention is based on the use of short-term energy endpoint detection of speech signal endpoint detection, to determine the start and end point of the speech signal, set x (�) through the pre-emphasis of the speech signal, then from the voice signal The n-th point short-term energy E_n windowed time is calculated as:

E n = Sigma m = - 8 8 [x (m) W (n - m) 2 = x 2 (n) * W 2 (n)

Second, extract voiceprint parameters extract voiceprint characteristic parameter that is extracted from the speaker's voice signal can reflect the personality of its identity parameters. The present invention is a 39-dimensional perception of linear prediction coefficient (Perception Linear Predictive Coefficients, PLP), the specific extraction process: first speech after the above steps do a short pretreatment fast Fourier transform, and then use the 24-step son Mel triangle filter bank with the frequency of critical analysis, and then to get the energy spectrum of the speech loudness curve pre-emphasis to simulate the auditory perception system sensitivity frequency energy spectrum coming by the following formula, etc. loudness weighting curve.

E (omega) = (omega 2 56.8 x 10 6) omega 4 (omega 2 6.3 x 10 6) 2 (omega 2 0.38 x 10 9) (omega 6 9.58 x 10 26)

Then use the index as a power function of the energy of 1/15 compression, inverse discrete Fourier transform to do next, calculate M-order all-

态的GMM模型中第m个高斯分量的权重，均值向量和协方差 矩阵。

三、语言模型训练

目前主流语音识别所用到的语言模型绝大部分是n-元(n-gram)统计语言模 型，是将词串的生成过程表示为一个词间跳转的n-1阶马尔可夫过程。训练和评价n-gram语言模型的指标为下式的模型复杂度(Perplexity)：

H = - min K - ∞ 1 K log 2 (P (w 1,..., w K)) ≈ - 1 K Sigma k = 1 K log 2 (P (w k | w k - 1, w k - 2,..., w k - N +!))

其中w_1,...,w_K为含有K个词的词序列，N为语言模型的元数，本发明中取 值为2。从理论上来看，语言模型对词串的期望复杂度应尽可能小，这表示语言模型在对给定的历史词串进行当前词预测时，能拥有更高的确信度。实际上语 言模型的训练准则就是使其对训练集合中句子的复杂度最小化。

四、语音解码

在给定声学模型和语言模型的情况下，解码器的好坏与语音识别的速度直 接相关，从而直接决定了用户的体验。虽然本发明用来唤醒电视所用的词语有限，由它们构成的解码网络并不大，但本发明仍然推荐采用基于加权有限状态 转换(WFST)的Viterbi解码技术，它可以将声学、语言模型和发音字典进行有 效的整合，对输入的语音信号PLP特征参数在词网络中进行搜索和匹配，从而 找到统计意义上最匹配的词串作为最终的识别结果。

上述语音识别单元中的声学模型和语言模型均可采用目前主流的开源工具 HTK或者Kaldi提前训练好并存储起来，在进行语音唤醒时，仅需调用存储好的 声学和语言模型，离线对采集到的用户语音进行PLP特征参数提取和解码即可。 实验证明，本发明方法具有训练时间短，声纹和语音识别速度都很快，且识别 率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云 端服务器，使用方便且保证了家庭信息的安全性。另外，本发明在电视关机时， 电视整体处于掉电状态，即电视机主控制器、语音识别单元、声纹识别单元、 信息存储单元均断电，只有语音采集单元实时采集用户的语音信号消耗电能， 而该模块耗电非常少，大大节省了电能消耗。

综上所述，本发明提供的基于声纹识别和语音识别进行个性化电视语音唤 醒的实现方法，能够在电视处于关机状态时，用户直接说类似于"hello电视" 这

pole model with dubin algorithm. Finally, the model coefficients are further converted to Cepstral PLP obtained characteristic parameters. According to the results of the collected endpoint detection signal is a signal to voice signal, such as noise or silence frame corresponding to the characteristic parameter discarded.

Third, voiceprint model training

The present invention is a user voiceprint model training uses a total change factor modeling techniques, the goal of this technology is to change the channel information space speaker identity information and ambient noise, the microphone to transmit voice signals contained etc. At the same time modeling. To fix a user's voice, after the processes of steps one and two, with the speaker and the average channel correlation Gaussian mixture model GMM ultra-vector S is expressed as follows:

S = m Tv e where m is the mean supervectors universal background model, T matrix for the total change in space, e represents the residual, v vector is the current target speaker (the user) and total change factors associated channel, which is a low-dimensional vector, representing the current speaker's identity information that the user's voiceprint model. e and v are subject to Gaussian distribution:

e ~ N (0, I) v ~ N (0, S) where S is a diagonal covariance matrix. For the purposes of the present invention, in order to obtain v, requires prior wake scene using approximate voice data collection and TV voice, use Expectation Maximization algorithm (EM) training to give general background model, gave a total change of space and residual vector. For any television user's voice, you can use the trained T, e and m and Gaussian mixture model S constructed by the user's voice, determine the total change factor for each user, that voiceprint model.

Fourth, voiceprint matching

The present invention is based on the use of the voice recognition system total change factors, the overall change in the use of two factor v_1 and v_2 to measure between the degree of matching voice contained two speaker voiceprint information, the formula is as follows: d (v 1, v 2) = <v 1, v 2> | | v 1 | | ⋅ | | v 2 | | cosine distance

样的唤醒词，电视机就可以实现直接开机并立即显示自己最喜爱的电视频道。 本发明的实现方法还可应用于电视机的用户个性化自动语音换台，同时可被移 植到通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

If d is greater than a predetermined threshold is considered two paragraphs from the same speaker's voice, voiceprint matches, otherwise the match fails. The results of research show that the probability of linear discriminant analysis by discriminating factor in the total change after more structured, can further improve the performance of voiceprint identification, but the present invention is primarily for home TV Wake, the number of family members is not much, generally between 1-7, so the above voiceprint recognition algorithm based on total variation factors are sufficient to ensure the accuracy of the present invention in terms of voiceprint identification.

Speech recognition algorithm

The following describes the present invention relates to a voice recognition unit used

Since the present invention is mainly used for wake-up and TV channel switching, when it comes to the user and a wake-up word indicates the number of words in the information channels are very small, so the speech recognition algorithm used in the speech recognition unit just one Based on a simple set of GMM-HMM (Gaussian mixture model based on hidden Markov model) identification system can be, and training and recognition processes are off-line method, which can greatly improve the speech recognition rate. Construction of a GMM-HMM speech recognition system includes an acoustic characteristic parameter extraction, acoustic model and language model training, speech decoding process. Specific steps are as follows:

First, extract the acoustic feature parameters

Since the perception of identity information and semantic information containing both linear prediction coefficient speaker, so the sound pattern recognition algorithm to the introduction of the PLP parameter extraction method is also used for voice recognition, but the dimension of the PLP for 52-dimensional parameter, Meanwhile, in order to enhance the ability to distinguish between acoustic features you need to be 52-dimensional parameter Heteroscedasticity PLP linear discriminant analysis (Heteroscedastic Linear Discriminant Analysis, HLDA) then used to carry out the acoustic modeling and decoding.

Second, the acoustic model training

Training

acoustic model using the maximum likelihood estimation, need to use EM expectation maximization algorithm and Baum-Welch algorithm, the former is used to dealing with incomplete data hidden Markov model brought down by the state series of hidden variables training issues and the latter for the efficient accumulation of statistics from the training data, so as to effectively extract model parameters to update the information they need. For the application of the present invention, during the acoustic model training, state the number of HMM take 5, each state GMM model using a Gaussian mixture containing eight components, said Gaussian mixture density of the j state following expression: b j (x) = Sigma m = 1 M c jm p jm (x) where Sigma m = 1 M c jm = 1, c jm greater than or equal to 0 p jm (x) = 1 (2 pi) D / 2 | Sigma jm | 1/2 exp {- 1 2 (x - mu jm) T Sigma jm - 1 (x - mu jm)}

M = 8 indicates the number of Gaussian mixture, x is the characteristic parameter to any speech frame, D is the dimension of features, c_{jm}, micro $_{jm}$ and S_{jm} j state for the GMM model in the m-th power of the Gaussian component weight, mean vector and covariance matrix.

