

(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
2019年2月14日 (14.02.2019)

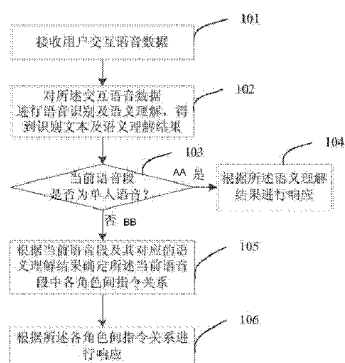


(10) 国际公布号
WO 2019/029352 A1

- (51) 国际专利分类号:
G10L 15/06 (2013.01)
- (21) 国际申请号: PCT/CN2018/096705
- (22) 国际申请日: 2018年7月23日 (23.07.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710676203.6 2017年8月9日 (09.08.2017) CN
- (71) 申请人: 科大讯飞股份有限公司 (IFLYTEK CO., LTD.) [CN/CN]; 中国安徽省合肥市高新开发区望江西路666号, Anhui 230088 (CN)。
- (72) 发明人: 李锐 (LI, Rui); 中国安徽省合肥市高新开发区望江西路666号, Anhui 230088 (CN)。 陈志刚 (CHEN, Zhigang); 中国安徽省合肥市高新开发区望江西路666号, Anhui 230088 (CN)。 王智国 (WANG, Zhiguo); 中国安徽省合肥市高新开发区望江西路666号, Anhui 230088 (CN)。 胡国平 (HU, Guoping); 中国安徽省合肥市高新开发区望江西路666号, Anhui 230088 (CN)。
- (74) 代理人: 深圳市威世博知识产权代理事务所 (普通合伙) (CHINA WISPRO INTELLECTUAL PROPERTY LLP.); 中国广东省深圳市南山区高新区粤兴三道8号中国地质大学产学研基地中地大楼A806, Guangdong 518057 (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, 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。

(54) Title: INTELLIGENT VOICE INTERACTION METHOD AND SYSTEM

(54) 发明名称: 一种智能语音交互方法及系统



- 101 Receive user interaction voice data
- 102 Perform voice recognition and semantic understanding on the interaction voice data to obtain recognized text and a semantic understanding result
- 103 Whether the current voice segment is voice of a single person?
- 104 Respond according to the semantic understanding result
- 105 Determine an instruction relationship between roles in the current voice segment according to the current voice segment and the corresponding semantic understanding result
- 106 Respond according to the instruction relationship between the roles
- AA Yes
- BB No

(57) Abstract: Disclosed in the present invention are an intelligent voice interaction method and system. The method comprises: receiving user interaction voice; performing voice recognition and semantic understanding on the interaction voice to obtain recognized text and a semantic understanding result; determining whether the current voice segment is voice of a single person; if yes, responding according to the semantic understanding result; otherwise, determining an instruction relationship between roles in the current voice segment according to the current voice segment and the corresponding semantic understanding result, and then responding according to the instruction relationship between the roles. The present invention can improve the responding accuracy rate in a human-machine interaction environment in which multiple persons participate, and improve the user experience.

(57) 摘要: 本发明公开了一种智能语音交互方法及系统, 该方法包括: 接收用户交互语音; 对所述交互语音进行语音识别及语义理解, 得到识别文本及语义理解结果; 确定当前语音段是否为单人语音; 如果是, 则根据所述语义理解结果进行响应; 否则, 根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系, 然后根据所述各角色间指令关系进行响应。本发明可以提高多人参与的人机交互环境下响应的正确率, 提升用户体验。

WO 2019/029352 A1

(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))。

一种智能语音交互方法及系统

【技术领域】

本发明涉及语音信号处理及自然语言理解领域，具体涉及一种智能语音交互方法及系统。

5

【背景技术】

随着人工智能技术的不断进步，人机语音交互也取得了长足的发展，各种语音助手APP和人机交互机器人大肆兴起，随之人们对自然便捷的人机交互渴望也达到了空前的高度。现有的人机交互方法多是基于端点检测技术确定用户有效交互语音，再对所述交互语音进行识别、语义理解，最后系统针对语义理解结果做出相应的响应。然而，人机交互常存在多人参与交互的情况，在该种情况下，不同角色的语音，可能是相互间的干扰、也有可能是补充或者是不同的交互指令，但现有的人机交互方法，会将多人的语音数据同作为一条语音指令数据进行识别、语义理解，最后进行响应，最终可能导致一次错误的交互。

15

【发明内容】

本发明实施例提供一种智能语音交互方法及系统，以避免在有多人参与的交互场景下，产生错误的理解及响应。

为此，本发明提供如下技术方案：

20

一种智能语音交互方法，所述方法包括：

接收用户交互语音数据；

对所述交互语音数据进行语音识别及语义理解，得到识别文本及语义理解结果；

确定当前语音段是否为单人语音；

如果是，则根据所述语义理解结果进行响应；

25

否则，根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系，然后根据所述各角色间指令关系进行响应。

优选地，所述方法还包括：预先构建说话人转折点判断模型，所述说话人转折点判断模型的构建过程包括：

确定说话人转折点判断模型的拓扑结构；

30

收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注；

利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数；

所述确定当前语音段是否为单人语音包括：

对于当前语音段中的每帧语音，提取其频谱特征；

35

将提取的频谱特征输入所述说话人转折点判断模型，根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点；

如果当前语音段中有至少一帧语音有转折点，则确定当前语音段不是单人语音；否则，确定当前语音段是单人语音。

优选地，所述如果当前语音段中有至少一帧语音有转折点，则确定当前语音段不是单人语音；否则，确定当前语音段是单人语音，包括：

如果当前语音段中有连续多帧语音均有转折点，则确定当前语音段不是单人语音；否则，确定当前语音段是单人语音。

5 优选地，所述根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系包括：

从当前语音段及其对应的语义理解结果中提取指令关联特征；
根据所述指令关联特征确定当前语音段中各角色间指令关系。

10 优选地，所述指令关联特征包括：声学特征和语义相关度特征；所述声学特征包括以下任意一种或多种：语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的夹角，所述关系夹角是指语音段所属声源与主麦克风连线与水平线之间的夹角；所述语义相关度特征为语义相关度值；

所述从当前语音段及其对应的语义理解结果中提取指令关联特征包括：

从当前语音段中提取所述声学特征；

15 根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值。

优选地，所述方法还包括：预先构建语义相关度模型，所述语义相关度模型的构建过程包括：

确定语义相关度模型的拓扑结构；

20 收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行语义相关度标注；

提取所述训练数据的语义相关特征；

利用所述语义相关特征及标注信息训练得到语义相关度模型；

所述根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值包括：

从当前语音段对应的语义理解结果中提取语义相关特征；

25 将所述语义相关特征输入所述语义相关度模型，根据所述语义相关度模型的输出得到当前语音段的语义相关度值。

优选地，所述语义相关特征包括：交互语音数据对应的文本词向量、交互语音数据中的用户指令涉及的业务类型。

30 优选地，所述方法还包括：预先构建指令关联识别模型，所述指令关联识别模型的构建过程包括：

确定指令关联识别模型的拓扑结构；

收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行角色间关联关系标注；

提取所述训练数据的指令关联特征；

35 利用所述指令关联特征及标注信息训练得到指令关联识别模型；

所述根据所述指令关联特征确定当前语音段中各角色间指令关系包括：

将所述指令关联特征输入所述指令关联识别模型，根据所述指令关联识别模型的输出得到当前语音段中各角色间指令关系。

优选地，所述各角色间指令关系包括：干扰、补充和独立。

40 一种智能语音交互系统，所述系统包括：

接收模块，用于接收用户交互语音数据；

语音识别模块，用于对所述交互语音数据进行语音识别，得到识别文本；

语义理解模块，用于对所述识别文本进行语义理解，得到语义理解结果；

判断模块，用于判断当前语音段是否为单人语音；

5 响应模块，用于在所述判断模块判断当前语音段是单人语音后，对所述语义理解结果进行响应；

指令关系识别模块，用于在所述判断模块判断当前语音段不是单人语音后，根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系；

10 所述响应模块，还用于根据所述指令关系识别模块确定的各角色间指令关系进行响应。

优选地，所述系统还包括：说话人转折点判断模型构建模块，用于预先构建说话人转折点判断模型；所述说话人转折点判断模型构建模块包括：

第一拓扑结构确定单元，用于确定说话人转折点判断模型的拓扑结构；

15 第一数据收集单元，用于收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注；

第一参数训练单元，用于利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数；

所述判断模块包括：

频谱特征提取单元，用于对于当前语音段中的每帧语音，提取其频谱特征；

20 转折点确定单元，用于将提取的频谱特征输入所述说话人转折点判断模型，根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点；

判断单元，用于在当前语音段中有至少一帧语音有转折点时，确定当前语音段不是单人语音；否则，确定当前语音段是单人语音。

25 优选地，所述判断单元具体用于在当前语音段中有连续多帧语音均有转折点时，确定当前语音段不是单人语音；否则，确定当前语音段是单人语音。

优选地，所述指令关系识别模块包括：

指令关联特征提取单元，用于从当前语音段及其对应的语义理解结果中提取指令关联特征；

30 指令关系确定单元，用于根据所述指令关联特征确定当前语音段中各角色间指令关系。

优选地，所述指令关联特征包括：声学特征和语义相关度特征；所述声学特征包括以下任意一种或多种：语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的夹角，所述关系夹角是指语音段所属声源与主麦克风连线与水平线之间的夹角；所述语义相关度特征为语义相关度值；

