

一种数字人动作编排方法及装置

本申请要求于2023年07月31日提交中国国家知识产权局、申请号为2023109570705、
申请名称为“一种数字人动作编排方法及装置”的中国专利申请的优先权，其全部内容通
5 过引用结合在本申请中。

技术领域

本说明书一个或多个实施例涉及虚拟数字人领域，尤其涉及一种数字人动作编排方法
及装置。

10

背景技术

现实世界中，人在说话的过程中通常会伴随着一些肢体动作，而人的很多下意识动作
通常和语义以及环境存在很大的关联。肢体语言作为日常沟通交流中的一种非语言信号
(nonverbal signals)，是人与人之间交流的重要组成部分，当演讲者配合了适当的手势时，可
15 以提高听众理解的准确度。

在实际业务中，数字人在进行播报、讲解、直播、交互的时候，都需要通过肢体动作
来辅助语言完成信息的传达。与此同时，在目前常见的数字人应用场景中，肢体动作通常
都是和语音/文本同时发生的，因此，如何在数字人播报的同时呈现符合语义的动作，对于
自然的数字人交互以及良好的用户体验至关重要。

20

发明内容

本说明书一个或多个实施例描述了一种数字人动作编排方法及装置，旨在给数字人播
报的文本匹配合适的动作。

第一方面，提供了一种数字人动作编排方法，包括：
25 获取目标文本对应的第一词语序列，以及获取第一动作集合，所述第一动作集合中包
含若干动作，任一动作具有预设的触发词以及对应的动作持续时长；

至少根据所述触发词，将所述第一词语序列中的各个词语与所述第一动作集合进行匹
配，得到第一匹配结果，其中包含各个词语对应的备选动作集合；

根据各个词语在所述目标文本中的位置，以及所述动作持续时长，从所述第一匹配结
30 果中确定一组在时间维度上没有重叠的动作序列，得到第一动作序列，所述第一动作序列
用于指示所述数字人在播报所述目标文本时做出相应的动作。

在一种可能的实施方式中，将所述第一词语序列中的各个词语与所述第一动作集合进

行匹配，得到第一匹配结果，包括：

针对任意目标词语，将该目标词语与所述触发词进行匹配，得到第一候选动作集合：

根据该目标词语的语义表征与所述第一动作集合中动作的动作表征的相似度，得到第二候选动作集合；

- 5 根据所述第一候选动作集合与所述第二候选动作集合，确定目标词语对应的备选动作集合。

在一种可能的实施方式中，根据该目标词语的语义表征与所述第一动作集合中动作的动作表征的相似度，得到第二候选动作集合，包括：

使用第一文本编码器对所述目标词语进行编码，得到目标语义表征；

- 10 获取使用动作编码器对所述第一动作集合中的任一动作进行编码得到的动作表征；

将动作表征与所述目标语义表征的相似度大于第一阈值的动作，归入第二候选动作集合。

在一种可能的实施方式中，所述第一文本编码器用于对第一语言的文本进行编码；所述动作编码器预先与适用于第二语言的第二文本编码器联合训练；所述第一文本编码器通

- 15 过以下方式训练得到：

使用所述第一文本编码器对第一语言的第一训练语料进行编码，得到第一语义表征；

使用所述第二文本编码器对第二语言的第二训练语料进行编码，得到第二语义表征，所述第二训练语料与所述第一训练语料具有相同的语义；

- 20 通过最大化所述第一语义表征和第二语义表征之间的相似度，调整所述第一文本编码器中的参数的值。

在一种可能的实施方式中，根据所述第一候选动作集合与所述第二候选动作集合，确定目标词语对应的备选动作集合，包括：

根据所述第一候选动作集合与第二候选动作集合的并集，确定备选动作集合；或

根据所述第一候选动作集合与第二候选动作集合的交集，确定备选动作集合。

- 25 在一种可能的实施方式中，所述第一词语序列中包括第一词语和第二词语；从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，得到第一动作序列，包括：

