

(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
2022 年 11 月 17 日 (17.11.2022)



(10) 国际公布号
WO 2022/237665 A1

(51) 国际专利分类号:
G10L 13/02 (2013.01) *G10L 25/63* (2013.01)
G10L 13/04 (2013.01)

(21) 国际申请号: PCT/CN2022/091348

(22) 国际申请日: 2022 年 5 月 7 日 (07.05.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110523097.4 2021年5月13日 (13.05.2021) CN

(71) 申请人: 北京有竹居网络技术有限公司(BEIJING
YOUZHUJU NETWORK TECHNOLOGY CO. LTD.)
[CN/CN]; 中国北京市平谷区林荫北街13号信
息大厦802室, Beijing 101299 (CN)。

(72) 发明人: 吴鹏飞(WU, Pengfei); 中国北京市海
淀区知春路63号中国卫星通信大厦今日头
条小邮局, Beijing 100086 (CN)。 潘俊杰(PAN,

Junjie); 中国北京市海淀区知春路63号中国
卫星通信大厦今日头条小邮局, Beijing 100086
(CN)。 马泽君(MA, Zejun); 中国北京市海淀区
知春路63号中国卫星通信大厦今日头
条小邮局, Beijing 100086 (CN)。

(74) 代理人: 北京品源专利代理有限公司(BEYOND
ATTORNEYS AT LAW); 中国北京市海淀区莲花
池东路39号西金大厦6层, Beijing 100036 (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, IQ, IR, IS,
IT, JM, 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,

(54) Title: SPEECH SYNTHESIS METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 语音合成方法、装置、电子设备和存储介质

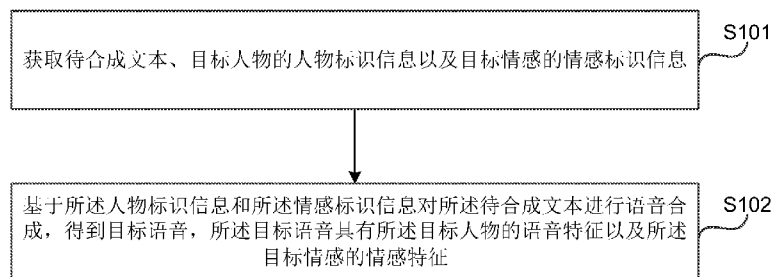


图 1

- S101 Acquire text to be synthesized, person identification information of a target person, and emotion identification information of a target emotion
- S102 Perform speech synthesis on the text on the basis of the person identification information and the emotion identification information to obtain target speech, wherein the target speech has speech features of the target person and emotion features of the target emotion

(57) Abstract: A speech synthesis method and apparatus, an electronic device, and a storage medium. The method comprises: acquiring text to be synthesized, person identification information of a target person, and emotion identification information of a target emotion (S101); and performing speech synthesis on the text on the basis of the person identification information and the emotion identification information to obtain target speech, wherein the target speech has speech features of the target person and emotion features of the target emotion (S102).

(57) 摘要: 一种语音合成方法、装置、电子设备和存储介质。该方法包括: 获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息(S101); 基于人物标识信息和情感标识信息对待合成文本进行语音合成, 得到目标语音, 目标语音具有目标人物的语音特征以及目标情感的情感特征(S102)。

SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, 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))。

语音合成方法、装置、电子设备和存储介质

本申请要求在2021年5月13日提交中国专利局、申请号为202110523097.4的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本公开实施例涉及计算机技术领域，例如涉及一种语音合成方法、装置、电子设备和存储介质。

背景技术

在语音合成领域中，特别是在生成有声书时，情感迁移是一项具有极大实用价值的技术。在生成有声书时，如果能够实现已授权使用的同一说话人不同语音之间的情感迁移，则说话人只需要录制部分带情感的语音即可实现该说话人带有不同情感的语音合成；如果能够实现已授权使用的不同说话人之间的情感迁移，则可以将具有情感演绎能力的说话人的语音中的情感迁移至情感演绎能力较差的说话人中，实现对情感演绎能力较差的说话人带有不同情感的语音的合成，可以直接根据已授权使用的某一说话人的已有语音生成其采用与小说情景相符的情感播报小说中的相应语句的有声书。

然而，相关技术仅能实现已授权使用的同一说话人带有不同情感的语音的合成，即实现同一说话人之间的情感迁移，而无法实现已授权使用的不同说话人之间带有不同情感的语音的合成，在生成有声书时，需要先录制已授权使用的说话人带有与小说情景相符的情感的语音，才能生成其采用该情感播报小说中的相应语句的有声书，而在已授权使用的说话人无法采用与小说情景相符的情感进行演绎或者无法获取到说话人带有该情感的语音时，是无法生成该说话人采用该情感播报小说中的相应语句的有声书的，导致有声书可供用户选择的说话人较为单一，无法满足人们的需求。

发明内容

本公开实施例提供一种语音合成方法、装置、电子设备和存储介质，以实现不同说话人带有不同情感的语音的合成。

第一方面，本公开实施例提供了一种语音合成方法，包括：

获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

第二方面，本公开实施例还提供了一种语音合成装置，包括：

获取模块，设置为获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

合成模块，设置为基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

第三方面，本公开实施例还提供了一种电子设备，包括：

至少一个处理器；

存储器，设置为存储至少一个程序，

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如本公开实施例所述的语音合成方法。

第四方面，本公开实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本公开实施例所述的语音合成方法。

附图说明

贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的，原件和元素不一定按照比例绘制。

图1为本公开实施例提供的一种语音合成方法的流程示意图；

图2为本公开实施例提供的另一种语音合成方法的流程示意图；

图3为本公开实施例提供的一种语音合成模型的结构示意图；

图4为本公开实施例提供的一种语音合成模型训练时的模型结构示意图；

图5为本公开实施例提供的一种语音合成装置的结构框图；

图6为本公开实施例提供的一种电子设备的结构示意图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过多种形式来实现，而且不应该被解释为限于这里阐述的实施例，相反提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

应当理解，本公开的方法实施方式中记载的多个步骤可以按照不同的顺序执行，和/或并行执行。此外，方法实施方式可以包括附加的步骤和/或省略执行示出的步骤。本公开的范围在此方面不受限制。

本文使用的术语“包括”及其变形是开放性包括，即“包括但不限于”。术语“基于”是“至少部分地基于”。术语“一个实施例”表示“至少一个实施例”；术语“另一实施例”表示“至少一个另外的实施例”；术语“一些实施例”表示“至少一些实施例”。其他术语的相关定义将在下文描述中给出。