35 所述指令关联特征提取单元包括：

声学特征提取子单元，用于从当前语音段中提取所述声学特征；

语义相关度特征提取子单元，用于根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值。

40 优选地，所述系统还包括：语义相关度模型构建模块，用于预先构建语义相关度模型；所述语义相关度模型构建模块包括：

第二拓扑结构确定单元，用于确定语义相关度模型的拓扑结构；

第二数据收集单元，用于收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行语义相关度标注；

语义相关特征提取单元，用于提取所述训练数据的语义相关特征；

5 第二训练单元，用于利用所述语义相关特征及标注信息训练得到语义相关度模型；
所述语义相关度特征提取子单元，具体用于从当前语音段对应的语义理解结果中提取语义相关特征；将所述语义相关特征输入所述语义相关度模型，根据所述语义相关度模型的输出得到当前语音段的语义相关度值。

10 优选地，所述语义相关特征包括：交互语音数据对应的文本词向量、交互语音数据中的用户指令涉及的业务类型。

优选地，所述系统还包括：指令关联识别模型构建模块，用于预先构建指令关联识别模型；所述指令关联识别模型构建模块包括；

第三拓扑结构确定单元，用于确定指令关联识别模型的拓扑结构；

15 第三数据收集单元，收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行角色间关联关系标注；

指令关联特征提取单元，用于提取所述训练数据的指令关联特征；

第三训练单元，用于利用所述指令关联特征及标注信息训练得到指令关联识别模型；

所述指令关系确定单元，具体用于将所述指令关联特征输入所述指令关联识别模型，根据所述指令关联识别模型的输出得到当前语音段中各角色间指令关系。

20 优选地，所述各角色间指令关系包括：干扰、补充和独立。

一种智能语音交互设备，其中，包括相互连接的处理器和存储器；

所述存储器用于存储程序指令；

所述处理器用于运行所述程序指令以执行：

接收用户交互语音数据；

25 对所述交互语音数据进行语音识别及语义理解，得到识别文本及语义理解结果；

确定当前语音段是否为单人语音；

如果是，则根据所述语义理解结果进行响应；

否则，根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系，然后根据所述各角色间指令关系进行响应。

30 优选地，所述处理器还用于：预先构建说话人转折点判断模型，所述说话人转折点判断模型的构建过程包括：

确定说话人转折点判断模型的拓扑结构；

收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注；

利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数；

35 所述确定当前语音段是否为单人语音包括：

对于当前语音段中的每帧语音，提取其频谱特征；

将提取的频谱特征输入所述说话人转折点判断模型，根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点；

40 如果当前语音段中有至少一帧语音有转折点，则确定当前语音段不是单人语音；否则，确定当前语音段是单人语音；

所述处理器执行的所述根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系包括：

从当前语音段及其对应的语义理解结果中提取指令关联特征；

根据所述指令关联特征确定当前语音段中各角色间指令关系。

5 在另一实施例中，上述处理器用于实现上述任一智能语音交互方法。

本发明实施例提供的智能语音交互方法及系统，针对多人参与的交互场景的特点，对接收到的用户交互语音数据，判断是否为单人语音；如果不是，则通过对交互数据进行更细致准确的分析，得到多人参与交互情况下各角色指令间关系，根据各角色指令间关系合理地做出交互响应，从而解决了传统语音交互方案因未考虑多人参与交互情况所
10 带来的用户意图理解错误、系统交互响应错误的问题，有效地提高了用户体验。

【附图说明】

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例中所需使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明中记载的一些
15 实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

图1是本发明实施例智能语音交互方法的流程图；

图2是本发明实施例中构建说话人转折点判断模型的流程图；

图3是本发明实施例中说话人转折点判断模型的时序示意图；

图4是本发明实施例中构建语义相关度模型的流程图；

20 图5是本发明实施例中语义相关度模型的拓扑结构示意图；

图6是本发明实施例中构建指令关联识别模型的流程图；

图7是本发明实施例智能语音交互系统的结构示意图；

图8是本发明实施例中指令关系识别模块的一种具体结构示意图；

图9是本发明实施例中语音段与主麦克风的关系夹角的一种示意图；

25 图10是本发明实施例中语音段与主麦克风的关系夹角的另一种示意图；

图11是本发明实施例中智能语音交互系统的另一结构示意图。

【具体实施方式】

为了使本技术领域的人员更好地理解本发明实施例的方案，下面结合附图和实施方式
30 对本发明实施例作进一步的详细说明。

现有的语音交互系统中，仅根据端点检测技术确定一条条用户语音指令，并未考虑存在多人说话的情形，因此一轮交互指令中的后半句可能是前半句的干扰，或者是前半句的一个补充，或者是完全独立的两个子指令，此时如果不加以区分，有可能会得到错误的指令，进而会导致系统做出错误的响应，影响用户体验。针对这一情况，本发明实施例提供一种智能语音交互方法，针对多人参与的交互场景的特点，通过对交互语音数据
35 进行更细致准确的分析判断，得到多人参与交互情况下各角色指令间关系，并根据各角色指令间关系合理地做出交互响应。

如图1所示，是本发明实施例智能语音交互方法的流程图，包括以下步骤：

步骤 101, 接收用户交互语音数据。

具体地, 可以基于现有端点检测技术对音频流进行检测, 得到音频流中的有效语音, 作为用户的交互语音。所述端点检测技术需要设定停顿时长阈值 eos (通常为 0.5s-1s), 如果语音停顿时间大于所述停顿时长阈值, 则将音频流切断, 将该段语音作为有效的用户交互语音。

步骤 102, 对所述交互语音数据进行语音识别及语义理解, 得到识别文本及语义理解结果。

所述语音识别可以实时进行, 即实时识别出截止到当前时刻用户所说的内容。具体地, 由声学模型和语言模型构成解码网络, 解码网络包含截止到当前时刻, 所有候选的识别结果路径, 从当前时刻选取解码得分最大的识别结果路径作为当前时刻的识别结果。接收到新的用户交互语音数据后, 重新选取得分最大的识别结果路径, 并更新之前的识别结果。

对语音识别结果进行语义理解可以采用现有技术, 比如, 基于语法规则的语义理解、基于本体知识库的语义理解、基于模型的语义理解等, 对此本发明不做限定。

步骤 103, 确定当前语音段是否为单人语音。如果是, 则执行步骤 104; 否则, 执行步骤 105。

在确定当前语音段是否为单人语音时, 可以采用现有技术, 比如, 多说话人识别技术等。

步骤 104, 根据所述语义理解结果进行响应。

具体响应方式比如可以是生成响应文本, 并将响应文本反馈给用户, 或者是对所述语义理解结果的一个具体动作, 对此本发明实施例不做限定。如果是响应文本, 可以通过语音播报的方式将所述响应文本反馈给用户; 如果是一个具体操作, 可以将该操作的结果呈现给用户。

步骤 105, 根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系。

具体地, 可以首先从当前语音段及其对应的语义理解结果中提取指令关联特征; 然后根据所述指令关联特征确定当前语音段中各角色间指令关系。

步骤 106, 根据所述各角色间指令关系进行响应。

具体地, 可以根据各角色间指令关系及预先设定的响应策略做出响应, 如后半段是对前半段的干扰则只响应前半段意图、后半段是对前半段的补充则响应整句意图、前后半段独立(即重启新一轮对话)则只响应后半段意图。

进一步地, 在上述步骤 103 中, 确定当前语音段是否为单人语音时, 本发明实施例还可以采用基于说话人转折点判断模型的方法。具体地, 可以预先构建说话人转折点判断模型, 基于该说话人转折点判断模型来确定当前语音段是否为单人语音。

如图 2 所示, 是本发明实施例中说话人转折点判断模型的构建流程, 包括以下步骤:

步骤 201, 确定说话人转折点判断模型的拓扑结构。

所述说话人转折点判断模型的拓扑结构可以采用神经网络, 如 DNN(深度神经网络)、RNN(循环神经网络)、CNN(卷积神经网络)等, 以 BiLSTM(双向长短期记忆网络)为例, 充分考虑到 BiLSTM 既能利用历史信息, 又能利用未来信息的优势, 可以更好地进行说话人转折点判断。

说话人转折点判断模型的拓扑结构主要包括输入层、隐层和输出层，其中输入层的输入为每帧语音的频谱特征，如 39 维的 PLP (Perceptual Linear Predictive, 感知线性预测) 特征；隐层比如包含 2 层；输出层有 2 个节点，为是否有转折点判断的 2 维向量，有转折点为 1，没有转折点则为 0。

5 图 3 示出了说话人转折点判断模型的时序示意图，其中， $F_1 \sim F_t$ 表示输入层节点输入的频谱特征向量， $h_1 \sim h_t$ 为隐层各节点的输出向量。

