



(12) 发明专利申请

(10) 申请公布号 CN 112201224 A

(43) 申请公布日 2021.01.08

(21) 申请号 202011073402.6

(22) 申请日 2020.10.09

(71) 申请人 北京分音塔科技有限公司
地址 100089 北京市海淀区中关村大街28-1号6层659

(72) 发明人 张明 关磊

(74) 专利代理机构 北京细软智谷知识产权代理有限公司 11471
代理人 岳凤羽

(51) Int. Cl.
G10L 15/00 (2013.01)
G10L 17/06 (2013.01)
G10L 13/033 (2013.01)
G06F 40/58 (2020.01)

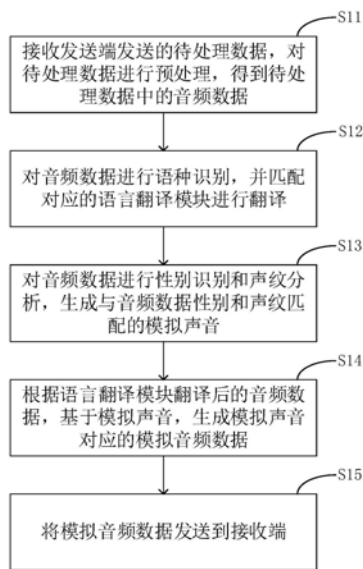
权利要求书2页 说明书6页 附图2页

(54) 发明名称

用于即时通话同声翻译的方法、设备及系统

(57) 摘要

本申请涉及一种用于即时通话同声翻译的方法、设备及系统,本申请中,在接收发送端发送的待处理数据后,由于待处理数据可能是音频数据,也可能是视频数据,所以需要对待处理数据进行预处理,得到待处理数据中的音频数据。对音频数据进行语种识别,并匹配对应的语言翻译模块进行翻译,本申请中,通过对音频数据进行性别识别和声纹分析,生成与音频数据性别和声纹匹配的模拟声音,通过模拟声音来模拟讲话者的声音向接收端发送语言翻译模块翻译后的音频数据,使收听者既能清楚明白的了解通话内容,也能对讲话者的声音和性别有一定的了解。



1. 一种用于即时通话同声翻译的方法,其特征在于,包括:
接收发送端发送的待处理数据,对所述待处理数据进行预处理,得到所述待处理数据中的音频数据;
对所述音频数据进行语种识别,并匹配对应的语言翻译模块进行翻译;
对所述音频数据进行性别识别和声纹分析,生成与所述音频数据性别和声纹匹配的模拟声音;
根据所述语言翻译模块翻译后的所述音频数据,基于所述模拟声音,生成所述模拟声音对应的模拟音频数据;
将所述模拟音频数据发送到接收端。
2. 根据权利要求1所述的方法,其特征在于,所述对所述待处理数据进行预处理,具体包括:
若所述待处理数据为视频数据,则对所述视频数据进行音频流分离,获取所述视频数据中的音频数据。
3. 根据权利要求2所述的方法,其特征在于,所述对所述待处理数据进行预处理,具体还包括:
对所述音频数据进行放大和滤波处理。
4. 根据权利要求1所述的方法,其特征在于,还包括:
对所述音频数据进行性别识别和声纹分析,在预先建立的声纹数据库中获取与所述音频数据性别和声纹匹配的模拟声音。
5. 根据权利要求1所述的方法,其特征在于,还包括:
在确定通话循环进行时,将每次通话循环中产生的音频数据进行训练,生成与所述音频数据性别和声纹匹配的优化模拟声音。
6. 根据权利要求5所述的方法,其特征在于,还包括:
在每次通话循环结束后,将当前模拟声音替换为上次通话循环中生成的与所述音频数据性别和声纹匹配的优化模拟声音。
7. 根据权利要求2所述的方法,其特征在于,还包括:
若所述待处理数据为视频数据,基于所述语言翻译模块翻译后的所述音频数据,生成所述模拟声音对应的文字数据;
将所述文字数据发送到所述接收端。
8. 一种用于即时通话同声翻译的设备,其特征在于,包括:处理器和存储器;
所述处理器与存储器通过通信总线相连接;
其中,所述处理器,用于调用并执行所述存储器中存储的程序;
所述存储器,用于存储程序,所述程序至少用于执行权利要求1-7任一项所述的用于即时通话同声翻译的方法。
9. 一种用于即时通话同声翻译的系统,其特征在于,包括:第一终端和第二终端;
所述第一终端和所述第二终端无线通信连接;
所述第一终端或所述第二终端内设置有如权利要求7所述的用于即时通话同声翻译的设备。
10. 一种用于即时通话同声翻译的系统,其特征在于,包括:第一终端,第二终端和云端

