

语音实时翻译方法、系统、设备以及存储介质

技术领域

本发明涉及在线翻译技术领域，具体地说，涉及一种语音实时翻译方法、系统、设备以及存储介质。

背景技术

随着国际性交流的增加，使用不同语种的语言沟通越来越频繁。交流过程中，因语言不同需要依赖交流双方的语言技能或要依靠人工口语翻译才能实现交流，给交流双方带来了不便。因此，通过使用翻译设备或穿戴设备配合耳机，并借助机器人语音，来实现实时在线翻译将成为最便捷的方式。

通话时，机器人语音翻译，负责将说话方的内容翻译为聆听方语种的语音进行播放。但翻译内容可能与说话方本意存在一定的误差，且说话内容和实时翻译语音会存在一定的时间差，而说话方不懂翻译语音内容，所以说话方无法及时获悉刚才说的某些内容被聆听方具体获悉的时间点和正确性，即是否正确的已经被听到，导致说话方无法确定自己说的话的传达时效性和正确性，需要说话方有意识地停顿，等待语音翻译全部完成后猜测聆听方是否已经听到以及理解正确。因此无法保证交流双方的沟通效率和正确性。

发明内容

针对现有技术中的问题，本发明的目的在于提供一种语音实时翻译方法、系统、设备以及存储介质，提高了跨语言翻译的准确率以及跨语言交流双方的沟通效率。

为实现上述目的，本发明提供了一种语音实时翻译方法，所述方法包括以下步骤：

- 对第一用户输入的第一语音信息进行语音识别，获得第一文本数据；
- 对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音；
- 基于所述第一文本数据，获得对应于第一语种类别的第一合成语音；

将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出。

可选地，所述对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音，包括：

对基于第一语音信息语音识别得到的第一文本数据进行翻译，获得对应于第二语种类别的第二文本数据；

基于所述第二文本数据，获得对应于第二语种类别的第二合成语音。

可选地，所述基于所述第一文本数据，获得对应于第一语种类别的第一合成语音，包括：

计算所述第二合成语音完成播放的预计耗时，得到第一预估时长；

基于所述第一预估时长，获得关于所述第二合成语音的第二时间戳信息；

基于所述第一文本数据进行语音合成，获得第一合成语音；并将所述第一合成语音对应的第一时间戳与所述第二时间戳信息进行同步。

可选地，所述方法还包括：

当接收到所述第一用户对所述第一合成语音的正反馈信息后，继续采集第一用户输入的下一段第一语音消息；

当接收到所述第一用户对所述第一合成语音的负反馈信息后，采集第一用户重新输入的所述第一语音信息，以及向第二用户输出关于所述第一用户重新输入的提示信息。

可选地，所述方法还包括：

分别基于所述第一文本数据和所述第二文本数据，提取各自的兴趣标签和对应的时间戳信息；

依据所述时间戳信息，将对应的兴趣标签分别标注于所述第一文本数据和所述第二文本数据。

可选地，在所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出之前，所述方法包括：

分别采集第一用户和第二用户的声纹信息；

基于所述声纹信息，分别识别得到第一用户和第二用户的身份信息；

基于所述第一用户和第二用户的身份信息，确定第一用户和第二用户之间的关

系类型；

基于第一用户和第二用户之间的关系类型，以及预设身份关系兴趣库，对所述兴趣标签进行筛选，将保留的兴趣标签作为目标兴趣标签；

所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出，包括：

对所述第一合成语音以及所述第二合成语音中与所述目标兴趣标签关联的语音进行语音提示。

可选地，所述对第一用户输入的第一语音信息进行语音识别，获得第一文本数据，包括：

获取关于所述第一语音信息的前文数据；

对第一用户输入的第一语音信息进行语音识别，获得初始文本数据；

利用所述前文数据对所述初始文本数据进行修正，得到第一文本数据。

可选地，在所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出之前，所述方法包括：

当检测到用于输出语音的一对耳机分别由不同用户佩戴时，控制所述耳机工作于第一状态；当检测到第一用户和第二用户各佩戴一对耳机时，控制所述耳机工作于第二状态；

其中，在第一状态下，两个耳机分别输出不同的语音，被一用户佩戴的一耳机作为该用户的语音输出通道；在第二状态下，一对耳机中的两个输出相同的语音，被一用户佩戴的一对耳机作为该用户的语音输出通道。

可选地，所述一对耳机设有 UWB 通信模块；在所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出之前，所述方法包括：