Third, the language model training mainstream speech recognition language models currently used by the vast majority of n- yuan (n-gram) language model, the word sequence generation process is expressed as one word jumps between n-1-order Markov process. Training and Evaluation n-gram language model indicators for the formula complexity of the model (Perplexity):

H = - min K - 8 1 K log 2 (P (w 1,..., w K)) ≈ - 1 K Sigma k = 1 K log 2 (P (wk | wk - 1, wk - 2,..., wk -! N)) where w_1 ,..., w_K as word sequence contains K-word, N is metadata language model, the present invention, a value of 2. From the theoretical point of view, the complexity of the language model to the desired word strings should be as small as possible, which means that the language model in the history of a given word string current word prediction, to have a higher degree of certainty. In fact, the language model training guidelines is to make it to the training set sentence complexity is minimized.

Fourth, the speech decoding

In the case of a given acoustic model and language model, the quality and speed of speech recognition is directly related to the decoder, which directly determines the user's experience. Although the present invention is used to wake up terms used in television is limited, they constitute a decoding network is not large, but the present invention is still recommended based on weighted finite state transition (WFST) of Viterbi decoding technology, it can be acoustic, language model and pronunciation Dictionary for effective integration of the voice signal input characteristic parameters PLP search and matching network in words, in order to find the statistical sense of the word that best matches the string as the final recognition result. the speech recognition unit of acoustic models and language models can be with the current mainstream open source tools HTK or Kaldi good training and stored in advance, during a voice wake, just call the stored good acoustic and language models offline the collected user speech PLP feature extraction and decoding can be. Experiments show that the method of the present invention has the training time is short, voiceprint and voice recognition speed very quickly, and the recognition rate characteristics. Voiceprint and voice recognition are just off-line training and testing can be completed without the need to send to the cloud server, easy to use and ensure the safety of family information. Further, the present invention when the TV off, the TV is powered down overall, the TV host controller, a voice recognition unit, voiceprint identification unit, an information storage unit are off, only the voice collecting unit collecting the user's voice signal in real-time consumption energy, and the module power consumption is very low, saving power consumption.

In summary, the implementation personalize Television Voice wake voiceprint recognition and speech recognition based on the present invention provides, it is possible when the TV is turned off, users just say like "hello television" this wake-up word TV can be turned on and direct display your favorite TV channels immediately. Implementation method of the present invention can also be used in TV user personalization automatic speech for Taiwan, and can be ported to run on general-purpose high-speed processor or DSP chips, widely used in smart home related fields.

Citing: CN103442290 A ; CN103595869 A ; CN103632669 A ; CN103646646 A ;

Cited by: CN104821168 A ; CN104821168 B ; CN105049882 A ; CN105049882 B ; CN105137768 A ; CN105141768 A ; CN105206271 A ; CN105355195 A ; CN105374355 A ; CN105609103 A ; CN105719650 A ; CN105791935 A ; CN105915988 A ; CN105959806 A ; CN105979376 A ; CN106098068 A ; CN107015781 A ; CN108320753 A ; US9799335 BB ; WO17092189 A1 ; WO17217978 A1 ; WO19051668 A1 ;

Non patent citations: NAJIM DEHAK, ETC: "Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification",

(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 104575504 A
(43)申请公布日 2015.04.29

(21)申请号 201410840544.9

(22)申请日 2014.12.24

(71)申请人 上海师范大学
地址 200234 上海市徐汇区桂林路100号

(72)发明人 龙艳花 叶宏 魏浩然

(74)专利代理机构 上海申新律师事务所 31272
代理人 周云

(51)Int.Cl.
G10L 17/04(2013.01)
G10L 15/26(2006.01)
H04N 21/441(2011.01)
H04N 21/439(2011.01)

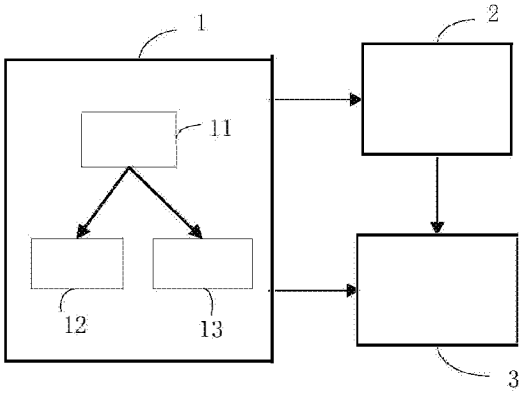
权利要求书3页 说明书7页 附图1页

(54)发明名称

采用声纹和语音识别进行个性化电视语音唤醒的方法

(57)摘要

本发明采用声纹和语音识别进行个性化电视语音唤醒的方法，具体指一种通过声纹识别进行电视机使用用户的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒，涉及声纹识别和语音识别技术。本发明构成系统包括语音控制系统(1)，信息存储单元(2)和电视机主控制器(3)以电信号方式连接组成。本发明具有训练时间短，声纹和语音识别速度都很快，且识别率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。还可应用于电视机的用户个性化自动语音换台，同时可被移植到通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。



CN 104575504 A

Petitioner Exhibit 1004, Page 917 of 1518

1.一种采用声纹和语音识别进行个性化电视语音唤醒的方法，通过采用声纹识别和语音识别系统，对用户的身份进行监测，实现个性化电视语音唤醒，其特征在于，包括如下步骤：

A. 个性化信息注册

采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别出用户最喜爱的电视频道信息；

B. 信息存储

将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电视机信息存储单元；

C. 语音唤醒

用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可根据声纹识别和语音识别系统对唤醒词的声纹和语音识别结果从电视机信息存储单元中调出用户的个性化信息，自动完成打开电视并自动切换到用户最喜爱的电视频道。

2.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述步骤A、步骤B，在用户作第一次操作时完成，以后只需操作步骤C，即可完成其个性化语音电视的唤醒；

当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进行用户个性化信息注册，对信息存储单元中的电视频道信息进行更新，此时无需进行声纹建模，仅需进行声纹匹配和语音识别。

3.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法的系统，其特征在于，所述系统包括用于用户个性化信息注册、唤醒词识别和声纹匹配的语音控制系统(1)，用于接收并存储从语音控制系统(1)中传递来的已训练好的用户声纹模型和已识别出的电视频道信息的信息存储单元(2)和接收从语音控制系统(1)和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操作的电视机主控制器(3)组成，以电信号方式连接，其中：

所述语音控制系统(1)由语音采集单元(11)，语音识别单元(12)和声纹识别单元(13)组成，其中：

所述语音采集单元(11)用于完成用户语音信号的采集和处理；

所述语音识别单元(12)用于识别用户所说的电视频道信息和唤醒词；

所述声纹识别单元(13)用于从用户语音中提取出代表其身份信息的声纹信息进行声纹建模和声纹匹配。

4.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特征在于，所述步骤A个性化信息注册，具体实现步骤为：

A1. 开启语音控制系统(1)，使其处于“用户注册”状态；

A2. 若用户为首次进行个性化信息注册，执行该步骤，否则转向步骤A6开始执行，用户对着电视机的语音采集单元(11)说话，说话内容为任意文本，持续时间至少30秒；

A3. 语音采集单元(11)采集输入的说话人语音信号，并对其进行A/D转换，滤波处理；

A4. 声纹识别单元(13)对经步骤A3处理后的语音信号作预处理，包括加窗分帧、预加重、端点检测，提取声纹特征参数，然后调用信息存储单元(2)中已有的声纹模型进行声纹匹配，若匹配成功，转向步骤A7；反之，进入下一步操作；

Petitioner Exhibit 1004, Page 918 of 1518

A5. 调用声纹识别单元(13)对从步骤A4中提取的特征参数进行声纹模型的训练,将训练好的声纹模型存储到信息存储单元(2);

A6. 用户对着电视机的语音采集单元(11)说出自己最喜欢的电视频道信息,如频道代码“3”代表湖南卫视,则说出数字“3”,转向步骤A3重新开始;