需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

图1为本公开实施例提供的一种语音合成方法的流程示意图。该方法可以由语音合成装置执行，其中，该装置可以由软件和/或硬件实现，可配置于电子设备中，典型的，可以配置在手机或平板电脑中。本公开实施例提供的语音合成方法适用于合成已授权使用的不同人物

带有不同情感的语音的场景。如图 1 所示，本实施例提供的语音合成方法可以包括：

S101、获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息。

其中，待合成文本可以理解为欲合成其对应的语音的文本，其在用户授权之后进行获取。目标人物可以为待合成文本的已授权使用的播讲人，即欲合成具有其语音特征的已授权使用的语音的人物。目标情感可以为已授权使用的目标人物在播讲待合成文本（或待合成文本中的某一条或多条语句）时所采用的情感，如高兴、中立、伤心或生气等。相应的，目标人物的人物标识信息可以为用于唯一标识待合成文本的播讲人的信息，如播讲人的人物名称、人物 ID 或人物编码等；目标情感的情感标识信息可为用于唯一标识待合成文本的播讲人在播讲待合成文本时所采用的情感的信息，如该情感的情感名称、情感 ID 或情感编码等。目标人物的人物标识信息和目标情感的情感标识信息可以由用户需要对待合成文本进行语音合成时输入，也可以由待合成文本的发布者或目标语音的提供商预先进行设置。

在一个示例性的场景中，已授权使用的用户在欲合成一段语音时，可以输入该语音所对应的待合成文本，并选择或输入该段语音的已授权使用的播讲人的人物标识信息以及已授权使用的该段语音中所应携带的情感的情感标识信息；相应的，电子设备可以获取用户输入的文本作为待合成文本、获取用户所选择或输入的人物标识信息作为目标人物的人物标识信息，并获取用户所选择或输入的情感标识信息作为目标情感的情感标识信息。

在另一个示例性的场景中，用户在阅读某一待合成文本（如文章等）时，若欲收听该待合成文本的语音，则可以输入或选择待合成文本的已授权使用的播放人的人物标识信息以及已授权使用的语音中所携带的情感的情感标识信息；相应的，电子设备可以获取用户所选择或输入的人物标识信息作为目标人物的人物标识信息，并获取用户所选择或输入的情感标识信息作为目标情感的情感标识信息。

在又一个示例性的场景中，小说供应商可以预先设置其所向用户提供的小说中的每条语句所应携带的情感；从而，用户在欲通过收听语音的方式阅读小说时，可以设置该小说中的每个人物对应的已授权使用的播讲人；相应的，电子设备可以获取用户所设置的小说中每个人物对应的已授权使用的播讲人，可以依次将小说文本中的每条语句所对应的已授权使用的播讲人的人物标识信息作为该条语句的目标人物的人物标识信息，将该条语句所对应的情感的情感标识信息作为目标情感的情感标识信息，以合成小说中的每条语句对应的语音。或者，有声书开发商在欲生成某一部小说的有声书时，可以设置小说中每条语句的播讲人以及每条语句所应携带的情感；从而，电子设备在接收到有声书开发商生成有声书的触发操作或者在接收到用户收听小说的有声书的触发操作时，可以依次将小说文本中的每条语句所对应的播讲人的人物标识信息作为该条语句的目标人物的人物标识信息，将该条语句所对应的情感的情感标识信息作为目标情感的情感标识信息，以合成小说中的每条语句所对应的语音。

S102、基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

其中，目标语音可以为对待合成文本（或待合成文本中的一条或多条语句）进行语音合成得到的目标语音，该目标语音具有目标语音特征和目标情感的情感特征，即该目标语音可以为已经授权的目标人物采用用户已经授权使用的目标情感对待合成文本（或待合成文本中的一条或多条语句）进行播讲得到的语音。

在本实施例中，在获取到待合成文本、目标人物的人物标识信息和目标情感的情感标识

信息之后，可以依据该目标人物的人物标识信息和目标情感的情感标识信息对待合成文本进行语音合成，如首先确定待合成文本的文本特征信息，依据目标人物的人物标识信息确定目标人物的语音特征信息（如语音特征向量），并依据目标情感的情感标识信息确定目标情感的情感特征信息（如情感特征向量），然后基于该文本特征信息、语音特征信息和情感特征信息生成待合成文本的目标语音，即合成目标人物以目标情感播讲的目标语音。在此，所确定的目标人物的语音特征信息可以为目标人物授权使用的语音中的语音特征信息，所生成的由目标人物进行播报的目标语音中所携带的目标情感可以为目标人物授权携带的情感。

在一个实施方式中，所述基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，包括：通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列；将所述语音频谱序列输入至声码器中，以对所述语音频谱序列进行语音合成，得到目标语音。

其中，待合成文本的语音频谱序列可以为待合成的目标语音的频谱序列，其可以为目标语音的梅尔频谱序列，以确保基于该频谱序列合成的目标语音能够更符合人的听觉习惯。

在上述实施方式中，可以通过预先训练得到的语音合成模型生成待合成文本的语音频谱序列并通过声码器将该语音频谱序列转换为目标语音。示例性的，可以将待合成文本、目标人物的人物标识信息和目标情感的情感标识信息输入至预先训练得到的语音合成模型中，通过预先训练得到的语音合成模型确定待合成文本的文本特征信息、目标人物的语音特征信息和目标情感的情感特征信息，根据该文本特征信息、该语音特征信息以及该情感特征信息生成待合成文本的语音批频谱序列，并将该语音频谱序列输入至声码器中，通过该声码器生成待合成文本的目标语音。

在一个实施方式中，所述待合成文本为小说文本，所述获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息，包括：按照所述待合成文本中的至少一条待合成语句的排列顺序，依次将每一条待合成语句确定为当前语句，并获取所述当前待合成语句的当前人物标识信息和当前情感标识信息；所述基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，包括：基于所述当前人物标识信息和所述当前情感标识信息对所述当前待合成语句进行语音合成，得到所述当前语句的目标语音。

其中，当前待合成语句可以为当前时刻需要合成的待合成文本中的语句；相应的，当前人物标识信息可以为当前待合成语句的目标人物的人物标识信息，即当前待合成语句的已授权使用的播讲人的人物标识信息；当前情感标识信息可以为当前待合成语句对应的目标情感的情感标识信息，即当前待合成语句所应携带的情感的情感标识信息。

在本实施例中，由于小说中包含多个人物的对话和旁白等语句，不同语句所对应的已授权使用的播讲人和/或情感可能是不相同的，因此，当待合成文本为小说文本时，可以逐句确定对待合成文本中的每条语句对应的目标人物和目标情感并进行语音合成。