分别从所述第一词语和第二词语对应的备选动作集合中确定第一动作和第二动作；

根据时间戳预测模型确定所述第一词语和第二词语在时间轴上的开始位置；

- 30 根据所述开始位置和动作持续时长，判断所述第一动作和第二动作在时间轴上是否有重叠；若有，则将第一动作和第二动作之一从所述第一动作序列中排除。

在一种可能的实施方式中，将第一动作和第二动作之一从所述第一动作序列中排除，包括：

将开始位置在后的词语所对应的动作，从所述所述第一动作序列中排除。

在一种可能的实施方式中，所述第一动作和第二动作还具有对应的优先级；将第一动作5 和第二动作之一从所述第一动作序列中排除，包括：

将优先级较低的动作从所述所述第一动作序列中排除。

在一种可能的实施方式中，还包括：

获取第二动作集合，所述第二动作集合中包含若干通用动作，任一通用动作具有对应的动作持续时长；

10 提取所述目标文本中词性为目标词性的词语，得到第二词语序列；

对所述第二词语序列中的任一词语，从所述第二动作集合中随机选择对应的通用动作，得到第二匹配结果；

根据第二词语序列中各个词语在所述目标文本中的位置，以及各个词语对应的通用动作的动作持续时长，从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列15 中，得到第二动作序列，使得所述第二动作序列中的各个动作在时间维度上没有重叠。

在一种可能的实施方式中，所述目标词性包括以下中的一种或多种：数量词、名动词、动副词、普通动词、时间、方位词。

在一种可能的实施方式中，对所述第二词语序列中的任一词语，从所述第二动作集合中随机选择对应的通用动作，得到第二匹配结果，包括：

20 从所述第二动作集合中随机选择通用动作，使得所述第二词语序列中相邻的词语之间对应不同的通用动作，得到第二匹配结果。

在一种可能的实施方式中，所述第二词语序列中包括第三词语；从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列中，得到第二动作序列，包括：

根据时间戳预测模型确定所述第三词语在时间轴上的开始位置；

25 根据所述第二匹配结果确定所述第三词语对应的第三动作；

根据所述开始位置和第三动作的动作持续时长，判断所述第三动作和所述第一动作序列中的任一动作在时间轴上是否有重叠，若没有，则将第三动作添加到所述第一动作序列中。

第二方面，提供了一种数字人动作编排装置，包括：

30 第一获取单元，配置为，获取目标文本对应的第一词语序列，以及获取第一动作集合，

所述第一动作集合中包含若干动作，任一动作具有预设的触发词以及对应的动作持续时长；

第一匹配单元，配置为，至少根据所述触发词，将所述第一词语序列中的各个词语与
所述第一动作集合进行匹配，得到第一匹配结果，其中包含各个词语对应的备选动作集合；

5 第一确定单元，配置为，根据各个词语在所述目标文本中的位置，以及所述动作持续
时长，从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，得到第一动作
序列，所述第一动作序列用于指示所述数字人在播报所述目标文本时做出相应的动作。

在一种可能的实施方式中，还包括：

第二获取单元，配置为，获取第二动作集合，所述第二动作集合中包含若干通用动作，
任一通用动作具有对应的动作持续时长；

10 词语提取单元，配置为，提取所述目标文本中词性为目标词性的词语，得到第二词语
序列；

第二匹配单元，配置为，对所述第二词语序列中的任一词语，从所述第二动作集合中
随机选择对应的通用动作，得到第二匹配结果；

15 第二确定单元，配置为，根据第二词语序列中各个词语在所述目标文本中的位置，以
及各个词语对应的通用动作的动作持续时长，从所述第二匹配结果中确定一组动作序列并
插入到所述第一动作序列中，得到第二动作序列，使得所述第二动作序列中的各个动作在
时间维度上没有重叠。

第三方面，提供了一种计算机可读存储介质，其上存储有计算机程序，当所述计算机
程序在计算机中执行时，令计算机执行第一方面的方法。

20 第四方面，提供了一种计算设备，包括存储器和处理器，其中，所述存储器中存储有
可执行代码，所述处理器执行所述可执行代码时，实现第一方面的方法。

本说明书实施例提出的一种数字人动作编排方法及装置，基于文本与动作预设的触发
词的匹配，以及文本语义编码与动作编码的匹配，为文本匹配合适的动作。同时，为了保
证不同动作之间的连贯性，避免数字人长时间静默，本说明书实施例还在相邻动作之间插
25 入合适的通用动作，保证动作序列的多样性和连贯性。

