

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 4 月 16 日 (16.04.2020)



(10) 国际公布号
WO 2020/073839 A1

- (51) 国际专利分类号:
G10L 15/02 (2006.01)
- (21) 国际申请号: PCT/CN2019/108828
- (22) 国际申请日: 2019 年 9 月 29 日 (29.09.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811186019.4 2018 年 10 月 11 日 (11.10.2018) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 曹元斌(CAO, Yuanbin); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团

- 法务部, Zhejiang 311121 (CN)。 张智超(ZHANG, Zhichao); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 风翮(FENG, He); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 王刚(WANG, Gang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,

(54) **Title:** VOICE WAKE-UP METHOD, APPARATUS AND SYSTEM, AND ELECTRONIC DEVICE

(54) 发明名称: 语音唤醒方法、装置、系统及电子设备

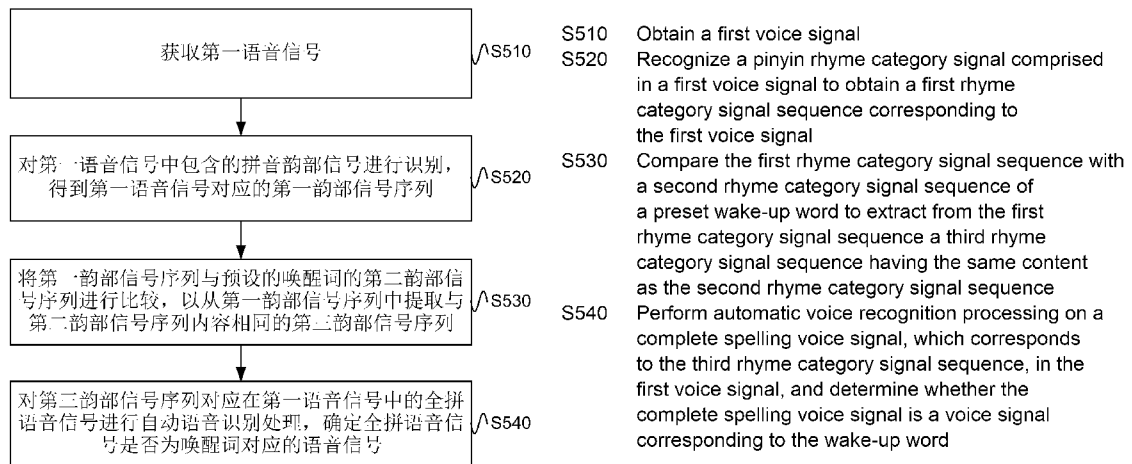


图 5

(57) **Abstract:** A voice wake-up method, apparatus and system, and an electronic device. The method comprises: obtaining a first voice signal (S510); recognizing a pinyin rhyme category signal comprised in a first voice signal to obtain a first rhyme category signal sequence corresponding to the first voice signal (S520); comparing the first rhyme category signal sequence with a second rhyme category signal sequence of a preset wake-up word to extract from the first rhyme category signal sequence a third rhyme category signal sequence having the same content as the second rhyme category signal sequence (S530); performing automatic voice recognition processing on a complete spelling voice signal, which corresponds to the third rhyme category signal sequence, in the first voice signal, and determining whether the complete spelling voice signal is a voice signal corresponding to the wake-up word (S540). The method can quickly and accurately recognize the wake-up word and improve a waken speed of a device.

CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种语音唤醒方法、装置、系统及电子设备, 该方法包括: 获取第一语音信号 (S510); 对第一语音信号中包含的拼音韵部信号进行识别, 得到第一语音信号对应的第一韵部信号序列 (S520); 将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较, 以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列 (S530); 对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理, 确定全拼语音信号是否为唤醒词对应的语音信号 (S540)。该方法能够快速、准确地识别唤醒词, 提高设备的被唤醒速度。

语音唤醒方法、装置、系统及电子设备

本申请要求 2018 年 10 月 11 日递交的申请号为 201811186019.4、发明名称为“语音唤醒方法、装置、系统及电子设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本说明书涉及计算机技术领域，尤其涉及一种语音唤醒方法、装置、系统及电子设备。

10 背景技术

随着人工智能相关应用的越来越深入发展，语音识别技术作为智能化设备的基本的交互方式，扮演着越来越重要的角色。语音识别技术涉及到很多方面，其中包括通过语音指令来唤醒设备、对设备的操作进行控制、与设备进行人机对话以及针对多个设备的语音指令控制等。高效和准确的语音识别技术以及快捷便利的唤醒模式，是智能化设备

15

的重要的发展方向。

目前，自定义唤醒的主要性能瓶颈在于端上（终端设备）计算资源有限，核心部分的分类器对语音特征所分的类别数直接影响到唤醒的速度和准确率。传统的拼音粒度的分类策略是取常用汉字的全拼做分类，带声调的 1200 多个，去掉声调的 400 多个，可以达到 80%左右的准确率。但是，要想达到更高的准确率，需要提高端上计算性能和完善

20

发明内容

本发明提供了一种语音唤醒方法、装置、系统及电子设备，能够快速、准确的识别唤醒词，提高设备的被唤醒速度。

25

为达到上述目的，本发明的实施例采用如下技术方案：

第一方面，提供了一种语音唤醒方法，包括：

获取第一语音信号；

对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；

30

将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述

第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；

对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

第二方面，提供了另一种语音唤醒方法，包括：

5 获取第一语音信号；

对所述第一语音信号中包含的元音信号进行识别，得到所述第一语音信号对应的第一元音信号序列；

将所述第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从所述第一元音信号序列中提取与所述第二元音信号序列内容相同的第三元音信号序列；

10 对所述第三元音信号序列对应所述第一语音信号中的全量语音信号进行自动语音识别处理，确定所述全量语音信号是否为所述唤醒词对应的语音信号。

第三方面，提供了一种语音唤醒装置，包括：

信号获取模块，用于获取第一语音信号；

15 信号识别模块，用于对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；

信号比较模块，用于将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；

20 语音识别模块，用于对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

第四方面，提供了另一种语音唤醒装置，包括：

信号获取模块，用于获取第一语音信号；

25 信号识别模块，用于对所述第一语音信号中包含的元音信号进行识别，得到所述第一语音信号对应的第一元音信号序列；

信号比较模块，用于将所述第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从所述第一元音信号序列中提取与所述第二元音信号序列内容相同的第三元音信号序列；

30 语音识别模块，用于对所述第三元音信号序列对应所述第一语音信号中的全量语音信号进行自动语音识别处理，确定所述全量语音信号是否为所述唤醒词对应的语音信

号。

第五方面，提供了一种语音唤醒系统，包括：

终端，用于获取第一语音信号；对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；将所述第三韵部信号序列对应的全拼语音信号发送至服务器；

所述服务器，用于对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

第六方面，提供了一种语音唤醒方法，包括：

终端获取第一语音信号；对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；将所述第三韵部信号序列对应的全拼语音信号发送至服务器；

所述服务器对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

第七方面，提供了一种电子设备，包括：

存储器，用于存储程序；

处理器，耦合至所述存储器，用于执行所述程序，以用于：

获取第一语音信号；

对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；

将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；

对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

第八方面，提供了另一种电子设备，包括：

存储器，用于存储程序；

处理器，耦合至所述存储器，用于执行所述程序，以用于：

获取第一语音信号；

对所述第一语音信号中包含的元音信号进行识别，得到所述第一语音信号对应的第一元音信号序列；

将所述第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从所述

5 第一元音信号序列中提取与所述第二元音信号序列内容相同的第三元音信号序列；

对所述第三元音信号序列对应所述第一语音信号中的全量语音信号进行自动语音识别处理，确定所述全量语音信号是否为所述唤醒词对应的语音信号。

本发明提供了一种语音唤醒方法、装置、系统及电子设备，在获取到待识别的第一语音信号后，先对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；然后，将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；最后，对第三韵部信号序列对应第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号，进而识别出第一语音信号中是否包含唤醒词。本方案采用先对待识别的语音信号中的韵部信号与唤醒词的韵部

10 进行比对，提取出待识别语音信号中韵部信号与唤醒词韵部相同的语音信号部分，然后针对该部分语音信号再整体通过自动语音识别处理以确定其中是否包含唤醒词，从而实现快速、准确的识别唤醒词，提高设备的被唤醒速度。