步骤 202，收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注。

步骤 203，利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数。

10 模型参数的具体训练方法可采用现有技术，如 BPTT (反向传播) 算法，在此不再详细描述。

相应地，基于上述说话人转折点判断模型，在确定当前语音段是否为单人语音时，可以从当前语音段的每帧语音中提取相应的频谱特征，将提取的频谱特征输入所述说话人转折点判断模型，根据模型输出即可确定每帧语音中是否有转折点，如果有转折点，
15 则表明转折点前后是不同的说话人的语音，相应地，如果当前语音段中有一帧语音有转折点，则确定当前语音段不是单人语音。当然，为了避免误判，还可以在当前语音段中有连续多帧 (比如连续 5 帧) 语音均有转折点时，才确定当前语音段不是单人语音，否则，确定当前语音段是单人语音。

前面提到，在确定当前语音段中各角色间指令关系时，可以先从当前语音段及其对应的语义理解结果中提取指令关联特征，然后根据所述指令关联特征确定当前语音段中
20 各角色间指令关系。

所述指令关联特征包括：声学特征和语义相关度特征；其中，所述声学特征包括以下任意一种或多种：语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的关系夹角，所述关系夹角是指语音段所属声源与主麦克风连线与水平线之间的夹角，如图
25 9 和图 10 所示，分别针对线性麦克风和环形麦克风阵列，示出了语音段所属声源与主麦克风连线与水平线之间的夹角 θ 。这些声学特征可以根据当前语音段得到。所述语义相关度特征可以用 0-1 之间的数值来表示，即语义相关度值，具体可以根据当前语音段对应的语义理解结果及预先构建的语义相关度模型来确定。

如图 4 所示，是本发明实施例中构建语义相关度模型的流程图，包括以下步骤：

30 步骤 401，确定语义相关度模型的拓扑结构；

所述语义相关度模型的拓扑结构可以采用神经网络，比如以 DNN 为例，如图 5 所示，文本词向量经过卷积及线性变换层之后得到低阶词向量特征，然后与业务类型特征进行拼接，送入 DNN 回归网络，最终输出一个 0-1 之间的语义相关度值。

35 步骤 402，收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行语义相关度标注；

步骤 403，提取所述训练数据的语义相关特征；

所述语义相关特征包括用户交互语音数据对应的文本词向量、用户指令涉及的业务类型。其中，文本词向量的提取可以采用现有技术，比如利用已知的字嵌入 (word embedding) 矩阵，提取识别文本中每个词的词向量 (如 50 维)，然后再将前后两个语音
40 片段的词向量进行拼接，形成一个固定长度的向量，不够的补 0，如总计 $50 \times 20 = 1000$

维。用户指令涉及的业务类型，比如可以是：闲聊、订票、天气、导航、音乐、乱说构成的6维向量。

步骤404，利用所述指令关联特征及标注信息训练得到指令关联识别模型

进一步地，在本发明实施例中，语音段中各角色间指令关系的确定也可以采用基于预先训练的模型来实现，即预先训练指令关联识别模型，将提取的指令关联特征输入该模型，根据模型的输出得到当前语音段中各角色间指令关系。

如图6所示，是本发明实施例中构建指令关联识别模型的流程图，包括以下步骤：

步骤601，确定指令关联识别模型的拓扑结构；

所述指令关联识别模型可以采用神经网络模型，以DNN为例，其模型拓扑结构主要包括输入层、隐层、输出层，其中输入层各节点分别输入相应的声学特征和语义相关度特征，比如可以优选上述三个声学特征，则输入层有4个节点；隐层同于常见的DNN隐层，一般取3-7层；输出层为3个节点，分别输出三种指令关联关系，即干扰、补充和独立。

步骤602，收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行角色间关联关系标注；

角色间关联关系即：干扰、补充和独立这三种关系。

步骤603，提取所述训练数据的指令关联特征；

所述指令关联特征即前面提到的声学特征和语义相关度特征；所述声学特征包括：语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的夹角；所述语义相关度特征为语义相关度值，具体可以从所述训练数据的每个语音段及对应的语义理解结果中提取，语义相关度特征的提取可以采用基于语义相关度模型的方式，具体过程可参照前面的描述，在此不再赘述。

步骤604，利用所述指令关联特征及标注信息训练得到指令关联识别模型。

模型的具体训练方法可采用现有技术，在此不再详细描述。

基于该指令关联识别模型，在确定当前语音段中各角色间指令关系时，可以将当前语音段及其对应的语义理解结果中提取的指令关联特征输入所述指令关联识别模型，根据所述指令关联识别模型的输出即可得到当前语音段中各角色间指令关系。

本发明实施例提供的智能语音交互方法，针对多人参与的交互场景的特点，对接收到的用户交互语音数据，判断是否为单人语音；如果不是，则通过对交互数据进行更细致准确的分析，得到多人参与交互情况下各角色指令间关系，根据各角色指令间关系合理地做出交互响应，从而解决了传统语音交互方案因未考虑多人参与交互情况所带来的用户意图理解错误、系统交互响应错误的问题，有效地提高了用户体验。

相应地，本发明实施例还提供一种智能语音交互系统，如图7所示，是该系统的一种结构示意图，该系统包括以下各模块：

接收模块701，用于接收用户交互语音数据；

语音识别模块702，用于对所述交互语音数据进行语音识别，得到识别文本；

语义理解模块703，用于对所述识别文本进行语义理解，得到语义理解结果；

判断模块704，用于判断当前语音段是否为单人语音；

响应模块705，用于在所述判断模块704判断当前语音段是单人语音后，对所述语义理解结果进行响应；

指令关系识别模块 706, 用于在所述判断模块 704 判断当前语音段不是单人语音后, 根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系;

相应地, 在该实施例中, 所述响应模块 705, 还用于根据所述指令关系识别模块 706 确定的各角色间指令关系进行响应。

5 也就是说, 在当前语音是单人语音的情况下, 响应模块 705 直接语义理解结果进行响应, 否则根据语义识别结果中各角色间指令关系进行响应。如后半段是对前半段的干扰则只响应前半段意图、后半段是对前半段的补充则响应整句意图、前后半段独立(即重启新一轮对话)则只响应后半段意图, 从而避免了在有多人参与交互的情况下响应错误的问题, 提高了用户体验。

10 需要说明的是, 上述判断模块 704 在判断当前语音段是否为单人语音时, 可以采用现有技术, 比如, 多说话人识别技术等; 也可以采用基于模型的方式, 比如, 由说话人转折点判断模型构建模块预先构建说话人转折点判断模型, 所述说话人转折点判断模型构建模块可以作为本发明系统的一部分, 也可以独立于本发明系统, 对此本发明实施例不做限定。

15 如前面所述, 所述说话人转折点判断模型可以采用深层神经网络, 如 DNN、RNN、CNN 等, 所述说话人转折点判断模型构建模块的一种具体结构可以包括以下各单元:

第一拓扑结构确定单元, 用于确定说话人转折点判断模型的拓扑结构;

第一数据收集单元, 用于收集大量包含多人参与的交互语音数据, 并对所述交互语音数据进行转折点标注;

20 第一参数训练单元, 用于利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数。

相应地, 基于该说话人转折点判断模型, 上述判断模块 704 的一种具体结构可以包括以下各单元:

频谱特征提取单元, 用于对于当前语音段中的每帧语音, 提取其频谱特征;

25 转折点确定单元, 用于将提取的频谱特征输入所述说话人转折点判断模型, 根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点;

判断单元, 用于在当前语音段中有至少一帧语音有转折点时, 确定当前语音段不是单人语音; 否则, 确定当前语音段是单人语音。

30 上述指令关系识别模块 706 具体可以从当前语音段及其对应的语义理解结果中提取指令关联特征, 然后利用这些特征确定当前语音段中各角色间指令关系。如图 8 所示, 所述指令关系识别模块 706 的一种具体结构包括: 指令关联特征提取单元 761 和指令关系确定单元 762, 其中: 所述指令关联特征提取单元 761 用于从当前语音段及其对应的语义理解结果中提取指令关联特征; 所述指令关系确定单元 762 用于根据所述指令关联特征确定当前语音段中各角色间指令关系。

35 所述指令关联特征包括: 声学特征和语义相关度特征; 所述声学特征包括以下任意一种或多种: 语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的夹角; 所述语义相关度特征为语义相关度值。相应地, 所述指令关联特征提取单元可以包括以下各子单元:

40 声学特征提取子单元, 用于从当前语音段中提取所述声学特征, 具体可以采用现有技术;

语义相关度特征提取子单元，用于根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值，具体可以采用基于模型的方式，比如，由语义相关度模型构建模块预先构建语义相关度模型。

所述语义相关度模型构建模块的一种具体结构包括以下各单元：

5 第二拓扑结构确定单元，用于确定语义相关度模型的拓扑结构；

第二数据收集单元，用于收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行语义相关度标注；

语义相关特征提取单元，用于提取所述训练数据的语义相关特征；

第二训练单元，用于利用所述语义相关特征及标注信息训练得到指令关联识别模型。

10 相应地，基于上述语义相关度模型，所述语义相关度特征提取子单元可以首先从当前语音段对应的语义理解结果中提取语义相关特征；然后将所述语义相关特征输入所述语义相关度模型，根据所述语义相关度模型的输出即可得到当前语音段的语义相关度值。

