

(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
2017 年 8 月 3 日 (03.08.2017)



(10) 国际公布号
WO 2017/128991 A1

- (51) 国际专利分类号:
G10L 15/26 (2006.01) *H04L 12/58* (2006.01)
- (21) 国际申请号: PCT/CN2017/071382
- (22) 国际申请日: 2017 年 1 月 17 日 (17.01.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201610052305.6 2016 年 1 月 26 日 (26.01.2016) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 英属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: 及
- (71) 申请人 (仅对美国): 鄢志杰 (YAN, Zhijie) [CN/CN]; 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: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (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, 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))。

(54) Title: INSTANT COMMUNICATION METHOD AND INSTANT COMMUNICATION SYSTEM BASED ON VOICE RECOGNITION

(54) 发明名称: 一种基于语音识别的即时通信方法和即时通信系统

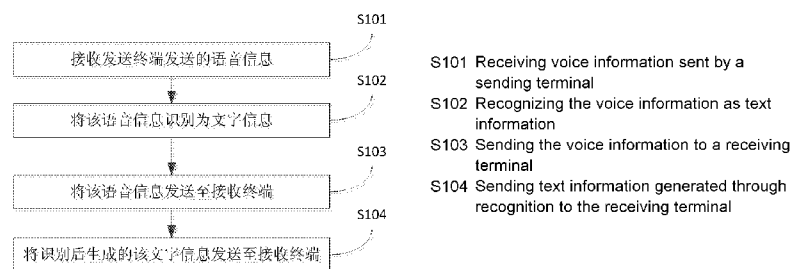


图1

(57) Abstract: An instant communication method and an instant communication system based on voice recognition. The instant communication method comprises: receiving voice information sent by a sending terminal (S101); performing voice recognition on the voice information to generate text information (S102); sending the voice information to a receiving terminal (S103); and sending the text information to the receiving terminal (S104). The instant communication method and system overcome the obstacle that a receiving terminal cannot listen to voice information after receiving same on certain occasions, thereby avoiding the problem of privacy leaks of a user.

(57) 摘要: 一种基于语音识别的即时通信方法和即时通信系统。即时通信方法包括: 接收发送终端发送的语音信息 (S101); 将语音信息进行语音识别, 生成文字信息 (S102); 将语音信息发送至接收终端 (S103); 以及将文字信息发送至接收终端 (S104)。即时通信方法和系统克服了某些场合下接收终端收到语音信息后无法收听的障碍, 避免了使用者隐私泄露的问题。

WO 2017/128991 A1

一种基于语音识别的即时通信方法和即时通信系统

技术领域

本申请涉及即时通信技术领域，尤其涉及一种基于语音识别的即时通信方法和即时通信系统。

背景技术

通过手机或平板电脑的社交 app 对讲聊天是很多软件常用的便利功能，例如腾讯的微信、阿里的钉钉、支付宝、淘宝等都具备这样的功能。目前这类功能主要的实现方式是发送终端通过语音方式录制自己的留言，接受方点接收到的信息，通过听筒或外放收听。

这类功能在方便发送终端的同时，对接收终端实际带来了一定障碍。主要缺点在于：接收终端无法像文字信息一样一目了然的看到信息内容，需要点按再将手机或平板拿到耳边用听筒收听，或是用手机或平板的扬声器外放，这在很多场合（例如会议中、或旁边有其他人），这是非常不便的，也可能存在隐私泄露的问题。

发明内容

鉴于上述问题，提出了本申请实施例以便提供一种克服上述问题或者至少部分地解决上述问题的基于语音识别的即时通信方法和即时通信系统。

为解决上述问题，本申请公开一种基于语音识别的即时通信方法，包括：

接收发送终端发送的语音信息；

将该语音信息进行语音识别，生成文字信息；

将该语音信息发送至接收终端；以及

将该文字信息发送至接收终端。

本申请另一实施例提出一种基于语音识别的即时通信方法，包括：

5 录制语音信息并发送至服务器；

接收经过识别该语音信息生成的文字信息，并显示该文字信息；

在接收到纠正操作指令后，进入编辑文字信息的界面；

显示编辑后文字信息，并将编辑后文字信息发送至服务器。

本申请再一实施例提出一种基于语音识别的即时通信方法，包括：

10 接收服务器发送的语音信息；

接收服务器发送的识别该语音信息后生成的文字信息；

显示并标记该文字信息。

本申请一实施例提出一种基于语音识别的即时通信系统，其特征在于，包括：

15 语音信息接收模块，用于接收发送终端发送的语音信息；

文字信息生成模块，用于将该语音信息进行语音识别，生成文字信息；

第一发送模块，用于将该语音信息发送至接收终端；以及

第二发送模块，用于将该文字信息发送至接收终端。

20 本申请另一实施例提出一种基于语音识别的即时通信系统，包括：

语音信息录制发送模块，用于录制语音信息并发送至服务器；

文字信息接收显示模块，用于接收经过识别该语音信息生成的文字信息，并显示该文字信息；

编辑模块，用于在接收到纠正操作指令后，进入编辑文字信息的界面；

显示发送模块，用于显示编辑后文字信息，并将编辑后文字信息发送至服务器。

5 本申请再一实施例提出一种基于语音识别的即时通信系统，包括：

语音信息获取模块，用于接收服务器发送的语音信息；

文字信息获取模块，用于接收服务器发送的识别该语音信息后生成的文字信息；

文字信息显示标记模块，用于显示并标记该文字信息。

10 本申请实施例至少具有以下优点：

本申请实施例提出的基于语音识别的即时通信方法和即时通信系统中，通过语音识别功能，将语音信息和文字信息均发送至接收终端，克服了接收终端获得信息的障碍，方便了使用者的使用，避免了隐私泄露的问题。

15

附图说明

图 1 是本申请第一实施例的基于语音识别的即时通信方法的流程图。

图 2 是本申请第二实施例的基于语音识别的即时通信方法的流程图。

图 3 是本申请第三实施例的基于语音识别的即时通信方法的流程图。

20 图 4 是本申请第四实施例的基于语音识别的即时通信方法的流程图。

图 5 是对应于本申请第一实施例的基于语音识别的即时通信方法的即时通信系统的方框图。

图 6 是对应于本申请第二实施例的基于语音识别的即时通信方法的即时通信系统的方框图。

图 7 是对应于本申请第三实施例的基于语音识别的即时通信方法的即时通信系统的方框图。

图 8 是对应于本申请第四实施例的基于语音识别的即时通信方法的即时通信系统的方框图。

5 具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员所获得的所有其他实施例，都属于本申请保护的范围。