示例性的，在对小说文本进行语音合成时，可以首先将小说文本的第一条语句确定为当前语句，获取当前语句的当前人物标识信息和当前情感标识信息，依据该当前人物标识信息和当前情感标识信息对当前待合成语句进行语音合成，得到当前语句的目标语音，将小说文本中位于当前待合成语句之后且与当前待合成语句相邻的下一条待合成语句确定为当前待合成语句，并返回执行获取当前语句的当前人物标识信息和当前情感标识信息的操作，直至不存在下一待合成语句为止，由此，即可实现对待合成的小说文本的语音合成，得到待合成的

小说文本的有声书。

本实施例提供的语音合成方法，获取待合成文本、目标人物的人物标识信息和目标情感的情感标识信息，并基于该人物标识信息和该情感标识信息对待合成文本进行语音合成，得到具有目标人物的语音特征以及目标情感的情感特征的目标语音。本实施例通过采用上述技术方案，可以在已经授权的情况下，实现不同人物具有不同情感的语音的合成，从而，在授权后，仅根据播讲人的任一语音，即能生成其采用与小说情景相符的情感播报小说中的相应语句的有声书，无需已授权使用的播讲人再采用该情感进行演绎，能够提供更多可供选择的有声书播讲人，满足人们收听有声书时的不同需求。

图2为本公开实施例提供的另一种语音合成方法的流程示意图。本实施例中的方案可以与上述实施例中的一个或多个示例方案组合。例如，所述通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列，包括：确定所述待合成文本的文本音素序列；将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列。

例如，在所述得到目标语音之后，还包括：播放所述目标语音。

相应的，如图2所示，本实施例提供的语音合成方法可以包括：

S201、获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息。

S202、确定所述待合成文本的文本音素序列。

其中，音素为根据语音的自然属性划分得到的最小语音单元，相应的，待合成文本的文本音素序列可以为待合成文本的最小语音单元序列。

例如，在获取到待合成文本后，可以对待合成文本进行音素提取，得到待合成文本的文本音素序列。

在本实施例中，可以将用于提取待合成文本的文本音素序列的功能模块独立于语音合成模型设置，在合成待合成文本的语音时，首先由该功能模块提取待合成文本的文本音素序列，并将该功能模块提取得到的待合成文本的文本音素序列输入至语音合成模型中进行语音合成，以降低语音合成模型的复杂程度。

可以理解的是，本实施例也可以将用于提取待合成文本的文本音素序列的功能模块嵌入到语音合成模型中，并在合成待合成文本的语音时，直接将待合成文本输入至语音合成模型中，由语音合成模型获取待合成文本的文本音素序列。

S203、将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列。

在本实施例中，语音合成模型可设置为根据待合成文本的文本音素序列、已授权使用的目标人物的人物标识信息和已授权使用的目标情感的情感标识信息确定待合成文本的语音频谱序列，即语音合成模型的输入为待合成文本的文本音素序列、目标人物的人物标识信息和目标情感的情感标识信息，输出为待合成文本的语音频谱序列。

在一个实施方式中，所述语音合成模型包括文本编码器、高维映射模块、情感标志层、注意力模块和解码器，所述文本编码器的输出端和所述情感标志层的输出端分别与所述注意力模块的输入端相连，所述高维映射模块的输出端和所述注意力模块的输出端分别与所述解码器的输入端相连。

在上述实施方式中，如图3所示，语音合成模型可以包括文本编码器30、高维映射模块31、情感标志层32、注意力模块33和解码器34，其中，文本编码器30的输出端可以与注意力模块33的输入端相连，设置为根据待合成文本的文本音素序列确定待合成文本的文本特征信息，如确定待合成文本的文本特征向量，并将待合成文本的文本特征信息输入至注意力模块33中；高维映射模块31的输出端可以与注意力模块33的输入端相连，设置为根据已授权使用的目标人物的人物标识信息确定目标人物的语音特征向量，如将已授权使用的目标人物的人物标识信息映射为目标人物的语音特征向量，并将该语音特征向量输入至注意力模块33或解码器34中（图3中以将语音特征向量输入至解码器34中为例）；情感标志层32的输出端可以与注意力模块33的输入端相连，设置为根据已授权使用的目标情感的情感标识信息确定目标情感的情感特征向量；注意力模块33的输入端可以与解码器34的输入端相连，设置为与解码器共同根据文本编码器30所输入的文本特征向量、高维映射模块31所输入的语音特征向量以及情感标志层32所输入的情感特征向量生成待合成文本的音频频谱序列。

在上述实施方式中，所述将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列，可以包括：采用所述文本编码器对所述文本音素序列进行编码，以得到所述待合成文本的文本特征向量；采用所述高维映射模块对所述人物标识信息进行高维映射，以得到所述待合成文本的人物特征向量；采用所述情感标志层确定与所述情感标识信息对应的情感特征向量，作为所述待合成文本的情感特征向量；将所述文本特征向量和所述情感特征向量输入至所述注意力模块中，并将所述注意力模块输出的中间向量以及所述人物特征向量输入至所述解码器中，以得到所述待合成文本的音频频谱序列。其中，中间向量可以理解为注意力模块在对所接收到的文本标识信息、人物标识信息和情感标识信息进行处理后输出的向量。

示例性的，在获取到待合成文本的文本音素序列、已授权使用的目标人物的人物标识信息和已授权使用的目标情感的情感标识信息之后，可以首先将该文本音素序列输入至语音合成模型的文本编码器中，通过该文本编码器确定待合成文本的文本特征向量；将该人物标识信息输入至语音合成模型的高维映射模块中，通过该高维映射模块确定目标人物的语音特征向量；并将该情感标识信息输入至语音合成模型的情感标志层中，通过该情感标志层确定目标情感的情感特征向量。然后将该文本特征向量、语音特征向量和情感特征向量输入至语音合成模型的注意力模块中，得到注意力模块输出的中间向量。最后，将该中间向量输入至语音合成模型的解码器中，并获取该解码器输出的音频频谱序列，作为待合成文本的音频频谱序列。

可以理解的是，在上述实施方式中，可以将文本编码器输出的文本特征向量和情感标志层输出的情感特征向量直接输入至注意力模块中；也可以将文本编码器输出的文本特征向量和情感标志层输出的情感特征向量组合为一个向量，如将文本编码器输出的文本特征向量和情感编码器输出的情感特征向量进行拼或相加，并将组合得到的向量输入至注意力模块中，如图3所示。此外，解码器可以按照语音帧逐帧合成目标语音，并在得到当前时刻的语音帧对应的音频频谱序列后，除对外输出该音频频谱序列之外，还可以将当前时刻的语音帧对应的音频频谱序列输入到注意力模块中，以作为注意力模块在确定下一时刻的下一语音帧对应的中间向量时的输入；相应的，注意力模块可以根据当前时刻的文本特征向量、语音特征向量、情感标志层以及解码器在上一时刻输出的上一语音帧的音频频谱序列确定中间向量。