所述一对耳机基于 UWB 通信模块，实时侦测两个耳机之间的距离；

当两个耳机之间的距离大于第一预设阈值时，控制所述耳机工作于第一状态；当两个耳机之间的距离小于第一预设阈值时，控制所述耳机工作于第二状态。

可选地，所述方法还包括：

根据第一合成语音中未播放的音频帧，以及已采集生成第一语音信息且未转换为第一合成语音的音频帧，组合形成音频调节库；

基于所述音频调节库中的音频帧数量，调节第一合成语音的播放语速；

其中，所述第一合成语音对应的播放语速随着所述音频调节库中的音频帧数量的增大而增大，且随着所述音频调节库中的音频帧数量的减小而减小。

本发明还提供了一种语音实时翻译系统，用于实现上述语音实时翻译方法，所述系统包括：

第一文本数据生成模块，对第一用户输入的第一语音信息进行语音识别，获得第一文本数据；

第二合成语音生成模块，对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音；

第一合成语音生成模块，基于所述第一文本数据，获得对应于第一语种类别的第一合成语音；

合成语音播放模块，将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出。

本发明还提供了一种语音实时翻译设备，包括：

处理器；

存储器，其中存储有所述处理器的可执行程序；

其中，所述处理器配置为经由执行所述可执行程序来执行上述任意一项语音实时翻译方法的步骤。

本发明还提供了一种计算机可读存储介质，用于存储程序，所述程序被处理器执行时实现上述任意一项语音实时翻译方法的步骤。

本发明与现有技术相比，具有以下优点及突出性效果：

本发明提供的语音实时翻译方法、系统、设备以及存储介质对说话方的说话内容，基于语音识别得到的第一文本数据，合成第一合成语音，在向聆听方播放第二合成语音的同时，向说话方同步播放第一合成语音，使得识别语音过程中出现错误导致翻译内容时产生的误差，能够及时被说话方知晓，提高了跨语言翻译的准确率以及跨语言交流双方的沟通效率；使得在跨语言交流中沟通更加顺畅。

附图说明

通过阅读参照以下附图对非限制性实施例所作的详细描述，本发明的其它特征、目的和优点将会变得更明显。

图 1 为本发明一实施例公开的一种语音实时翻译方法的示意图；
图 2 为本发明一实施例公开的语音实时翻译方法所涉及的应用场景示意图；
图 3 为本发明一实施例公开的语音实时翻译方法所涉及的另一应用场景示意图；
图 4 为本发明一实施例公开的语音实时翻译方法所涉及的另一应用场景示意图；
图 5 为本发明另一实施例公开的一种语音实时翻译方法的示意图；
图 6 为本发明一实施例公开的一种语音实时翻译方法同步工作原理示意图；
图 7 为本发明另一实施例公开的一种语音实时翻译方法的示意图；
图 8 为本发明另一实施例公开的一种语音实时翻译方法的示意图；
图 9 为本发明另一实施例公开的一种语音实时翻译方法的示意图；
图 10 为本发明一实施例公开的一种语音实时翻译系统的结构示意图；
图 11 为本发明一实施例公开的一种语音实时翻译设备的结构示意图。。

具体实施方式

现在将参考附图更全面地描述示例实施方式。然而，示例实施方式能够以多种形式实施，且不应被理解为限于在此阐述的实施方式。相反，提供这些实施方式使得本发明将全面和完整，并将示例实施方式的构思全面地传达给本领域的技术人员。在图中相同的附图标记表示相同或类似的结构，因而将省略对它们的重复描述。

如图 1 所示，本发明一实施例公开了一种语音实时翻译方法，本实施例可适用于双方跨语言实时进行语音交流时进行语音识别和翻译的情况，该方法可以由在线实时语音翻译装置来执行，该装置可以由硬件和/或软件来实现，该装置可以配置在任意终端或网元中。通过在线实时语音翻译，可以为使用不同语言的用户双方实时提供翻译服务，使双方进行顺利沟通。该方法具体包括如下步骤：

S110，对第一用户输入的第一语音信息进行语音识别，获得第一文本数据。

参考图 2，在上述操作中，第一用户 21 和第二用户 22 进行跨语种的语言交流，第一用户 21 可以采用基于第一语种的第一语音输入第一语音信息，在线语音翻译装置可以对其进行语音识别，得到第一文本数据。对第一用户 21 发出的语音进行收音可以通过第一用户 21 佩戴的耳机实现。