A7. 语音识别单元(12)对语音采集单元(11)采集到的语音进行识别,将识别出的电视频道代码存储到信息存储单元(2)中并与声纹匹配成功的声纹模型进行关联;

A8. 退出个性化信息注册。

5.如权利要求1所述的采用声纹和语音识别进行个性化电视语音唤醒的方法,其特征在于,所述步骤C语音唤醒,具体实现步骤为:

C1. 当语音控制系统(1)处于“语音唤醒”开启状态时,用户对着电视机语音采集单元(11)说出唤醒词;

C2. 语音采集单元(11)采集输入的说话人语音信号,并对其进行A/D转换,滤波处理;

C3. 语音识别单元(12)对上述处理过的语音进行识别,将识别出的内容与预先设定的唤醒词进行匹配,以判断采集到的输入语音信号是否为电视机的语音唤醒指令,若匹配成功则将唤醒操作指令传递给电视机主控制器(3),否则退出唤醒系统;

C4. 声纹识别单元(13)将处理过的语音与信息存储单元(2)中的声纹模型进行声纹匹配,若匹配成功,则将与用户声纹模型关联的电视频道信息和唤醒操作指令输出给电视机主控制器(3),否则退出唤醒系统;

C5. 电视机主控制器(3)在成功收到语音控制系统(1)传递来的电视频道信息和唤醒操作指令后,执行电视开机操作并自动切换到对应用户喜爱的电视频道上。

6.如权利要求4所述的采用声纹和语音识别进行个性化电视语音唤醒的方法,其特征在于,所述步骤A5中的声纹模型的训练方法为总变化因子建模技术。

7.如权利要求6所述的采用声纹和语音识别进行个性化电视语音唤醒的方法,其特征在于,所述总变化因子建模技术,给定一段用户的语音,与说话人及信道相关的高斯混合模型GMM的均值超向量S表示如下:

$$S = m + Tv + \varepsilon$$

其中,m为通用背景模型的均值超向量,T矩阵为总变化空间,ε为残差,v矢量是与当前目标说话人(用户)及信道相关的总变化因子,它代表着当前说话人的身份信息,即用户的声纹模型。

8.如权利要求7所述的采用声纹和语音识别进行个性化电视语音唤醒的方法,其特征在于,所述ε和v均服从高斯分布:

$$\varepsilon \sim N(0, I)$$

$$v \sim N(0, \Sigma)$$

其中,Σ为对角协方差矩阵,为得到v,需要事先采用与电视机语音唤醒场景近似的语音数据集合,利用期望最大化算法训练得到通用背景模型,得出总变化空间和残差向量;对于任意电视用户的语音,就可利用已训练好的T,ε和m以及由用户语音构建的高斯混合模型S,求出每个用户的总变化因子,即声纹模型。

9.如权利要求4,5所述的采用声纹和语音识别进行个性化电视语音唤醒的方法,其特征在于,所述声纹匹配,使用的是基于总变化因子的声纹识别系统,采用两个总变化因子 v_1

Petitioner Exhibit 1004, Page 919 of 1518

和 v_2之间的余弦距离来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

$$d(\mathbf{v}_1, \mathbf{v}_2) = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \cdot \|\mathbf{v}_2\|}$$

如果 d 值大于预先设定的门限值则认为这两段语音来自同一个说话人，声纹匹配成功，否则匹配失败。

10. 如权利要求 3 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法的系统，其特征在于，所述语音识别单元 (12) 声学模型训练时，隐马尔可夫模型的状态数取 5，每个状态使用一个含 8 个混合高斯成分的高斯混合模型 GMM 表示。

11. 如权利要求 3 所述的采用声纹和语音识别进行个性化电视语音唤醒的方法的系统，其特征在于，所述语音识别单元 (12) 的语音解码，采用将声学、语言模型和发音字典进行有效的整合，对输入的语音信号 PLP 特征参数在词网络中进行搜索和匹配，从而找到统计意义上最匹配的词串作为最终识别结果的基于加权有限状态转换的 Viterbi 解码。

Petitioner Exhibit 1004, Page 920 of 1518

采用声纹和语音识别进行个性化电视语音唤醒的方法

技术领域

[0001] 本发明涉及声纹识别和语音识别技术，通过声纹识别进行电视机使用用户的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒。

背景技术

[0002] 近年来，随着智能语音处理技术、互联网和云计算技术的快速发展，目前市面上出现的智能电视大部分都开始支持简单的语音操作，如进行音量调节，资源搜索等，同时电视语音助手等应用也开始出现。但由于中国各地的方言种类多、同音字多、各地方言的语音语调的不同、周围环境噪声等因素造成语音识别的准确率很低，电视无法执行用户语音命令的现象时常出现。因此，对消费者来说，目前的语音电视在一定程度上并不能满足其各项需求。同时，这些人机语音交互功能大多需要通过云端计算来完成，存在一定的安全隐患。

[0003] 现在几乎每个家庭都拥有一台或几台电视机，但一台电视通常会被几位家庭成员使用，特别是放置在客厅的电视，而每位家庭成员对电视节目频道的喜好却不尽相同，这时在打开电视后通常需要用遥控器对电视节目频道进行遍历切换以找到自己最喜欢的频道，非常不方便且繁琐。

发明内容

[0004] 为解决现有技术的不足，本发明提供一种采用语音进行个性化电视唤醒的实现方法，能够对用户语音采用声纹识别和语音识别，实现对用户的身份进行监测并使用非常简单的唤醒词就能唤醒电视打开其最喜欢的电视频道。

[0005] 众所周知，声纹识别又称说话人识别，是生物识别技术的一种，声纹识别和语音识别均属于智能语音处理技术的范畴。具体是将说话人语音信号中的声纹特征参数与训练好的声纹模型进行模式匹配从而自动进行说话人身份确认或辨认的技术。语音识别是通过模式识别和语言理解将语音信号中包含的文字信息提取出来的技术，主要包括特征提取、声学模型和语言模型训练、解码技术等。

[0006] 本发明具体实现步骤如下：

[0007] A. 个性化信息注册

[0008] 采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别出用户最喜爱的电视频道信息。

[0009] B. 信息存储

[0010] 将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电视机信息存储单元。

[0011] C. 语音唤醒

[0012] 用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可根据声纹识别和语音识别系统对唤醒词的声纹和语音识别结果从电视机信息存储单元中调出用户的个性

Petitioner Exhibit 1004, Page 921 of 1518

化信息，自动完成打开电视并自动切换到用户最喜爱的电视频道。

[0013] 以上所述的采用声纹和语音识别进行个性化电视语音唤醒的方法，其特点是，所述步骤A、步骤B，在用户作第一次操作时完成，以后只需操作步骤C，即可完成其个性化语音电视的唤醒。

[0014] 当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进行用户个性化信息注册，对信息存储单元中的电视频道信息进行更新，此时无需进行声纹建模，仅需进行声纹匹配和语音识别。

[0015] 由于本发明使用了声纹识别技术，很容易对不同用户建立起其对应的声纹密钥（声纹模型），通过不同用户声纹密钥的唯一性实现对其喜爱的电视频道信息进行关联和保护，进而实现具有用户个性化的电视语音唤醒功能。

[0016] 本发明与现有技术相比，具有以下特点：

[0017] 第一、用户只需张嘴说简单的唤醒词，如“hello电视”就能自动打开电视并立即切换到自己最喜欢的电视台节目频道，无需再手动开机遥控，给用户带来智能、时尚的应用体验。

[0018] 第二、个性化注册时的声纹识别性能与所说的文本和使用的语言无关，方便使用。

[0019] 第三、用声纹作为密码，简单的电视频道代码和唤醒词语音作为交互方式，因此语音识别系统仅需实现简单的关键词识别即可，识别速度快，且声纹识别和语音识别仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。

[0020] 第四、可设置多位用户的个性化语音开机信息，除唤醒功能外，同时也可用于个性化自动语音换台。

[0021] 第五、本发明的算法实现模块具有很好的可移植性和可扩展性，可被移植到有线电视的机顶盒、通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