在本实施例中，语音合成模型能够生成已授权使用的不同人物采用不同情感对待合成文本进行播讲的语音，即当用户（或供应商）所选择或设置的目标人物和/或目标情感不相同，本实施例所采用的语音合成模型能够生成不同的目标语音。该语音合成模型在训练时的模型结构如图4所示，可以在情感标志层32的输入端连接注意力层35，在情感标志层32的输出端连接情感分类器36，在注意力层35的输入端连接参考编码器37，并在参考编码器37的输出端连接人物分类器38，此时，语音合成模型的训练过程可以为：

a、获取至少一个说话人已经授权使用的语音作为语音样本，并获取每个已经授权使用的语音样本所对应的文本音素序列。其中，每个语音样本中至少包含一条具有某一情感的情感语音。

b、提取每个已经授权使用的语音样本的音频频谱序列。

c、分别将每个已经授权使用的语音样本的文本音素序列输入至文本编码器中，将已经授权使用的语音样本的原始音频频谱序列输入至参考编码器中，并将相应语音样本对应的说话人的人物标识信息输入至高维映射模块中，并获取人物分类器输出的人物标识信息、情感分类器输出的情感标识信息以及解码器的输出音频频谱序列，以通过对抗训练的方式对语音合成模型进行训练，并通过反向传播算法对语音合成模型进行优化。

其中，反向传播算法可包括3个优化损失函数：输出音频频谱序列相较于原始音频频谱序列的重构误差（如最小均方差）、情感分类器输出的情感标识信息与音频语音中的真实情感之间的交叉熵损失以及人物分类器所输出的人物标识信息与音频语音对应的真实人物标识信息之间的误差。

d、重复迭代步骤c，直至模型收敛为止，如直至上述优化损失函数的值小于或等于预设误差阈值为止，或者，直至重复迭代次数达到预设次数阈值为止等。

S204、将所述语音频谱序列输入至声码器中，以对所述语音频谱序列进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

在本实施例中，在获取到语音合成模型输出的待合成文本的语音频谱序列后，可以将该文语音频谱序列输入到声码器中，通过声码器将该语音频谱序列转换为目标语音。其中，该声码器可以为任选的声码器，可以为预先训练得到的、与语音合成模型匹配的声码器，以提高目标语音的合成效果，此时，相应的，在训练语音合成模型时，也可以对语音合成模型所连接的声码器进行训练，以使得声码器能够合成效果较好的目标语音。

S205、播放所述目标语音。

在本实施例中，在合成目标语音后，还可以在用户已经授权播放的情况下，播放该目标语音，以便于用户进行收听。示例性的，可以在声码器合成目标语音之后，即播放用户已经授权播放的该目标语音，如在用户终端合成并播放目标语音；也可以在声码器合成目标语音后，存储该目标语音，并在接收到针对该目标语音的播放请求时，再播放该目标语音，如在服务端合成并存储待合成文本的目标语音，并在接收到某一用户终端发送的、针对该目标语音的播放请求时，将该目标语音发送给该用户终端，以通过该用户终端播放该目标语音。

本实施例提供的语音合成方法，获取待合成文本的文本音素序列，通过语音合成模型根据待合成文本的文本音素序列、目标人物的人物标识信息和目标情感的情感标识信息生成待合成文本的语音频谱序列，通过声码器将该语音频谱序列合成为目标语音，并播放该目标语音，能够在基于用户的授权实现不同人物带有不同情感的语音的合成前提下，提高语音的

合成效果，进而提高用户收听有声书的体验。

图5为本公开实施例提供的一种语音合成装置的结构框图。该装置可以由软件和/或硬件实现，可配置于电子设备中，典型的，可以配置在手机或平板电脑中，可通过执行语音合成方法对文本进行语音合成。如图5所示，本实施例提供的语音合成装置可以包括：获取模块501和合成模块502，其中，

获取模块501，设置为获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

合成模块502，设置为基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

本实施例提供的语音合成装置，通过获取模块获取待合成文本、目标人物的人物标识信息和目标情感的情感标识信息，并通过合成模块基于该人物标识信息和该情感标识信息对待合成文本进行语音合成，得到具有目标人物的语音特征以及目标情感的情感特征的目标语音。本实施例通过采用上述技术方案，可以在已经授权的情况下，实现不同人物具有不同情感的语音的合成，从而，在授权后，仅根据播讲人的任一语音，即能生成其采用与小说情景相符的情感播报小说中的相应语句的有声书，无需已授权使用的播讲人再采用该情感进行演绎，能够提供更多可供选择的有声书播讲人，满足人们收听有声书时的不同需求。

在上述方案中，所述合成模块502可以包括：频谱确定单元，设置为通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列；语音合成单元，设置为将所述语音频谱序列输入至声码器中，以对所述语音频谱序列进行语音合成，得到目标语音。

在上述方案中，所述频谱确定单元可以包括：音素获取子单元，设置为确定所述待合成文本的文本音素序列；频谱确定子单元，设置为将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列。

在上述方案中，所述语音合成模型可以包括文本编码器、高维映射模块、情感标志层、注意力模块和解码器，所述文本编码器的输出端和所述情感标志层的输出端分别与所述注意力模块的输入端相连，所述高维映射模块的输出端和所述注意力模块的输出端分别与所述解码器的输入端相连。

在上述方案中，所述频谱确定子单元可以设置为：采用所述文本编码器对所述文本音素序列进行编码，以得到所述待合成文本的文本特征向量；采用所述高维映射模块对所述人物标识信息进行高维映射，以得到所述待合成文本的人物特征向量；采用所述情感标志层确定与所述情感标识信息对应的情感特征向量，作为所述待合成文本的情感特征向量；将所述文本特征向量和所述情感特征向量输入至所述注意力模块中，并将所述注意力模块输出的中间向量以及所述人物特征向量输入至所述解码器中，以得到所述待合成文本的音频频谱序列。

例如，本实施例提供的语音合成装置还可以包括：语音播放模块，设置为在所述得到目标语音之后，播放所述目标语音。

在上述方案中，所述待合成文本可以为小说文本，所述获取模块501可以设置为：按照所述待合成文本中的至少一条待合成语句的排列顺序，依次将每一条待合成语句确定为当前

语句，并获取所述当前待合成语句的当前人物标识信息和当前情感标识信息；所述合成模块 502 可以设置为：基于所述当前人物标识信息和所述当前情感标识信息对所述当前待合成语句进行语音合成，得到所述当前语句的目标语音。