本实施例将第一用户 21 作为跨语言沟通刚开始时的说话方，将第二用户 22 作为沟通刚开始时的聆听方，因此在沟通刚开始时，该步骤采集的是第一用户 21 输入的第一语音信息，然后对该用户的语音信息进行识别。

由于沟通是双向的，所以在后续持续的沟通时，该步骤也将采集第二用户 22 输入的第二语音信息，然后对第二语音信息进行识别。其中，上述在线实时语音翻译装置可以为图 2 中的翻译设备 23，也可以为服务器。本申请对此不作限制。

另一方面，当在线实时语音翻译装置为翻译设备时，可以不限于一台翻译设备。比如参考图 3，第一用户 21 和第二用户 22 可以各自使用一台翻译设备，比如第一用户 21 使用第一翻译设备 24，第二用户 22 使用第二翻译设备 25。第一翻译设备 24 和第二翻译设备 25 之间进行数据传输和通信。不同语种之间的文本翻译可以由第一翻译设备 24 或者第二翻译设备 25 完成，第一翻译设备 24 和第二翻译设备 25 还用于完成各自对应用户的语音合成。

需要说明的是，上述图 2 和图 3 示出的是面对面实时翻译场景。本实施例还可以应用于远距离在线翻译场景。图 4 示出的是一种远距离在线翻译场景示意图。该图示中，第一用户 21 和第二用户 22 分别基于笔记本电脑进行在线翻译。

S120，对上述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音。具体而言，在线语音翻译一般涉及两个环节：第一个环节是进行语音识别，即将第一用户输入的第一语种语音识别为文字信息；第二个环节是将文字信息基于翻译语料库进行翻译，再生成第二语种的语音信息或文字信息，提供给第二用户。因此，在一些实施例中，步骤 S120 包括：

S121，对上述第一文本数据进行翻译，获得对应于第二语种类别的第二文本数据。

S122，基于上述第二文本数据，获得对应于第二语种类别的第二合成语音。上述步骤即为将第一语种的第一文本数据翻译为第二语种的第二文本数据，然后基于第二文本数据进行语音合成。

S130，基于上述第一文本数据，获得对应于第一语种类别的第一合成语音。其中，上述第一合成语音是用于让此时作为说话方的第一用户，能够通过本方的机器人获知翻译语音的播放进度，以及及时发现语音识别过程中出现的错误，从而可以让说话方准确的获得翻译内容的正确性和传达给对方的时间，从而让语言不同的

交流方能实现及时纠正错误和提升互动性。

为了让说话方的机器人与聆听方的机器人播放内容保持相同进度,就需要将第一合成语音和第二合成语音进行时间戳同步。于是,在一些实施例中,参考图 5,步骤 S130 包括:

S131, 计算上述第二合成语音完成播放的预计耗时, 得到第一预估时长。

S132, 基于上述第一预估时长, 获得关于上述第二合成语音的第二时间戳信息。以及

S133, 基于上述第一文本数据进行语音合成, 获得第一合成语音; 并将上述第一合成语音对应的第一时间戳与上述第二时间戳信息进行同步。

该实施例利用上述步骤对第二语种语音合成耗时进行预估, 从而基于此进行时间戳同步。具体实施时, 示例性地, 可以将两个合成语音各自划分为多个音频段落, 然后将第一合成语音的音频段落与第二合成语音对应的音频段落进行时间戳同步, 由此实现两个合成语音的时间戳同步。

这样本申请即实现在本方语种内容的合成语音和对方语种内容的合成语音之间建立同步机制, 实现两方的播放内容保持相同进度。

S140, 将上述第一合成语音向第一用户输出, 以及同步将上述第二合成语音向第二用户输出。

本实施例将两个合成语音同步播放, 利于保证播放内容保持相同进度, 可以让说话方准确的获得翻译内容的正确性和传达给对方的时间, 从而让语言不同的交流方能实现及时纠正错误和提升互动性。

并且, 同步播放时, 说话方可以持续继续说话, 也可以暂时不说话, 聆听该第一合成语音。若说话方继续说话, 那么将该第一合成语音的播放音量设置为不高于说话方发音音量的第四预设阈值, 这样可以不干扰说话方的正常表达, 又能让说话方通过耳机听见所说内容。