附图说明

为了更清楚地说明本说明书披露的多个实施例的技术方案，下面将对实施例描述中所
需要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说明书披露的多
30 个实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这
些附图获得其它的附图。

- 图 1 示出根据一个实施例的数字人动作编排方法的实施场景示意图；
- 图 2 示出根据一个实施例的数字人动作编排方法的流程图；
- 图 3 示出根据一个实施例的文本与动作匹配方法的流程图；
- 图 4 示出根据一个实施例的数字人动作粗排阶段的示意图；
- 5 图 5 示出根据一个实施例的精排阶段的时间轴示意图；
- 图 6 示出根据一个实施例的数字人动作编排装置的示意性框图。

具体实施方式

- 下面结合附图，对本说明书提供的方案进行描述。
- 10 数字人又称虚拟形象，是运用数字技术创造出来的，与人类形象接近的数字化人物形象。在实际业务中，数字人在进行播报、讲解、直播、交互的时候，都需要通过肢体动作来辅助语言完成信息的传达。在目前常见的数字人应用场景中，肢体动作通常都是和语音/文本同时发生的，因此，基于数字人说话时语义与动作一致性的需求，需要一种动作编排方案，使得数字人在播报的同时呈现符合语义的动作。
- 15 图 1 示出根据一个实施例的数字人动作编排方法的实施场景示意图。在图 1 的示例中，输入数据包括数字人要播报的文本数据，以及给数字人编排动作所使用的候选动作集。方案整体分为粗排、精排、重排三个阶段。
- 在粗排阶段，通过计算基于规则匹配的动作集合以及基于语义相似度匹配的动作集合的交集/并集，得到备选动作集合。其中，基于规则的匹配主要考虑输入文本分词之后的短
- 20 语是否与候选动作集中动作配置的触发词一致，将匹配的动作添加到基于规则匹配的动作集合中，同时在用户指定的文本位置插入特定的动作；基于语义相似度的匹配则考虑文本分词之后的文本语义表征与候选动作集中动作表征的相似度，将相似度大于指定阈值的动作添加到基于语义相似度匹配的动作集合中。
- 在粗排阶段只考虑了文本与动作的匹配度，但是由于候选动作集中的动作都是有一定
- 25 持续时长的，而输入文本每个字在数字人播报的输出音频中也有起始和结束时间。因此在精排阶段，为了避免不同动作出现在时间轴相同的位置（即产生动作重叠），会对动作和文本进行时间轴对齐，即先按照动作起始的位置以及配置的优先级对备选动作集合中的动作进行排序，然后保留优先级高且互相不重叠的动作，删除其他冲突动作。
- 为了保证不同动作之间的连贯性，避免数字人长时间静默，在重排阶段，对输入文本
- 30 进行句法结构解析，获取指定语法结构的短语，从输入的通用动作集合中选择合适的通用动作插入到指定短语的位置，并基于动作配置的优先级进行排序，保证动作序列的多样性

和连贯性。

以下结合具体的实施例，描述上述数字人动作编排方法的具体实施步骤。图 2 示出根据一个实施例的数字人动作编排方法的流程图，所述方法的执行主体可以为任何具有计算、处理能力的平台或服务器或设备集群等。

5 在步骤 202，获取目标文本对应的第一词语序列，以及获取第一动作集合，所述第一动作集合中包含若干动作，任一动作具有预设的触发词以及对应的动作持续时长。

第一词语序列可以通过目标文本进行分词得到，序列中的词语根据实际需求可以是单字、单词、短语或它们的组合。第一词语序列中的各个词语还包含一个自身在目标文本中的位置信息，例如索引信息。第一动作集合中包含若干与文本和环境存在关联的动作。动作的触发词用于指示数字人在播报相应的词语时做出相应的动作。

10 在一个实施例中，为了降低整体的计算开销，考虑到人在说话时通常会在特定词性的词语做动作的习惯，根据词性将第一词语序列中的低频动作词性的词语移除，只保留词性为普通名词、形容词、普通动词、副形词、动副词、名形词、名动词、副词中的一种或多种的词语。