附图说明

[0022] 图1为本发明采用声纹和语音识别进行个性化电视语音唤醒的方法系统组成框图；

[0023] 图2为本发明实施例用户个性化信息注册系统流程图；

[0024] 图3为本发明实施例个性化语音唤醒电视的流程图。

具体实施方式

[0025] 以下结合附图和实施例对本发明作进一步的描述

[0026] 本发明构建了一套采用声纹和语音识别进行个性化电视语音唤醒的系统（如图1所示），该系统包括用于用户个性化信息注册、唤醒词识别和声纹匹配的语音控制系统1，用于接收并存储从语音控制系统1中传递来的已训练好的用户声纹模型和已识别出的电视频道信息的信息存储单元2和接收从语音控制系统1和信息存储单元中传递的控制信息，负责执行个性化电视语音唤醒操作的电视机主控制器3组成，以电信号方式连接。其中：

[0027] 所述语音控制系统1由语音采集单元11，语音识别单元12和声纹识别单元13组成，其中：

Petitioner Exhibit 1004, Page 922 of 1518

[0028] 所述语音采集单元 11 用于完成用户语音信号的采集和处理；

[0029] 所述语音识别单元 12 用于识别用户所说的电视频道信息和唤醒词；

[0030] 所述声纹识别单元 13 用于从用户语音中提取出代表其身份信息的声纹信息进行声纹建模和声纹匹配。

[0031] 本发明的语音控制系统有二个状态按钮，即“用户注册”和“语音唤醒”。当用户需要进行个性化信息注册时开启“用户注册”，平时均处于“语音唤醒”状态，这样有利于节省电能。

[0032] 下面结合附图分别介绍用户个性化信息注册过程和自动语音唤醒的具体实施方式

[0033] 用户首次进行个性化信息注册时（如图 2 所示）实施例，需对电视机说话二次，第一次为任意文本的语音，用于声纹建模，第二次为用户自己最喜欢的电视频道信息，用于将电视频道信息与用户声纹模型进行关联。具体步骤如下：

[0034] A1. 开启语音控制系 1，使其处于“用户注册”状态；

[0035] A2. 若用户为首次进行个性化信息注册，执行该步骤，否则转向步骤 A6 开始执行，用户对着电视机的语音采集单元 11 说话，说话内容为任意文本，持续时间至少 30 秒；

[0036] A3. 语音采集单元 11 采集输入的说话人语音信号，并对其进行 A/D 转换，滤波处理；

[0037] A4. 声纹识别单元 13 对经步骤 A3 处理后的语音信号作预处理，包括加窗分帧、预加重、端点检测，提取声纹特征参数，然后调用信息存储单元 2 中已有的声纹模型进行声纹匹配，若匹配成功，转向步骤 A7；反之，进入下一步操作；

[0038] A5. 调用声纹识别单元 13 对从步骤 A4 中提取的特征参数进行声纹模型的训练，将训练好的声纹模型存储到信息存储单 2；

[0039] A6. 用户对着电视机的语音采集单元 11 说出自己最喜欢的电视频道信息，如频道代码“3”代表湖南卫视，则说出数字“3”，转向步骤 A3 重新开始；

[0040] A7. 语音识别单元 12 对语音采集单元 11 采集到的语音进行识别，将识别出的电视频道代码存储到信息存储单元 2 中并与声纹匹配成功的声纹模型进行关联；

[0041] A8. 退出个性化信息注册。

[0042] 当用户的电视频道喜好发生改变时，可通过重新进行个性化信息注册来更新信息存储单元 2 中的电视频道信息，此时只需对着电视机的语音采集单元 11 说出电视频道信息即可，语音控制系统 1 会自动从上述的步骤 A6 开始执行，无需对用户声纹重新建模。

[0043] 自动进行个性化语音唤醒电视（如图 3 所示）实施例的具体步骤如下：

[0044] C1. 当语音控制系统 1 处于“语音唤醒”开启状态时，用户对着电视机语音采集单元 11 说出唤醒词；

[0045] C2. 语音采集单元 11 采集输入的说话人语音信号，并对其进行 A/D 转换，滤波处理；

[0046] C3. 语音识别单元 12 对上述处理过的语音进行识别，将识别出的内容与预先设定的唤醒词进行匹配，以判断采集到的输入语音信号是否为电视机的语音唤醒指令，若匹配成功则将唤醒操作指令传递给电视机主控制器 3，否则退出唤醒系统；

[0047] C4. 声纹识别单元 13 将处理过的语音与信息存储单元 2 中的声纹模型进行声纹匹

Petitioner Exhibit 1004, Page 923 of 1518

配，若匹配成功，则将与用户声纹模型关联的电视频道信息和唤醒操作指令输出给电视机主控制器 3，否则退出唤醒系统；

[0048] C5. 电视机主控制器 3 在成功收到语音控制系统 1 传递来的电视频道信息和唤醒操作指令后，执行电视开机操作并自动切换到对应用户喜爱的电视频道上。

[0049] 上述提到的预先设定的特定语音唤醒词可为任意设定的词组，如“打开电视机”、“开电视”、“打开电视”、“hello 电视”、“hello TV”等等。语音采集单元 11 可为麦克风或其他语音采集设备，用于感知周边的语音信号。

[0050] 本发明涉及到的声纹识别单元 13 所采用的识别算法为基于总变化因子（又称为 i-Vector）的说话人确认方法，其具体构建步骤如下：

[0051] 一、语音信号的预处理

[0052] 语音信号的预处理包括对语音信号进行预加重、加窗分帧和端点检测。对于经过 A/D 转换后的数字语音信号，预加重使用如下的一阶数字滤波器来提升语音信号的高频分量：

[0053] $H(z) = 1-0.97z^{-1}$

[0054] 然后采用如下式的汉明窗窗函数对语音信号进行平滑，并在加窗过程中进行语音信号的分帧操作，其中 N 为窗的长度。

[0055] $W(n) = 0.54+0.46\cos[2\pi n/(N-1)]$

[0056] 对于采样率为 8000Hz 的语音，分帧时使用的帧长为 25 毫秒，帧移为 10 毫秒。本发明采用的是基于短时能量的端点检测法对语音信号进行端点检测，从而判断语音信号的起始点和终止点，设 x(·) 为通过预加重后的语音信号，则表示从语音信号的第 n 个点开始加窗时的短时能量 E_n 计算公式为：

[0057] $$E_n = \sum_{m=-\infty}^{\infty}\left[x(m)W(n-m)\right]^2 = x^2(n) * W^2(n)$$

[0058] 二、声纹特征参数的提取

[0059] 声纹特征参数的提取即从说话人的语音信号中提取出能反映其身份个性的参数。本发明使用的是 39 维的感知线性预测系数（Perception Linear Predictive Coefficients，PLP），具体提取过程为：首先对上述步骤一预处理过后的语音做短时快速傅里叶变换，然后使用 24 阶子带的梅尔三角滤波器组进行临界频率分析，接着对得到的语音能量谱进行等响度曲线预加重用以模拟人耳的听觉感知系统对频率的敏感度，即将能量谱用如下式中的等响度曲线进行加权。

[0060] $$E(\omega) = \frac{(\omega^2 + 56.8\times10^6)\omega^4}{(\omega^2 + 6.3\times10^6)^2 \times (\omega^2 + 0.38\times10^9)(\omega^6 + 9.58\times10^{26})}$$

[0061] 然后使用指数为 1/15 的幂函数对能量进行压缩，接下来做离散傅里叶反变换，用 dubin 算法计算 M 阶全极点模型，最后对得到的模型系数进一步转换为倒谱系数得到 PLP 特征参数。根据端点检测结果对采集到的信号中为语音信号之外的信号，如噪声或静音帧对应的特征参数直接丢弃。

[0062] 三、声纹模型训练

[0063] 本发明中的用户声纹模型训练采用的是总变化因子建模技术，该技术的目标在于

Petitioner Exhibit 1004, Page 924 of 1518

对语音信号中含有的说话人身份信息和周围环境噪声、麦克风传输等信道信息变化空间进行同时建模。给定一段用户的语音，经过上述步骤一和二的处理后，其与说话人及信道相关的高斯混合模型 GMM 的均值超向量 S 表示如下：