由于实时在线翻译分为现场面对面和非现场远距离翻译两种场景, 因此具体实施时, 在现场面对面翻译场景下, 两个用户可以分别佩戴一副耳机中的一个, 即共享一副耳机。也可以分别佩戴一副耳机。当各戴一个耳机时, 可以将第一合成语音通过左声道耳机向第一用户输出, 同步将上述第二合成语音通过右声道耳机向第二用户输出。在非现场远距离翻译场景下, 两个用户分别佩戴一副耳机。

图6为同步播放合成语音的原理示意图。参考图6，A用户为第一用户，A的语音人是指用于向A用户播放第一合成语音的机器人。B用户为第二用户，B的语音人是指用于向B用户播放第二合成语音的机器人。

S_x（包括S₁、S₂与S₃）为第一用户的原声。S₁、S₂与S₃可以分别表示A用户发出的第一段、第二段与第三段语音。S_x'为A的语音人根据S_x识别为第一语种的内容后合成的机器语音，即识别后的合成的语音。S₁'、S₂'与S₃'分别表示向A用户播放的，且分别与第一段语音S₁对应的第一语种合成语音、与第二段语音S₂对应的第一语种合成语音以及与第三段语音S₃对应的第一语种合成语音。

TS_x'为B的语音人根据S_x识别内容翻译成第二语种的文本后，合成的机器语音，即翻译后的语音。那么，TS₁'、TS₂'以及TS₃'分别表示向B用户播放的，且分别与第一段语音S₁对应的第二语种合成语音、与第二段语音S₂对应的第二语种合成语音以及与第三段语音S₃对应的第二语种合成语音。

在本申请的另一实施例中，公开了另一种语音实时翻译方法。参考图7，该方法在上述图1对应实施例的基础上，还包括步骤：

S150，判断是否接收到第一用户对上述第一合成语音的正反馈信息。若是，则执行步骤S160。否则执行步骤S170。

S160，当接收到第一用户对上述第一合成语音的正反馈信息后，继续采集第一用户输入的下一段第一语音消息。

S170，采集第一用户重新输入的上述第一语音信息，以及向第二用户输出关于上述第一用户重新输入的提示信息。

具体而言，也即当第一用户确认翻译准确之后，继续采集第一用户输入的下一段第一语音消息。当第一用户确认翻译错误之后，提示第一用户重新表达内容并重新采集，并告知第二用户刚才播放的第二合成语音存在错误，将会重新播放。

关于判断翻译是否准确的实现方式，在其他实施例中，还可以通过当第一合成语音完成播放后的一预设时间段内，未接收到第一用户的预设发声语音时，就确定翻译准确。否则，就确定翻译存在错误。从而让语言不同的交流方能实现及时纠正错误和提升互动性。

在一些实施例中，在上述图1对应实施例的基础上，步骤S110包括：

对第一语音信息进行声纹侦测，仅对属于第一用户发声的对应音频段落进行语

音识别，得到第一文本数据。也即对不属于第一用户发声的对应音频段落，不进行语音识别。

步骤 S120 包括：对第一语音信息进行声纹侦测，仅对属于第一用户发声的对应音频段落进行翻译，得到第二合成语音。也即，翻译过程中，当侦测到不属于第一用户发声的音频段落时，暂停翻译。当侦测到属于第一用户发声的音频段落时，继续进行翻译，直至完成翻译。