15 在步骤 204，至少根据所述触发词，将所述第一词语序列中的各个词语与所述第一动作集合进行匹配，得到第一匹配结果，其中包含各个词语对应的备选动作集合。

在一个实施例中，实现步骤 204 的流程如图 3 所示，图 3 示出根据一个实施例的文本与动作匹配方法的流程图。

20 在步骤 302，针对任意目标词语，将该目标词语与所述第一动作集合中动作对应的触发词进行匹配，得到第一候选动作集合。

在一个实施例中，还会获取用户在指定的文本位置插入特定的动作的用户自定义动作集合，并将用户自定义动作集合添加到第一候选动作集合中。

在步骤 304，根据该目标词语的语义表征与所述第一动作集合中动作的动作表征的相似度，得到第二候选动作集合。

25 在一个实施例中，使用第一文本编码器对目标词语进行编码，得到目标语义表征；获取使用动作编码器对第一动作集合中的任一动作进行编码得到的动作表征；将动作表征与目标语义表征的相似度大于第一阈值的动作，归入第二候选动作集合。

通过对第一文本编码器进行训练，使得第一文本编码器生成的语义表征和动作编码器生成的动作表征位于同一个表征空间中，进而可以计算两个编码器生成的表征之间的相似

30 度。

在一个实施例中，会存在已经预先经过联合训练的第二文本编码器和动作编码器，使得二者生成的表征位于同一个表征空间中。第二文本编码器与第一文本编码器所编码的语言种类不同，例如第一文本编码器用于对第一语言的文本进行编码，第二文本编码器用于对第二语言的文本进行编码。此时，通过获取第一语言与第二语言的平行训练语料，即具有相同语义的两种语言的训练语料，使用第一文本编码器对第一语言的第一训练语料进行编码，得到第一语义表征，使用第二文本编码器对第二语言的第二训练语料进行编码，得到第二语义表征，第一训练语料与第二训练语料具有相同的语义，然后通过最大化第一语义表征和第二语义表征之间的相似度，调整第一文本编码器中的参数的值，使得第一文本编码器生成的语义表征和动作编码器生成的动作表征也位于同一个表征空间中。

在一个更加具体的实施例中，使用的第一语言为中文，第二语言为英文，第一文本编码器为中文的 BERT (Bidirectional Encoder Representation from Transformers) 编码器，第二文本编码器和动作编码器分别为 motionCLIP 的文本编码器和动作编码器。motionCLIP 在 CLIP (Contrastive Language-Image Pre-training, 对比文本-图像预训练模型) 的文本编码器和图像编码器的基础上训练了动作编码器，使得动作编码器可以与 CLIP 的文本编码器的表征空间对齐。由于 motionCLIP 的文本编码器只支持对英文进行编码，当数字人播报的文本为中文时，则需要再训练一个支持中文的编码器，使其与 motionCLIP 的动作编码器的表征空间进行对齐。通过使用中英文平行训练语料，最大化英文 motionCLIP 文本编码器得到的英文表征与中文 BERT 编码器得到的中文表征之间的余弦相似度，对中文 BERT 编码器进行微调 (fine-tuning)，使得中文 BERT 编码器的表征空间与 motionCLIP 的动作编码器的表征空间对齐。在其他实施例中，还可以使用其它相似度计算方法计算英文 motionCLIP 文本编码器得到的英文表征与中文 BERT 编码器得到的中文表征之间的相似度，这里不做限定。

在步骤 306，根据所述第一候选动作集合与所述第二候选动作集合，确定目标词语对应的备选动作集合。

通过对第一候选动作集合与第二候选动作集合求交集或者并集，确定备选动作集合。也可以根据实际需要，对上述两个集合进行其他集合运算，或者从任一集合中选取一部分动作组成备选动作集合，这里不做限定。

应当说明的是，在一些实施例中，由于第一动作集合中的动作种类可能较少，所以无法做到对第一词语序列中任意一个词语都匹配到对应的动作的情况，此时，存在一些词语，它们对应的备选动作集合为空集。

在其他实施例中，还可以使用其它匹配方法将第一词语序列中的各个词语与第一动作集合进行匹配，例如，可以基于统计学统计人在说某些特定词语时最常出现的动作，根据统计频率对词语和动作进行匹配，这里不做限定。

步骤 202 和步骤 204 可以对应图 1 中的粗排阶段，粗排阶段的示意图可以如图 4 所示。

5 然后，回到图 2，在步骤 206，根据各个词语在所述目标文本中的位置，以及所述动作持续时长，从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，得到第一动作序列，所述第一动作序列用于指示所述数字人在播报所述目标文本时做出相应的动作。