上述说明仅是本申请技术方案的概述，为了能够更清楚了解本申请的技术手段，而可依照说明书的内容予以实施，并且为了让本申请的上述和其它目的、特征和优点能够

20 更明显易懂，以下特举本申请的具体实施方式。

附图说明

通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本申请的

25 的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：

图 1 为语音唤醒的基本流程逻辑示意图；

图 2 为语音唤醒的基本流程中端上唤醒引擎的处理逻辑示意图；

图 3 为本发明实施例的唤醒引擎的处理逻辑示意图；

图 4 为本发明实施例的语音唤醒系统结构图；

30 图 5 为本发明实施例的语音唤醒方法流程图一；

图 6 为本发明实施例的语音唤醒方法流程图二；

图 7 为本发明实施例的韵部分类训练方法流程图一；

图 8 为本发明实施例的韵部分类训练方法流程图二；

图 9 为本发明实施例的语音唤醒装置结构图一；

5 图 10 为本发明实施例的语音唤醒装置结构图二；

图 11 为本发明实施例的韵部分类训练装置结构图一；

图 12 为本发明实施例的韵部分类训练装置结构图二；

图 13 为本发明实施例的语音唤醒方法流程图三；

图 14 为本发明实施例的电子设备的结构示意图一；

10 图 15 为本发明实施例的电子设备的结构示意图二。

具体实施方式

下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例
15 所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

如图 1 所示，为语音唤醒的基本流程，语音设备接受到语音信号后，先对语音信号进行信号处理（主要包括降噪、回声消除）和特征抽取，从而将原始输入音频信号转换成端上（终端）唤醒引擎可以识别的特征（即语音的频谱信号）；然后将特征输入至唤
20 醒引擎进行唤醒词的比对识别；当唤醒词命中，会继续指示服务端执行后续指令，如播放歌曲、相声等。

在图 1 所示的语音唤醒的基本流程中，端上唤醒引擎可以认为是执行唤醒的核心部分。如图 2 所示，该端上唤醒引擎主要包括两部分：分类器和后处理部分。

一、分类器，用于将连续语音特征转换成不同的类别，这一部分计算，往往是所有
25 唤醒工作最消耗计算资源的部分，通常神经网络的最后一层输出的分类数，直接决定了整个网络的计算规模。传统的基于隐马尔科夫模型-深度神经网络（Hidden Markov Model-Deep Neural Network, HMM-DNN），采用音速（phone）的概率密度函数（Probability Density Function, PDF）进行建模，而要想达到生产可用状态，需要至少 6000~8000 个分类；采用拼音做分类，也需要 1200~400 多个分类。

30 二、后处理，唤醒词的检验都存在后处理部分，传统方法检测整个词，可以在对经

分类器输出的语音进行平滑处理（smooth）后采用动态时间规整算法（Dynamic Time Warping, DTW）识别该语音与唤醒词是否相同；也可以采用自动语音识别（Automatic Speech Recognition, ASR）技术对语音是否命中唤醒词进行识别。

采用上述的端上唤醒引擎中的分类器，由于最终使用的分类数较多致使分类网络庞大，需要端上配置较高的计算性能。

本发明实施例改善了现有技术中分类网络庞大导致的需要在端上配置较高计算资源才能准确、快速执行语音唤醒的缺陷，其核心思想是，将执行语音唤醒的核心部分拆分成两次唤醒词的识别过程。第一次唤醒词识别过程在终端上完成，该过程只对唤醒词的拼音韵部进行分类识别，完成对待识别的语音信号的初步识别过程。然后，将初步筛选出的与唤醒词韵部信号相同的韵部信号所对应的全量语音信号发送云端，由云端再次对该语音信号整体进行识别，确定该语音信号是否命中唤醒词。

如图 3 所示，为本发明实施例的唤醒引擎的处理逻辑示意图，涉及设备端（如智能音箱等可接收和识别语音的终端）以及云端（设置有服务器）两个执行语音唤醒的主体。

在设备端上，待识别的语音信号首先经过第一次唤醒词识别，该识别过程只对语音信号的韵部信号通过预先训练生成的分类器进行韵部分类识别；然后，将识别出的韵部信号序列通过后处理与唤醒词的韵部进行比对，以判断语音信号中是否命中唤醒词的韵部，并将命中唤醒词韵部的全量语音信号传送到云端。

在云端，待识别的语音信号为韵部信号与唤醒词韵部相同的全量语音信号，对这些语音信号进行第二次唤醒词识别（二次检验），该识别过程是对语音信号整体进行识别，例如采用 ASR 技术，对该语音信号是否命中唤醒词进行识别。

基于上述语音唤醒的方案思想，图 4 为本发明实施例提供的语音唤醒系统结构图。

如图 4 所示，该系统包括终端 410 和服务器 420，其中：

终端 410 包括：

信号获取模块，用于获取第一语音信号，该第一语音信号例如为中文语音信号；

信号识别模块，用于对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；

信号比较模块，用于将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；

服务器 420 包括：

语音识别模块，用于对第三韵部信号序列对应第一语音信号中的全拼语音信号进

行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号。

下面通过多个实施例来进一步说明本申请的技术方案。

实施例一

基于上述语音唤醒的方案思想，如图 5 所示，其为本发明实施例示出的语音唤醒方法流程图一，该方法的执行主体可以为部署在图 4 中终端 410 和服务器 420 中的模块。其中，步骤 S510~S530 可在端上（终端）执行，步骤 S540 可在云端（服务器）中执行。如图 5 所示，该语音唤醒方法包括如下步骤：

S510，获取第一语音信号，该第一语音信号例如为中文语音信号。

其中，第一语音信号可以为通过语音设备接收的语音信号，通过对该语音信号进行唤醒词的识别，以进一步唤醒目标设备。

S520，对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列。

本步骤中，将拼音的声部和韵部分开：如 tian->t, ian; mao-> m, ao。汉语的日常对话中，拼音声部（简称“声部”）如 t、m 往往都是爆破音，从语音信号的特征频谱图上看，声部就是一个短暂的尖峰或低谷，基本所有的延长音，都在拼音韵部（简称“韵部”）。传统的三音素（triphone）建模中，往往需要结合前、后 phone，才能达到不错的识别准确率，本方案在端上计算优先的情况下，去掉对声部的识别，只是对第一语音信号中包含的韵部信号进行识别，从而得到第一语音信号对应的第一韵部信号序列。该第一韵部信号序列中包含时间序列，以及位于时间序列上每个时刻点的韵部信号。

S530，将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列。

传统的唤醒词识别方法是检测整个词，本方案为了减少端上的计算量，只在端上识别每个字的韵部，即将上述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列。

例如，假设预设的唤醒词为“你好”，对应的第二韵部信号序列为“i , ä o”，那么在第一韵部信号序列中的信号序列为“i , ä o”的，都可作为第三韵部信号序列。

S540，对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号。

在实际应用场景中，由于端上只验证了韵部，如：“好”，“老”，“考”，由于拥有同样韵部，都可顺利通过分类，并可能作为韵部识别唤醒词的初步结果。因此，在

云端上需要执行二次校验，对第三韵部信号序列对应在第一语音信号中的全拼语音信号（包含声部信号）进行自动语音识别 ASR 处理，以确定全拼语音信号是否为唤醒词对应的语音信号。

本环节的二次校验就是过滤掉声部与唤醒词声部不一样的语音信号部分，这么做的好处是端上过滤到了绝大部分的非唤醒词，而云端只需要做最后校验就可以识别到真正的唤醒词，如此平衡了端上和服务端的计算，既可以有很高的准确率，同时不会有由于端上模型过大带来的高延时。

本发明提供的语音唤醒方法，在获取到待识别的第一语音信号后，先对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；然后，将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；最后，对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号，进而识别出第一语音信号中是否包含唤醒词。本方案采用先对待识别的语音信号中的韵部信号与唤醒词的韵部进行比对，提取出待识别语音信号中韵部信号与唤醒词韵部相同的语音信号部分，然后针对该部分语音信号再整体通过自动语音识别处理以确定其中是否包含唤醒词，从而实现快速、准确的识别唤醒词，提高设备的被唤醒速度。