这样利于保证跨语言沟通时实时翻译的准确性。

在一些实施例中，在上述图 1 对应实施例的基础上，在步骤 S140 之前，该方法还包括：

分别实时采集第一用户和第二用户的环境语音数据。

对环境语音数据进行声纹侦测。

当侦测到对应第一用户的环境语音数据存在其他人的声纹信息时，向第一用户发送第一预设提示信息。该第一预设提示信息用于语音提醒第一用户周围有外人，注意保护隐私。

当侦测到对应第二用户的环境语音数据存在其他人的声纹信息时，暂停对第一语音信息进行翻译，并向第二用户发送第二预设提示信息。

直至在第二用户侧未侦测到其他人的声纹信息时，继续对第一语音信息进行翻译。

这样利于对第一用户和第二用户的隐私保护。示例性地，上述步骤可以位于步骤 S130 和步骤 S140 之间。

在一些实施例中，在上述图 1 对应实施例的基础上，该方法在步骤 S140 之前，还包括步骤：

分别基于上述第一文本数据和上述第二文本数据，提取各自的兴趣标签和对应的时间戳信息。

依据上述时间戳信息，将对应的兴趣标签分别标注于上述第一文本数据和上述第二文本数据。

示例性地，上述步骤可以位于步骤 S130 和步骤 S140 之间。

步骤 S140 包括：在播放第一合成语音以及第二合成语音时，对兴趣标签关联的语音进行语音提示。

其中，上述兴趣标签和时间戳具有一一对应的关系。利用该兴趣标签可以方便用户在回听语音时，可以快速找到重点内容。上述兴趣标签可以以文本或者语音的形式进行展示。

本申请对反复提及的内容或说话音调或语义分析内容重点，敏感隐私，又或聆听方临时无法专注倾听但又想后续可回顾内容时，主动启动兴趣标签标注功能，记录时间戳和对应的兴趣标签并存储。当用户回看兴趣标签时，可以展示出通话翻译的兴趣点，可跳转查看兴趣点前后内容，或重听兴趣点前后的交谈历史，提升用户体验。还可以依靠兴趣标签使得交流两方关注到重点内容，避免被错误理解或者忽视，利于提高跨语言交流体验。

在一些实施例中，在上述实施例的基础上，该方法在步骤 S140 之前，还包括步骤：

分别采集第一用户和第二用户的声纹信息。

基于所述声纹信息，分别识别得到第一用户和第二用户的身份信息。

基于所述第一用户和第二用户的身份信息，确定第一用户和第二用户之间的关系类型。

基于第一用户和第二用户之间的关系类型，以及预设身份关系兴趣库，对上述提取的兴趣标签进行筛选，保留下来的作为目标兴趣标签。

示例性地，上述步骤可以位于步骤 S130 和步骤 S140 之间。

步骤 S140 包括：在播放第一合成语音以及第二合成语音时，对目标兴趣标签关联的语音进行语音提示。

其中，第一用户和第二用户之间的关系类型的确定，可以依据预设身份关系数据库实现。

在一些实施例中，在上述图 1 对应实施例的基础上，步骤 S110 包括：

S111，获取关于上述第一语音信息的前文数据。

S112，对第一用户输入的第一语音信息进行语音识别，获得初始文本数据。

S113，利用前文数据对初始文本数据进行修正，得到第一文本数据。

具体而言，获取用户在发声第一语音信息之前发声的其他数据，基于这些数据对识别的文本进行修正，利于提高翻译的准确性，进而提高跨语言交流时的顺畅度。比如，当前文数据对同一语音翻译为第一词语，而该步骤将该语音翻译为第二词语

时，此时就将第二词语修正为第一词语。

在一些实施例中，在上述图 1 对应实施例的基础上，参考图 8，在步骤 S140 之前，还包括步骤：

S180，当检测到用于输出语音的一对耳机分别由不同用户佩戴时，控制上述耳机工作于第一状态；当检测到第一用户和第二用户各佩戴一对耳机时，控制上述耳机工作于第二状态。

其中，在第一状态下，两个耳机分别输出不同的语音，被用户佩戴的一耳机作为该用户的语音输出通道。在第二状态下，一对耳机中的两个输出相同的语音，被用户佩戴的一对耳机作为该用户的语音输出通道。

示例性地，上述步骤可以位于步骤 S130 和步骤 S140 之间。

步骤 S140 替换为步骤 S141：将第一合成语音基于第一用户佩戴的耳机向第一用户输出，同步将第二合成语音基于第二用户佩戴的耳机向第二用户输出。

具体而言，比如一副耳机中的两个耳机分别被两个用户佩戴时，即第一用户佩戴其中一个，第二用户佩戴另一个，此时可以控制该副耳机工作于面对面翻译状态。当每个用户各佩戴一副耳机时，控制该副耳机工作于远距离翻译状态。面对面翻译状态时，需要将对应该用户的合成语音，编码为该用户佩戴的耳机对应的声道音频。比如，当第一用户佩戴一副耳机中的左声道耳机时，就将第一合成语音编码为左声道音频。相应地，将第二合成语音编码为右声道音频。

将第一用户佩戴的耳机作为第一语音输出通道，将第二用户佩戴的耳机作为第二语音输出通道。将第一合成语音基于第一语音输出通道向第一用户输出，同步将第二合成语音基于第二语音输出通道向第二用户输出。