10 可以使用多种方式确定一组在时间维度上没有重叠的动作序列。在一个实施例中，针对各个词语，从其对应的备选动作集合选择一个动作作为该词语的对应动作。第一词语序列中各个词语的对应动作可以依次排列成初始动作序列。然后，根据各个动作在时间轴上的重叠情况，从初始动作序列中剔除一些动作，将其排除在第一动作序列之外，最后形成没有重叠的第一动作序列。

15 下面以两个词语为例进行举例说明。具体的，所述第一词语序列中包括第一词语和第二词语，分别从第一词语和第二词语对应的备选动作集合中确定第一动作和第二动作，然后根据时间戳预测模型确定第一词语和第二词语在时间轴上的开始位置。时间戳预测模型可以是文本转语音模型 TTS (Text To Speech) 中的一个子模块，也可以是独立的模型，这里不做限定。在得到两个词语的开始位置后，分别将两个动作的开始位置设置为对应词语的开始位置，根据开始位置和动作持续时长，判断第一动作和第二动作在时间轴上是否有重叠，若有重叠，则将第一动作和第二动作之一从初始动作序列中剔除，即从第一动作序列排除。例如，可以保留开始位置在前的词语所对应的动作，排除开始位置在后的动作；又或者，当第一动作和第二动作还具有对应的优先级时，则保留优先级较高的动作，将优先级较低的动作从第一动作序列排除。使用本实施例中的方法可以处理任意两个在时间轴上有重叠的词语。

25 在另一个实施例中，首先如前所述将各个词语对应的动作按照时间开始顺序进行排序，得到动作序列 M。当动作具有预设的优先级时，则同时按照时间开始顺序和优先级进行排序，得到动作序列 M。然后对第一动作序列进行多轮迭代更新，每轮从动作序列 M 中选择一个动作添加到第一动作序列中，然后从动作序列 M 中移除多个动作，重复多轮迭代，直到动作序列 M 为空，其中，任一轮迭代包括：

30 选择动作序列 M 中排在最前面的动作，将其添加到第一动作序列中；

根据该动作的开始位置，将其放置在时间轴上；

根据该动作的开始位置和动作持续时长，选择动作序列 M 中所有与该动作在时间轴上有重叠的动作，并将它们移除出动作序列 M，然后执行下一轮迭代。

通过本实施例中的方法可以确定一组在时间维度上没有重叠的动作序列。

5 在一个具体的例子中，如图 5 所示的精排阶段的时间轴示意图，包括文本序列和动作序列，词语和动作的持续时长由对应矩形的长度表示。输入的文本为：“你好，我是一个数字人。”，对其进行分词后，得到词语序列“你好”、“，”、“我”、“是”、“一个”、“数字人”、“。”，使用如前所述的方法为词语序列中的词语匹配相应的动作，并确定动作对应的开始位置，组成动作序列 M。

10 然后，首先，从动作序列 M 中选择最前面的动作：动作 1，将动作 1 添加到第一动作序列中，并放置在时间轴上，可以看到，动作 1 的持续时长覆盖了词语“你好”、“，”和“我”，这些词语对应的动作会与动作 1 重叠，于是把它们对应的动作从动作序列 M 中移除。接下来，从动作序列 M 中剩余的动作中选择最前面的动作：动作 2，将动作 2 添加到第一动作序列中，并放置在时间轴上，动作 2 的持续时长覆盖了词语“是”和“一个”，
15 于是把它们对应的动作从动作序列 M 中移除。然后从动作序列 M 中剩余的动作中选择最前面的动作：动作 3，将动作 3 添加到第一动作序列中，并放置在时间轴上，并将词语“数字人”、“。”对应的动作从动作序列 M 中移除。此时动作序列 M 为空，结束动作编排迭代，得到第一动作序列：动作 1、动作 2、动作 3。

在其他实施例中，还可以使用其它方法从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，只要没有重叠即可。例如，将动作持续时长的中点放置于对应词语的开始位置，也就是说，当数字人还没有播报该词语时，便已经开始做出该词语对应的动作。然后依照这种放置方式确定一组在时间维度上没有重叠的动作序列。