需要说明的是，上述所述语义相关度模型构建模块可以作为本发明系统的一部分，也可以独立于本发明系统，对此本发明实施例不做限定。

15 上述指令关系确定单元 762 具体可以采用基于模型的方式来确定当前语音段中各角色间指令关系，比如，由指令关联识别模型构建模块预先构建指令关联识别模型。

所述指令关联识别模型构建模块的一种具体结构包括以下各单元：

第三拓扑结构确定单元，用于确定指令关联识别模型的拓扑结构；

20 第三数据收集单元，收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行角色间关联关系标注；

指令关联特征提取单元，用于提取所述训练数据的指令关联特征；

第三训练单元，用于利用所述指令关联特征及标注信息训练得到指令关联识别模型。

25 相应地，基于上述指令关联识别模型指令关系确定单元 762 可以将所述指令关联特征输入所述指令关联识别模型，根据所述指令关联识别模型的输出即可得到当前语音段中各角色间指令关系。

30 本发明实施例提供的智能语音交互系统，针对多人参与的交互场景的特点，对接收到的用户交互语音数据，判断是否为单人语音；如果不是，则通过对交互数据进行更细致准确的分析，得到多人参与交互情况下各角色指令间关系，根据各角色指令间关系合理地做出交互响应，从而解决了传统语音交互方案因未考虑多人参与交互情况所带来的用户意图理解错误、系统交互响应错误的问题，有效地提高了用户体验。本发明智能语音交互系统可以应用于各种人机交互设备或装置中，对交互环境适应性强，响应准确率高。

本发明实施例还提供另一种文本行识别系统，如图 11 所示，是本发明实施例智能语音交互系统的另一结构示意图。

35 在该实施例中，该系统包括相互连接的处理器 111 和存储器 112。该存储器 112 用于存储程序指令，而且还可用于存储处理器 111 在处理过程中的数据。处理器 111 用于运行该程序指令以执行上述实施例中的智能语音交互方法。

40 具体地，该智能语音交互系统可以为机器人、手机、电脑等任意具有信息处理能力的设备。处理器 111 还可以称为 CPU（Central Processing Unit，中央处理单元）。处理器 111 可能是一种集成电路芯片，具有信号的处理能力。处理器 111 还可以是通用处

理器、数字信号处理器（DSP）、专用集成电路（ASIC）、现成可编程门阵列（FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

5 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。而且，以上所描述的系统实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，10 即可以理解并实施。

以上对本发明实施例进行了详细介绍，本文中应用了具体实施方式对本发明进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及装置；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本发明的限制。

权利要求书

1. 一种智能语音交互方法，其中，所述方法包括：
接收用户交互语音数据；
- 5 对所述交互语音数据进行语音识别及语义理解，得到识别文本及语义理解结果；
确定当前语音段是否为单人语音；
如果是，则根据所述语义理解结果进行响应；
否则，根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系，然后根据所述各角色间指令关系进行响应。
- 10 2. 根据权利要求 1 所述的方法，其中，所述方法还包括：预先构建说话人转折点判断模型，所述说话人转折点判断模型的构建过程包括：
确定说话人转折点判断模型的拓扑结构；
收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注；
利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数；
- 15 所述确定当前语音段是否为单人语音包括：
对于当前语音段中的每帧语音，提取其频谱特征；
将提取的频谱特征输入所述说话人转折点判断模型，根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点；
如果当前语音段中有至少一帧语音有转折点，则确定当前语音段不是单人语音；否
20 则，确定当前语音段是单人语音。
3. 根据权利要求 2 所述的方法，其中，所述如果当前语音段中有至少一帧语音有转折点，则确定当前语音段不是单人语音；否则，确定当前语音段是单人语音，包括：
如果当前语音段中有连续多帧语音均有转折点，则确定当前语音段不是单人语音；
否则，确定当前语音段是单人语音。
- 25 4. 根据权利要求 1 所述的方法，其中，所述根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系包括：
从当前语音段及其对应的语义理解结果中提取指令关联特征；
根据所述指令关联特征确定当前语音段中各角色间指令关系。
5. 根据权利要求 4 所述的方法，其中，所述指令关联特征包括：声学特征和语义相
30 关度特征；所述声学特征包括以下任意一种或多种：语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的的关系夹角，所述关系夹角是指语音段所属声源与主麦克风连线与水平线之间的夹角；所述语义相关度特征为语义相关度值；
所述从当前语音段及其对应的语义理解结果中提取指令关联特征包括：
从当前语音段中提取所述声学特征；
35 根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值。
6. 根据权利要求 5 所述的方法，其中，所述方法还包括：预先构建语义相关度模型，所述语义相关度模型的构建过程包括：
确定语义相关度模型的拓扑结构；
收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行语义
40 相关度标注；

提取所述训练数据的语义相关特征；

利用所述语义相关特征及标注信息训练得到语义相关度模型；

所述根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值包括：

从当前语音段对应的语义理解结果中提取语义相关特征；

5 将所述语义相关特征输入所述语义相关度模型，根据所述语义相关度模型的输出得到当前语音段的语义相关度值。

7. 根据权利要求 6 所述的方法，其中，所述语义相关特征包括：交互语音数据对应的文本词向量、交互语音数据中的用户指令涉及的业务类型。

10 8. 根据权利要求 4 所述的方法，其中，所述方法还包括：预先构建指令关联识别模型，所述指令关联识别模型的构建过程包括；

确定指令关联识别模型的拓扑结构；

收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行角色间关联关系标注；

提取所述训练数据的指令关联特征；

15 利用所述指令关联特征及标注信息训练得到指令关联识别模型；

所述根据所述指令关联特征确定当前语音段中各角色间指令关系包括：

将所述指令关联特征输入所述指令关联识别模型，根据所述指令关联识别模型的输出得到当前语音段中各角色间指令关系。

20 9. 根据权利要求 4 所述的方法，其中，所述各角色间指令关系包括：干扰、补充和独立。

10. 一种智能语音交互系统，其中，所述系统包括：

接收模块，用于接收用户交互语音数据；

语音识别模块，用于对所述交互语音数据进行语音识别，得到识别文本；

语义理解模块，用于对所述识别文本进行语义理解，得到语义理解结果；

25 判断模块，用于判断当前语音段是否为单人语音；

响应模块，用于在所述判断模块判断当前语音段是单人语音后，对所述语义理解结果进行响应；

指令关系识别模块，用于在所述判断模块判断当前语音段不是单人语音后，根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系；

30 所述响应模块，还用于根据所述指令关系识别模块确定的各角色间指令关系进行响应。

11. 根据权利要求 10 所述的系统，其中，所述系统还包括：说话人转折点判断模型构建模块，用于预先构建说话人转折点判断模型；所述说话人转折点判断模型构建模块包括：

35 第一拓扑结构确定单元，用于确定说话人转折点判断模型的拓扑结构；

第一数据收集单元，用于收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注；

第一参数训练单元，用于利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数；

40 所述判断模块包括：

频谱特征提取单元，用于对于当前语音段中的每帧语音，提取其频谱特征；

转折点确定单元，用于将提取的频谱特征输入所述说话人转折点判断模型，根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点；

5 判断单元，用于在当前语音段中有至少一帧语音有转折点时，确定当前语音段不是单人语音；否则，确定当前语音段是单人语音。

12. 根据权利要求 11 所述的系统，其中，所述判断单元具体用于在当前语音段中有连续多帧语音均有转折点时，确定当前语音段不是单人语音；否则，确定当前语音段是单人语音。

13. 根据权利要求 10 所述的系统，其中，所述指令关系识别模块包括：

10 指令关联特征提取单元，用于从当前语音段及其对应的语义理解结果中提取指令关联特征；

指令关系确定单元，用于根据所述指令关联特征确定当前语音段中各角色间指令关系。

14. 根据权利要求 13 所述的系统，其中，所述指令关联特征包括：声学特征和语义
15 相关度特征；所述声学特征包括以下任意一种或多种：语音段的平均音量大小、语音段的信噪比、语音段与主麦克风的关系夹角，所述关系夹角是指语音段所属声源与主麦克风连线与水平线之间的夹角；所述语义相关度特征为语义相关度值；

所述指令关联特征提取单元包括：

声学特征提取子单元，用于从当前语音段中提取所述声学特征；

20 语义相关度特征提取子单元，用于根据当前语音段对应的语义理解结果确定当前语音段的语义相关度值。

15. 根据权利要求 14 所述的系统，其中，所述系统还包括：语义相关度模型构建模块，用于预先构建语义相关度模型；所述语义相关度模型构建模块包括：

第二拓扑结构确定单元，用于确定语义相关度模型的拓扑结构；

25 第二数据收集单元，用于收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行语义相关度标注；

语义相关特征提取单元，用于提取所述训练数据的语义相关特征；

第二训练单元，用于利用所述语义相关特征及标注信息训练得到语义相关度模型；

30 所述语义相关度特征提取子单元，具体用于从当前语音段对应的语义理解结果中提取语义相关特征；将所述语义相关特征输入所述语义相关度模型，根据所述语义相关度模型的输出得到当前语音段的语义相关度值。