在一些实施例中，在上述实施例的基础上，上述耳机设有 UWB(Ultra Wide Band, 超宽带无线载波通信)通信模块。步骤 S180 包括：

上述一对耳机基于 UWB 通信模块，实时侦测两个耳机之间的距离。

当两个耳机之间的距离大于第一预设阈值时，控制上述耳机工作于第一状态；当两个耳机之间的距离小于第一预设阈值时，控制上述耳机工作于第二状态。

具体而言，当两个耳机之间的距离大于第一预设阈值时，说明一副耳机中的两个耳机分别被两个用户佩戴。当两个耳机之间的距离小于第一预设阈值时，说明一副耳机由同一用户佩戴。

示例性地，上述第一预设阈值可以为 25cm。在其他实施例中，耳机之间的距离也可以通过耳机中内设的超声波信号模块实现侦测，本申请对此不作限制。

在一些实施例中，在上述图 1 对应实施例的基础上，参考图 9，该方法包括步骤：

S190，根据第一合成语音中未播放的音频帧，以及已采集生成第一语音信息且未转换为第一合成语音的音频帧，组合形成音频调节库。

S200，基于上述音频调节库中的音频帧数量，调节第一合成语音的播放语速。其中，上述第一合成语音对应的播放语速随着上述音频调节库中的音频帧数量的增大而增大，且随着上述音频调节库中的音频帧数量的减小而减小。在一些实施例中，也可以同步对第二合成语音的播放语速进行调节。

这样可以避免比如当用户已经说到第 S6 个音频段落时，耳机里还在回放第 S3 个音频段落，那样将不利于说话方及时发现语音识别过程中出现的错误，从而让语言不同的交流方能实现及时纠正错误。

在一些实施例中，在上述图 1 对应实施例的基础上，在上述步骤 S110 之前，该方法还包括：

当检测到第一用户仅佩戴一副耳机中的一个，以及该对耳机中两个耳机的距离大于第一预设阈值时，采集第一用户所处环境的环境声数据，并从上述环境声数据中提取出多个语音数据。

确定上述多个语音数据对应的语种类别，作为第二语种类别。

基于该副耳机中的另一个向外界环境中播放提示信息；上述提示信息用于提示第二用户佩戴该副耳机中的另一个。

上述步骤 S140 包括：

在检测到第二用户佩戴第二个耳机后，将第二合成语音通过上述第二个耳机向第二用户输出。

这样可以便于在第一用户处于一个陌生语种的环境中，比如去另一个国家旅游时，与第二用户需要面对面翻译的场景下，帮助与第二用户快速建立沟通，利于提高跨语言沟通的顺畅度。

需要说明的是，本申请中公开的上述所有实施例可以进行自由组合，组合后得到的技术方案也在本申请的保护范围之内。

如图 10 所示，本发明一实施例还公开了一种语音实时翻译系统 10，该系统包括：

第一文本数据生成模块 101，对第一用户输入的第一语音信息进行语音识别，获得第一文本数据。

第二合成语音生成模块 102，对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音。

第一合成语音生成模块 103，基于所述第一文本数据，获得对应于第一语种类别的第一合成语音。

合成语音播放模块 104，将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出。

可以理解的是，本发明的语音实时翻译系统还包括其他支持语音实时翻译系统运行的现有功能模块。图 10 显示的语音实时翻译系统仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

本实施例中的语音实时翻译系统用于实现上述的语音实时翻译的方法，因此对于语音实时翻译系统的具体实施步骤可以参照上述对语音实时翻译的方法的描述，此处不再赘述。

本发明一实施例还公开了一种语音实时翻译设备，包括处理器和存储器，其中存储器存储有所述处理器的可执行程序；处理器配置为经由执行可执行程序来执行上述语音实时翻译方法中的步骤。图 11 是本发明公开的语音实时翻译设备的结构示意图。下面参照图 11 来描述根据本发明的这种实施方式的电子设备 600。图 11 显示的电子设备 600 仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

如图 11 所示，电子设备 600 以通用计算设备的形式表现。电子设备 600 的组件可以包括但不限于：至少一个处理单元 610、至少一个存储单元 620、连接不同平台组件（包括存储单元 620 和处理单元 610）的总线 630、显示单元 640 等。

其中，存储单元存储有程序代码，程序代码可以被处理单元 610 执行，使得处理单元 610 执行本说明书上述语音实时翻译方法部分中描述的根据本发明各种示例性实施方式的步骤。例如，处理单元 610 可以执行如图 1 中所示的步骤。