服务器；

所述云端服务器分别与所述第一终端和所述第二终端无线通信连接；
所述云端服务器内设置有如权利要求7所述的翻译设备。

用于即时通话同声翻译的方法、设备及系统

技术领域

[0001] 本申请涉及通信技术领域,尤其涉及一种用于即时通话同声翻译的方法、设备及系统。

背景技术

[0002] 目前的电话通话主要是原生音频传输,在跨国跨域中的沟通中,特别是商贸、旅游领域,经常会涉及多种语言的沟通,双方对话者难以完全明白对方的意思表达,一般只能通过一些基础的语言知识,结合各种词典书本、翻译软件等辅助进行有限的沟通,影响了沟通效率。

发明内容

[0003] 为至少在一定程度上克服相关技术中存在的问题,本申请提供一种用于即时通话同声翻译的方法、设备及系统。

[0004] 本申请的方案如下:

[0005] 根据本申请实施例的第一方面,提供一种用于即时通话同声翻译的方法,包括:

[0006] 接收发送端发送的待处理数据,对所述待处理数据进行预处理,得到所述待处理数据中的音频数据;

[0007] 对所述音频数据进行语种识别,并匹配对应的语言翻译模块进行翻译;

[0008] 对所述音频数据进行性别识别和声纹分析,生成与所述音频数据性别和声纹匹配的模拟声音;

[0009] 根据所述语言翻译模块翻译后的所述音频数据,基于所述模拟声音,生成所述模拟声音对应的模拟音频数据;

[0010] 将所述模拟音频数据发送到接收端。

[0011] 优选的,在本申请一种可实现的方式中,所述对所述待处理数据进行预处理,具体包括:

[0012] 若所述待处理数据为视频数据,则对所述视频数据进行音频流分离,获取所述视频数据中的音频数据。

[0013] 优选的,在本申请一种可实现的方式中,所述对所述待处理数据进行预处理,具体还包括:

[0014] 对所述音频数据进行放大和滤波处理。

[0015] 优选的,在本申请一种可实现的方式中,还包括:

[0016] 对所述音频数据进行性别识别和声纹分析,在预先建立的声纹数据库中获取与所述音频数据性别和声纹匹配的模拟声音。

[0017] 优选的,在本申请一种可实现的方式中,还包括:

[0018] 在确定通话循环进行时,将每次通话循环中产生的音频数据进行训练,生成与所述音频数据性别和声纹匹配的优化模拟声音。

- [0019] 优选的,在本申请一种可实现的方式中,还包括:
- [0020] 在每次通话循环结束后,将当前模拟声音替换为上次通话循环中生成的与所述音频数据性别和声纹匹配的优化模拟声音。
- [0021] 优选的,在本申请一种可实现的方式中,还包括:
- [0022] 若所述待处理数据为视频数据,基于所述语言翻译模块翻译后的所述音频数据,生成所述模拟声音对应的文字数据;
- [0023] 将所述文字数据发送到所述接收端。
- [0024] 根据本申请实施例的第二方面,提供一种用于即时通话同声翻译的设备,包括:处理器和存储器;
- [0025] 所述处理器与存储器通过通信总线相连接;
- [0026] 其中,所述处理器,用于调用并执行所述存储器中存储的程序;
- [0027] 所述存储器,用于存储程序,所述程序至少用于执行以上任一项所述的用于即时通话同声翻译的方法。
- [0028] 根据本申请实施例的第三方面,提供一种用于即时通话同声翻译的系统,包括:第一终端和第二终端;
- [0029] 所述第一终端和所述第二终端无线通信连接;
- [0030] 所述第一终端或所述第二终端内设置有如以上所述的用于即时通话同声翻译的设备。
- [0031] 根据本申请实施例的第四方面,提供一种用于即时通话同声翻译的系统,包括:第一终端,第二终端和云端服务器;
- [0032] 所述云端服务器分别与所述第一终端和所述第二终端无线通信连接;
- [0033] 所述云端服务器内设置有如以上所述的翻译设备。
- [0034] 本申请提供的技术方案可以包括以下有益效果:本申请中,将对话双方定义为发送端和接收端,即讲话者为发送端,收听者为接收端,双方在交换讲话者和收听者身份时,发送端和接收端也随之变化。本申请中,在接收发送端发送的待处理数据后,由于待处理数据可能是音频数据,也可能是视频数据,所以需要对待处理数据进行预处理,得到待处理数据中的音频数据。对音频数据进行语种识别,并匹配对应的语言翻译模块进行翻译,本申请中,通过对音频数据进行性别识别和声纹分析,生成与音频数据性别和声纹匹配的模拟声音,通过模拟声音来模拟讲话者的声音向接收端发送语言翻译模块翻译后的音频数据,使收听者既能清楚明白的了解通话内容,也能对讲话者的声音和性别有一定的了解。
- [0035] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本申请。