实施例二

如图 6 所示，为本发明实施例的语音唤醒方法流程图二。在上一实施例中所示方法的基础上，增加了预处理环节，并对步骤 S520 和 S530 进行了细化。如图 6 所示，该语音唤醒方法包括如下步骤：

S610，获取第一语音信号，该第一语音信号例如为中文语音信号。

步骤 S610 与步骤 S510 内容相同。

S620，对第一语音信号进行去噪的预处理。

在获取到第一语音信号之后，可对第一语音信号进行降噪、回声消除等预处理，以最大程度保留第一语音信号中的有效信号比例。

S630，获取预处理后的第一语音信号的特征频谱。

其中，所谓特征频谱指在进行分类识别或分类训练时，待处理的语音信号需要被转换为满足一定特征要求的频谱信号。

比如，在对第一语音信号进行分类识别时，将待识别的第一语音信号转换为频谱信

号后，将音频按固定时间长度切成如 20ms 左右的帧频谱信号，以作为后续分类识别时的特征频谱。

S640，对第一语音信号的特征频谱采用韵部分类器进行分类计算，得到第一语音信号对应的第一韵部信号序列。

- 5 其中，韵部分类器可为预先训练生成的语音分类模型，但该语音分类模型只对语音信号中的韵部信号进行分类，并输出相应韵部信号的序列值。

步骤 S630~S640 为上述步骤 S520 的细化。

进一步地，可采用如图 7 所示的韵部分类训练方法，训练生成上述韵部分类器，方法包括：

- 10 S710，获取用于模型训练的语音信号的特征频谱。

S720，对特征频谱中的拼音韵部信号进行标注。

通常，相同韵部由于在发音过程中受到声部信号的影响，它们在特征频谱中表现的形态也会不完全相同。通过有监督的学习，可以快速、准确地锁定不同韵部所对应的韵部信号的特征形态。

- 15 S730，以已标注的拼音韵部信号作为训练样本，采用神经网络算法以及连接时序分类的联合模型算法训练生成韵部分类器。

训练过程主要包含两个处理环节，一个是如何准确对不同韵部的特征频谱信号进行韵部分类；另一个则是如何将分好类别的韵部放置到语音信号中的正确位置。

- 20 在解决这两个问题时，可以采用神经网络算法实现对不同韵部的特征频谱信号进行准确的韵部分类，并结合连接时序分类（Connectionist Temporal Classification, CTC）算法，以锁定分好类别的韵部在语音信号中的正确位置。采用这两种模型算法进行联合建模，以基于训练样本训练生成韵部分类器。

进一步地，还可采用如图 8 所示的韵部分类训练方法，训练生成上述韵部分类器，方法包括：

- 25 S810，获取用于模型训练的语音信号的特征频谱。

S820，对特征频谱中的拼音韵部信号进行标注。

通常，相同韵部由于在发音过程中受到声部信号的影响，它们在特征频谱中表现的形态也会不完全相同。通过有监督的学习，可以快速、准确地锁定不同韵部所对应的韵部信号的特征形态。

- 30 S830，以已标注的拼音韵部信号作为训练样本，采用隐马尔科夫模型以及深度神经

网络的联合模型算法训练生成韵部分类器。

训练过程主要包含两个处理环节，一个是如何准确对不同韵部的特征频谱信号进行韵部分类；另一个则是如何将分好类别的韵部放置到语音信号中的正确位置。

在解决这两个问题时，也可以采用隐马尔科夫模型（HMM-DNN）两种模型算法进行联合建模，以基于训练样本训练生成韵部分类器。

与现有技术不同的是，本方案中的分类器，为针对拼音韵部进行分类的韵部分类器。

S650，采用动态时间规整算法将第一韵部信号序列与预设的唤醒词的第二韵部信号序列按时序对应进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列。

10 在将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较时，可采用动态时间规整（Dynamic Time Warping, DTW）算法将比对的两个信号序列进行位置对齐，然后按时序对应进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列。

15 S660，对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号。

步骤 S660 与步骤 S540 内容相同。

本发明提供的语音唤醒方法，在实施例一的基础上进行了方法拓展：

首先，在获取到第一语音信号之后，对第一语音信号进行预处理，以最大程度保留第一语音信号中的有效信号比例。

20 其次，通过预先训练的韵部分类器，对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列，以实现快速识别。在训练韵部分类器时，采用神经网络算法以及连接时序分类的联合模型算法进行训练建模，或者采用隐马尔科夫模型以及深度神经网络的联合模型算法进行训练建模，以保证训练出的韵部分类器的准确度。

25 最后，采用动态时间规整算法对第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以快速准确的得到第三韵部信号序列。

实施例三

如图 9 所示，为本发明实施例的语音唤醒装置结构图一，该语音唤醒装置可设置在图 4 所示的语音唤醒装置系统中，用于执行如图 5 所示的方法步骤，其包括：

30 信号获取模块 910，用于获取第一语音信号；

信号识别模块 920，用于对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；

信号比较模块 930，用于将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；

语音识别模块 940，用于对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号。

进一步地，如图 10 所示，上述语音唤醒装置中，信号识别模块 920 可包括：

特征获取单元 101，用于获取第一语音信号的特征频谱；

10 信号识别单元 102，用于对第一语音信号的特征频谱采用韵部分类器进行分类计算，得到第一语音信号对应的第一韵部信号序列。

进一步地，在图 10 所示的语音唤醒装置中，还可包括：

预处理模块 103，用于对第一语音信号进行去噪的预处理。

进一步地，上述信号比较模块 930 具体可用于，

15 采用动态时间规整算法将第一韵部信号序列与预设的唤醒词的第二韵部信号序列按时序对应进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列。

图 10 所示语音唤醒装置可用于执行图 6 所示的方法步骤。

进一步地，如图 11 所示，上述语音唤醒装置中，还可包括：

20 第一频谱获取模块 111，用于获取用于模型训练的语音信号的特征频谱；

第一信号标注模块 112，用于对特征频谱中的拼音韵部信号进行标注；

第一训练模块 113，用于以已标注的拼音韵部信号作为训练样本，采用神经网络算法以及连接时序分类的联合模型算法训练生成韵部分类器。

进一步地，如图 12 所示，上述语音唤醒装置中，还可包括：

25 第二频谱获取模块 121，用于获取用于模型训练的语音信号的特征频谱；

第二信号标注模块 122，用于对所述特征频谱中的拼音韵部信号进行标注；

第二训练模块 123，用于以所述已标注的拼音韵部信号作为训练样本，采用隐马尔科夫模型以及深度神经网络的联合模型算法训练生成所述韵部分类器。

图 11、图 12 所示装置可用于对应执行图 7、图 8 所示的方法步骤。

30 本发明提供的语音唤醒装置，在获取到待识别的第一语音信号后，先对第一语音信

号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；然后，将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；最后，对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号

5 是否为唤醒词对应的语音信号，进而识别出第一语音信号中是否包含唤醒词。本方案采用先对待识别的语音信号中的韵部信号与唤醒词的韵部进行比对，提取出待识别语音信号中韵部信号与唤醒词韵部相同的语音信号部分，然后针对该部分语音信号再整体通过自动语音识别处理以确定其中是否包含唤醒词，从而实现快速、准确的识别唤醒词，提高设备的被唤醒速度。

10 进一步地，在获取到第一语音信号之后，对第一语音信号进行预处理，以最大程度保留第一语音信号中的有效信号比例。

进一步地，通过预先训练的韵部分类器，对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列，以实现快速识别。在训练韵部分类器时，采用神经网络算法以及连接时序分类的联合模型算法进行训练建模，或者采用

15 隐马尔科夫模型以及深度神经网络的联合模型算法进行训练建模，以保证训练出的韵部分类器的准确度。

进一步地，采用动态时间规整算法对第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以快速准确的得到第三韵部信号序列。

实施例四

20 基于上述语音唤醒的方案思想，如图 13 所示，其为本发明实施例示出的语音唤醒方法流程图三，该方法的执行主体可以为部署在图 4 中终端 410 和服务器 420 中的模块。其中，步骤 S131~133 可在端上（终端）执行，步骤 S134 可在云端（服务器）中执行。如图 13 所示，该语音唤醒方法包括如下步骤：