10 本申请的核心思想之一在于，提出一种即时通信方法和即时通信系统，使用语音识别将语音信息进行识别，并通过服务器将文字信息直接显示在发送终端和接收终端的屏幕上，方便了接收终端接收信息，克服了某些场合下接收终端收到语音信息后无法收听的障碍，避免了使用者隐私泄露的问题。

15

第一实施例

本申请第一实施例提出一种基于语音识别的即时通信方法，如图 1 所示为本申请第一实施例的基于语音识别的即时通信方法的流程图。本申请第一实施例中的即时通信方法应用于服务器，包括如下步骤：

20 S101，接收发送终端发送的语音信息；

在这一步骤中，发送终端可以在即时通信界面（例如聊天界面）录制语音信息，录音完成之后松开该标记或按钮，则录制完成。之后，发送终端将语音信息通过网络发送至服务器。

S102，将该语音信息识别为文字信息；

25 在这一步骤中，服务器接受到该方发送的语音信息之后，通过语音识别技术，将该语音信息识别为文字信息。语音识别技术是本领域常用的

技术，在此并不赘述。

S103，将该语音信息发送至接收终端；

在这一步骤中，服务器将步骤 S101 中接收到的语音信息发送至接收终端。

- 5 值得注意的是，步骤 S103 可以与步骤 S102 同时执行或先后执行，当先后执行时，步骤 S102 和步骤 S103 的步骤顺序并不特别限定。

S104，将识别后生成的该文字信息发送至接收终端；

- 10 在这一步骤中，服务器将经过语音识别处理后生成的文字信息发送给接收终端。优选地，在这一步骤中，服务器在发送文字信息的同时发送指定标记，用于区别由语音信息转成的文字信息和发送方直接以文字方式输入的文字信息。

值得注意的是，当步骤 S103 在步骤 S102 之后执行时，步骤 S104 可以与步骤 S103 同时执行，或者步骤 S104 可以先于或后于步骤 S103 执行，本申请并不特别限定。

- 15 在一实施例中，可以先执行步骤 S103，将步骤 S101 中收到的语音信息发送至接收终端，再执行步骤 S102，将语音信息经过语音识别生成文字信息，之后执行步骤 S104，将识别后生成的文字信息发送至接收终端；在另一实施例中，可以先执行步骤 S102，将步骤 S101 中收到的语音信息进行语音识别生成文字信息，再同时或先后执行步骤 S103 和步骤
20 S104，将语音信息和识别后生成的文字信息发送至接收终端。

- 综上所述，本申请第一实施例提出一种基于语音识别的即时通信方法，将语音信息通过识别生成文字信息，通过服务器将语音信息和文字信息均发送至接收终端。该实施例提供的即时通信方法方便了接收终端接收信息，克服了某些场合下接收终端收到语音信息后无法收听的障
25 碍，避免了使用者隐私泄露的问题。

第二实施例

本申请第二实施例提出一种基于语音识别的即时通信方法，如图 2 所示为本申请第二实施例的基于语音识别的即时通信方法的流程图。本申请第一实施例中的即时通信方法应用于服务器，包括如下步骤：

- 5 S201，接收发送终端发送的语音信息；
- S202，将该语音信息识别为文字信息；
- S203，将该语音信息发送至接收终端；
- S204，将识别后生成的该文字信息发送至接收终端；

10 上述步骤 S201 至 S204 与第一实施例中的步骤 S101 至步骤 S104 相同或相似，在此并不赘述。

 在一优选实施例中，在步骤 S202 之后，该方法还可以包括

 S205，将识别后生成的该文字信息发送至发送终端；

 在这一步骤中，服务器将在步骤 S202 中生成的文字信息发送至发送终端。

15 其中，步骤 S205、步骤 S204 和步骤 S203 的执行顺序并不限制，三者可以同时执行，或者以任意顺序先后执行，本申请并不特别限制。

 另外，在步骤 S202 之后，所述方法还可以包括：

 S206，将识别后生成的该文字信息储存于数据库；

20 在这一步骤中，服务器将识别后生成的文字信息发送至与服务器连接的数据库中备用。这一步骤 S206 可以与步骤 S203 至 S205 中的任一者同时或以任意顺序先后执行，本申请并不特别限制。

 在步骤 S202 之后，所述方法还可以包括：

 S207，将辅助错误纠正信息发送至发送终端；

 这一步骤可以与步骤 S203 至 S205 中的任一者同时或以任意顺序先后

执行，本申请并不特别限制。优选地，步骤 S207 可以与步骤 S205 同时执行，即在将识别后生成的文字信息发送至发送终端的同时，将错误辅助纠正信息同时发送至发送终端，供发送终端修改识别后的文字信息。

在语音识别过程中，将会产生词图（word graph）及识别词多候选信息，在步骤 S207 中，可以根据词图里的信息，使用算法，推荐备选的纠错词给使用者点选。这些错误信息通过回传发送终端，可以辅助更高效的对识别文本进行错误纠正。例如，当发送终端的使用者选择错误纠正、并点击识别错误的某字词时，可通过辅助纠正信息得到该字词的其他候选字词，并显示在虚拟键盘上，使用者可通过点击正确候选高效的进行错误纠正。具体地，举例来说，使用者说：“我要买黄色的”，语音识别错误识别成“我要买红色的”，当使用者点击“红色”这个词时，算法可根据词图信息，提示出“黄色”这个第二候选供使用者点选。用户点击“黄色”，即完成了替换纠错的操作，非常简单快捷。

之后，所述方法还可以包括：

步骤 S208，接收发送终端发出的编辑后文字信息，并发送至接收终端；

在这一步骤中，当发送终端的使用者完成纠正后，发送终端将编辑后文字信息发送至服务器，服务器接收该编辑后文字信息，并发送至接收终端。

优选地，在步骤 S208 之后，本申请还可以包括：

步骤 S209，将编辑后文字信息发送至数据库。

在这一步骤中，被纠正过的自动语音识别结果价值很高、尤为重要，它提示了：1）服务器未能完全正确地识别该语音信息；2）该语音信息的正确文字信息已由用户通过纠正给出。对这类编辑后文字信息，可以利用语音识别系统的训练算法，记录识别错误的文字内容、所对应的语音内容和正确的语音内容，避免此后再犯类似错误。这类错误纠正数据对语音识别系统自我进化的功能比其他数据所不可比拟的。