本公开实施例提供的语音合成装置可执行本公开任意实施例提供的语音合成方法，具备执行语音合成方法相应的功能模块和有益效果。未在本实施例中详尽描述的技术细节，可参见本公开任意实施例所提供的语音合成方法。

下面参考图 6，其示出了适于用来实现本公开实施例的电子设备（例如终端设备）600 的结构示意图。本公开实施例中的终端设备可以包括但不限于诸如移动电话、笔记本电脑、数字广播接收器、PDA（个人数字助理）、PAD（平板电脑）、PMP（便携式多媒体播放器）、车载终端（例如车载导航终端）等等的移动终端以及诸如数字 TV、台式计算机等等的固定终端。图 6 示出的电子设备仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

如图 6 所示，电子设备 600 可以包括处理装置（例如中央处理器、图形处理器等）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储装置 606 加载到随机访问存储器（RAM）603 中的程序而执行多种适当的动作和处理。在 RAM 603 中，还存储有电子设备 600 操作所需的多种程序和数据。处理装置 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

通常，以下装置可以连接至 I/O 接口 605：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 606；包括例如液晶显示器（LCD）、扬声器、振动器等的输出装置 607；包括例如磁带、硬盘等的存储装置 606；以及通信装置 609。通信装置 609 可以允许电子设备 600 与其他设备进行无线或有线通信以交换数据。虽然图 6 示出了具有多种装置的电子设备 600，但是应理解的是，并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。

根据本公开的实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括承载在非暂态计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中，该计算机程序可以通过通信装置 609 从网络上被下载和安装，或者从存储装置 606 被安装，或者从 ROM 602 被安装。在该计算机程序被处理装置 601 执行时，执行本公开实施例的方法中限定的上述功能。

需要说明的是，本公开上述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——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。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开中，计算机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。而在本公开中，计算机可读信号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，其中承载了计算机可读的程序代码。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。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、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。计算机可读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：电线、光缆、RF（射频）等等，或者上述的任意合适的组合。

在一些实施方式中，客户端、服务器可以利用诸如 HTTP（HyperText Transfer Protocol，超文本传输协议）之类的任何当前已知或未来研发的网络协议进行通信，并且可以与任意形式或介质的数字数据通信（例如，通信网络）互连。通信网络的示例包括局域网（“LAN”），广域网（“WAN”），网际网（例如，互联网）以及端对端网络（例如，ad hoc 端对端网络），以及任何当前已知或未来研发的网络。

上述计算机可读介质可以是上述电子设备中所包含的；也可以是单独存在，而未装配入该电子设备中。

上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该电子设备执行时，使得该电子设备：获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的操作的计算机程序代码，上述程序设计语言包括但不限于面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开多种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。其中，模块的名称在某种情况下并不构成对该单元本身的限定。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、片上系统（SOC）、复杂可编程逻辑设备（CPLD）等等。

在本公开的上下文中，机器可读介质可以是有形的介质，其可以包含或存储以供指令执

行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备，或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦除可编程只读存储器（EPROM或快闪存储器）、光纤、便捷式紧凑盘只读存储器（CD-ROM）、光学储存设备、磁储存设备、或上述内容的任何合适组合。

根据本公开的一个或多个实施例，示例1提供了一种语音合成方法，包括：

获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

根据本公开的一个或多个实施例，示例2根据示例1所述的方法，所述基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，包括：

通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列；

将所述语音频谱序列输入至声码器中，以对所述语音频谱序列进行语音合成，得到目标语音。

根据本公开的一个或多个实施例，示例3根据示例2所述的方法，所述通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列，包括：

确定所述待合成文本的文本音素序列；

将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列。

根据本公开的一个或多个实施例，示例4根据示例3所述的方法，所述语音合成模型包括文本编码器、高维映射模块、情感标志层、注意力模块和解码器，所述文本编码器的输出端和所述情感标志层的输出端分别与所述注意力模块的输入端相连，所述高维映射模块的输出端和所述注意力模块的输出端分别与所述解码器的输入端相连。

根据本公开的一个或多个实施例，示例5根据示例4所述的方法，所述将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列，包括：

采用所述文本编码器对所述文本音素序列进行编码，以得到所述待合成文本的文本特征向量；

采用所述高维映射模块对所述人物标识信息进行高维映射，以得到所述待合成文本的人物特征向量；

采用所述情感标志层确定与所述情感标识信息对应的情感特征向量，作为所述待合成文本的情感特征向量；

将所述文本特征向量和所述情感特征向量输入至所述注意力模块中，并将所述注意力模块输出的中间向量以及所述人物特征向量输入至所述解码器中，以得到所述待合成文本的音频频谱序列。

根据本公开的一个或多个实施例，示例6根据示例1-5任一所述的方法，在所述得到目标语音之后，还包括：

播放所述目标语音。

根据本公开的一个或多个实施例，示例7根据示例1-5任一所述的方法，所述待合成文本为小说文本，所述获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息，包括：

按照所述待合成文本中的至少一条待合成语句的排列顺序，依次将每一条待合成语句确定为当前语句，并获取所述当前待合成语句的当前人物标识信息和当前情感标识信息；

所述基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，包括：

基于所述当前人物标识信息和所述当前情感标识信息对所述当前待合成语句进行语音合成，得到所述当前语句的目标语音。

根据本公开的一个或多个实施例，示例8提供了一种语音合成装置，包括：

获取模块，设置为获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

合成模块，设置为基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

根据本公开的一个或多个实施例，所述合成模块包括：

频谱确定单元，设置为通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列；

语音合成单元，设置为将所述语音频谱序列输入至声码器中，以对所述语音频谱序列进行语音合成，得到目标语音。

根据本公开的一个或多个实施例，所述频谱确定单元包括：

音素获取子单元，设置为确定所述待合成文本的文本音素序列；

频谱确定子单元，设置为将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列。

根据本公开的一个或多个实施例，所述语音合成模型包括文本编码器、高维映射模块、情感标志层、注意力模块和解码器，所述文本编码器的输出端和所述情感标志层的输出端分别与所述注意力模块的输入端相连，所述高维映射模块的输出端和所述注意力模块的输出端分别与所述解码器的输入端相连。