S131，获取第一语音信号。

25 本步骤中，对第一语音信号的语言类型不做限定，例如可以为中文、英文、日文等。该第一语音信号可以为通过语音设备接收的语音信号，通过对该语音信号进行唤醒词的识别，以进一步唤醒目标设备。

S132，对第一语音信号中包含的元音信号进行识别，得到第一语音信号对应的第一元音信号序列。

30 自然语音按发音学范畴划分，可为元音和辅音，比如在中文中，元音对应为拼音中

的韵部、辅音对应为拼音中的声部；又比如在英文中，包含 5 个元音：a、e、i、o、u，21 个辅音；又比如在日文中，包含 5 个元音，以“あ・い・う・え・お”这五个假名来表示，音韵学上，其发音接近[a][i][u][e][o]，辅音方面，有清音-“か・さ・た・な・は・ま・や・ら・わ行”的辅音、浊音-“が・ざ・だ・ば行”的辅音、半浊音-“ぱ行”的辅音。对任一语言类型的第一语音信号中包含的元音信号进行识别，均可以得到第一语音信号对应的第一元音信号序列。例如，当第一语音信号为中文语音信号时，第一语音信号对应的第一元音信号序列可以为图 5 所示方法中的第一韵部信号。

S133，将第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从第一元音信号序列中提取与第二元音信号序列内容相同的第三元音信号序列。

10 例如，当处理对象为中文语音信号时，可以执行如步骤 S530 的内容，将上述第一韵部信号与唤醒词的第二韵部信号序列进行比较，从而从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列。

S134，对第三元音信号序列对应在第一语音信号中的全量语音信号进行自动语音识别处理，确定全量语音信号是否为唤醒词对应的语音信号。

15 其中，第三元音信号序列对应在第一语音信号中的全量语音信号为第三元音信号序列对应在第一语音信号中的区间范围内所有的语音信号。当语音信号为中文语音信号时，该全量语音信号即为上述第三韵部信号序列对应在第一语音信号中的全拼语音信号。

进一步地，根据第一语音信号所属语言类型的不同，上述第一语音信号中包含的元音信号可以具体为第一语音信号所属语言类型所包含的单音节中元音对应的语音信号。

20 例如，当第一语音信号为中文语音信号时，第一语音信号中包含的元音信号即为中文所包含的单字中韵部对应的语音信号。

实施例五

本发明实施例提供一种语音唤醒装置，该语音唤醒装置可包含图 9 所示的所有模块，用于执行图 13 所示的方法步骤，其包括：

25 信号获取模块 910，用于获取第一语音信号；

信号识别模块 920，用于对第一语音信号中包含的元音信号进行识别，得到第一语音信号对应的第一元音信号序列；

信号比较模块 930，用于将第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从第一元音信号序列中提取与第二元音信号序列内容相同的第三元音信号序列；

30

语音识别模块 940，用于对第三元音信号序列对应在第一语音信号中的全量语音信号进行自动语音识别处理，确定全量语音信号是否为唤醒词对应的语音信号。

进一步地，上述第一语音信号中包含的元音信号可以为第一语音信号所属语言类型所包含的单音节中元音对应的语音信号。

- 5 例如，第一语音信号所属语言类型可包括：中文、英文、日文等。当第一语音信号为中文语音信号时，本实施例中的语音唤醒装置可以执行如图 5 所示的方法步骤。

实施例六

本实施例提供了一种语音唤醒系统，包括：

- 10 终端，用于获取第一语音信号，该第一语音信号例如为中文语音信号；对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；将第三韵部信号序列对应的全拼语音信号发送至服务器；

- 15 服务器，用于对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号。

相应的，基于上述语音唤醒系统，本实施例还提供了一种语音唤醒方法，即从终端和服务端两侧的执行流程对语音唤醒方法进行描述。该方法包括：

- 20 终端获取第一语音信号，该第一语音信号例如为中文语音信号；对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵部信号序列；将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；将第三韵部信号序列对应的全拼语音信号发送至服务器；

服务器对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为唤醒词对应的语音信号。

- 25 通过将整个唤醒流程拆分为两部分：第一部分在终端侧通过识别第一语音信号中的韵部信号，对唤醒词进行初次识别；第二部分在服务器侧通过对初次识别所提炼的韵部信号对应的全拼语音信号进行自动语音识别，从而完成整个语音信号是否命中唤醒词的识别过程。该方法使得整个唤醒过程在终端和服务器的计算量达到平衡，减少了终端的计算压力，提高了整个语音唤醒过程的执行效率。

- 30 实施例七

前面实施例三描述了一种语音唤醒装置的整体架构，该装置的功能可借助一种电子设备实现完成，如图 14 所示，其为本发明实施例的电子设备的结构示意图，具体包括：存储器 141 和处理器 142。

存储器 141，用于存储程序。

5 除上述程序之外，存储器 141 还可被配置为存储其它各种数据以支持在电子设备上的操作。这些数据的示例包括用于在电子设备上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。

存储器 141 可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁
10 存储器，快闪存储器，磁盘或光盘。

处理器 142，耦合至存储器 141，用于执行存储器 141 中的程序，以用于：

获取第一语音信号；

对第一语音信号中包含的拼音韵部信号进行识别，得到第一语音信号对应的第一韵
15 部信号序列；

将第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从第一韵部信号序列中提取与第二韵部信号序列内容相同的第三韵部信号序列；

对第三韵部信号序列对应在第一语音信号中的全拼语音信号进行自动语音识别处理，确定全拼语音信号是否为所述唤醒词对应的语音信号。

20 上述的具体处理操作已经在前面实施例中进行了详细说明，在此不再赘述。

进一步，如图 14 所示，电子设备还可以包括：通信组件 143、电源组件 144、音频组件 145、显示器 146 等其它组件。图 14 中仅示意性给出部分组件，并不意味着电子设备只包括图 14 所示组件。

通信组件 143 被配置为便于电子设备和其他设备之间有线或无线方式的通信。电子设备可以接入基于通信标准的无线网络，如 WiFi，2G 或 3G，或它们的组合。在一个示例性实施例中，通信组件 143 经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，通信组件 143 还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

30 电源组件 144，为电子设备的各种组件提供电力。电源组件 144 可以包括电源管理

系统，一个或多个电源，及其他与为电子设备生成、管理和分配电力相关联的组件。

音频组件 145 被配置为输出和/或输入音频信号。例如，音频组件 145 包括一个麦克风（MIC），当电子设备处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器 141 或
5 经由通信组件 143 发送。在一些实施例中，音频组件 145 还包括一个扬声器，用于输出音频信号。

显示器 146 包括屏幕，其屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。触摸传感器可以
10 不仅感测触摸或滑动动作的边界，而且还检测与触摸或滑动操作相关的持续时间和压力。

实施例八

前面实施例五描述了一种语音唤醒装置的整体架构，该装置的功能可借助一种电子设备实现完成，如图 15 所示，其为本发明实施例的电子设备的结构示意图，具体包括：存储器 151 和处理器 152。

15 存储器 151，用于存储程序。

除上述程序之外，存储器 151 还可被配置为存储其它各种数据以支持在电子设备上的操作。这些数据的示例包括用于在电子设备上操作的任何应用程序或方法的指令，联系人数据，电话簿数据，消息，图片，视频等。

存储器 151 可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，如
20 静态随机存取存储器（SRAM），电可擦除可编程只读存储器（EEPROM），可擦除可编程只读存储器（EPROM），可编程只读存储器（PROM），只读存储器（ROM），磁存储器，快闪存储器，磁盘或光盘。

处理器 152，耦合至存储器 151，用于执行存储器 151 中的程序，以用于：

获取第一语音信号；

25 对第一语音信号中包含的元音信号进行识别，得到第一语音信号对应的第一元音信号序列；

将第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从第一元音信号序列中提取与第二元音信号序列内容相同的第三元音信号序列；

对第三元音信号序列对应在第一语音信号中的全量语音信号进行自动语音识别处
30 理，确定全量语音信号是否为唤醒词对应的语音信号。

上述的具体处理操作已经在前面实施例中进行了详细说明，在此不再赘述。

进一步，如图 15 所示，电子设备还可以包括：通信组件 153、电源组件 154、音频组件 155、显示器 156 等其它组件。图 15 中仅示意性给出部分组件，并不意味着电子设备只包括图 15 所示组件。