附图说明

- [0036] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。
- [0037] 图1是本申请一个实施例提供的一种用于即时通话同声翻译的方法的流程示意图;
- [0038] 图2是本申请一个实施例提供的一种用于即时通话同声翻译的设备的结果示意

图；

[0039] 图3是本申请一个实施例提供的一种用于即时通话同声翻译的系统的结果示意图；

[0040] 图4是本申请一个另实施例提供的一种用于即时通话同声翻译的系统的结果示意图。

[0041] 附图标记:处理器-21;存储器-22;第一终端-31;第二终端-32;云端服务器-33。

具体实施方式

[0042] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本申请相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本申请的一些方面相一致的装置和方法的例子。

[0043] 一种用于即时通话同声翻译的方法,参照图1,包括:

[0044] S11:接收发送端发送的待处理数据,对待处理数据进行预处理,得到待处理数据中的音频数据;

[0045] 本申请中,将对话双方定义为发送端和接收端,即讲话者为发送端,收听者为接收端,双方在交换讲话者和收听者身份时,发送端和接收端也随之变化。

[0046] 发送端和接收端的身份随着通话的进行随时互换。

[0047] 由于待处理数据可能是音频数据,也可能是视频数据,所以需要对待处理数据进行预处理,得到待处理数据中的音频数据,对待处理数据进行预处理,具体包括:

[0048] 若待处理数据为视频数据,则对视频数据进行音频流分离,获取视频数据中的音频数据。

[0049] 进一步的,对音频数据进行放大和滤波处理。

[0050] S12:对音频数据进行语种识别,并匹配对应的语言翻译模块进行翻译;

[0051] S13:对音频数据进行性别识别和声纹分析,生成与音频数据性别和声纹匹配的模拟声音;

[0052] 声纹,是用电声学仪器显示的携带言语信息的声波频谱。现代科学研究表明,声纹不仅具有特定性,而且有相对稳定性的特点。声纹也可以理解为讲话者的声线。

[0053] S14:根据语言翻译模块翻译后的音频数据,基于模拟声音,生成模拟声音对应的模拟音频数据;

[0054] S15:将模拟音频数据发送到接收端。

[0055] 本申请中,通过对音频数据进行性别识别和声纹分析,生成与音频数据性别和声纹匹配的模拟声音,通过模拟声音来模拟讲话者的声音向接收端发送语言翻译模块翻译后的音频数据,使收听者既能清楚明白的了解通话内容,也能对讲话者的声音和性别有一定的了解。

[0056] 一些实施例中的用于即时通话同声翻译的方法,还包括:

[0057] 对音频数据进行性别识别和声纹分析,在预先建立的声纹数据库中获取与音频数据性别和声纹匹配的模拟声音。

[0058] 本申请中可以支持私人定制模拟声音。用户可以通过定制自己的模拟声音,将定

制的模拟声音存储到预先建立的声纹数据库中进行保存,以便日后使用时更加方便。

[0059] 本申请中还可以将新用户通话过程中产生的新的模拟声音存储到声纹数据库中,以便该用户下次使用时直接进行匹配。

[0060] 一些实施例中的用于即时通话同声翻译的方法,还包括:

[0061] 在确定通话循环进行时,将每次通话循环中产生的音频数据进行训练,生成与音频数据性别和声纹匹配的优化模拟声音。

[0062] 将每一次发送端切换到接收端,再切换回发送端的过程作为一次通话循环。在确定通话循环进行时,将每次通话循环中产生的音频数据进行训练,生成与音频数据性别和声纹匹配的优化模拟声音,即在长时间通话中,通过不断的收取用户的语音数据,对模拟声音进行训练,使模拟声音与用户的匹配度越来越高。