根据本公开的一个或多个实施例，所述频谱确定子单元设置为：

采用所述文本编码器对所述文本音素序列进行编码，以得到所述待合成文本的文本特征向量；

采用所述高维映射模块对所述人物标识信息进行高维映射，以得到所述待合成文本的人物特征向量；

采用所述情感标志层确定与所述情感标识信息对应的情感特征向量，作为所述待合成文本的情感特征向量；

将所述文本特征向量和所述情感特征向量输入至所述注意力模块中，并将所述注意力模

块输出的中间向量以及所述人物特征向量输入至所述解码器中，以得到所述待合成文本的音频频谱序列。

根据本公开的一个或多个实施例，所述语音合成装置还包括：

语音播放模块，设置为在所述得到目标语音之后，播放所述目标语音。

根据本公开的一个或多个实施例，所述待合成文本为小说文本，所述获取模块设置为：

按照所述待合成文本中的至少一条待合成语句的排列顺序，依次将每一条待合成语句确认为当前语句，并获取所述当前待合成语句的当前人物标识信息和当前情感标识信息；

所述合成模块设置为：基于所述当前人物标识信息和所述当前情感标识信息对所述当前待合成语句进行语音合成，得到所述当前语句的目标语音。

根据本公开的一个或多个实施例，示例 9 提供了一种电子设备，包括：

一个或多个处理器；

存储器，设置为存储一个或多个程序，

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如示例 1-7 中任一所述的语音合成方法。

根据本公开的一个或多个实施例，示例 10 提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如示例 1-7 中任一所述的语音合成方法。

以上描述仅为本公开的示例实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开中所涉及的公开范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述公开构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

此外，虽然采用特定次序描绘了多种操作，但是这不应理解为要求这些操作以所示出的特定次序或以顺序次序执行来执行。在一定环境下，多任务和并行处理可能是有利的。同样地，虽然在上文论述中包含了若干具体实现细节，但是这些不应被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实施例中。相反地，在单个实施例的上下文中描述的多种特征也可以单独地或以任何合适的子组合的方式实现在多个实施例中。

权利要求书

1、一种语音合成方法，包括：

获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

2、根据权利要求1所述的方法，其中，所述基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，包括：

通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列；

将所述语音频谱序列输入至声码器中，以对所述语音频谱序列进行语音合成，得到目标语音。

3、根据权利要求2所述的方法，其中，所述通过预先训练得到的语音合成模型基于所述人物标识信息和所述情感标识信息确定所述待合成文本的语音频谱序列，包括：

确定所述待合成文本的文本音素序列；

将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列。

4、根据权利要求3所述的方法，其中，所述语音合成模型包括文本编码器、高维映射模块、情感标志层、注意力模块和解码器，所述文本编码器的输出端和所述情感标志层的输出端分别与所述注意力模块的输入端相连，所述高维映射模块的输出端和所述注意力模块的输出端分别与所述解码器的输入端相连。

5、根据权利要求4所述的方法，其中，所述将所述文本音素序列、所述人物标识信息和所述情感标识信息输入至预先训练得到的语音合成模型中，并获取所述语音合成模型输出的语音频谱序列，包括：

采用所述文本编码器对所述文本音素序列进行编码，以得到所述待合成文本的文本特征向量；

采用所述高维映射模块对所述人物标识信息进行高维映射，以得到所述待合成文本的人物特征向量；

采用所述情感标志层确定与所述情感标识信息对应的情感特征向量，作为所述待合成文本的情感特征向量；

将所述文本特征向量和所述情感特征向量输入至所述注意力模块中，并将所述注意力模块输出的中间向量以及所述人物特征向量输入至所述解码器中，以得到所述待合成文本的音频频谱序列。

6、根据权利要求1-5任一所述的方法，在所述得到目标语音之后，还包括：

播放所述目标语音。

7、根据权利要求1-5任一所述的方法，其中，所述待合成文本为小说文本，所述小说文本包括至少一条待合成语句，所述获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息，包括：

按照所述至少一条待合成语句的排列顺序，依次将每条待合成语句确定为当前语句，并获取所述当前待合成语句的当前人物标识信息和当前情感标识信息；

所述基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到

目标语音，包括：

基于所述当前人物标识信息和所述当前情感标识信息对所述当前待合成语句进行语音合成，得到所述当前语句的目标语音。

8、一种语音合成装置，包括：

获取模块，设置为获取待合成文本、目标人物的人物标识信息以及目标情感的情感标识信息；

合成模块，设置为基于所述人物标识信息和所述情感标识信息对所述待合成文本进行语音合成，得到目标语音，所述目标语音具有所述目标人物的语音特征以及所述目标情感的情感特征。