[0064] $S = m + Tv + \varepsilon$

[0065] 其中 m 为通用背景模型的均值超向量，T 矩阵为总变化空间，ε 表示残差，v 矢量是与当前目标说话人（用户）及信道相关的总变化因子，它是一个低维矢量，代表着当前说话人的身份信息，即用户的声纹模型。ε 和 v 均服从高斯分布：

[0066] $\varepsilon \sim N(0, I)$

[0067] $v \sim N(0, \Sigma)$

[0068] 其中 Σ 为对角协方差矩阵。对于本发明而言，为了得到 v，需要事先采用与电视机语音唤醒场景近似的语音数据集合，利用期望最大化算法 (EM) 训练得到通用背景模型，得出总变化空间和残差向量。对于任意电视用户的语音，就可利用已训练好的 T，ε 和 m 以及由用户语音构建的高斯混合模型 S，求出每个用户的总变化因子，即声纹模型。

[0069] 四、声纹匹配

[0070] 本发明使用的基于总变化因子的声纹识别系统中，采用两个总变化因子 v_1 和 v_2 之间的余弦距离来度量两段语音中所含说话人声纹信息的匹配程度，其公式如下：

[0071] $$d(\mathbf{v}_1, \mathbf{v}_2) = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \cdot \|\mathbf{v}_2\|}$$

[0072] 如果 d 值大于预先设定的门限值则认为这两段语音来自同一个说话人，声纹匹配成功，否则匹配失败。有研究结果显示，通过概率线性判别分析规整后的总变化因子的鉴别能力更强，可进一步提升声纹识别的性能，但由于本发明主要适用于家庭的电视唤醒，家庭成员数目并不多，一般为 1 到 7 之间，因此上述基于总变化因子的声纹识别算法已足以保证本发明在声纹识别方面的准确率。

[0073] 下面介绍本发明中涉及到的语音识别单元所采用的语音识别算法

[0074] 由于本发明主要用于电视唤醒和频道切换，用户使用时涉及到的唤醒词和用于表示电视频道信息的词语数量都非常少，因此语音识别单元中采用的语音识别算法只需一套简单的基于 GMM-HMM（基于高斯混合模型的隐马尔可夫模型）识别系统即可，且训练和识别过程均采用离线方法，这样能够大大提高语音识别的速度。构建一套 GMM-HMM 语音识别系统主要包括声学特征参数的提取，声学模型和语言模型训练，语音解码过程。具体实现步骤如下：

[0075] 一、声学特征参数的提取

[0076] 由于感知线性预测系数中同时含有说话人的身份信息和语义信息，因此上述介绍的声纹识别算法中的 PLP 参数提取方法同样用于语音识别，但 PLP 参数的维数为 52 维，同时为了增强声学特征的区分能力，需要将 52 维的 PLP 参数进行异方差线性判别分析 (Heteroscedastic Linear Discriminant Analysis, HLDA) 后再用于声学模型建模和解码。

[0077] 二、声学模型训练

[0078] 声学模型的训练采用最大似然估计准则，需要用到期望最大化算法 EM 和

Petitioner Exhibit 1004, Page 925 of 1518

Baum-Welch 算法，前者用来处理隐马尔可夫模型中由状态序列隐变量带来的不完全数据下的训练问题，而后者用于从训练数据中高效的累积统计量，从而有效抽取模型参数更新所需要的信息。对于本发明的应用，在进行声学模型训练时，隐马尔可夫模型的状态数取 5，每个状态使用一个含 8 个混合高斯成分的 GMM 模型表示，第 j 状态的高斯混合密度表达式如下：

[0079] $$b_j(x)=\sum_{m=1}^{M}c_{jm}p_{jm}(x)$$

[0080] 其中 $$\sum_{m=1}^{M}c_{jm}=1,\quad c_{jm}\geq 0$$

[0081] $$p_{jm}(x)=\frac{1}{(2\pi)^{D/2}\left|\Sigma_{jm}\right|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu_{jm})^{T}\Sigma_{jm}^{-1}(x-\mu_{jm})\right\}$$

[0082] M = 8 表示高斯混合数目，x 为任意语音帧的特征参数，D 为特征的维数，c_{jm}，μ_{jm} 和 Σ_{jm} 为第 j 状态的 GMM 模型中第 m 个高斯分量的权重，均值向量和协方差矩阵。

[0083] 三、语言模型训练

[0084] 目前主流语音识别所用到的语言模型绝大部分是 n- 元 (n-gram) 统计语言模型，是将词串的生成过程表示为一个词间跳转的 n-1 阶马尔可夫过程。训练和评价 n-gram 语言模型的指标为下式的模型复杂度 (Perplexity)：

[0085] $$H=-\lim_{K\to\infty}\frac{1}{K}\log_2\left(P(w_1,\cdots,w_K)\right)\approx-\frac{1}{K}\sum_{k=1}^{K}\log_2\left(P(w_k\mid w_{k-1},w_{k-2},\cdots,w_{k-N+1})\right)$$

[0086] 其中 $w_1,\cdots,w_K$ 为含有 K 个词的词序列，N 为语言模型的元数，本发明中取值为 2。从理论上来看，语言模型对词串的期望复杂度应尽可能小，这表示语言模型在对给定的历史词串进行当前词预测时，能拥有更高的确信度。实际上语言模型的训练准则就是使其对训练集合中句子的复杂度最小化。

[0087] 四、语音解码

[0088] 在给定声学模型和语言模型的情况下，解码器的好坏与语音识别的速度直接相关，从而直接决定了用户的体验。虽然本发明用来唤醒电视所用的词语有限，由它们构成的解码网络并不大，但本发明仍然推荐采用基于加权有限状态转换 (WFST) 的 Viterbi 解码技术，它可以将声学、语言模型和发音字典进行有效的整合，对输入的语音信号 PLP 特征参数在词网络中进行搜索和匹配，从而找到统计意义上最匹配的词串作为最终的识别结果。

[0089] 上述语音识别单元中的声学模型和语言模型均可采用目前主流的开源工具 HTK 或者 Kaldi 提前训练好并存储起来，在进行语音唤醒时，仅需调用存储好的声学和语言模型，离线对采集到的用户语音进行 PLP 特征参数提取和解码即可。实验证明，本发明方法具有训练时间短，声纹和语音识别速度都很快，且识别率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。另外，本发明在电视关机时，电视整体处于掉电状态，即电视机主控制器、语音识别单元、声纹识别单元、信息存储单元均断电，只有语音采集单元实时采集用户的语音信号消耗电能，而

Petitioner Exhibit 1004, Page 926 of 1518

该模块耗电非常少，大大节省了电能消耗。

[0090] 综上所述，本发明提供的基于声纹识别和语音识别进行个性化电视语音唤醒的实现方法，能够在电视处于关机状态时，用户直接说类似于“hello 电视”这样的唤醒词，电视机就可以实现直接开机并立即显示自己最喜爱的电视频道。本发明的实现方法还可应用于电视机的用户个性化自动语音换台，同时可被移植到通用的DSP 高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