综上所述，本申请第二实施例提出一种基于语音识别的即时通信方法，将语音信息通过识别生成文字信息，通过服务器将语音信息和文字信息均发送至接收终端，并将文字信息发送至发送终端，在发送给发送终端之后提供辅助修改信息，利用该信息可以让发送终端的使用者能够
5 高效地修改。该实施例提供的即时通信方法方便了接收终端接收信息，克服了某些场合下接收终端收到语音信息后无法收听的障碍，避免了使用者隐私泄露的问题，同时进一步保证了接收终端接收到信息的准确性。

10 第三实施例

本申请第三实施例提出一种基于语音识别的即时通信方法，如图 3 所示为本申请第三实施例的基于语音识别的即时通信方法的流程图。本申请第三实施例中的即时通信方法应用于信息的发送终端，包括如下步骤：

15 S301，录制语音信息并发送至服务器；

在这一步骤中，发送终端可以在即时通信界面（例如聊天界面）录制语音信息，例如按住输入框的指定标记或按钮不放，则开始录音，录音完成之后松开该标记或按钮，则录制完成。在录制完成之后，该即时通信界面可以默认为直接发送，或者发送终端点击另一标记或按钮，将信息通过网络发送至服务器。
20

S302，接收经过服务器识别该语音信息后的生成文字信息，并显示该文字信息；

在这一步骤中，服务器将发送终端发送的语音信息进行语音识别生成文字信息并回传给发送终端，发送终端接收识别后的文字信息，并进行显示。例如在聊天界面，发送终端在步骤 S301 中将录制好的语音信息发送给服务器，在此步骤 S302 中，发送终端可在同一聊天界面中接收服务器回传的识别该语音信息后生成的文字信息，并显示于该聊天界面。
25

S303, 在接收到纠正操作指令后, 开启错误纠正界面, 进入编辑文字信息的界面;

在这一步骤中, 当发送终端的使用者认为语音识别后生成的文字信息的内容与语音信息不一致, 则可以通过发出纠正操作指令开启错误纠正界面。例如, 纠正操作指令可以为使用者长按该文字信息, 发送终端即接收该指令并开启错误纠正界面, 进入编辑文本状态, 同时该纠正界面可以显示虚拟键盘或者手写键盘等输入界面, 供使用者纠正错误。使用者可以通过虚拟键盘等对文字信息进行增、删等操作。

之后, 本方法还可以包括:

10 S304, 显示编辑后文字信息, 并将编辑后文字信息发送至服务器。

在这一步骤中, 发送终端的使用者编辑之后的编辑后文字信息已显示在发送端, 该文字信息同时由发送终端上传至服务器中, 由该服务器发送至接收方并进行同步显示, 本申请不再赘述。

在一优选实施例中, 步骤 S302 之后还可以包括:

15 步骤 S302a, 接收服务器发送的辅助修改信息;

在这一步骤中, 将在语音识别过程中产生的词图 (word graph) 及识别词多候选信息发送至发送终端, 可以辅助发送终端使用者更高效的对识别文本进行错误纠正。

在步骤 S303 中, 该错误纠正界面不仅可以显示文本信息进入编辑状态、虚拟键盘或者手写键盘等输入界面, 同时可以显示步骤 S302a 中服务器发送的辅助修改信息, 例如, 当服务器认为语音识别之后生成的文本信息中某一句话或某一个词不符合语法构成, 则可以在该句或该词的下方加上虚线下划线, 同时在发送终端显示界面的其他位置 (例如输入界面) 显示服务器发送来的辅助修改信息中包含的多个候选词, 供使用者点
20 选正确的候选词。或者, 当发送方选择错误纠正、并点击识别错误的某
25 字词时, 可通过辅助纠正信息得到该字词的其他候选字词, 并显示在虚拟键盘上, 用户可通过点击正确候选高效的进行错误纠正。

在一优选实施例中，步骤 S302 之后还包括：

S302b，在接收到播放语音信息指令后，播放语音信息；

在该步骤中，若发送终端的使用者通过点击所显示的文字信息等方式发出播放语音信息指令，则发送终端可以通过听筒或扬声器播放在步
5 骤 3101 中录制的语音信息。

综上所述，本申请第三实施例提出一种基于语音识别的即时通信方法，将语音信息通过识别生成文字信息，并提供错误纠正功能，可以让发送终端的使用者能够修改识别后的文字信息。该实施例提供的即时通信方法方便了接收终端接收信息，克服了某些场合下接收终端收到语音
10 信息后无法收听的障碍，避免了使用者隐私泄露的问题，同时保证了接收终端接收到信息的准确性。

优选地，本申请第三实施例还可以接收服务器发出的辅助修改信息，可以让使用者高效地修改文本信息，进一步提高了信息的准确性和及时性。

第四实施例

本申请第四实施例提出一种基于语音识别的即时通信方法，如图 4 所示为本申请第四实施例的基于语音识别的即时通信方法的流程图。本申请第四实施例中的即时通信方法应用于信息的接收终端，包括如下步
20 骤：

S401，接收服务器发送的语音信息；

在这一步骤中，发送终端录制语音信息并发送至服务器，在由服务器将该语音信息发送至接收终端；

S402，接收服务器发送的识别该语音信息后生成的文字信息；

在这一步骤中，服务器将该语音信息经过语音识别生成文字信息之后，发送至接收终端，接收终端接收经过识别生成的这一文字信息。

值得注意的是，步骤 S401 和步骤 S402 可以同时或先后执行，即接收终端可以同时或先后接收语音信息和生成的文字信息，本申请并不特别限制。优选地，服务器将语音信息转成文字信息之后，再将语音信息和文字信息同时发送给接收终端，接收终端同时接收该语音信息和该文字信息。

S403，显示并标记该文字信息；

在这一步骤中，接收终端可以将该文字信息显示于即时通信的界面上。由于该文字信息是由语音信息经过识别后生成，为了将其区别于发送方直接以文字输入的文字信息，可以对该文字信息进行标记，例如通过设置特别的底色、字体、标记特别的字符（例如“语音识别”或“ASR”）来区分普通文字信息和语音识别的文字信息。

在标记该文字信息中，一种可能的方式是，当接收终端接收到语音信息和对应于该语音信息的文字信息，则接收终端将该文字信息进行标记，使之区别于服务器发来的由发送终端直接以文字形式输入的文字信息；另一种可能的方式是，服务器在发送该文字信息时同时发送标记，该标记与该文字信息同时显示于接收终端的显示界面上。在这一种情况下，步骤 S402 之后还包括：