5 通信组件 153 被配置为便于电子设备和其他设备之间有线或无线方式的通信。电子设备可以接入基于通信标准的无线网络，如 WiFi，2G 或 3G，或它们的组合。在一个示例性实施例中，通信组件 153 经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中，通信组件 153 还包括近场通信（NFC）模块，以促进短程通信。例如，在 NFC 模块可基于射频识别（RFID）技术，红外数据协会（IrDA）
10 技术，超宽带（UWB）技术，蓝牙（BT）技术和其他技术来实现。

电源组件 154，为电子设备的各种组件提供电力。电源组件 154 可以包括电源管理系统，一个或多个电源，及其他与为电子设备生成、管理和分配电力相关联的组件。

音频组件 155 被配置为输出和/或输入音频信号。例如，音频组件 155 包括一个麦克风（MIC），当电子设备处于操作模式，如呼叫模式、记录模式和语音识别模式时，麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器 151 或
15 经由通信组件 153 发送。在一些实施例中，音频组件 155 还包括一个扬声器，用于输出音频信号。

显示器 156 包括屏幕，其屏幕可以包括液晶显示器（LCD）和触摸面板（TP）。如果屏幕包括触摸面板，屏幕可以被实现为触摸屏，以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。触摸传感器可以
20 不仅感测触摸或滑动动作的边界，而且还检测与触摸或滑动操作相关的持续时间和压力。

本领域普通技术人员可以理解：实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时，执行包括上述各方法实施例的步骤；而前述的存储介质包括：ROM、RAM、
25 磁碟或者光盘等各种可以存储程序代码的介质。

最后应说明的是：以上各实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述各实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技
30 术方案的范围。

权 利 要 求 书

1、一种语音唤醒方法，包括：

获取第一语音信号；

5 对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；

将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；

对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

10 2、根据权利要求 1 所述的方法，其中，所述对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列包括：

获取所述第一语音信号的特征频谱；

对所述第一语音信号的特征频谱采用韵部分类器进行分类计算，得到所述第一语音信号对应的所述第一韵部信号序列。

15 3、根据权利要求 2 所述的方法，其中，所述方法还包括：

获取用于模型训练的语音信号的特征频谱；

对所述特征频谱中的拼音韵部信号进行标注；

以已标注的拼音韵部信号作为训练样本，采用神经网络算法以及连接时序分类的联合模型算法训练生成所述韵部分类器。

20 4、根据权利要求 2 所述的方法，其中，所述方法还包括：

获取用于模型训练的语音信号的特征频谱；

对所述特征频谱中的拼音韵部信号进行标注；

以已标注的拼音韵部信号作为训练样本，采用隐马尔科夫模型以及深度神经网络的联合模型算法训练生成韵部分类器。

25 5、根据权利要求 1 所述的方法，其中，所述对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列之前还包括：

对所述第一语音信号进行去噪的预处理。

6、根据权利要求 1 所述的方法，其中，所述将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列包括：

30

采用动态时间规整算法将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列按时序对应进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列。

7、根据权利要求1所述的方法，其中，所述第一语音信号为中文语音信号。

5 8、一种语音唤醒方法，包括：

获取第一语音信号；

对所述第一语音信号中包含的元音信号进行识别，得到所述第一语音信号对应的第一元音信号序列；

10 将所述第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从所述第一元音信号序列中提取与所述第二元音信号序列内容相同的第三元音信号序列；

对所述第三元音信号序列对应所述第一语音信号中的全量语音信号进行自动语音识别处理，确定所述全量语音信号是否为所述唤醒词对应的语音信号。

9、根据权利要求8所述的方法，其中，所述第一语音信号中包含的元音信号为所述第一语音信号所属语言类型所包含的单音节中元音对应的语音信号。

15 10、一种语音唤醒装置，包括：

信号获取模块，用于获取第一语音信号；

信号识别模块，用于对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；

20 信号比较模块，用于将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；

语音识别模块，用于对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

25 11、一种语音唤醒装置，包括：

信号获取模块，用于获取第一语音信号；

信号识别模块，用于对所述第一语音信号中包含的元音信号进行识别，得到所述第一语音信号对应的第一元音信号序列；

30 信号比较模块，用于将所述第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从所述第一元音信号序列中提取与所述第二元音信号序列内容相同的第三

元音信号序列；

语音识别模块，用于对所述第三元音信号序列对应所述第一语音信号中的全量语音信号进行自动语音识别处理，确定所述全量语音信号是否为所述唤醒词对应的语音信号。

5 12、一种语音唤醒系统，包括：

终端，用于获取第一语音信号；对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；将所述第三韵部信号序列对应的全拼语音信号发送至服务器；

所述服务器，用于对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

13、一种语音唤醒方法，包括：

15 终端获取第一语音信号；对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；将所述第三韵部信号序列对应的全拼语音信号发送至服务器；