16. 根据权利要求 15 所述的系统，其中，所述语义相关特征包括：交互语音数据对应的文本词向量、交互语音数据中的用户指令涉及的业务类型。

35 17. 根据权利要求 13 所述的系统，其中，所述系统还包括：指令关联识别模型构建模块，用于预先构建指令关联识别模型；所述指令关联识别模型构建模块包括：

第三拓扑结构确定单元，用于确定指令关联识别模型的拓扑结构；

第三数据收集单元，收集大量包含多人参与的交互语音数据作为训练数据，并对所述训练数据进行角色间关联关系标注；

指令关联特征提取单元，用于提取所述训练数据的指令关联特征；

40 第三训练单元，用于利用所述指令关联特征及标注信息训练得到指令关联识别模型；

所述指令关系确定单元，具体用于将所述指令关联特征输入所述指令关联识别模型，根据所述指令关联识别模型的输出得到当前语音段中各角色间指令关系。

18. 根据权利要求 17 所述的系统，其中，所述各角色间指令关系包括：干扰、补充和独立。

- 5 19. 一种智能语音交互系统，其中，包括相互连接的处理器和存储器；
所述存储器用于存储程序指令；
所述处理器用于运行所述程序指令以执行：
接收用户交互语音数据；
对所述交互语音数据进行语音识别及语义理解，得到识别文本及语义理解结果；
10 确定当前语音段是否为单人语音；
如果是，则根据所述语义理解结果进行响应；
否则，根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系，然后根据所述各角色间指令关系进行响应。

- 15 20. 根据权利要求 19 所述的系统，其中，所述处理器还用于：预先构建说话人转折点判断模型，所述说话人转折点判断模型的构建过程包括：
确定说话人转折点判断模型的拓扑结构；
收集大量包含多人参与的交互语音数据，并对所述交互语音数据进行转折点标注；
利用所述交互语音数据及标注信息训练得到说话人转折点判断模型参数；
所述确定当前语音段是否为单人语音包括：
20 对于当前语音段中的每帧语音，提取其频谱特征；
将提取的频谱特征输入所述说话人转折点判断模型，根据所述说话人转折点判断模型的输出确定每帧语音是否有转折点；
如果当前语音段中有至少一帧语音有转折点，则确定当前语音段不是单人语音；否则，确定当前语音段是单人语音；
25 所述处理器执行的所述根据当前语音段及其对应的语义理解结果确定所述当前语音段中各角色间指令关系包括：
从当前语音段及其对应的语义理解结果中提取指令关联特征；
根据所述指令关联特征确定当前语音段中各角色间指令关系。