S402a，接收服务器发送的标记信息。

在这一步骤中，这一标记信息例如可以为设置特别的底色、字体、标记特别的字符（例如“语音识别”或“ASR”）等。

优选地，在步骤 S403 之后，该方法还可以包括：

S404，当接收到使用者的播放该语音信息的指令，播放该语音信息；

在这一实施例中，播放语音信息的指令可以为使用者点击该文字信息，当使用者点击所显示的文字信息，该接收终端通过听筒或扬声器播放步骤 S401 中接收到的语音信息；

优选地，在步骤 S403 之后，该方法还可以包括：

S405，接收服务器发送的编辑后文字信息，并显示编辑后文字信息；

在该步骤中，当发送终端对文字信息进行错误纠正后，发送终端将纠正后文本信息发送至服务器，由服务器发送至接收终端，接收终端接收该编辑后文字信息，并进行显示。优选地，接收终端可以用编辑后文字信息覆盖修改之前的文字信息。

综上所述，本申请第四实施例提出一种基于语音识别的即时通信方法，将语音信息通过识别生成文字信息，并提供错误纠正功能，可以让接收终端的使用者直接接收经过语音识别的文字信息，并能够明确该文字信息是由发送终端直接以文字形式发出还是经过语音识别后生成的文字信息。该实施例提供的即时通信方法方便了接收终端接收信息，克服了某些场合下接收终端收到语音信息后无法收听的障碍，避免了使用者隐私泄露的问题。

图 5 所示为对应于本发明第一实施例的基于语音识别的即时通信方法的即时通信系统，如图 5 所示，该实施例中的即时通信系统 500 包括如下模块：

语音信息接收模块 501，用于接收发送终端发送的语音信息；

文字信息生成模块 502，用于将该语音信息进行语音识别，生成文字信息；

第一发送模块 503，用于将该语音信息发送至接收终端；以及

第二发送模块 504，用于将该文字信息发送至接收终端。

图 6 所示为对应于本发明第二实施例的基于语音识别的即时通信方法的即时通信系统，如图 6 所示，在一优选实施例中，除了上述语音信息接

收模块 601、文字信息生成模块 602、第一发送模块 603、第二发送模块 604 之外，所述系统 600 还包括：

第三发送模块 605，用于将该文字信息发送至发送终端。

此外，所述系统 600 还包括：

- 5 信息收发模块 606，用于接收所述发送终端发出的编辑后文字信息，并发送至接收终端。

在一优选实施例中，所述系统还包括：

第一存储模块 607，将该文字信息储存于数据库。

在一优选实施例中，所述系统还包括：

- 10 第四发送模块 608，用于将辅助错误纠正信息发送至发送终端；以及
 信息收发模块 609，用于接收所述发送终端发出的编辑后文字信息，并发送至接收终端。

在一优选实施例中，所述系统还包括：

- 文字信息关联模块 610，用于将编辑后文字信息发送至数据库，并与
15 纠正前的所述文字信息关联。

在一优选实施例中，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词。

在一优选实施例中，所述指定字、词或句的词图和候选字词从所述数据库中获得。

- 20 在一优选实施例中，所述第一发送模块和所述第二发送模块同时执行，将所述将该语音信息和所述文字信息同时发送至接收终端。

图 7 所示为对应于本发明第三实施例的基于语音识别的即时通信方法的即时通信系统，如图 7 所示，该实施例中的即时通信系统 700 包括如下

模块:

语音信息录制发送模块 701, 用于录制语音信息并发送至服务器;

文字信息接收显示模块 702, 用于接收经过识别该语音信息生成的文字信息, 并显示该文字信息;

5 编辑模块 703, 用于在接收到纠正操作指令后, 进入编辑文字信息的界面;

显示发送模块 704, 用于显示编辑后文字信息, 并将编辑后文字信息发送至服务器。

在一优选实施例中, 所述系统还包括:

10 辅助修改信息接收模块 705, 用于接收服务器发送的辅助修改信息。

在一优选实施例中, 所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词, 所述候选字词显示在所述编辑文字信息的界面中。

在一优选实施例中, 所述编辑文字信息的界面包括输入界面。

15 在一优选实施例中, 所述系统还包括:

语音信息播放模块 706, 用于在接收到播放语音信息指令后, 播放语音信息。

在一优选实施例中, 所述播放语音信息指令通过使用者点击该文字信息生成。

20

图 8 所示为对应于本发明第四实施例的基于语音识别的即时通信方法的即时通信系统, 如图 8 所示, 该实施例中的即时通信系统 800 包括如下模块:

语音信息获取模块 801, 用于接收服务器发送的语音信息;

文字信息获取模块 802，用于接收服务器发送的识别该语音信息后生成的文字信息；

文字信息显示标记模块 803，用于显示并标记该文字信息。

在一优选实施例中，所述系统还包括：

- 5 标记信息获取模块 804，用于接收服务器发送的标记信息。

在一优选实施例中，所述文字信息获取模块和所述标记信息获取模块同时执行，将所述文字信息和所述标记信息同时获取。

在一优选实施例中，文字信息显示标记模块用于显示所述文字信息，利用所述标记信息对所述文字信息进行标记。

- 10 在一优选实施例中，所述系统还包括：

语音信息播放模块 805，用于当接收到使用者的播放该语音信息的指令，播放该语音信息。

在一优选实施例中，所述播放该语音信息的指令通过使用者点击该文字信息生成。

- 15 在一优选实施例中，所述系统还包括：

接收显示模块 806，用于接收服务器发送的编辑后文字信息，并显示该编辑后文字信息。

在一优选实施例中，所述编辑后文字信息以覆盖编辑前文字信息的方式显示。

- 20 对于装置实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

综上所述，本申请实施例提出的基于语音识别的即时通信方法和即时通信系统，至少具有以下优点：

(1) 本申请实施例提出的基于语音识别的即时通信方法和即时通信

系统中，通过语音识别功能，克服了接收终端获得信息的障碍，方便了使用者的使用，避免了隐私泄露的问题。

(2) 本申请实施例提出的基于语音识别的即时通信方法和即时通信系统中，通过错误修改功能，使得发送终端有机会纠正语音识别系统的错误；

(3) 本申请实施例提出的基于语音识别的即时通信方法和即时通信系统中，通过数据收集功能，获得真实识别错误数据以改进语音识别系统的性能。

(4) 本申请实施例提出的基于语音识别的即时通信方法和即时通信系统中，错误纠正的步骤方便发送终端进行错误纠正；

(5) 本申请实施例提出的基于语音识别的即时通信方法和即时通信系统中，信息标记的步骤方便接收终端辨识收到的信息是虚拟键盘输入还是语音信息；

(6) 本申请实施例提出的基于语音识别的即时通信方法和即时通信系统中，如果是语音信息，接收终端可以点选识别语音信息后生成的文字信息，对原始的语音信息进行回放。

本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

本领域内的技术人员应明白，本申请实施例的实施例可提供为方法、装置、或计算机程序产品。因此，本申请实施例可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请实施例可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