20 所述服务器对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

14、一种电子设备，包括：

存储器，用于存储程序；

处理器，耦合至所述存储器，用于执行所述程序，以用于：

25 获取第一语音信号；

对所述第一语音信号中包含的拼音韵部信号进行识别，得到所述第一语音信号对应的第一韵部信号序列；

将所述第一韵部信号序列与预设的唤醒词的第二韵部信号序列进行比较，以从所述第一韵部信号序列中提取与所述第二韵部信号序列内容相同的第三韵部信号序列；

30 对所述第三韵部信号序列对应所述第一语音信号中的全拼语音信号进行自动语

音识别处理，确定所述全拼语音信号是否为所述唤醒词对应的语音信号。

15、一种电子设备，包括：

存储器，用于存储程序；

处理器，耦合至所述存储器，用于执行所述程序，以用于：

5 获取第一语音信号；

对所述第一语音信号中包含的元音信号进行识别，得到所述第一语音信号对应的第一元音信号序列；

将所述第一元音信号序列与预设的唤醒词的第二元音信号序列进行比较，以从所述第一元音信号序列中提取与所述第二元音信号序列内容相同的第三元音信号序列；

10 对所述第三元音信号序列对应所述第一语音信号中的全量语音信号进行自动语音识别处理，确定所述全量语音信号是否为所述唤醒词对应的语音信号。

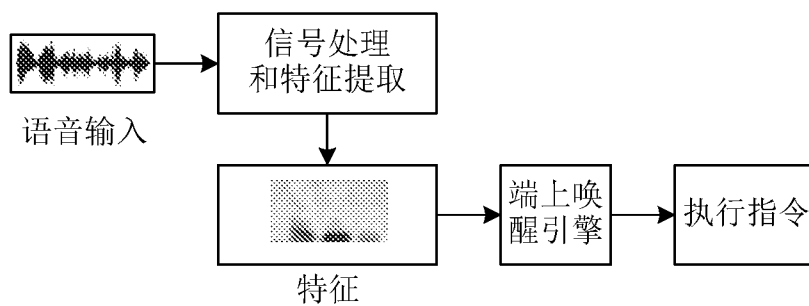


图 1

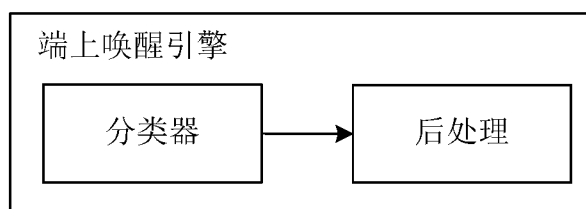


图 2

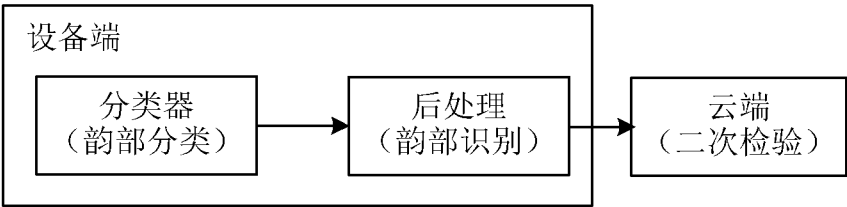


图 3

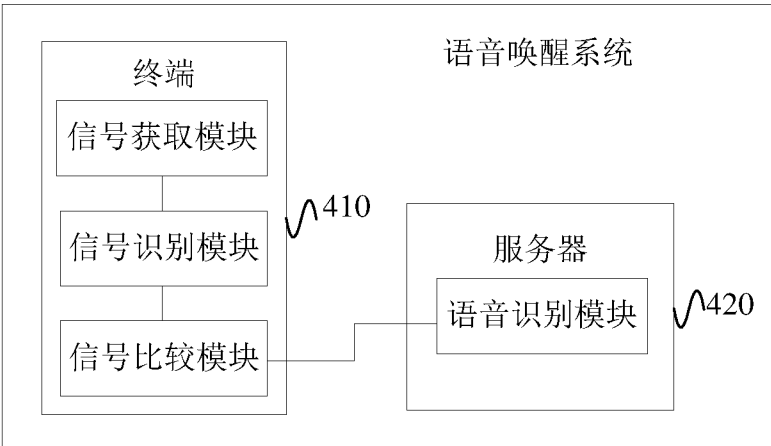


图 4

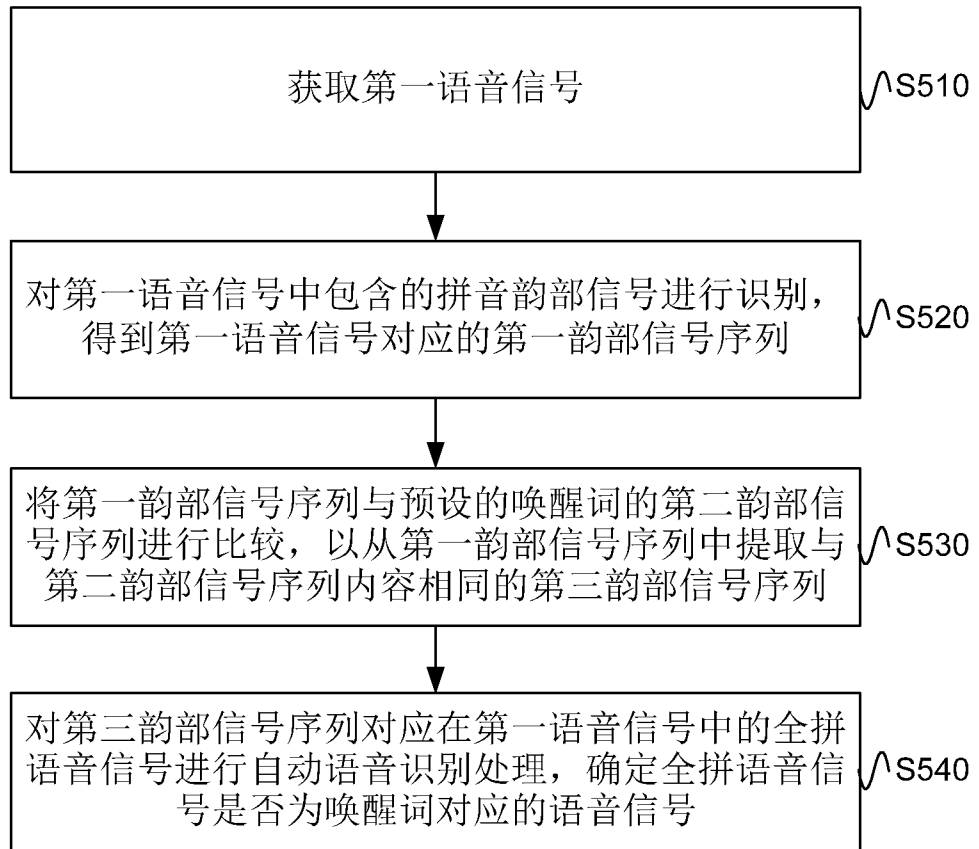


图 5

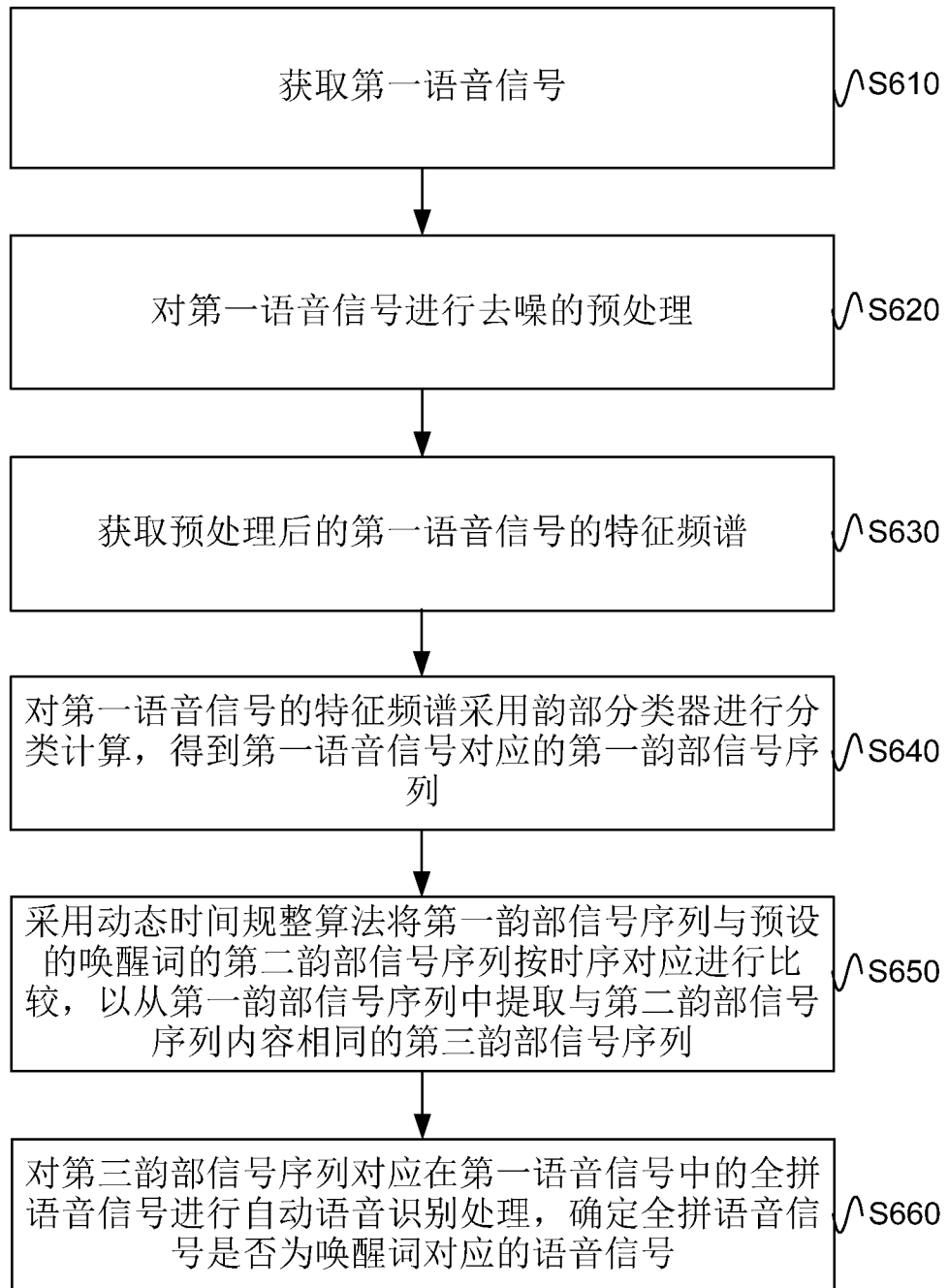


图 6

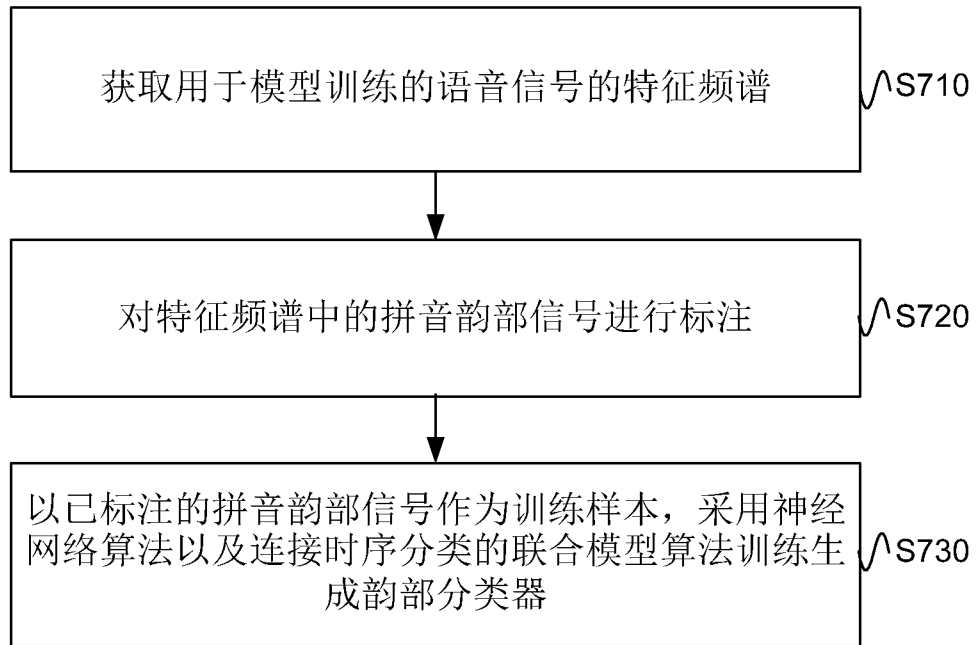


图 7

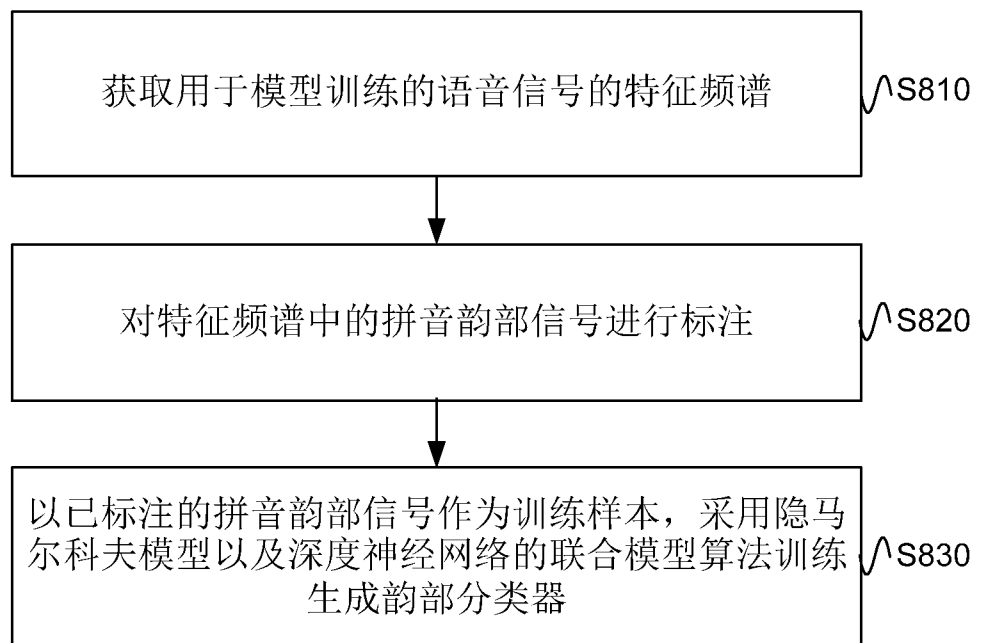


图 8

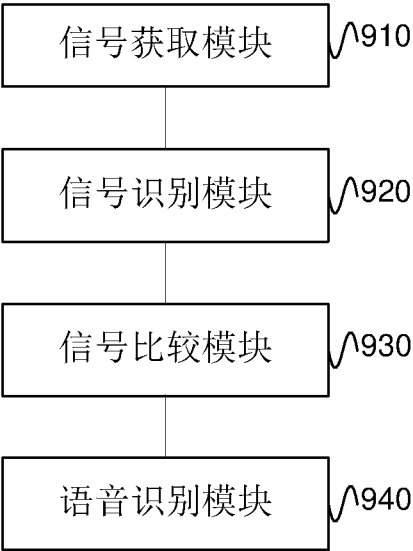


图 9

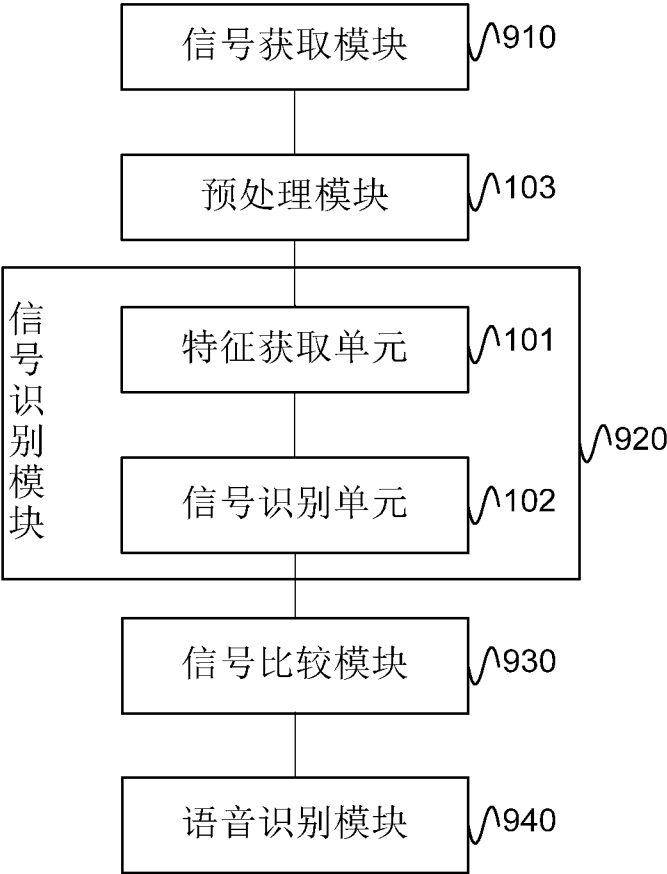


图 10

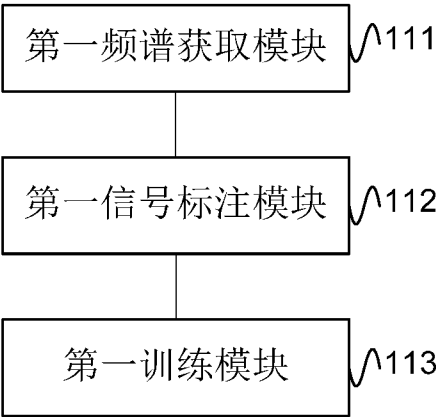


图 11

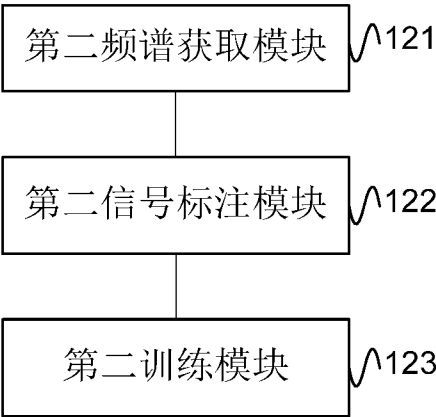


图 12

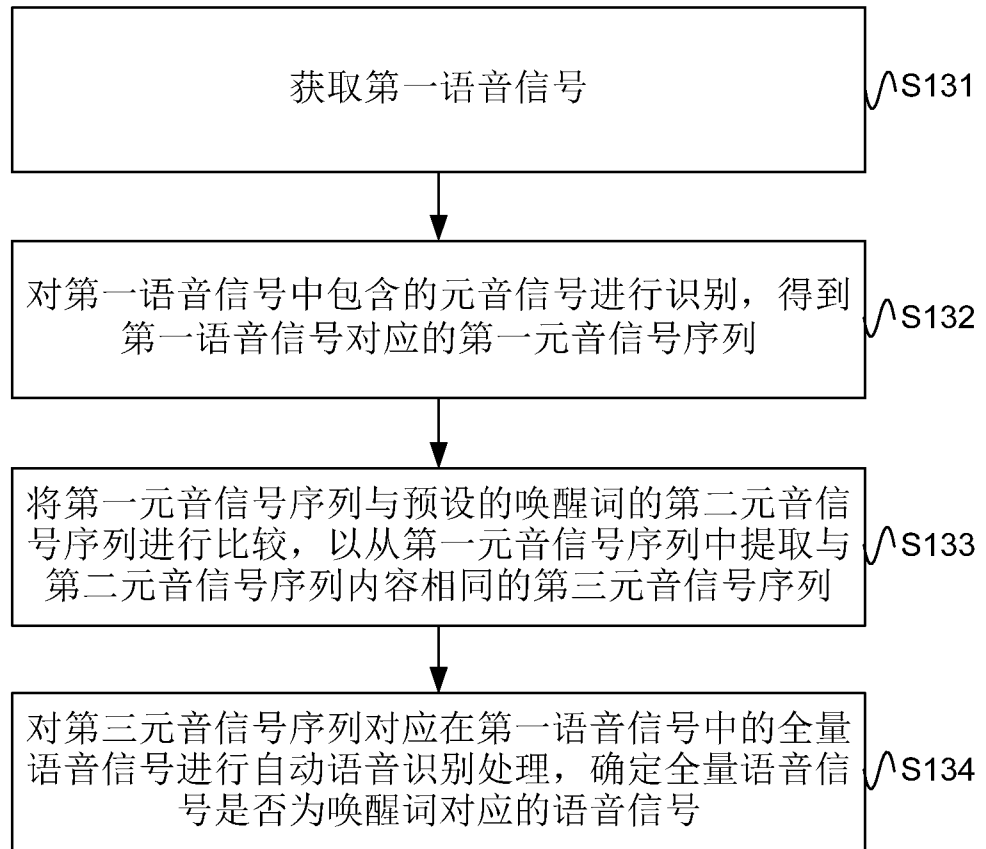


图 13

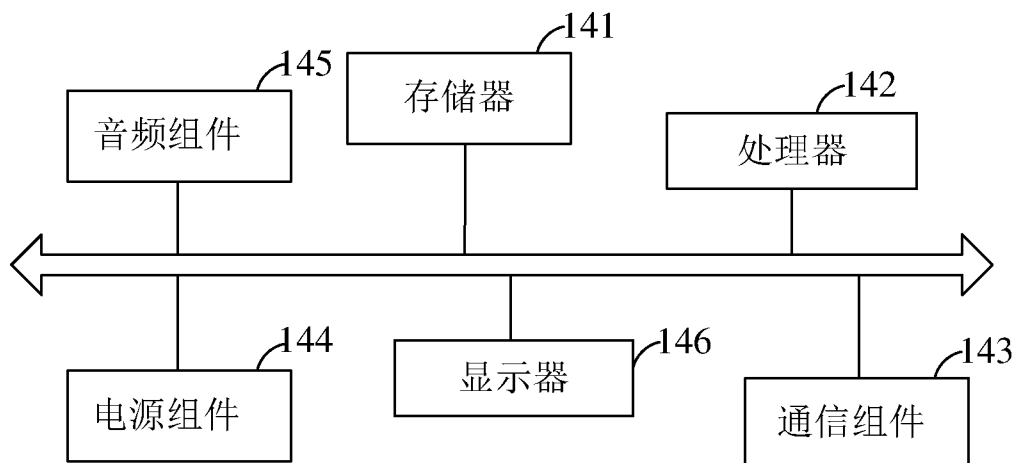


图 14

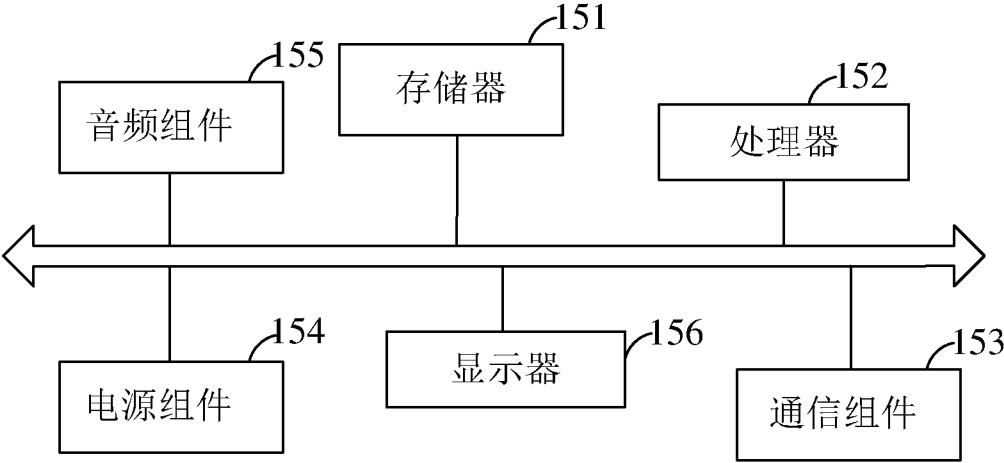


图 15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/108828