[0063] 进一步的:

[0064] 在每次通话循环结束后,将当前模拟声音替换为上次通话循环中生成的与音频数据性别和声纹匹配的优化模拟声音。在长时间通话中,使模拟声音与用户的匹配度越来越高。

[0065] 一些实施例中的用于即时通话同声翻译的方法,还包括:

[0066] 若待处理数据为视频数据,基于语言翻译模块翻译后的音频数据,生成模拟声音对应的文字数据;

[0067] 将文字数据发送到接收端。

[0068] 在待处理数据为视频数据时,还基于语言翻译模块翻译后的音频数据,生成模拟声音对应的文字数据,将文字数据发送到接收端进行显示,方便接收端的用户更清楚的了解通话内容。

[0069] 一种用于即时通话同声翻译的设备,参照图2,包括:处理器21和存储器22;

[0070] 处理器21与存储器22通过通信总线相连接;

[0071] 其中,处理器21,用于调用并执行存储器22中存储的程序;

[0072] 存储器22,用于存储程序,程序至少用于执行以上任一实施例中的用于即时通话同声翻译的方法。

[0073] 一种用于即时通话同声翻译的虚拟装置,包括:

[0074] 接收模块,用于接收发送端发送的待处理数据;

[0075] 处理模块,用于对待处理数据进行预处理,得到待处理数据中的音频数据;

[0076] 语种识别模块,用于对音频数据进行语种识别

[0077] 翻译模块,用于对音频数据进行翻译;

[0078] 声纹性别分析模块,用于对音频数据进行性别识别和声纹分析;

[0079] 模拟声音生成模块,用于生成与音频数据性别和声纹匹配的模拟声音;

[0080] 模拟音频数据生成模块,用于根据语言翻译模块翻译后的音频数据,基于模拟声音,生成模拟声音对应的模拟音频数据;

[0081] 发送模块,用于将模拟音频数据发送到接收端。

[0082] 一种用于即时通话同声翻译的系统,参照图3,包括:第一终端31和第二终端32;

[0083] 第一终端31和第二终端32无线通信连接。

[0084] 第一终端31或第二终端32内设置有如以上实施例中的用于即时通话同声翻译的

设备。

[0085] 本实施例中的即时通话同声翻译的系统,可以将以上实施例中的用于即时通话同声翻译的设备仅设置在第一终端31或第二终端32即可实现第一终端31和第二终端32之间的通话,不需要额外架设其他设备。

[0086] 一种用于即时通话同声翻译的系统,参照图4,包括:第一终端31,第二终端32和云端服务器33;

[0087] 云端服务器33分别与第一终端31和第二终端32无线通信连接;

[0088] 云端服务器33内设置有如以上实施例中的翻译设备。

[0089] 本实施例中的即时通话同声翻译的系统,通过架设云端服务器33的方式,将用于即时通话同声翻译的设备从第一终端31或第二终端32上分离出来,减少了第一终端31或第二终端32的电能消耗和开发难度。

[0090] 可以理解的是,上述各实施例中相同或相似部分可以相互参考,在一些实施例中未详细说明的内容可以参见其他实施例中相同或相似的内容。

[0091] 需要说明的是,在本申请的描述中,术语“第一”、“第二”等仅用于描述目的,而不能理解为指示或暗示相对重要性。此外,在本申请的描述中,除非另有说明,“多个”的含义是指至少两个。

[0092] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本申请的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本申请的实施例所属技术领域的技术人员所理解。

[0093] 应当理解,本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如,如果用硬件来实现,和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或他们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0094] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0095] 此外,在本申请各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机可读取存储介质中。

[0096] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0097] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不

一定指的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

[0098] 尽管上面已经示出和描述了本申请的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本申请的限制,本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