30

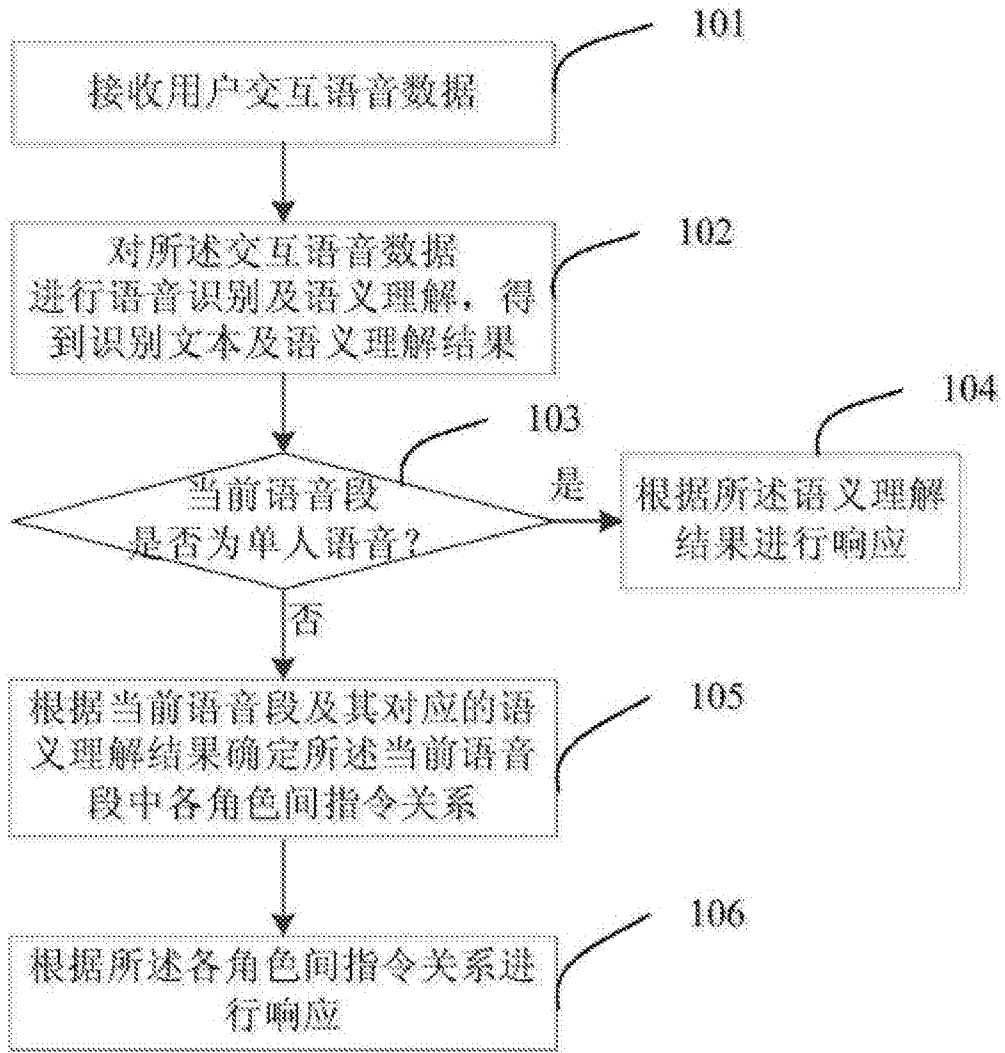


图 1

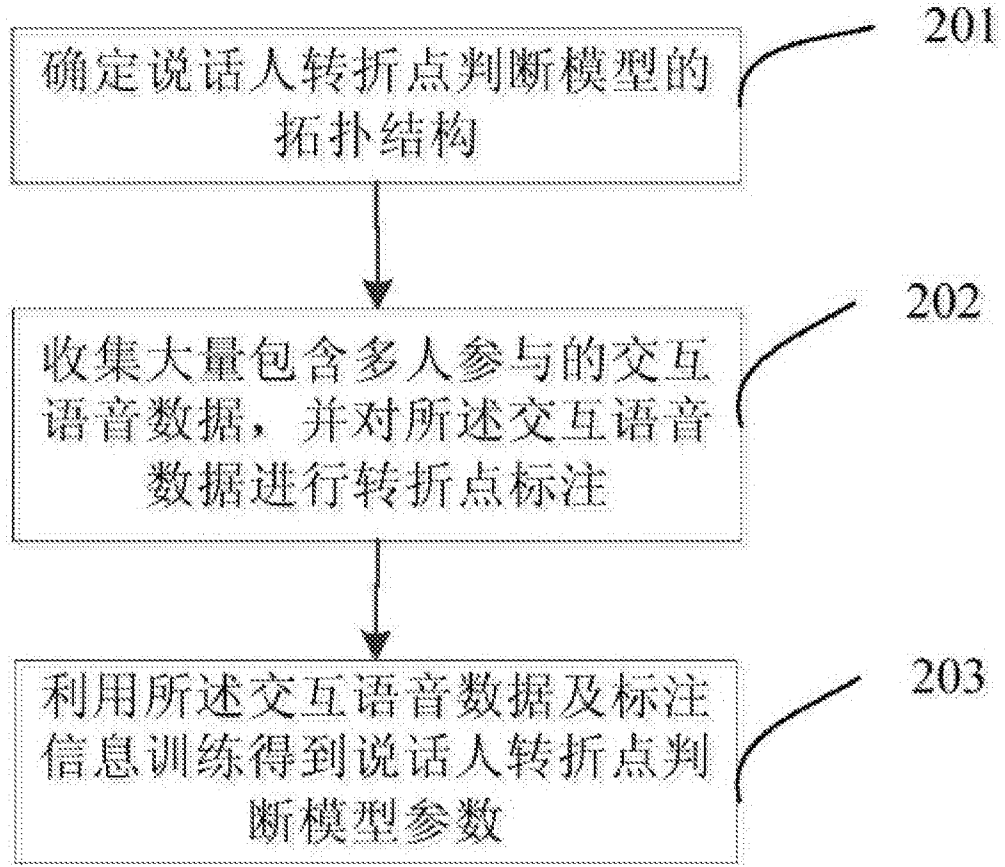


图 2

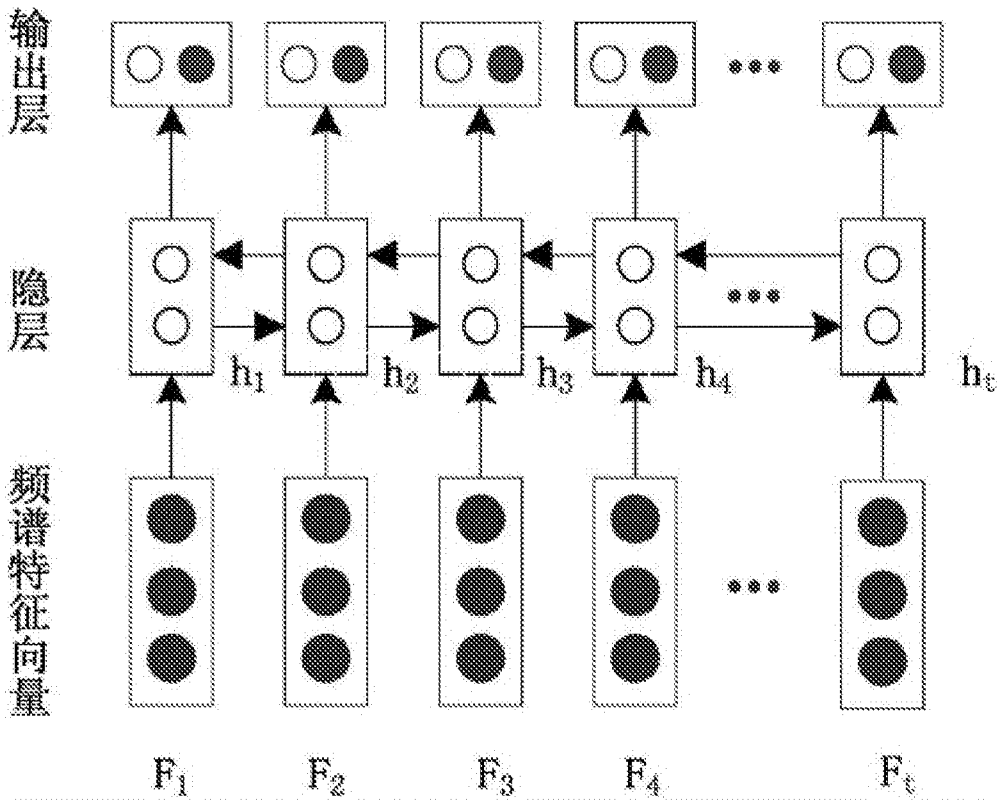


图 3

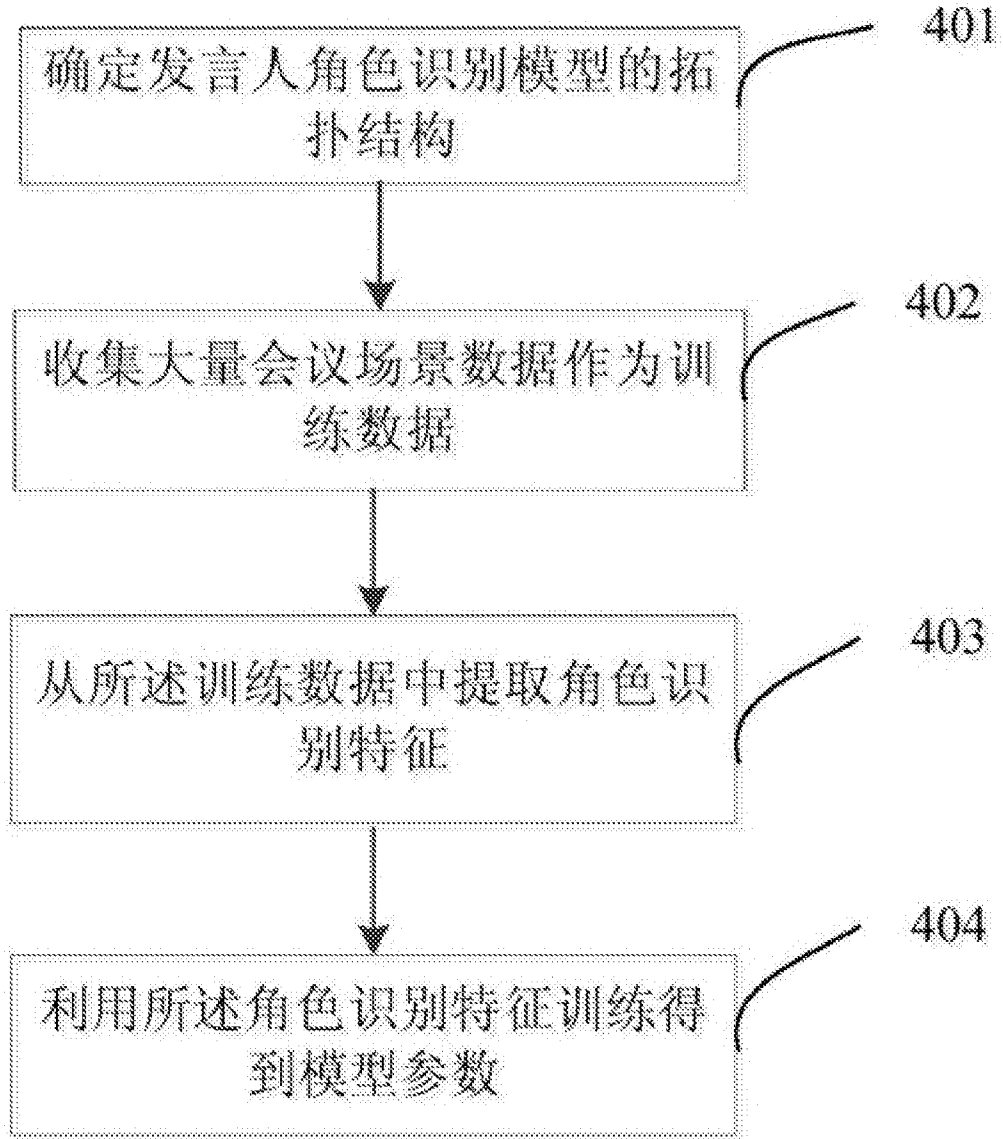


图 4

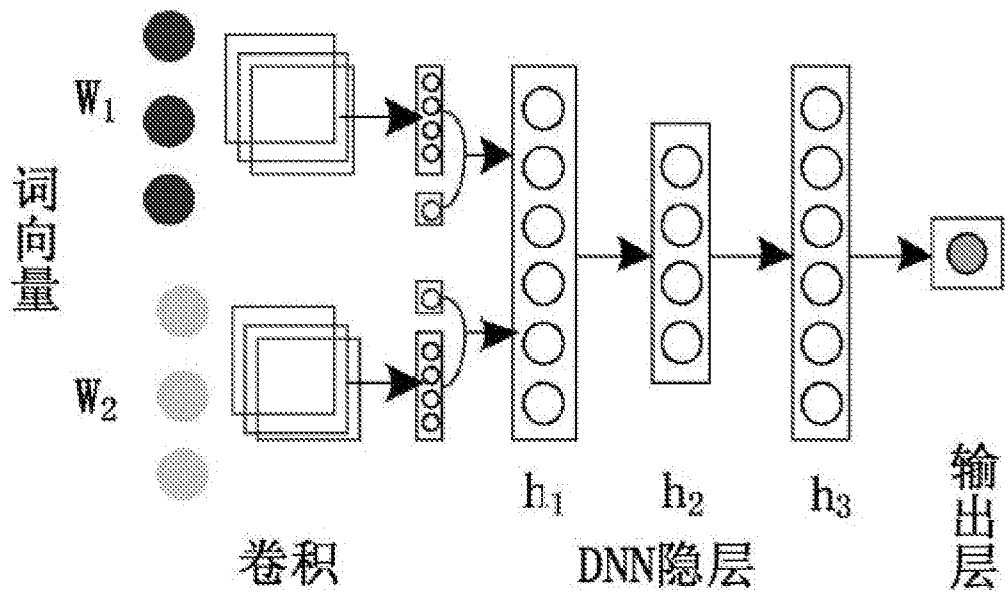


图 5

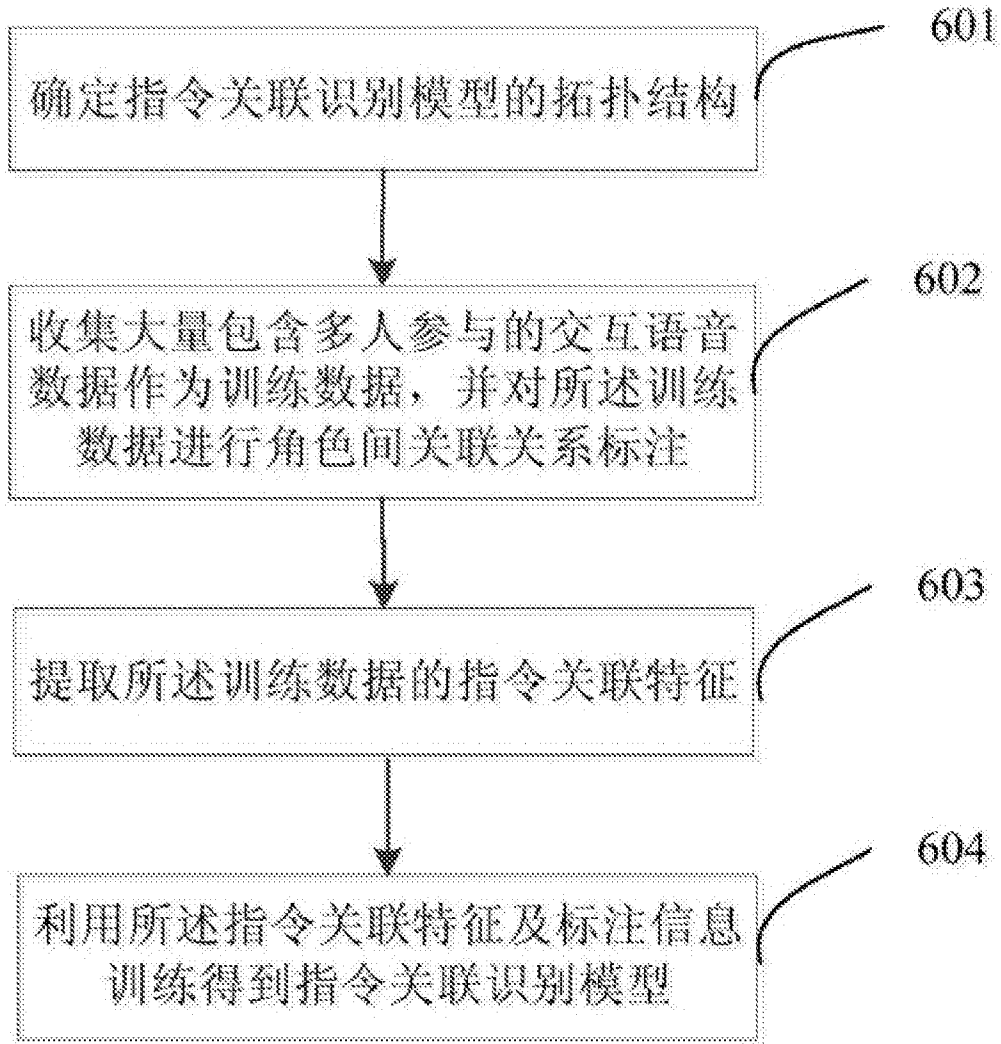


图 6

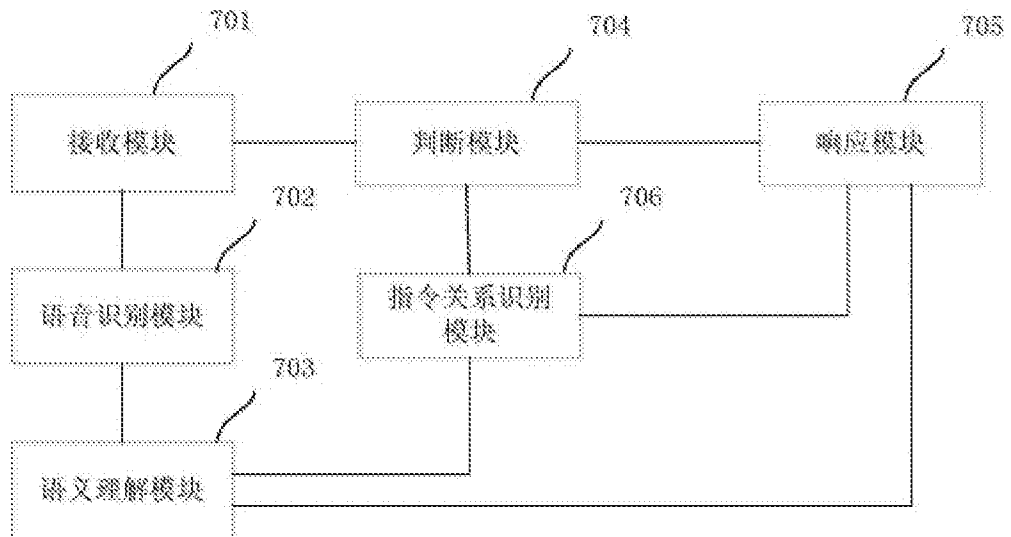


图 7

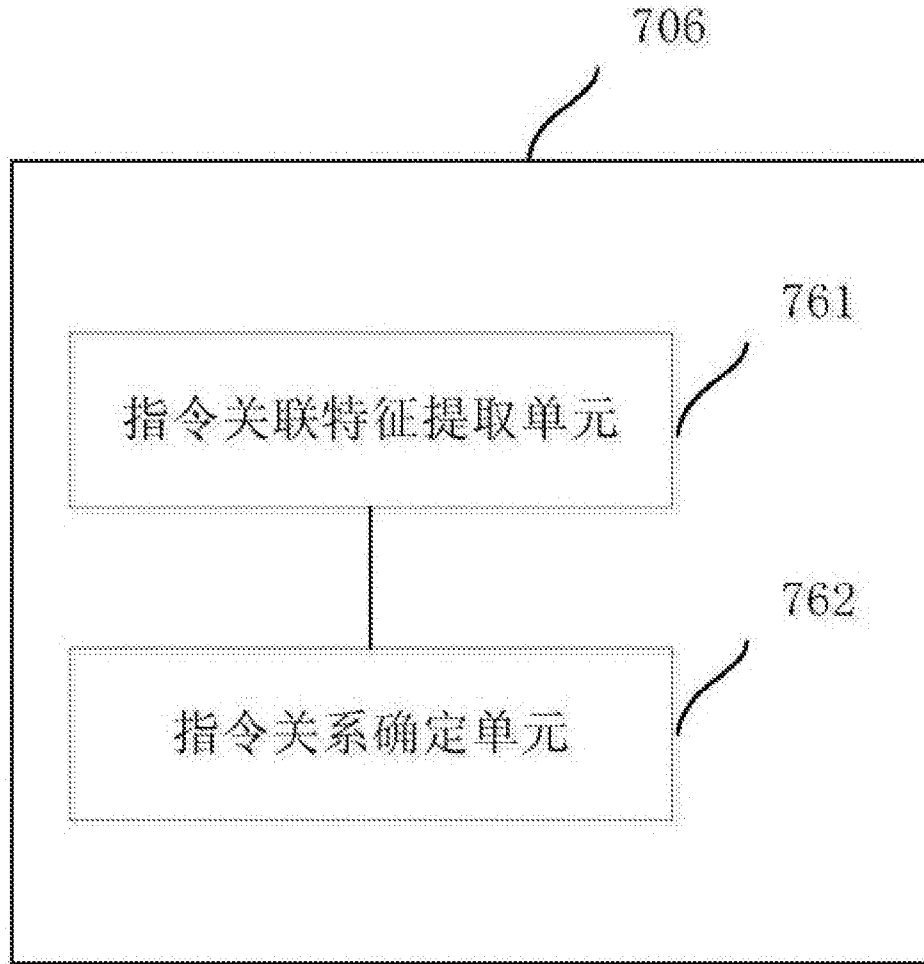


图 8

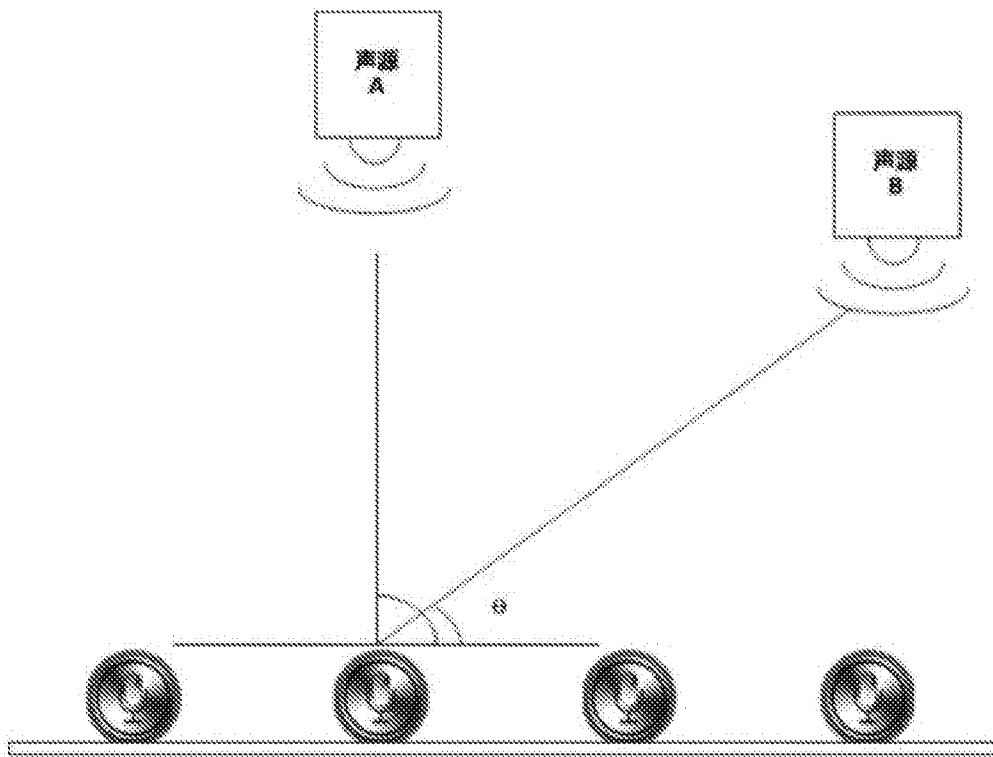


图 9

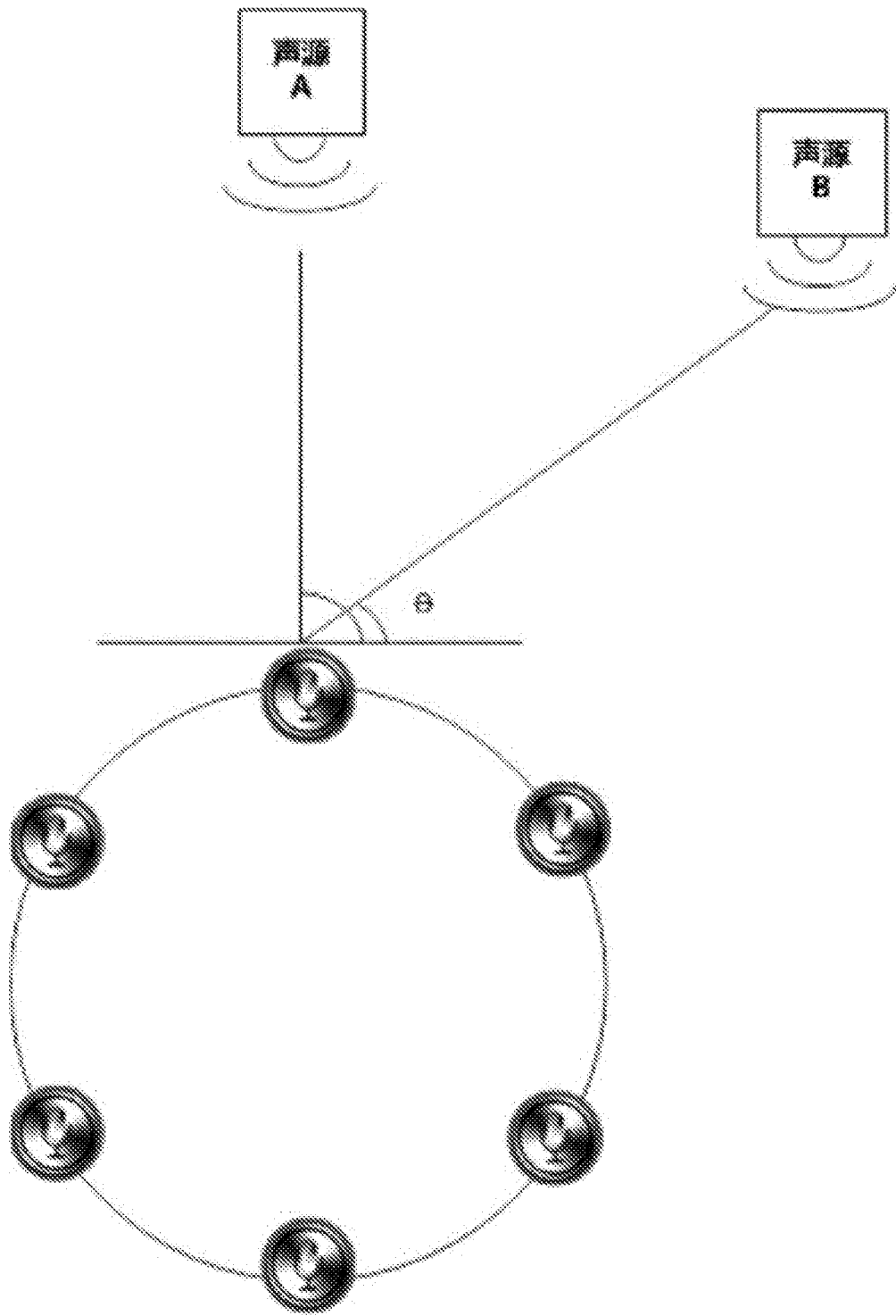


图 10

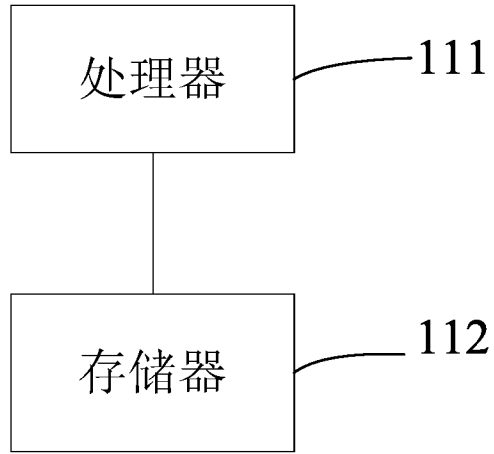


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/096705

A. CLASSIFICATION OF SUBJECT MATTER		
G10L 15/06(2013.01)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		
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)		
CNPAT, CNKI, WPI, EPODOC: 角色, 多人, 处理, 声音, 识别, 文本, 单人, 语音, 关系, 语段, 文字, 声学, 语义, 音频, 用户, audio?, role, manag+, identif+, text, single, speech??. sound?, voice?, semant+, phase?, stage?, paragraph?,		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 107437415 A (IFLYTEK CO., LTD.) 05 December 2017 (2017-12-05) claims 1-14	1-20
A	CN 104732969 A (HONGFUJIN PRECISION INDUSTRY (SHENZHEN) CO., LTD. ET AL.) 24 June 2015 (2015-06-24) description, paragraphs [0008]-[0024]	1-20
A	CN 102800315 A (SHANGHAI PATEO ELECTRONIC EQUIPMENT MANUFACTURING CO., LTD.) 28 November 2012 (2012-11-28) entire document	1-20
A	CN 104333956 A (LANGFANG POWER SUPPLY COMPANY OF STATE GRID JIBEI ELECTRIC POWER COMPANY ET AL.) 04 February 2015 (2015-02-04) entire document	1-20
A	US 2016379638 A1 (AMAZON TECHNOLOGIES, INC.) 29 December 2016 (2016-12-29) entire document	1-20