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15

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

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

DWPI; CNABS; CNTXT, IEEE: 唤醒, 激活, 苏醒, 关键词, 热词, 元音, 韵母, 韵部, 局部, 部分, 定位, vowel, part, phoneme, wake up, wake on, trig+, locat+, keyword+, spot+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2018151772 A1 (GOOGLE LLC.) 23 August 2018 (2018-08-23) description, paragraphs [0001]-[0006]	1-15
A	JP 08314490 A (JAPAN BROADCASTING CORPORATION) 29 November 1996 (1996-11-29) entire document	1-15
A	CN 107221325 A (ASUSTEK COMPUTER INC.) 29 September 2017 (2017-09-29) entire document	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

12 December 2019

Date of mailing of the international search report

06 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/108828

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2018151772	A1	23 August 2018	KR	20190109532	A	25 September 2019
				US	2018233150	A1	16 August 2018
				EP	3559944	A1	30 October 2019
				US	2019304465	A1	03 October 2019
				DE	202017106606	U1	14 February 2018
				US	10311876	B2	04 June 2019
				CN	110268469	A	20 September 2019
JP	08314490	A	29 November 1996	JP	H08314490	A	29 November 1996
CN	107221325	A	29 September 2017	TW	I620170	B	01 April 2018
				TW	201735008	A	01 October 2017
				US	2017278512	A1	28 September 2017