步骤 206 可以对应图 1 中的精排阶段。通过图 2 示出的数字人动作编排方法，已经可以得到一组在时间上没有冲突的数字人动作序列，满足动作编排的需求。但是，在一些实
25 施例中，并不是文本中的所有的词语都能够第一动作集合中找到匹配的动作，这就可能导致在粗排和精排产生的第一动作序列中，相邻动作之间存在长时间空白（即静默动作）的情况，导致在文本播报过程中数字人动作不够丰富、用户体验较差的问题。

为了提升动作序列的丰富度，在一些可能的实施方式中，还包括：

在步骤 208，获取第二动作集合，所述第二动作集合中包含若干通用动作，任一通用
30 动作具有对应的动作持续时长。

通用动作是指没有配置触发词的弱语义动作，这些动作配合任何播报文案都不会显得突兀，因此适合作为第一动作集合中强语义动作之间的承接动作，从而提升整个动作编排序列中动作的丰富程度。

在步骤 210，提取所述目标文本中词性为目标词性的词语，得到第二词语序列。

- 5 在一个实施例中，对目标文本进行句法解析，提取所述目标文本中词性为目标词性的词语，目标词性包括以下中的一种或多种：数量词、名动词、动副词、普通动词、时间、方位词。

在步骤 212，对所述第二词语序列中的任一词语，从所述第二动作集合中随机选择对应的通用动作，得到第二匹配结果。

- 10 在一个实施例中，为了使数字人不连续做出相同的通用动作，从所述第二动作集合中随机选择通用动作，使得所述第二词语序列中相邻的词语之间对应不同的通用动作，得到第二匹配结果。

- 在步骤 214，根据第二词语序列中各个词语在所述目标文本中的位置，以及各个词语对应的通用动作的动作持续时长，从所述第二匹配结果中确定一组动作序列并插入到所述
15 第一动作序列中，得到第二动作序列，使得所述第二动作序列中的各个动作在时间维度上没有重叠。

- 在一个实施例中，第二词语序列中包括第三词语。首先根据时间戳预测模型确定所述第三词语在时间轴上的开始位置，根据第二匹配结果确定第三词语对应的第三动作，然后根据所述开始位置和第三动作的动作持续时长，判断所述第三动作和所述第一动作序列中的任一动作在时间轴上是否有重叠，若没有，则将第三动作添加到所述第一动作序列中。
20

在其他实施例中，还可以使用其它方法从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列中，只要没有重叠即可。例如，将动作持续时长的中点放置于对应词语的开始位置，然后依照这种放置方式确定一组在时间维度上与第一动作序列没有重叠的动作序列并插入到第一动作序列中。

- 25 步骤 208 至步骤 214 可以对应图 1 中的重排阶段。通过步骤 208 至步骤 214，可以避免相邻动作之间存在长时间空白的情况。

- 本说明书一个或多个实施例提出基于预设动作库进行粗排-精排-重排的方式来确定数字人在播报文本时的动作序列。这种基于预设动作库的方式，能够保证候选动作的质量，从而避免输出动作序列中存在低质量动作和动作不连贯的问题。通过人工预设动作触发词
30 的方式来提升动作编排结果的语义一致性和可控性，同时度量文本表征和动作表征的推理

方法，进一步提升编排的动作序列与输入文本的语义一致性。针对预设动作库中动作存在一定的持续时长，导致编排的动作序列可能在音频时间轴上重叠的问题，提出了从词语开始位置向后扩展的时间轴对齐方法，从而在保证动作序列与文本语义一致性的基础上，避免动作在播放时重叠导致不可用的问题。考虑到基于语义匹配的方法得到的动作序列存在
5 的稀疏问题，提出了基于句法解析的方法在特定的句法结构中插入通用承接动作，从而提升编排的动作序列的丰富性。

根据另一方面的实施例，还提供一种数字人动作编排装置。图 6 示出根据一个实施例的数字人动作编排装置的示意性框图，该装置可以部署在任何具有计算、处理能力的设备、平台或设备集群中。如图 6 所示，该装置 600 包括：

10 第一获取单元 601，配置为，获取目标文本对应的第一词语序列，以及获取第一动作集合，所述第一动作集合中包含若干动作，任一动作具有预设的触发词以及对应的动作持续时长；