9、一种电子设备，包括：

至少一个处理器；

存储器，设置为存储至少一个程序，

当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如权利要求 1-7 中任一所述的语音合成方法。

10、一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求 1-7 中任一所述的语音合成方法。

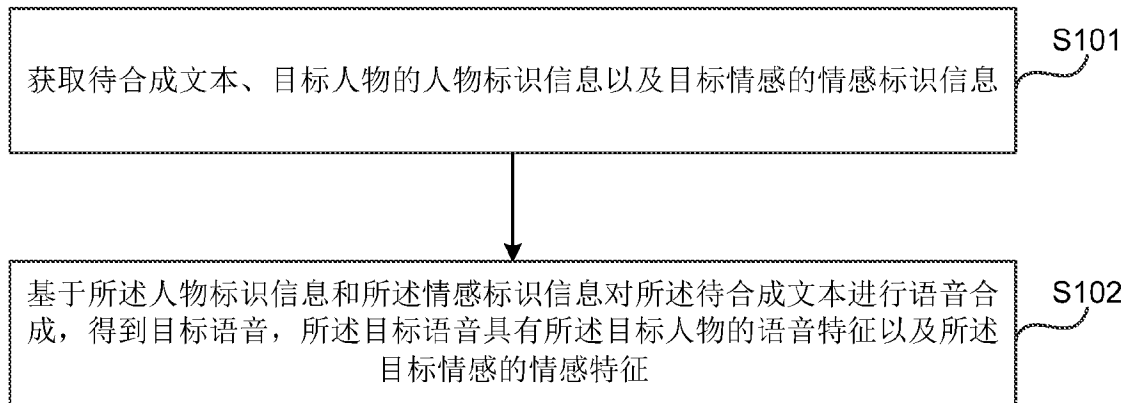


图 1

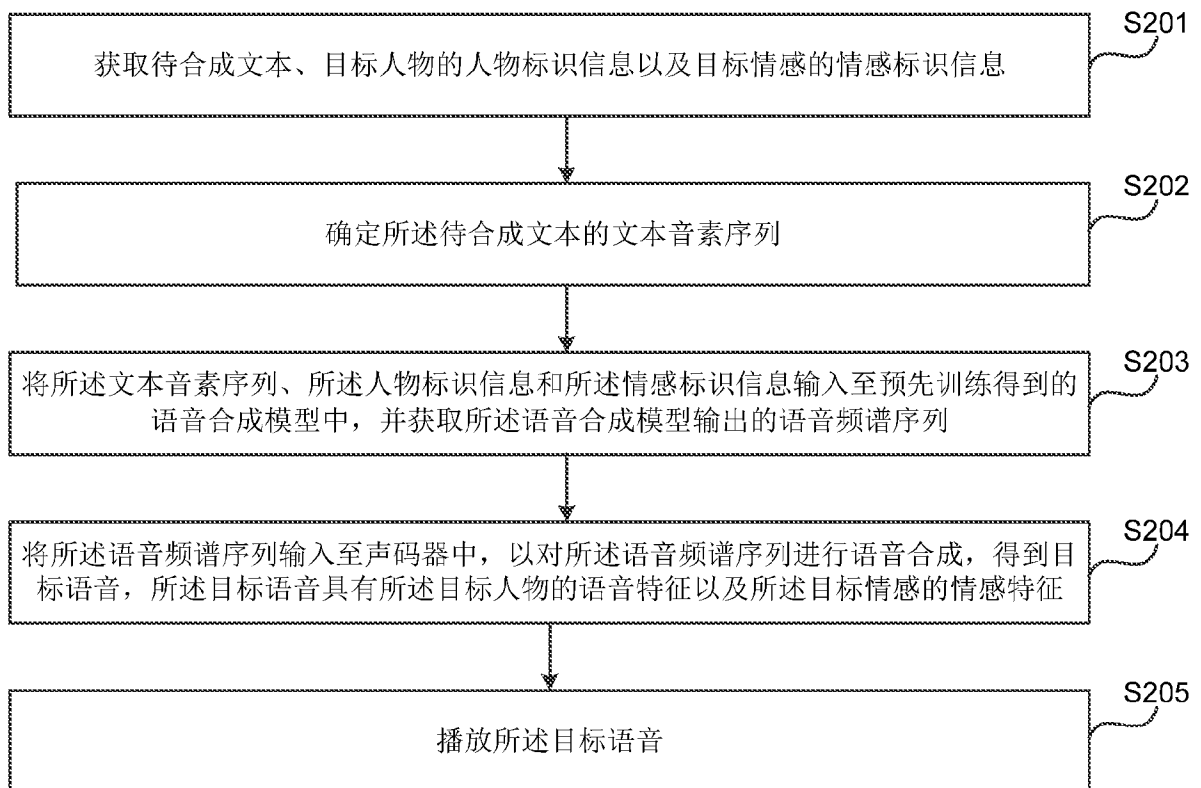


图 2

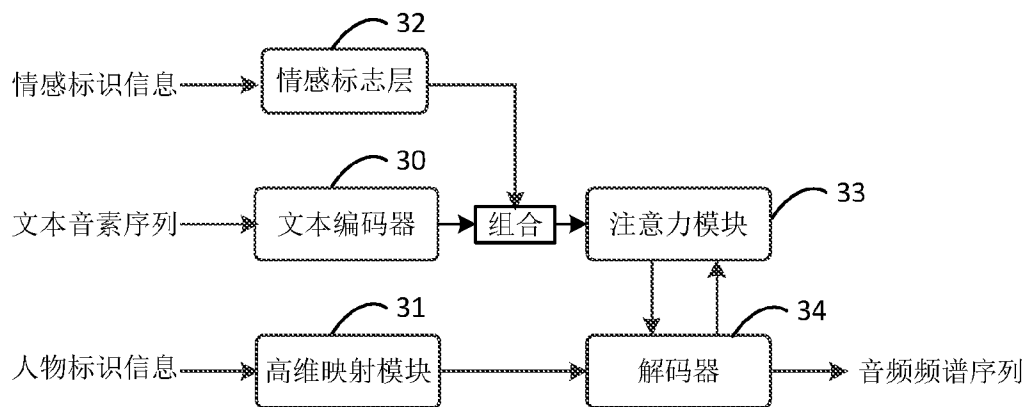


图 3

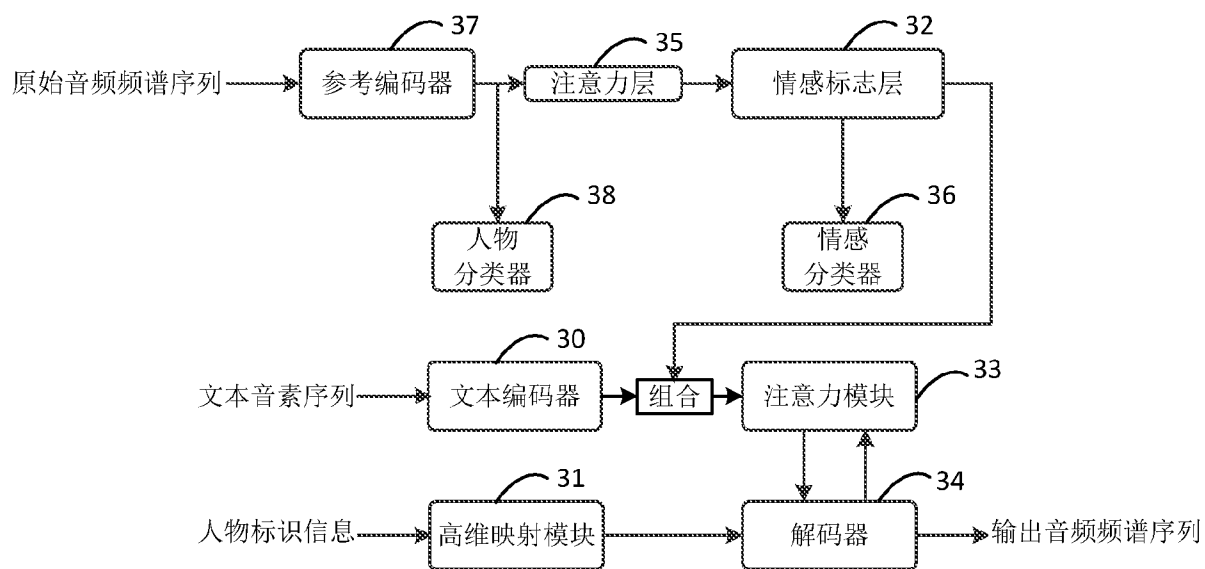


图 4

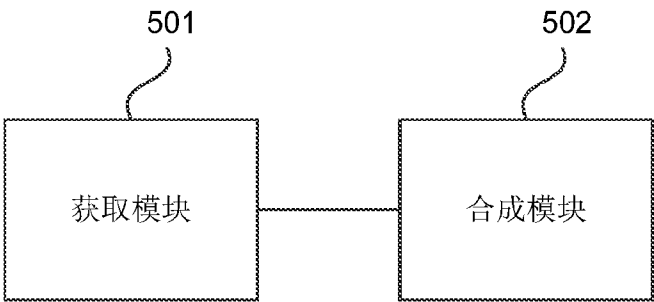


图 5

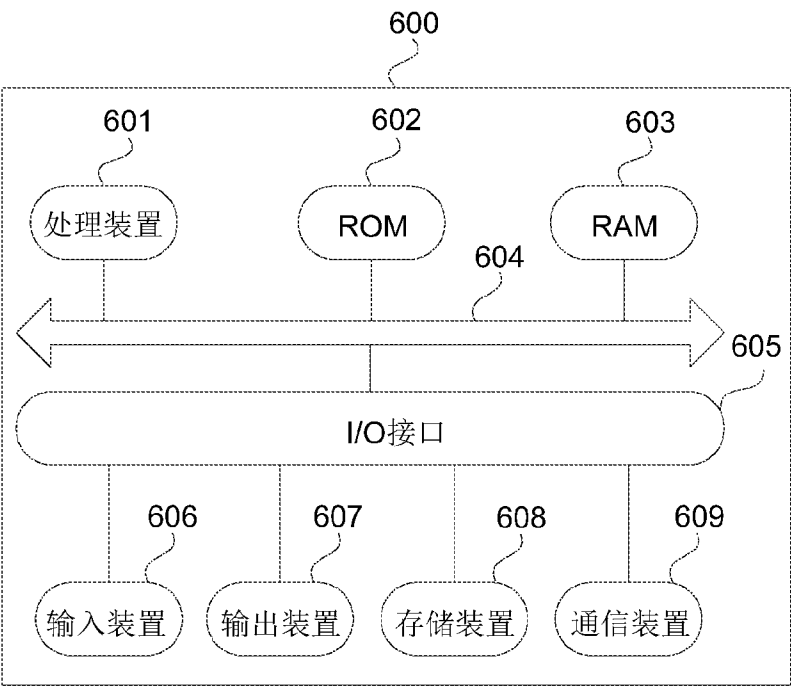


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/091348

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/02(2013.01)i; G10L 13/04(2013.01)i; G10L 25/63(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)

CNTXT, ENTXTC, CNABS, WPABSC: 语音, 合成, 小说, 文本, 人物, 角色, 标识, 标签, 情感, 情绪, 模型, 声码器, 编码, 解码, 注意力, 关注, 映射, speech, synthes+, text, identif+, label, character, emotion, encode+, decode+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113257218 A (BEIJING YOUZHUJU NETWORK TECHNOLOGY CO., LTD.) 13 August 2021 (2021-08-13) claims 1-10, description, paragraphs 1-146, and figures 1-6	1-10
X	WO 2020190054 A1 (HUMELO INC.) 24 September 2020 (2020-09-24) description, paragraphs 43-69 and 135-143	1-10
X	CN 111667811 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 15 September 2020 (2020-09-15) description, paragraphs 31-153	1, 6-10
X	US 2020035215 A1 (LG ELECTRONICS INC.) 30 January 2020 (2020-01-30) description, paragraphs 136-142, 169-215, 225-246, and 302	1, 6-10
A	CN 112349273 A (CTRIIP COMPUTER TECHNOLOGY (SHANGHAI) CO., LTD.) 09 February 2021 (2021-02-09) entire document	1-10
A	CN 112289299 A (BEIJING DA MI TECHNOLOGY CO., LTD.) 29 January 2021 (2021-01-29) entire document	1-10

☒ 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

23 June 2022

Date of mailing of the international search report

30 June 2022

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

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/091348

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2016329043 A1 (LG ELECTRONICS INC.) 10 November 2016 (2016-11-10) entire document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/091348

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113257218	A	13 August 2021	None			