				US	9805714	B2	31 October 2017

国际检索报告

国际申请号

PCT/CN2019/108828

A. 主题的分类

G10L 15/02 (2006.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G10L 15

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

DWPI; CNABS; CNTXT, IEEE: 唤醒, 激活, 苏醒, 关键词, 热词, 元音, 韵母, 韵部, 局部, 部分, 定位, vowel, part, phoneme, wake up, wake on, trig+, locat+, keyword+, spot+

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	WO 2018151772 A1 (GOOGLE LLC) 2018年 8月 23日 (2018 - 08 - 23) 说明书第1-6段	1-15
A	JP 08314490 A (JAPAN BROADCASTING CORP.) 1996年 11月 29日 (1996 - 11 - 29) 全文	1-15
A	CN 107221325 A (华硕电脑股份有限公司) 2017年 9月 29日 (2017 - 09 - 29) 全文	1-15

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 12月 12日

国际检索报告邮寄日期

2020年 1月 6日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

游晓梅

电话号码 (86-10)62089539

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/108828

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
WO	2018151772	A1	2018年 8月 23日	KR	20190109532	A	2019年 9月 25日
				US	2018233150	A1	2018年 8月 16日
				EP	3559944	A1	2019年 10月 30日
				US	2019304465	A1	2019年 10月 3日
				DE	202017106606	U1	2018年 2月 14日
				US	10311876	B2	2019年 6月 4日
				CN	110268469	A	2019年 9月 20日
JP	08314490	A	1996年 11月 29日	JP	H08314490	A	1996年 11月 29日
CN	107221325	A	2017年 9月 29日	TW	I620170	B	2018年 4月 1日
				TW	201735008	A	2017年 10月 1日
				US	2017278512	A1	2017年 9月 28日
				US	9805714	B2	2017年 10月 31日