在一个典型的配置中，所述计算机设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。内存可能包括计算机可读介质

中的非永久性存储器, 随机存取存储器(RAM) 和/或非易失性内存等形式, 如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信号存储。信号可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括, 但不限于相变内存 (PRAM)、静态随机存取存储器 (SRAM)、动态随机存取存储器 (DRAM)、其他类型的随机存取存储器 (RAM)、只读存储器 (ROM)、电可擦除可编程只读存储器 (EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带, 磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质, 10 可用于存储可以被计算设备访问的信号。按照本文中的界定, 计算机可读介质不包括非持续性的电脑可读媒体(transitory media), 如调制的数据信号和载波。

本申请实施例是参照根据本申请实施例的方法、终端设备(系统)、和 15 计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理终端设备的处理器以产生一个机器, 使得通过计算机或其他可编程数据处理终端设备的 20 处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理终端设备以特定方式工作的计算机可读存储器中, 使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品, 该指令装置实现在流程图 25 图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理终端设备上, 使得在计算机或其他可编程终端设备上执行一系列操作步骤以产生计算机实现的处理, 从而在计算机或其他可编程终端设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多

个方框中指定的功能的步骤。

尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

以上对本申请所提供的一种基于语音识别的即时通信方法和即时通信系统，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，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申请的限制。

权 利 要 求 书

1. 一种基于语音识别的即时通信方法，其特征在于，包括：

接收发送终端发送的语音信息；

5 将该语音信息进行语音识别，生成文字信息；

将该语音信息发送至接收终端；以及

将该文字信息发送至接收终端。

2. 如权利要求 1 所述的即时通信方法，其特征在于，在将该语音信息进行语音识别，生成文字信息之后，所述方法还包括：

10 将该文字信息发送至发送终端。

3. 如权利要求 2 所述的即时通信方法，其特征在于，在将该文字信息发送至发送终端之后，所述方法还包括：

接收所述发送终端发出的编辑后文字信息，并发送至接收终端。

4. 如权利要求 3 所述的即时通信方法，其特征在于，在将该语音信息
15 信息进行语音识别，生成文字信息之后，并在接收所述发送终端发出的编辑后文字信息，并发送至接收终端之前，所述方法还包括：

将辅助错误纠正信息发送至发送终端，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词。

5. 如权利要求 2 所述的即时通信方法，其特征在于，在将该语音信息
20 息进行语音识别，生成文字信息之后，所述方法还包括：

将该文字信息储存于数据库；

在将该语音信息进行语音识别，生成文字信息之后，所述方法还包括：

将辅助错误纠正信息发送至发送终端；

接收所述发送终端发出的编辑后文字信息，并发送至接收终端；

在接收发送终端发出的编辑后文字信息，并发送至接收终端之后，所述方法还包括：

5 将编辑后文字信息发送至数据库，并与纠正前的所述文字信息关联。

6. 如权利要求 5 所述的即时通信方法，其特征在于，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词，所述指定字、词或句的词图和候选字词从所述数据库中获得。

10 7. 一种基于语音识别的即时通信方法，其特征在于，包括：

录制语音信息并发送至服务器；

接收经过识别该语音信息生成的文字信息，并显示该文字信息；

在接收到纠正操作指令后，进入编辑文字信息的界面；

显示编辑后文字信息，并将编辑后文字信息发送至服务器。

15 8. 如权利要求 7 所述的即时通信方法，其特征在于，在接收经过识别该语音信息生成的文字信息，并显示该文字信息之后，所述方法还包括：

20 接收服务器发送的辅助修改信息，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词，所述候选字词显示在所述编辑文字信息的界面中。

9. 如权利要求 7 所述的即时通信方法，其特征在于，在接收经过识别该语音信息生成的文字信息，并显示该文字信息之后，所述方法还包括：

在接收到播放语音信息指令后，播放语音信息。

10. 如权利要求 9 所述的即时通信方法，其特征在于，所述播放语音信息指令通过使用者点击该文字信息生成。

11. 一种基于语音识别的即时通信方法，其特征在于，包括：

5 接收服务器发送的语音信息；

接收服务器发送的识别该语音信息后生成的文字信息；

显示并标记该文字信息。

12. 如权利要求 11 所述的即时通信方法，其特征在于，所述方法还包括：

10 接收服务器发送的标记信息。

13. 如权利要求 12 所述的即时通信方法，其特征在于，所述显示并标记该文字信息的步骤包括：

显示所述文字信息，利用所述标记信息对所述文字信息进行标记。

14. 如权利要求 11 所述的即时通信方法，其特征在于，所述显示并标记该文字信息的步骤之后，所述方法还包括：

当接收到使用者的播放该语音信息的指令，播放该语音信息，所述播放该语音信息的指令通过使用者点击该文字信息生成。

15. 如权利要求 11 所述的即时通信方法，其特征在于，在显示并标记该文字信息的步骤之后，所述方法还包括：

20 接收服务器发送的编辑后文字信息，并显示该编辑后文字信息。

16. 如权利要求 15 所述的即时通信方法，其特征在于，所述编辑后文字信息以覆盖编辑前文字信息的方式显示。

17. 一种基于语音识别的即时通信系统，其特征在于，包括：

语音信息接收模块，用于接收发送终端发送的语音信息；

文字信息生成模块，用于将该语音信息进行语音识别，生成文字信息；

第一发送模块，用于将该语音信息发送至接收终端；以及

5 第二发送模块，用于将该文字信息发送至接收终端。

18. 如权利要求 17 所述的即时通信系统，其特征在于，所述系统还包括：

第三发送模块，用于将该文字信息发送至发送终端。

10 19. 如权利要求 18 所述的即时通信系统，其特征在于，所述系统还包括：

信息收发模块，用于接收所述发送终端发出的编辑后文字信息，并发送至接收终端。

20. 如权利要求 19 所述的即时通信系统，其特征在于，所述系统还包括：

15 第四发送模块，用于将辅助错误纠正信息发送至发送终端，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词。