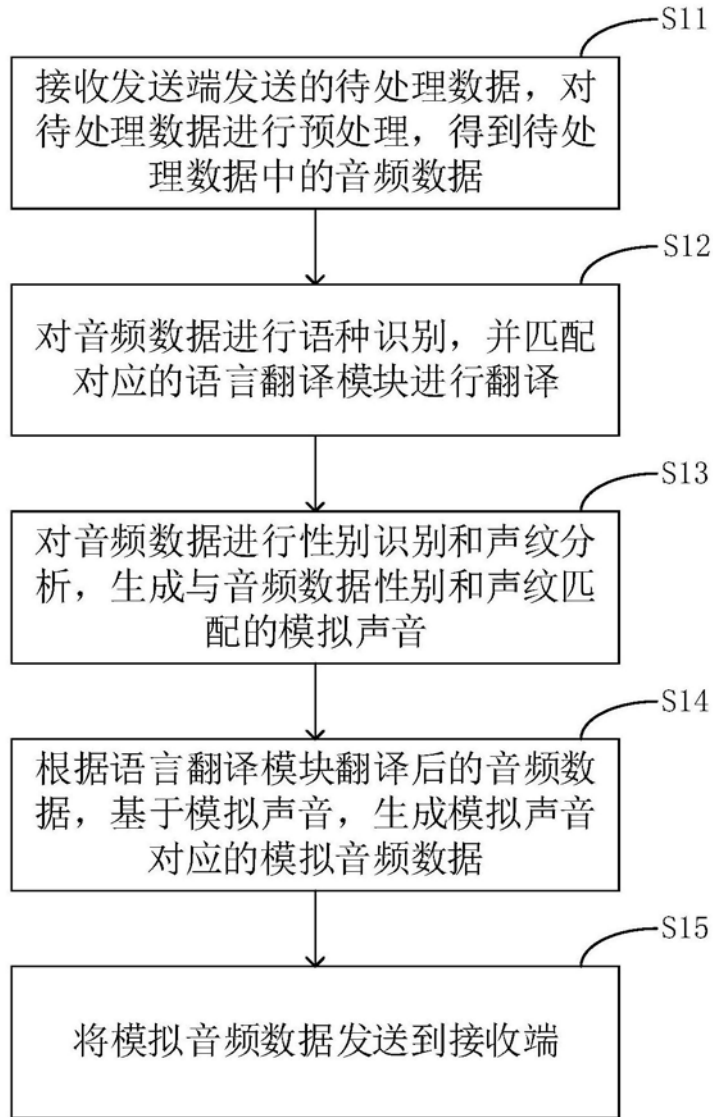


图1

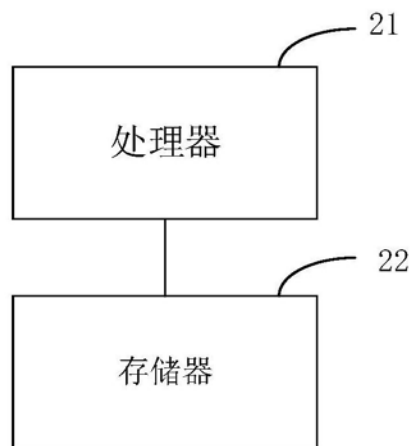


图2

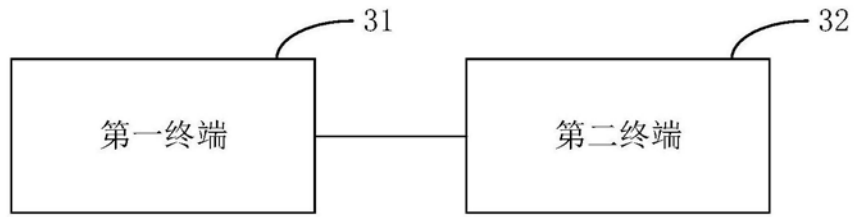


图3

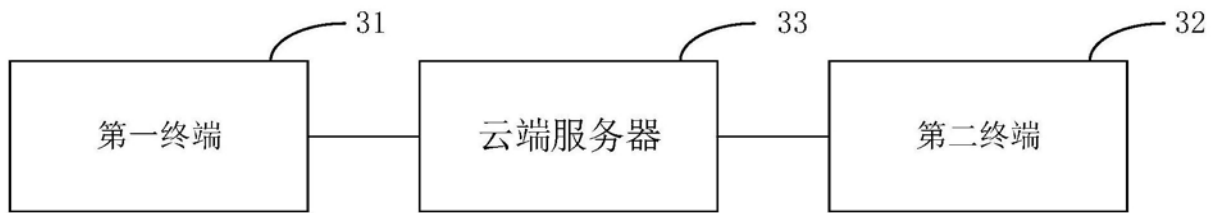


图4