WO	2020190054	A1	24 September 2020	KR	102057927	B1	20 December 2019
				KR	20200111609	A	29 September 2020
CN	111667811	A	15 September 2020	CN	111667811	B	07 September 2021
US	2020035215	A1	30 January 2020	KR	20190104941	A	11 September 2019
				US	11074904	B2	27 July 2021
CN	112349273	A	09 February 2021	None			
CN	112289299	A	29 January 2021	None			
US	2016329043	A1	10 November 2016	WO	2015111818	A1	30 July 2015
				KR	20150087023	A	29 July 2015

国际检索报告

国际申请号

PCT/CN2022/091348

A. 主题的分类

G10L 13/02(2013.01)i; G10L 13/04(2013.01)i; G10L 25/63(2013.01)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

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))

CNTXT, ENTXT, CNABS, WPABSC, 语音, 合成, 小说, 文本, 人物, 角色, 标识, 标签, 情感, 情绪, 模型, 声码器, 编码, 解码, 注意力, 关注, 映射, speech, syntheses, text, identif, label, character, emotion, encode, decode

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 113257218 A (北京有竹居网络技术有限公司) 2021年8月13日 (2021 - 08 - 13) 权利要求1-10, 说明书第1-146段, 附图1-6	1-10
X	WO 2020190054 A1 (HUMERO INC.) 2020年9月24日 (2020 - 09 - 24) 说明书第43-69、135-143段	1-10
X	CN 111667811 A (北京百度网讯科技有限公司) 2020年9月15日 (2020 - 09 - 15) 说明书第31-153段	1、6-10
X	US 2020035215 A1 (LG ELECTRONICS INC.) 2020年1月30日 (2020 - 01 - 30) 说明书第136-142、169-215、225-246、302段	1、6-10
A	CN 112349273 A (携程计算机技术上海有限公司) 2021年2月9日 (2021 - 02 - 09) 全文	1-10
A	CN 112289299 A (北京大米科技有限公司) 2021年1月29日 (2021 - 01 - 29) 全文	1-10
A	US 2016329043 A1 (LG ELECTRONICS INC.) 2016年11月10日 (2016 - 11 - 10) 全文	1-10

☐ 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

2022年6月23日

国际检索报告邮寄日期

2022年6月30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

授权官员

吴琼

电话号码 (86-10)62089510

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/091348

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	113257218	A	2021年8月13日	无			
WO	2020190054	A1	2020年9月24日	KR	102057927	B1	2019年12月20日
				KR	20200111609	A	2020年9月29日
CN	111667811	A	2020年9月15日	CN	111667811	B	2021年9月7日
US	2020035215	A1	2020年1月30日	KR	20190104941	A	2019年9月11日
				US	11074904	B2	2021年7月27日
CN	112349273	A	2021年2月9日	无			
CN	112289299	A	2021年1月29日	无			
US	2016329043	A1	2016年11月10日	WO	2015111818	A1	2015年7月30日
				KR	20150087023	A	2015年7月29日