21. 如权利要求 18 所述的即时通信系统，其特征在于，所述系统还包括：

20 第一存储模块，将该文字信息储存于数据库第四发送模块，用于将辅助错误纠正信息发送至发送终端；

信息收发模块，用于接收所述发送终端发出的编辑后文字信息，并发送至接收终端；

25 文字信息关联模块，用于将编辑后文字信息发送至数据库，并与纠正前的所述文字信息关联。

22. 如权利要求 21 所述的即时通信系统，其特征在于，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词，所述指定字、词或句的词图和候选字词从所述数据库中获得。

5 23. 一种基于语音识别的即时通信系统，其特征在于，包括：

语音信息录制发送模块，用于录制语音信息并发送至服务器；

文字信息接收显示模块，用于接收经过识别该语音信息生成的文字信息，并显示该文字信息；

10 编辑模块，用于在接收到纠正操作指令后，进入编辑文字信息的界面；

显示发送模块，用于显示编辑后文字信息，并将编辑后文字信息发送至服务器。

24. 如权利要求 23 所述的即时通信系统，其特征在于，所述系统还包括：

15 辅助修改信息接收模块，用于接收服务器发送的辅助修改信息，所述辅助错误纠正信息包括针对所述文字信息的指定字、词或句的词图和候选字词，所述候选字词显示在所述编辑文字信息的界面中。