Petitioner Exhibit 1004, Page 927 of 1518

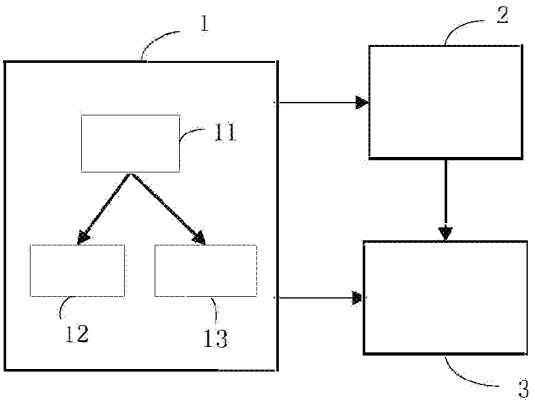


图 1

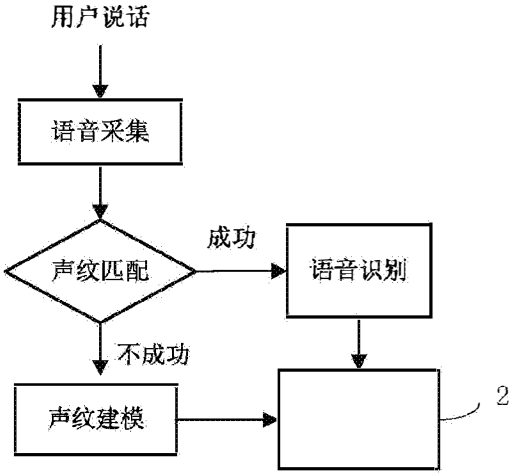


图 2

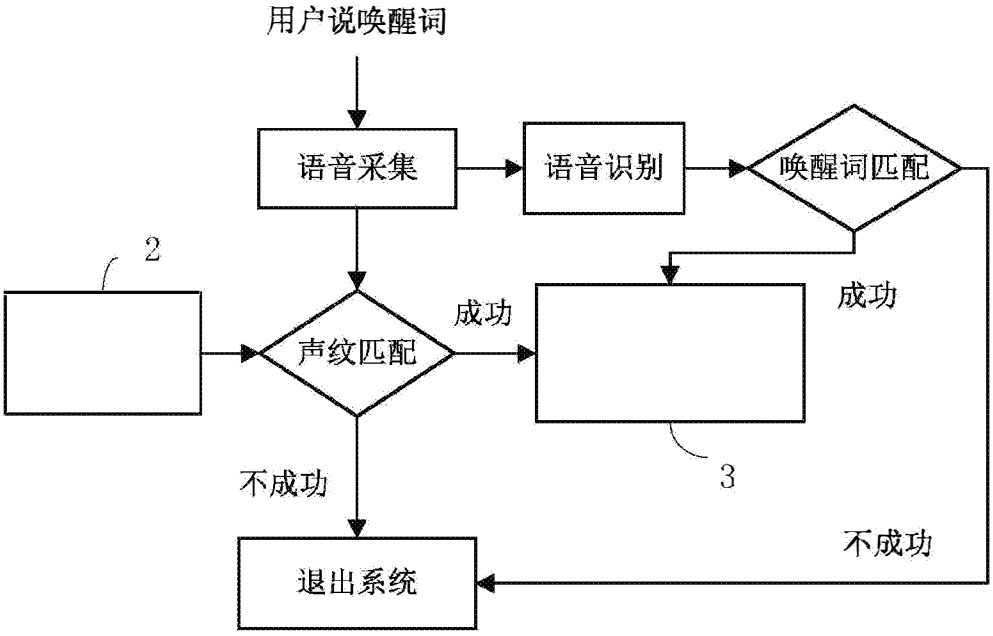


图 3

Petitioner Exhibit 1004, Page 928 of 1518

PATENT COOPERATION TREATY

ADVANCE E-MAIL

PCT

NOTIFICATION CONCERNING
TRANSMITTAL OF COPY OF INTERNATIONAL
PRELIMINARY REPORT ON PATENTABILITY
(CHAPTER I OF THE PATENT COOPERATION
TREATY)
(PCT Rule 44bis.1(c))

From the INTERNATIONAL BUREAU

To:

RUDOY, Daniel, G.
Wolf, Greenfield & Sacks. P.C.
600 Atlantic Avenue
Boston, MA 02210-2206
ETATS-UNIS D'AMERIQUE

Date of mailing *(day/month/year)*
27 December 2018 (27.12.2018)

Applicant's or agent's file reference
N0484.71178

IMPORTANT NOTICE

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/037495	15 June 2016 (15.06.2016)	

Applicant
NUANCE COMMUNICATIONS, INC.

The International Bureau transmits herewith a copy of the international preliminary report on patentability (Chapter I of the Patent Cooperation Treaty)

The International Bureau of WIP●
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No. +41 22 338 82 70

Authorized officer
Kihwan Moon

e-mail: pct.team1@wipo.int

Form PCT/IB/326 (January 2004)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY REPORT ON PATENTABILITY
(Chapter I of the Patent Cooperation Treaty)

(PCT Rule 44*bis*)

Applicant's or agent's file reference N0484.71178	**FOR FURTHER ACTION**	See item 4 below
International application No. PCT/US2016/037495	International filing date *(day/month/year)* 15 June 2016 (15.06.2016)	Priority date *(day/month/year)*
International Patent Classification (8th edition unless older edition indicated) See relevant information in Form PCT/ISA/237		
Applicant NUANCE COMMUNICATIONS, INC.		

1. This international preliminary report on patentability (Chapter I) is issued by the International Bureau on behalf of the International Searching Authority under Rule 44 *bis*.1(a).

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   In the attached sheets, any reference to the written opinion of the International Searching Authority should be read as a reference to the international preliminary report on patentability (Chapter I) instead.

3. This report contains indications relating to the following items:

☒	Box No. I	Basis of the report
☐	Box No. II	Priority
☐	Box No. III	Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
☐	Box No. IV	Lack of unity of invention
☒	Box No. V	Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
☐	Box No. VI	Certain documents cited
☐	Box No. VII	Certain defects in the international application
☐	Box No. VIII	Certain observations on the international application

4. The International Bureau will communicate this report to designated Offices in accordance with Rules 44*bis*.3(c) and 93*bis*.1 but not, except where the applicant makes an express request under Article 23(2), before the expiration of 30 months from the priority date (Rule 44*bis* .2).

	Date of issuance of this report 18 December 2018 (18.12.2018)
The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. +41 22 338 82 70	Authorized officer Kihwan Moon e-mail: pct.team1@wipo.int

Form PCT/IB/373 (January 2004)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/037495	15.06.2016	

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/22 ADD. G10L15/08 G10L17/22

Applicant
NUANCE COMMUNICATIONS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	see form PCT/ISA/210	Zimmermann, Elko Telephone No. +49 89 2399-0

Form PCT/ISA/237 (Cover Sheet) (January 2015)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-53
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-53
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-53
	No:	Claims	

2. Citations and explanations

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

CN 104 575 504 A is regarded as being the prior art closest to the subject-matter of claim 1, and discloses a system wherein an acoustic signal is processed in parallel by a speech recognizer and a voiceprint identification unit in order to determine whether a wake-up command was received from a registered user, see figure 3 and paragraphs [0050] to [0056] of the computer-generated translation annexed to this opinion, which correspond to paragraphs [0043] to [0049] of CN 104 575 504 A.

The subject-matter of claim 1 differs from this known system in that the acoustic signal is interpreted using information indicative of the speaker's identity and is therefore new (Article 33(2) PCT).

The problem to be solved by the present invention may be regarded as improving the robustness of the wake-up word recognition.

The solution to this problem proposed in claim 1 of the present application is considered as involving an inventive step (Article 33(3) PCT) for the following reasons: By using information indicative of the speaker's identity components of the speech recognizer like vocabulary or acoustic model can be adapted in order to improve the speech recognizer's ability to accurately determine whether a speaker uttered any one of the wake-up words. In the prior art at hand there is no hint at interpreting the acoustic signal using information indicative of the speaker's identity, and it is not considered to be obvious in view of the parallel processing illustrated in figure 3 of CN 104 575 504 A.

Claims 20 and 37 correspond to claim 1, and claims 2-19, 21-36 and 38-53 are dependent claims and as such also meet the requirements of the PCT with respect to novelty and inventive step.

PatentOrder

Publication number: CN105575395 A

Publication country: CHINA

Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION

Publication date: 20160511

Application number: CN201410541444

Application date: 20141014

Priority: CN201410541444 20141014 CN201410541444;

Assignee[std]: ZTE CORP ;

Inventor[std]: WANG JIE ;

International class[8]: G10L17/22; G10L15/22;

Family members: CN105575395 A

Title: VOICE WAKE-UP METHOD AND APPARATUS, TERMINAL, AND PROCESSING METHOD THEREOF

Title: 语音唤醒方法及装置、终端及其处理方法

Abstract:

The invention discloses a voice wake-up method and apparatus, a terminal, and a processing method thereof. The voice wake-up method comprises: a wake-up word voice signal sent out by a terminal user is collected and a wake-up word identification result and a vocal print feature of the wake-up word voice signal are obtained and stored; when a terminal is in a sleep state, a voice signal inputted externally is detected; a voice identification result and a vocal print feature of the voice signal are obtained; the voice identification result and the stored wake-up word identification result as well as a matching result of the vocal print feature and the stored vocal print feature are obtained; and according to the matching result, whether the terminal needs to be woken up is determined; and if so, the terminal is woken up. With the method, technical problems of low user wake-up rate, high wrong wake-up rate, and poor user experience according to the existing voice wake-up technology are solved.