存储单元 620 可以包括易失性存储单元形式的可读介质，例如随机存取存储单

元 (RAM) 6201 和/或高速缓存存储单元 6202, 还可以进一步包括只读存储单元 (ROM) 6203。

存储单元 620 还可以包括具有一组 (至少一个) 程序模块 6205 的程序/实用工具 6204, 这样的程序模块 6205 包括但不限于: 操作系统、一个或者多个应用程序、其它程序模块以及程序数据, 这些示例中的每一个或某种组合中可能包括网络环境的实现。

总线 630 可以为表示几类总线结构中的一种或多种, 包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

电子设备 600 也可以与一个或多个外部设备 700 (例如键盘、指向设备、蓝牙设备等) 通信, 还可与一个或者多个使得用户能与该电子设备 600 交互的设备通信, 和/或与使得该电子设备 600 能与一个或多个其它计算设备进行通信的任何设备 (例如路由器、调制解调器等等) 通信。这种通信可以通过输入/输出 (I/O) 接口 650 进行。并且, 电子设备 600 还可以通过网络适配器 660 与一个或者多个网络 (例如局域网 (LAN), 广域网 (WAN) 和/或公共网络, 例如因特网) 通信。网络适配器 660 可以通过总线 630 与电子设备 600 的其它模块通信。应当明白, 尽管图中未示出, 可以结合电子设备 600 使用其它硬件和/或软件模块, 包括但不限于: 微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储平台等。

本发明还公开了一种计算机可读存储介质, 用于存储程序, 所述程序被执行时实现上述语音实时翻译方法中的步骤。在一些可能的实施方式中, 本发明的各个方面还可以实现为一种程序产品的形式, 其包括程序代码, 当程序产品在终端设备上运行时, 程序代码用于使终端设备执行本说明书上述语音实时翻译方法中描述的根据本发明各种示例性实施方式的步骤。

如上所示, 该实施例的计算机可读存储介质的程序在执行时, 对说话方的说话内容, 基于语音识别得到的第一文本数据, 合成第一合成语音, 在向聆听方播放第二合成语音的同时, 向说话方同步播放第一合成语音, 使得识别语音过程中出现错误导致翻译内容时产生的误差, 能够及时被说话方知晓, 提高了跨语言翻译的准确率以及跨语言交流双方的沟通效率。

本发明一实施例公开了一种计算机可读存储介质。该存储介质是实现上述方法的程序产品，其可以采用便携式紧凑盘只读存储器（CD-ROM）并包括程序代码，并可以在终端设备，例如个人电脑上运行。然而，本发明的程序产品不限于此，在本文件中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM或闪存）、光纤、便携式紧凑盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。

计算机可读存储介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读存储介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。可读存储介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF 等等，或者上述的任意合适的组合。

可以以一种或多种程序设计语言的任意组合来编写用于执行本发明操作的程序代码，程序设计语言包括面向对象的程序设计语言—诸如 Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中，远程计算设备可以通过任意种类的网络，包括局域网（LAN）或广域网（WAN），连接到用户计算设备，或者，可以连接到外部计算设备（例如利用因特网服务提供商来通过因特网连接）。

本发明实施例提供的语音实时翻译方法、系统、设备以及存储介质对说话方的说话内容，基于语音识别得到的第一文本数据，合成第一合成语音，在向聆听方播

放第二合成语音的同时，向说话方同步播放第一合成语音，使得识别语音过程中出现错误导致翻译内容时产生的误差，能够及时被说话方知晓，提高了跨语言翻译的准确率以及跨语言交流双方的沟通效率；使得在跨语言交流中沟通更加顺畅。

以上内容是结合具体的优选实施方式对本发明所作的进一步详细说明，不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干简单推演或替换，都应当视为属于本发明的保护范围。

权利要求书

- 1.一种语音实时翻译方法，其特征在于，包括以下步骤：
对第一用户输入的第一语音信息进行语音识别，获得第一文本数据；
对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音；
基于所述第一文本数据，获得对应于第一语种类别的第一合成语音；
将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出。
- 2.如权利要求1所述的语音实时翻译方法，其特征在于，所述对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音，包括：
对基于第一语音信息语音识别得到的第一文本数据进行翻译，获得对应于第二语种类别的第二文本数据；
基于所述第二文本数据，获得对应于第二语种类别的第二合成语音。
- 3.如权利要求2所述的语音实时翻译方法，其特征在于，所述基于所述第一文本数据，获得对应于第一语种类别的第一合成语音，包括：
计算所述第二合成语音完成播放的预计耗时，得到第一预估时长；
基于所述第一预估时长，获得关于所述第二合成语音的第二时间戳信息；
基于所述第一文本数据进行语音合成，获得第一合成语音；并将所述第一合成语音对应的第一时间戳与所述第二时间戳信息进行同步。
- 4.如权利要求1所述的语音实时翻译方法，其特征在于，所述方法还包括：
当接收到所述第一用户对所述第一合成语音的正反馈信息后，继续采集第一用户输入的下一段第一语音消息；
当接收到所述第一用户对所述第一合成语音的负反馈信息后，采集第一用户重新输入的所述第一语音信息，以及向第二用户输出关于所述第一用户重新输入的提示信息。
- 5.如权利要求2所述的语音实时翻译方法，其特征在于，所述方法还包括：
分别基于所述第一文本数据和所述第二文本数据，提取各自的兴趣标签和对应的时间戳信息；
依据所述时间戳信息，将对应的兴趣标签分别标注于所述第一文本数据和所述