25. 如权利要求 23 所述的即时通信系统，其特征在于，所述系统还包括：

20 语音信息播放模块，用于在接收到播放语音信息指令后，播放语音信息。

26. 如权利要求 25 所述的即时通信系统，其特征在于，所述播放语音信息指令通过使用者点击该文字信息生成。

25 27. 一种基于语音识别的即时通信系统，其特征在于，包括：

语音信息获取模块，用于接收服务器发送的语音信息；

文字信息获取模块，用于接收服务器发送的识别该语音信息后生成的文字信息；

文字信息显示标记模块，用于显示并标记该文字信息。

5 28. 如权利要求 27 所述的即时通信系统，其特征在于，所述系统还包括：

标记信息获取模块，用于接收服务器发送的标记信息。

29. 如权利要求 28 所述的即时通信系统，其特征在于，文字信息显示标记模块用于显示所述文字信息，利用所述标记信息对所述文字信息
10 进行标记。

30. 如权利要求 27 所述的即时通信系统，其特征在于，所述系统还包括：

15 语音信息播放模块，用于当接收到使用者的播放该语音信息的指令，播放该语音信息，所述播放该语音信息的指令通过使用者点击该文字信息生成。

31. 如权利要求 27 所述的即时通信系统，其特征在于，所述系统还包括：

接收显示模块，用于接收服务器发送的编辑后文字信息，并显示该编辑后文字信息。

20 32. 如权利要求 31 所述的即时通信系统，其特征在于，所述编辑后文字信息以覆盖编辑前文字信息的方式显示。

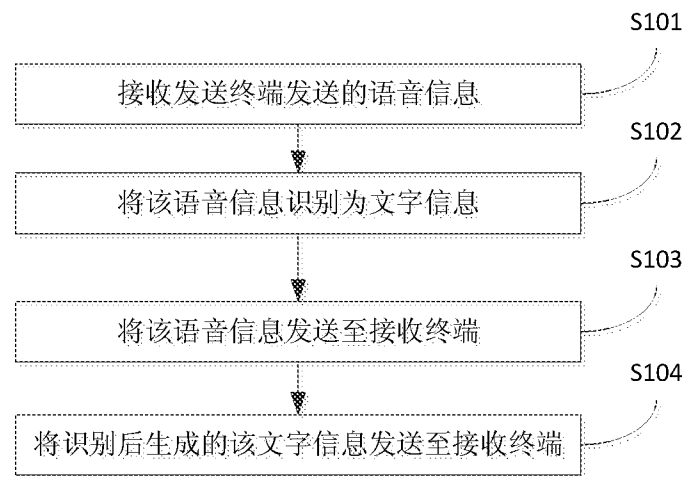


图1

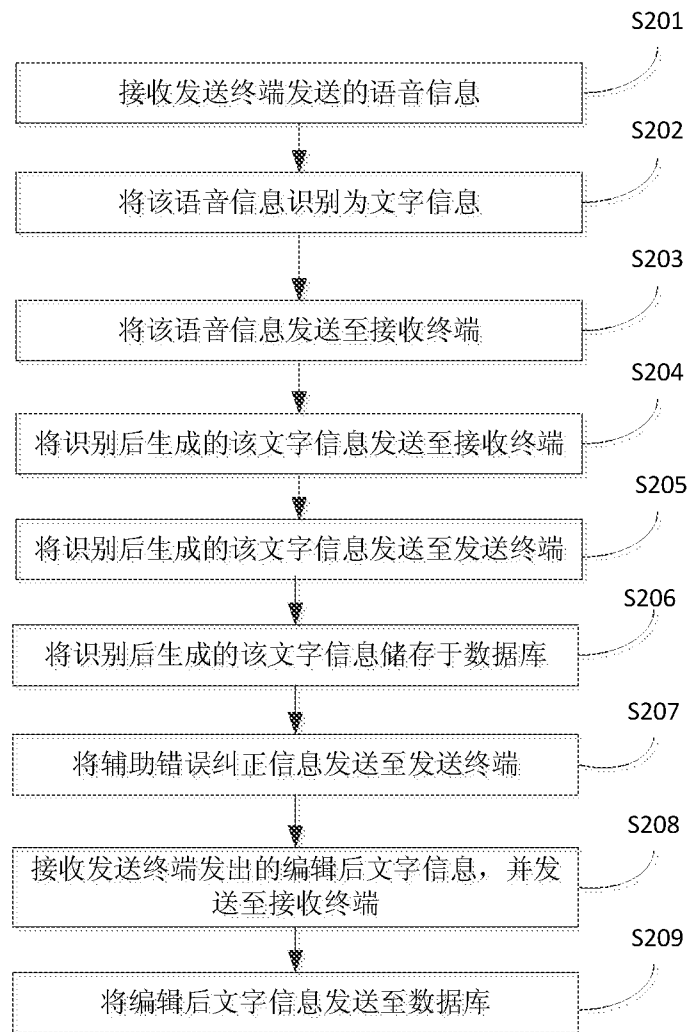


图2

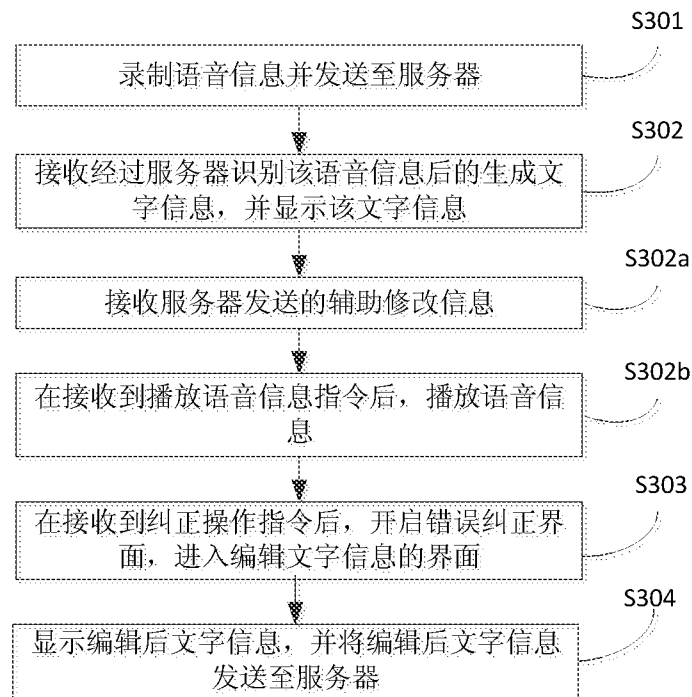


图3

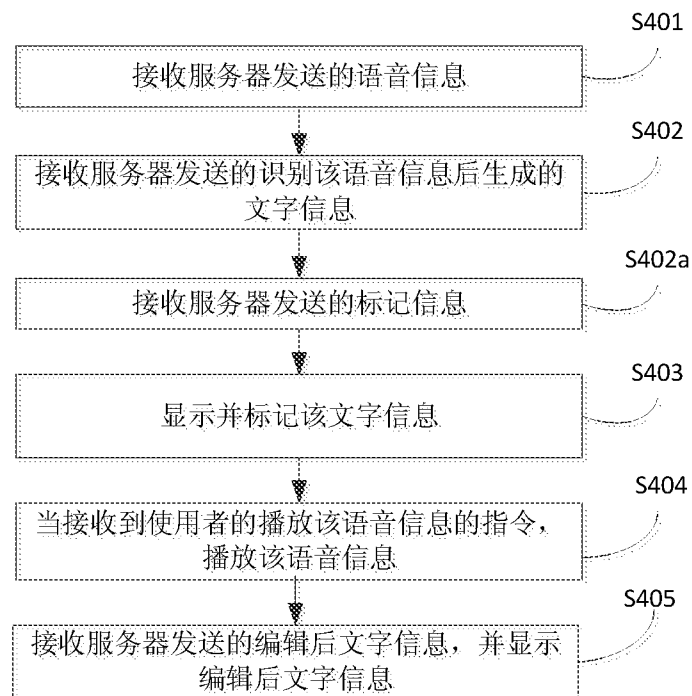


图4

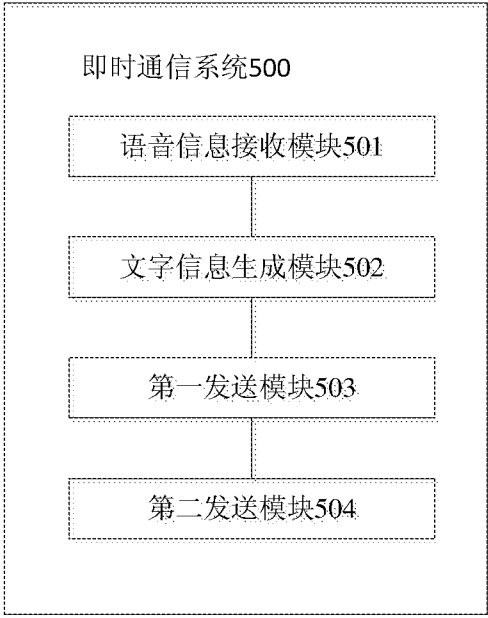


图5

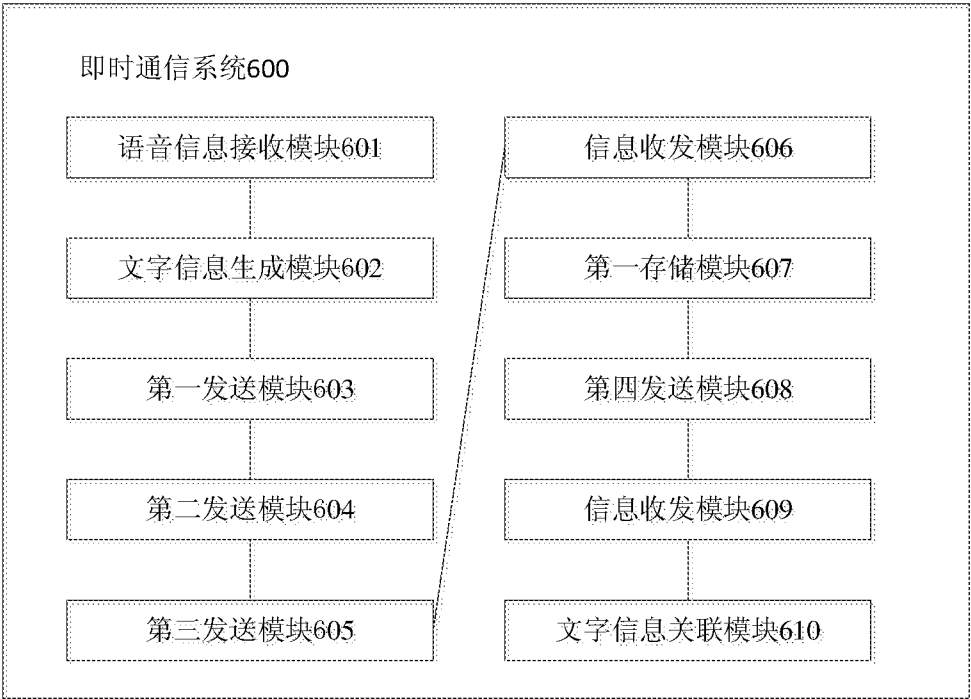


图6

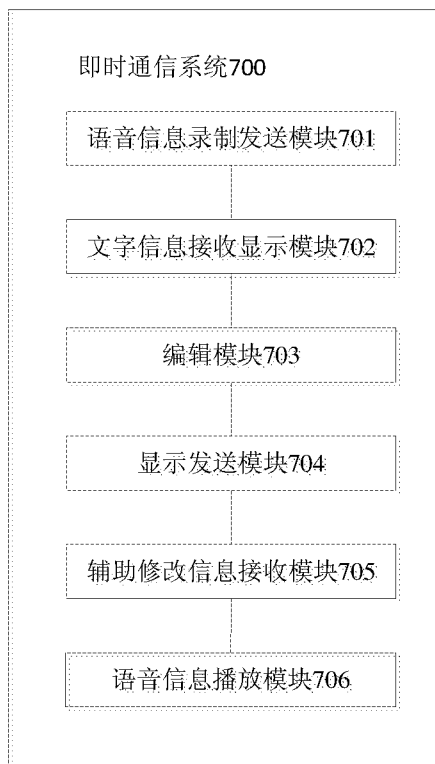


图7

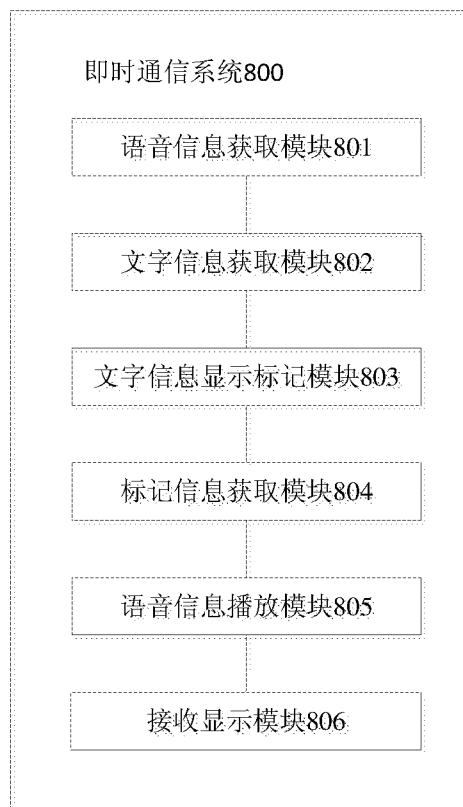


图8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/071382

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/26 (2006.01) i; H04L 12/58 (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; H04L

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)

CNABS, CNTXT, VEN, USTXT, WOTXT, EPTXT: instant messaging, or wechat or short message or qq, candidate, edit, IM, instant 3d communicat+, information, message, wechat, dialog+, voice, sound, speech, recogn+, text, letter, ASR, correct+, rectify+, revis+, amend+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105430208 A (XIAOMI TECHNOLOGY CO., LTD.), 23 March 2016 (23.03.2016), description, paragraphs [0128]-[0140], and figure 3	1, 17
X	CN 104732975 A (HUAWEI TECHNOLOGIES CO., LTD.), 24 June 2015 (24.06.2015), description, paragraphs [0096]-[0122], and figures 1-6	1-32
X	CN 103632670 A (QINGDAO INTWORK NETWORK TECHNOLOGY CO., LTD.), 12 March 2014 (12.03.2014), description, paragraphs [0024]-[0035], and figure 1	1, 17
A	CN 102946499 A (GUANGZHOU IFLY ZUNHONG INFORMATION TECHNOLOGY CO., LTD.), 27 February 2013 (27.02.2013), the whole document	1-32
A	CN 102710539 A (ZTE CORP.), 03 October 2012 (03.10.2012), the whole document	1-32
A	WO 2013184048 A1 (TELEFONAKTIEBOLAGET LM ERICSSON), 12 December 2013 (12.12.2013), the whole document	1-32