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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		Date of mailing of the international search report
20 September 2018		11 October 2018
Name and mailing address of the ISA/CN		Authorized officer
State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China		
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/096705

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107437415	A	05 December 2017	None			
CN	104732969	A	24 June 2015	None			
CN	102800315	A	28 November 2012	None			
CN	104333956	A	04 February 2015	None			
US	2016379638	A1	29 December 2016	WO	2016209924	A1	29 December 2016

<p>A. 主题的分类</p> <p>G10L 15/06 (2013.01) i</p> <p>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</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G10L</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNPAT, CNKI, WPI, EPODOC: 角色, 多人, 处理, 声音, 识别, 文本, 单人, 语音, 关系, 语段, 文字, 声学, 语义, 音频, 用户, audio?, role, manag+, identif+, text, single, speech??. sound?, voice?, semant+, phase?, stage?, paragraph?,</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 107437415 A (科大讯飞股份有限公司) 2017年 12月 5日 (2017 - 12 - 05) 权利要求1-14</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 104732969 A (鸿富锦精密工业深圳有限公司等) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0008]-[0024]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 102800315 A (上海博泰悦臻电子设备制造有限公司) 2012年 11月 28日 (2012 - 11 - 28) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 104333956 A (国网冀北电力有限公司廊坊供电公司等) 2015年 2月 4日 (2015 - 02 - 04) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2016379638 A1 (AMAZON TECHNOLOGIES, INC.) 2016年 12月 29日 (2016 - 12 - 29) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 107437415 A (科大讯飞股份有限公司) 2017年 12月 5日 (2017 - 12 - 05) 权利要求1-14	1-20	A	CN 104732969 A (鸿富锦精密工业深圳有限公司等) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0008]-[0024]段	1-20	A	CN 102800315 A (上海博泰悦臻电子设备制造有限公司) 2012年 11月 28日 (2012 - 11 - 28) 全文	1-20	A	CN 104333956 A (国网冀北电力有限公司廊坊供电公司等) 2015年 2月 4日 (2015 - 02 - 04) 全文	1-20	A	US 2016379638 A1 (AMAZON TECHNOLOGIES, INC.) 2016年 12月 29日 (2016 - 12 - 29) 全文	1-20
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 107437415 A (科大讯飞股份有限公司) 2017年 12月 5日 (2017 - 12 - 05) 权利要求1-14	1-20																		
A	CN 104732969 A (鸿富锦精密工业深圳有限公司等) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0008]-[0024]段	1-20																		
A	CN 102800315 A (上海博泰悦臻电子设备制造有限公司) 2012年 11月 28日 (2012 - 11 - 28) 全文	1-20																		
A	CN 104333956 A (国网冀北电力有限公司廊坊供电公司等) 2015年 2月 4日 (2015 - 02 - 04) 全文	1-20																		
A	US 2016379638 A1 (AMAZON TECHNOLOGIES, INC.) 2016年 12月 29日 (2016 - 12 - 29) 全文	1-20																		
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																				
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2018年 9月 20日</p>		<p>国际检索报告邮寄日期</p> <p>2018年 10月 11日</p>																		
<p>ISA/CN的名称和邮寄地址</p> <p>中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>受权官员</p> <p>张颖浩</p> <p>电话号码 (86-10)53961649</p>																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/096705

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)
CN	107437415	A	2017年 12月 5日	无			
CN	104732969	A	2015年 6月 24日	无			
CN	102800315	A	2012年 11月 28日	无			
CN	104333956	A	2015年 2月 4日	无			
US	2016379638	A1	2016年 12月 29日	WO	2016209924	A1	2016年 12月 29日