第一匹配单元 602，配置为，至少根据所述触发词，将所述第一词语序列中的各个词语与所述第一动作集合进行匹配，得到第一匹配结果，其中包含各个词语对应的备选动作
15 集合；

第一确定单元 603，配置为，根据各个词语在所述目标文本中的位置，以及所述动作持续时长，从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，得到第一动作序列，所述第一动作序列用于指示所述数字人在播报所述目标文本时做出相应的动作。

在一些可能的实施方式中，装置 600 还包括：

20 第二获取单元 604，配置为，获取第二动作集合，所述第二动作集合中包含若干通用动作，任一通用动作具有对应的动作持续时长；

词语提取单元 605，配置为，提取所述目标文本中词性为目标词性的词语，得到第二词语序列；

25 第二匹配单元 606，配置为，对所述第二词语序列中的任一词语，从所述第二动作集合中随机选择对应的通用动作，得到第二匹配结果；

第二确定单元 607，配置为，根据第二词语序列中各个词语在所述目标文本中的位置，以及各个词语对应的通用动作的动作持续时长，从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列中，得到第二动作序列，使得所述第二动作序列中的各个动作在时间维度上没有重叠。

30 根据另一方面的实施例，还提供一种计算机可读存储介质，其上存储有计算机程序，

当所述计算机程序在计算机中执行时，令计算机执行上述任一实施例所描述的方法。

根据再一方面的实施例，还提供一种计算设备，包括存储器和处理器，其中，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现上述任一实施例所描述的方法。

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。尤其，对于装置实施例而言，由于其基本相似于方法实施例，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。

10 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

15 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、物品或者装置不仅包括那些要素，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、物品或者装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括要素的过程、方法、物品或者装置中还存在另外的相同要素。

20 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

25 以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

权 利 要 求 书

1. 一种数字人动作编排方法，包括：
- 获取目标文本对应的第一词语序列，以及获取第一动作集合，所述第一动作集合中包含若干动作，任一动作具有预设的触发词以及对应的动作持续时长；
- 至少根据所述触发词，将所述第一词语序列中的各个词语与所述第一动作集合进行匹配，得到第一匹配结果，其中包含各个词语对应的备选动作集合；
- 根据各个词语在所述目标文本中的位置，以及所述动作持续时长，从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，得到第一动作序列，所述第一动作序列用于指示所述数字人在播报所述目标文本时做出相应的动作。
2. 根据权利要求1所述的方法，其中，将所述第一词语序列中的各个词语与所述第一动作集合进行匹配，得到第一匹配结果，包括：
- 针对任意目标词语，将该目标词语与所述触发词进行匹配，得到第一候选动作集合；
- 根据该目标词语的语义表征与所述第一动作集合中动作的动作表征的相似度，得到第二候选动作集合；
- 根据所述第一候选动作集合与所述第二候选动作集合，确定目标词语对应的备选动作集合。
3. 根据权利要求2所述的方法，其中，根据该目标词语的语义表征与所述第一动作集合中动作的动作表征的相似度，得到第二候选动作集合，包括：
- 使用第一文本编码器对所述目标词语进行编码，得到目标语义表征；
- 获取使用动作编码器对所述第一动作集合中的任一动作进行编码得到的动作表征；
- 将动作表征与所述目标语义表征的相似度大于第一阈值的动作，归入第二候选动作集合。
4. 根据权利要求3所述的方法，其中，所述第一文本编码器用于对第一语言的文本进行编码；所述动作编码器预先与适用于第二语言的第二文本编码器联合训练；所述第一文本编码器通过以下方式训练得到：
- 使用所述第一文本编码器对第一语言的第一训练语料进行编码，得到第一语义表征；
- 使用所述第二文本编码器对第二语言的第二训练语料进行编码，得到第二语义表征，所述第二训练语料与所述第一训练语料具有相同的语义；
- 通过最大化所述第一语义表征和第二语义表征之间的相似度，调整所述第一文本编码器中的参数的值。

5. 根据权利要求2所述的方法, 其中, 根据所述第一候选动作集合与所述第二候选动作集合, 确定目标词语对应的备选动作集合, 包括:

根据所述第一候选动作集合与第二候选动作集合的并集, 确定备选动作集合; 或

根据所述第一候选动作集合与第二候选动作集合的交集, 确定备选动作集合。

6. 根据权利要求1所述的方法, 其中, 所述第一词语序列中包括第一词语和第二词语; 从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列, 得到第一动作序列, 包括:

分别从所述第一词语和第二词语对应的备选动作集合中确定第一动作和第二动作;

根据时间戳预测模型确定所述第一词语和第二词语在时间轴上的开始位置;

根据所述开始位置和动作持续时长, 判断所述第一动作和第二动作在时间轴上是否有重叠; 若有, 则将第一动作和第二动作之一从所述第一动作序列中排除。

7. 根据权利要求6所述的方法, 其中, 将第一动作和第二动作之一从所述第一动作序列中排除, 包括:

将开始位置在后的词语所对应的动作, 从所述所述第一动作序列中排除。

8. 根据权利要求6所述的方法, 其中, 所述第一动作和第二动作还具有对应的优先级; 将第一动作和第二动作之一从所述第一动作序列中排除, 包括:

将优先级较低的动作从所述所述第一动作序列中排除。

9. 根据权利要求1所述的方法, 还包括:

获取第二动作集合, 所述第二动作集合中包含若干通用动作, 任一通用动作具有对应的动作持续时长;

提取所述目标文本中词性为目标词性的词语, 得到第二词语序列;

对所述第二词语序列中的任一词语, 从所述第二动作集合中随机选择对应的通用动作, 得到第二匹配结果;

根据第二词语序列中各个词语在所述目标文本中的位置, 以及各个词语对应的通用动作的动作持续时长, 从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列中, 得到第二动作序列, 使得所述第二动作序列中的各个动作在时间维度上没有重叠。

10. 根据权利要求9所述的方法, 其中, 所述目标词性包括以下中的一种或多种: 数量词、名动词、动副词、普通动词、时间、方位词。

11. 根据权利要求9所述的方法, 其中, 对所述第二词语序列中的任一词语, 从所述第二动作集合中随机选择对应的通用动作, 得到第二匹配结果, 包括:

从所述第二动作集合中随机选择通用动作，使得所述第二词语序列中相邻的词语之间对应不同的通用动作，得到第二匹配结果。

12. 根据权利要求 9 所述的方法，其中，所述第二词语序列中包括第三词语；从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列中，得到第二动作序列，包括：

根据时间戳预测模型确定所述第三词语在时间轴上的开始位置；

根据所述第二匹配结果确定所述第三词语对应的第三动作；

根据所述开始位置和第三动作的动作持续时长，判断所述第三动作和所述第一动作序列中的任一动作在时间轴上是否有重叠，若没有，则将第三动作添加到所述第一动作序列中。

13. 一种数字人动作编排装置，包括：

第一获取单元，配置为，获取目标文本对应的第一词语序列，以及获取第一动作集合，所述第一动作集合中包含若干动作，任一动作具有预设的触发词以及对应的动作持续时长；

第一匹配单元，配置为，至少根据所述触发词，将所述第一词语序列中的各个词语与所述第一动作集合进行匹配，得到第一匹配结果，其中包含各个词语对应的备选动作集合；

第一确定单元，配置为，根据各个词语在所述目标文本中的位置，以及所述动作持续时长，从所述第一匹配结果中确定一组在时间维度上没有重叠的动作序列，得到第一动作序列，所述第一动作序列用于指示所述数字人在播报所述目标文本时做出相应的动作。

14. 根据权利要求 13 所述的装置，还包括：

第二获取单元，配置为，获取第二动作集合，所述第二动作集合中包含若干通用动作，任一通用动作具有对应的动作持续时长；

词语提取单元，配置为，提取所述目标文本中词性为目标词性的词语，得到第二词语序列；

第二匹配单元，配置为，对所述第二词语序列中的任一词语，从所述第二动作集合中随机选择对应的通用动作，得到第二匹配结果；

第二确定单元，配置为，根据第二词语序列中各个词语在所述目标文本中的位置，以及各个词语对应的通用动作的动作持续时长，从所述第二匹配结果中确定一组动作序列并插入到所述第一动作序列中，得到第二动作序列，使得所述第二动作序列中的各个动作在时间维度上没有重叠。

15. 一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机

中执行时，令计算机执行权利要求 1-12 中任一项所述的方法。

16. 一种计算设备，包括存储器和处理器，其中，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-12 中任一项所述的方法。