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 05 April 2017 (05.04.2017)	Date of mailing of the international search report 26 April 2017 (26.04.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer HAN, Limin Telephone No.: (86-10) 62085768

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/071382

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105430208 A	23 March 2016	None	
CN 104732975 A	24 June 2015	None	
CN 103632670 A	12 March 2014	None	
CN 102946499 A	27 February 2013	CN 102946499 B	14 October 2015
CN 102710539 A	03 October 2012	None	
WO 2013184048 A1	12 December 2013	EP 2856745 A1	08 April 2015
		US 9538348 B2	03 January 2017
		EP 2856745 A4	13 January 2016
		US 2015148084 A1	28 May 2015

国际检索报告

国际申请号

PCT/CN2017/071382

A. 主题的分类

G10L 15/26(2006.01)i; H04L 12/58(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; H04L

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))

CNABS, CNTXT, VEN, USTXT, WOTXT, EPTXT: 即时通信, 即时通讯, or 微信 or 短信 or qq, 信息, 语音, 识别, 文字, 文本, 候选, 修改, 修正, 纠错, 校正, 编辑, IM, instant 3d communicat+, information, message, wechat, dialog+, voice, sound, speech, recogn+, text, letter, ASR, correct+, rectify+, revis+, amend+

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
PX	CN 105430208 A (小米科技有限责任公司) 2016年 3月 23日 (2016 - 03 - 23) 说明书第[0128]-[0140]段, 图3	1, 17
X	CN 104732975 A (华为技术有限公司) 2015年 6月 24日 (2015 - 06 - 24) 说明书第[0096]-[0122]段, 图1-6	1-32
X	CN 103632670 A (青岛英特沃克网络科技有限公司) 2014年 3月 12日 (2014 - 03 - 12) 说明书第[0024]-[0035]段, 图1	1, 17
A	CN 102946499 A (广州市讯飞博鸿信息技术有限公司) 2013年 2月 27日 (2013 - 02 - 27) 全文	1-32
A	CN 102710539 A (中兴通讯股份有限公司) 2012年 10月 3日 (2012 - 10 - 03) 全文	1-32
A	WO 2013184048 A1 (TELEFONAKTIEBOLAGET LM ERICSSON) 2013年 12月 12日 (2013 - 12 - 12) 全文	1-32

☐ 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

2017年 4月 5日

国际检索报告邮寄日期

2017年 4月 26日

ISA/CN的名称和邮寄地址

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

受权官员

韩黎敏

传真号 (86-10)62019451

电话号码 (86-10)62085768

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/071382

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	105430208	A	2016年 3月 23日	无			
CN	104732975	A	2015年 6月 24日	无			
CN	103632670	A	2014年 3月 12日	无			
CN	102946499	A	2013年 2月 27日	CN	102946499	B	2015年 10月 14日
CN	102710539	A	2012年 10月 3日	无			
WO	2013184048	A1	2013年 12月 12日	EP	2856745	A1	2015年 4月 8日
				US	9538348	B2	2017年 1月 3日
				EP	2856745	A4	2016年 1月 13日
				US	2015148084	A1	2015年 5月 28日

表 PCT/ISA/210 (同族专利附件) (2009年7月)