Abstract:

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

The present invention discloses a method and apparatus for voice, terminal and processing method to wake up. Voice wake-up method of the present invention includes: a wake-word speech signal acquisition end-user issue, get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and stored; the terminal is in a dormant state, the detection and external input speech signal; acquiring the speech recognition result voiceprint characteristics of the speech signal; obtaining the wake-up word speech

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。

recognition result of the recognition result and storage, as well as matches the voiceprint and voiceprint stored; according to the match results to determine whether the need to wake up the terminal, and if so, to wake the terminal; the voice of the present invention is a wake-up method can solve the prior art existing user voice wake wake low error rate and wake poor user experience technical problems.

Claims:

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：

.

1 A speech wake method comprising the steps of: acquisition wake-word speech signals sent by the end-user, access to the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

In the terminal is in a dormant state, detection and external input speech signal;

Being the speech signal a speech recognition result voiceprint feature;

Get the wake-up word speech recognition result of the recognition result and storage, as well as matches the voiceprint and voiceprint characteristic features stored;

According to the matching results to determine whether the need to wake up the terminal, and if so, to wake the terminal.

Before

2. The method of claim 1 or claim 2, characterized in that, after the detection of the external input speech signal, the speech recognition result to obtain the speech signal and a voiceprint characteristic, further They include:

Analyzing the voice signal meets a preset signal requirements, if, in the speech signal to obtain speech recognition results and a voiceprint characteristic.

.

3 The method as claimed in claim 2, characterized in that the determination whether the speech signal is a signal meets a preset requirement comprises: of the speech signal is processed to obtain user data and voice background noise data;

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；

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：

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.

4. The method as claimed in claim 3, characterized in that said voice data to said user data and said background noise and background noise models are preset to match steps include: 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 of the pre-match match points to obtain a second value;

According to the matching results to determine whether the speech signal meets a preset signal required steps 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 score is determined whether the speech signal is greater than a preset threshold, if it is determined to meet the requirements of a preset signal, and if not, it is determined that the signal does not meet the preset requirements.

Methods

5 as claimed in claim any one of claims 1-4, characterized in that the said matching results according to the procedure judges whether the wake-up terminal comprising:

When the speech recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the voiceprint of the speech signal stored speech wakeup word voiceprint match is found, determination need to wake up the terminal.

.

6 The method as claimed in one any of claim 2-4, wherein said determining step includes the required preset signal of said speech signal meets:

After the first multi-core audio processor core unit wakes up, the first nuclear unit judges that the

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；

speech signal meets the preset signal requirements;

Get the voiceprint features speech recognition result of the speech signal comprises:

When the first nuclear unit is judged to be the first nuclear unit of the multi-core audio processor in the wake of a second nuclear unit;

The second nuclear unit acquired by the speech recognition result of the speech signal voiceprint feature;

Get the wake of the speech recognition result of the word recognition result is stored, and the matching results of the step of the voiceprint stored voiceprint features include:

The second nuclear unit acquires wake word recognition result of the speech recognition result and storage, as well as matches the voiceprint characteristic with the stored voiceprint said;

According to the result of the match is determined whether the terminal needs to wake up, if the step of the terminal include wake-up:

The second nuclear unit is required to wake the terminal according to the matching result of the determination, if so, the second nuclear unit to wake the terminal.

.

7 The method as claimed in claim 6, characterized in that, further comprising:

When the first core unit judgment is NO, the first nuclear unit goes to sleep; and / or

When the second nuclear unit judgment is NO, the second nuclear unit into hibernation.

.

8 as claimed in claim 6, characterized in that said second core said terminal means comprises wake-up steps:

The second nuclear unit sends an interrupt event to wake the dormant terminal in the main processor.

.

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；

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

9 A terminal processing method, characterized by comprising the steps of:

Voice wake use claims 1-5 method according to any one of the wake-up terminal;

After the wake-up terminal, perform the corresponding operation according to preset voice recognition result of the speech signal.

.

10 The method as claimed in claim 9, characterized in that said terminal comprises the step of wake-up: generated based on the recognition result of the speech signal preset type of wake-up interrupt events to wake the terminal;

The default action steps speech recognition result of the speech signal include: type of interrupt event to perform the corresponding operation according to the preset wake.

.

11 A speech wake apparatus, characterized in that, 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.

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。

.

12 apparatus as claimed in claim 11, characterized by further comprising: a second judging module;

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.

.

13 The device of claim 12, characterized in that said 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.

14. Apparatus as claimed in claim 12 or 13, characterized in that said first acquisition module, a second acquisition module, the judging module and the first module wakeup integrated multi-core audio processor of the 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.

15. Apparatus according to any one of claim 11-13, wherein said first determining module is used for the speech recognition result when the speech signal and wake word stored speech recognition result matches, and the voiceprint of the speech signal stored speech wakeup word voiceprint

match is successful, it is determined that the terminal needs to wake up.

.

16 A terminal, characterized by comprising: a speech according to any one of the processing modules 11-13 and a wake-up apparatus as claimed in claim;

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.

17. The terminal as claimed in claim 16, wherein said speech wake-up means for generating a wake-up module wakeup preset type of the identification result of the speech signal interrupt event wake the terminal;

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

Description:

语音唤醒方法及装置、终端及其处理方法

技术领域

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

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。

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。

Voice wake-up method and apparatus, terminal and approach

FIELD

The present invention relates to the field of speech recognition technology, particularly to a method and apparatus for voice wake, terminal and processing methods.

BACKGROUND voice in mobile terminals has not only exist in a separate program applications, and slowly began to add as an independent input finger touch input, without touching the phone, remote control and more intelligent and friendly service to mobile phone users.

The Voice wake-up technology to provide users with complete hands-free voice commands remote control handset provides premise wake wake mainstream technology software, but the software running on premise phone system startup, is guaranteed to receive the user's voice commands anytime, anywhere voice engine needs to run in the background and has been listening, mobile phone system will not enter standby power-saving sleep state, these wake-up voice application

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。

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：

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。

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。

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。

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：

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。

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。

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

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。

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。

发明内容

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。

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：

technology for power consumption unusually strong. Speech technology currently in the market not only in technology application taking into account the functional-friendliness, but also concerned about the mobile phone power consumption, and to ask for too much is not the voice of the user to bother, so now most of the manufacturers wake up programs for speech recognition system use hibernation wakeup hardware keys or cell phone system is running software button clicks wake.

With the continuous improvement of technical personnel, has now appeared speech Low-power wake-up technology, the application of low-power sleep state in the mobile phone chip for direct use voice commands to wake the phone operating system, without requiring the user to manually operate the hardware and software.

However, the prior art voice low-power wake-up function, you can only use a fixed wake-word speech model generated by this preset fixed wake-acoustic model of the word. Specifically, existing voice low-power wake-up implementation functions, including the following:

Record number of large-volume "Selected wake word" crowd sound sample processing and acoustic features training and extraction.

The statistics estimated from a "fixed wake-up word" sound characteristics preset in the phone system.

The sound waveform individual user input and generate sound statistical characteristics for comparison. due to the low power performance audio chip arithmetic processing capacity is limited, the sample population statistics generated acoustic model will be in a separate individual users use the following questions:

1, the user's individual sound pattern difference is far below or above sampling statistical models feature extraction confidence interval, causing the user to wake low. If the sample size is too small or too simple or population characteristics, for example, all Mandarin, all adults, will result in sound Some specific groups, such as the voices of children, the elderly fail to wake up.

2, in noisy environments generate complex noise environment sampling statistical model similar likelihood curve, causing waking up rate. When

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。

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。

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：

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。

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：

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。

should wake wake chaos, causing the user problems.