第二文本数据。

6.如权利要求5所述的语音实时翻译方法，其特征在于，在所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出之前，所述方法包括：

分别采集第一用户和第二用户的声纹信息；

基于所述声纹信息，分别识别得到第一用户和第二用户的身份信息；

基于所述第一用户和第二用户的身份信息，确定第一用户和第二用户之间的关系类型；

基于第一用户和第二用户之间的关系类型，以及预设身份关系兴趣库，对所述兴趣标签进行筛选，将保留的兴趣标签作为目标兴趣标签；

所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出，包括：

对所述第一合成语音以及所述第二合成语音中与所述目标兴趣标签关联的语音进行语音提示。

7.如权利要求1所述的语音实时翻译方法，其特征在于，所述对第一用户输入的第一语音信息进行语音识别，获得第一文本数据，包括：

获取关于所述第一语音信息的前文数据；

对第一用户输入的第一语音信息进行语音识别，获得初始文本数据；

利用所述前文数据对所述初始文本数据进行修正，得到第一文本数据。

8.如权利要求1所述的语音实时翻译方法，其特征在于，在所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出之前，所述方法包括：

当检测到用于输出语音的一对耳机分别由不同用户佩戴时，控制所述耳机工作于第一状态；当检测到第一用户和第二用户各佩戴一对耳机时，控制所述耳机工作于第二状态；

其中，在第一状态下，两个耳机分别输出不同的语音，被一用户佩戴的一耳机作为该用户的语音输出通道；在第二状态下，一对耳机中的两个输出相同的语音，被一用户佩戴的一对耳机作为该用户的语音输出通道。

9.如权利要求8所述的语音实时翻译方法，其特征在于，所述一对耳机设有

UWB 通信模块；在所述将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出之前，所述方法包括：

所述一对耳机基于 UWB 通信模块，实时侦测两个耳机之间的距离；

当两个耳机之间的距离大于第一预设阈值时，控制所述耳机工作于第一状态；
当两个耳机之间的距离小于第一预设阈值时，控制所述耳机工作于第二状态。

10.如权利要求 1 所述的语音实时翻译方法，其特征在于，所述方法还包括：

根据第一合成语音中未播放的音频帧，以及已采集生成第一语音信息且未转换为第一合成语音的音频帧，组合形成音频调节库；

基于所述音频调节库中的音频帧数量，调节第一合成语音的播放语速；

其中，所述第一合成语音对应的播放语速随着所述音频调节库中的音频帧数量的增大而增大，且随着所述音频调节库中的音频帧数量的减小而减小。

11.一种语音实时翻译系统，用于实现如权利要求 1 所述的语音实时翻译方法，其特征在于，所述系统包括：

第一文本数据生成模块，对第一用户输入的第一语音信息进行语音识别，获得第一文本数据；

第二合成语音生成模块，对所述第一语音信息进行翻译，获得对应于第二语种类别的第二合成语音；

第一合成语音生成模块，基于所述第一文本数据，获得对应于第一语种类别的第一合成语音；

合成语音播放模块，将所述第一合成语音向第一用户输出，以及同步将所述第二合成语音向第二用户输出。

12.一种语音实时翻译设备，其特征在于，包括：

处理器；

存储器，其中存储有所述处理器的可执行程序；

其中，所述处理器配置为经由执行所述可执行程序来执行权利要求 1 至 10 中任意一项所述语音实时翻译方法的步骤。

13.一种计算机可读存储介质，用于存储程序，其特征在于，所述程序被处理器执行时实现权利要求 1 至 10 中任意一项所述语音实时翻译方法的步骤。