3, since the only pre-recorded "fixed wake-up word" by manufacturers to wake up the terminal, the user can not

Custom enough wake-up word, can not meet the needs of individual users, reducing the user experience.

Thus, although the prior art to achieve a certain power wake-up voice control, but there are a wake-up users is low, error rate, and wake up poor user experience technical problems.

Summary

The main technical problem to be solved by the present invention is to provide a method and apparatus for voice, terminal and approach wake, can solve the existing user voice technology exists wake wake low error rate and wake poor user experience technical problems.

To solve the above technical problem, the present invention provides a method for wake-up speech, comprising the steps of: acquisition wake-word speech signals sent by the end-user, access to the wake-up wake-up word recognition results voiceprint characteristic word speech signals and storing;

In the terminal is in a dormant state, detection and external input speech signal;

Being the speech signal a speech recognition result voiceprint feature;

Get the wake-up word speech recognition result of the recognition result and storage, as well as matches the voiceprint and voiceprint characteristic features stored;

According to the matching results to determine whether the need to wake up the terminal, and if so, to wake the terminal.

Before

Further, after detecting the speech signal the method of input, the speech recognition result to obtain a voiceprint characteristic of the speech signal, further comprising: determining whether the voice signal meets the requirements of a preset

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：

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。

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：

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；

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。

进一步地，所述方法还包括：

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。

进一步地，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。

signal, and if so, Get the speech signal a speech recognition result voiceprint features.

Furthermore, the judgment of the speech signal meets the requirements of preset signal comprises: 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.

Step

Further, the voice data and the user data of the background noise, respectively preset to match the background noise model include: 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 of the pre-match match points to obtain a second value;

According to the matching results to determine whether the speech signal meets a preset signal required steps 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 score is determined whether the speech signal is greater than a preset threshold, if it is determined to meet the requirements of a preset signal, and if not, it is determined that the signal does not meet the preset requirements.

Furthermore, according to the results to determine whether the steps need to wake up the terminal includes the matching:

When the speech recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the voiceprint of the speech signal stored speech wakeup word voiceprint match is found, determination need to wake up the terminal.

Furthermore, the judgment of the speech signal meets the requirements of preset signal comprises:

After the first multi-core audio processor core unit wakes up, the first nuclear unit determines

Petitioner Exhibit 1004, Page 944 of 1518

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：

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；

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。

进一步地，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。

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；

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。

进一步地，所述装置还包括：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

whether the speech signal is preset to meet the signaling requirements;

Get the voiceprint features speech recognition result of the speech signal comprises:

When the first nuclear unit is judged to be the first nuclear unit of the multi-core audio processor in the wake of a second nuclear unit;

The second nuclear unit acquired by the speech recognition result of the speech signal voiceprint feature;

Get the wake of the speech recognition result of the word recognition result is stored, and the matching results of the step of the voiceprint stored voiceprint features include:

The second nuclear unit acquires wake word recognition result of the speech recognition result and storage, as well as matches the voiceprint characteristic with the stored voiceprint said;

According to the result of the match is determined whether the terminal needs to wake up, if the step of the terminal include wake-up:

The second nuclear unit is required to wake the terminal according to the matching result of the determination, if so, the second nuclear unit to wake the terminal.

Further, the method further comprises:

When the first core unit judgment is NO, the first nuclear unit goes to sleep; and / or

When the second nuclear unit judgment is NO, the second nuclear unit into hibernation.

Further, the second nuclear unit to wake the terminal comprises:

The second nuclear unit sends an interrupt event to wake the dormant terminal in the main processor.

The same technique to solve the above problems, the present invention also provides a method for processing a terminal, comprising the steps of: using voice above the top five in any one of the wake-up wake up the terminal;

After the wake-up terminal, perform the corresponding operation according to preset voice

征。

进一步地，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。

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；

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。

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。

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；

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。

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；

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。

本发明的有益效果是：

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

recognition result of the speech signal.

Further, the wake of the terminal comprises: generate preset type of wake-up interrupt events to wake the terminal according to the recognition result of the speech signal;

The default action steps speech recognition result of the speech signal include: type of interrupt event to perform the corresponding operation according to the preset wake.

Similarly to solve the above technical problem, the present invention also provides a wake-speech apparatus 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.

Furthermore, the apparatus further comprises: a second judging module;

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;

Petitioner Exhibit 1004, Page 946 of 1518

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。

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。

附图说明

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；

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；

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；

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；

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；

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；

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；

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；

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；

the first module for obtaining said second determination means determines YES, obtaining a speech recognition result of the speech signal and a voiceprint characteristic.

Further, 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.

Furthermore, the first acquisition module, a second acquisition module, the judging module and the first module is integrated in the wake of the second core multicore audio processor unit; a second judgment module integrated the first multi-core audio processor core unit;

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.

Further, the first determining module is used when the speech recognition result of the speech signal with the stored wakeup word speech recognition result of the match is successful, and the voiceprint of the speech signal with the stored when the wake-up word voice voiceprint match is successful, it is determined that the terminal needs to wake up.

The same technique to solve the above problems, the present invention also provides a terminal, comprising a processing module and said 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.

Further, the voice wake-up means for generating a wake-up module preset type of wake-up interrupt wake-up event of the terminal based on the identification result of the speech signal;

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

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。

实施例一：

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：

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。

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。

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。

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。

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:

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；

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2优化声纹特征1，当唤醒词识别

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

Advantageous Effects

The present invention is:

The present invention provides a method and apparatus for voice, terminal and approach wake, wake-up can improve the user rate and reduce the error rate and enhance the user experience wake. In particular, the present invention is the voice wakeup method comprising: signal acquisition wake-word speech issued by the end user to get the wake-up wake-up word recognition results voiceprint characteristic word speech signals and stored; the terminal is in a sleep state, detection external input speech signal; obtaining a speech recognition result voiceprint characteristics of the speech signal; obtaining the wake-up word speech recognition result of the recognition result and storage, as well as matches the voiceprint features and characteristics of the voiceprint stored; according to the matching results to determine whether the need to wake up the terminal, and if so, to wake the terminal; the voice of the present invention, methods for waking up to their voice as a sound sample end-user, end-users get the sound characteristics of individuals (ie, the recognition result harmony pattern feature) as an audio sample characteristics; wake during voice because sound sample features voiceprint and voiceprint wakeup extraction (same sound the same user mode), to avoid the user's individual sound pattern difference in sound Sample user's sound pattern caused wake-up failed to provide a wake-up rate; in addition, the method of the present invention 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, reduces the error rate to wake; the method of the present invention can also provide custom wake-up word, to meet the needs of individual users, to enhance the user experience.

In addition, because 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.

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；

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；

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。

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。

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。

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。

步骤103：获取所述语音信号的语音识别结果和声纹特征。

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。

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。

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。

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。

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

Brief Description of the

Figure 1 embodiment of the present invention to a speech provided a wake-up schematic flow diagram of a method embodiment;

2 the present invention method for obtaining user-defined voice wake-word sample characteristics provide a flow diagram illustrating the embodiment;

3 embodiment of the present invention method for judging whether the speech signal example of a preset signal provides a schematic flow diagram meet the requirements;

Figure 4 is a schematic of one configuration of the invention to provide a voice Example wake implementation of the system;

Figure 5 embodiment of the present invention to a speech provided a wake-up system for voice wake schematic flow diagram embodiment;

6 a schematic diagram of the process according to a second embodiment of a terminal to treatment;

Figure 7 is a schematic view of a first configuration of the invention to provide a wake-up voice Example Three apparatus embodiment;

8 a schematic view of a second structure of the present invention to provide a wake-up voice Example Three apparatus embodiment;

FIG. 9 is a schematic view of the structure of the invention to provide a third voice Example Three wakeup device embodiment;

10 the present invention structure diagram of an embodiment of a terminal provided in the third embodiment.

DETAILED DESCRIPTION

The following figures present invention will be described in further detail by specific embodiments in combination.

Example one:

Taking into account the existing fixed-word speech wake wake wake-user rate method exists, waking up rate and poor user experience technical problems; the present embodiment provides a wake-up can improve end-user rate, reduce the