

(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
2019 年 12 月 19 日 (19.12.2019)



(10) 国际公布号
WO 2019/237708 A1

- (51) 国际专利分类号:
G10L 15/16 (2006.01)
- (21) 国际申请号: PCT/CN2018/122942
- (22) 国际申请日: 2018 年 12 月 22 日 (22.12.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810623961.6 2018年6月15日 (15.06.2018) CN
- (71) 申请人: 山东大学 (SHANDONG UNIVERSITY) [CN/CN]; 中国山东省济南市高新区舜华路 1500 号, Shandong 250101 (CN)。
- (72) 发明人: 甘甜 (GAN, Tian); 中国山东省济南市高新区舜华路 1500 号, Shandong 250101 (CN)。 马志鑫 (MA, Zhixin); 中国山东省济南市高新区

舜华路 1500 号, Shandong 250101 (CN)。 宋雪萌 (SONG, Xuemeng); 中国山东省济南市高新区舜华路 1500 号, Shandong 250101 (CN)。 聂礼强 (NIE, Liqiang); 中国山东省济南市高新区舜华路 1500 号, Shandong 250101 (CN)。

(74) 代理人: 济南圣达知识产权代理有限公司 (JINAN SHENGDA INTELLECTUAL PROPERTY AGENCY CO., LTD); 中国山东省济南市经十路 17703 号华特广场 B308 室, Shandong 250061 (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, 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,

(54) **Title:** INTERPERSONAL INTERACTION BODY LANGUAGE AUTOMATIC GENERATION METHOD AND SYSTEM BASED ON DEEP LEARNING

(54) 发明名称: 基于深度学习的人际交互肢体语言自动生成方法及系统

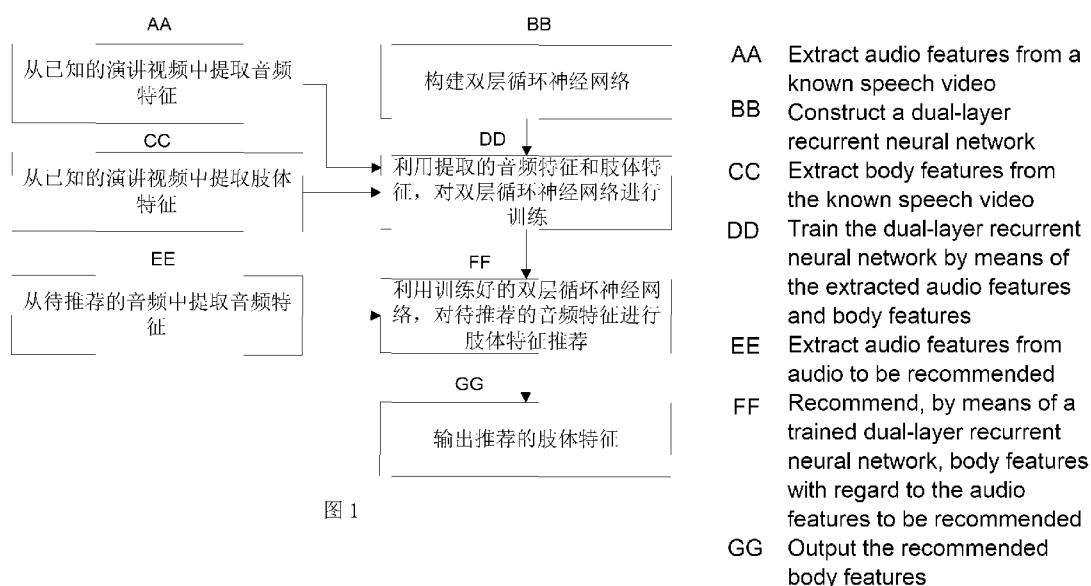


图 1

(57) **Abstract:** Disclosed are an interpersonal interaction body language automatic generation method and system based on deep learning. The method comprises: step (1): extracting, from a speech video, audio training features of a character within a time range that corresponds to each frame of image; and at the same time, extracting, from the speech video, body training features of the character in each frame of image, and establishing a one-to-one correlation between the audio training features and the body training features; step (2): constructing a dual-layer recurrent neural network model; and taking the audio training features as an input value of the dual-layer recurrent neural network model, taking the body training features as an output value of the dual-layer recurrent neural network model, and training the constructed dual-layer recurrent neural network model to obtain a trained dual-layer recurrent neural network model; step (3): extracting audio test features from audio that has a body action to be recommended; and step (4): inputting the audio test

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))。

features into the trained dual-layer recurrent neural network model to output recommended body test features.

(57) 摘要: 一种基于深度学习的人际交互肢体语言自动生成方法及系统, 包括: 步骤(1): 从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征; 同时, 从演讲视频中提取每一帧图像内人物的肢体训练特征, 建立音频训练特征与肢体训练特征之间的一一对应关系; 步骤(2): 构建双层循环神经网络模型; 将音频训练特征作为双层循环神经网络模型的输入值, 将肢体训练特征作为双层循环神经网络模型的输出值, 对已构建的双层循环神经网络模型进行训练, 得到训练好的双层循环神经网络模型; 步骤(3): 从待推荐肢体动作的音频中提取音频测试特征; 步骤(4): 将音频测试特征输入到训练好的双层循环神经网络模型中, 输出推荐的肢体测试特征。

基于深度学习的人际交互肢体语言自动生成方法及系统

技术领域

本发明涉及人工智能技术领域，特别是涉及基于深度学习的人际交互肢体语言自动生成方法及系统。

背景技术

演讲是给观众传递信息的最有效方式之一。除了演讲者的演说内容外，在其演讲过程中演讲者的肢体语言对信息传达也十分重要。通常，不同的演讲者在演讲过程中的表达方式有很大不同。其中，手势作为一种视觉感知，与演说一同表达演讲者的思想。在演讲过程中，手势不仅帮助演讲者丰富其语言内容的画面感，帮助传达情绪，有时还能传达言语无法表达的信息（例如，好的站姿和肢体动作能表现一个人的精神面貌与气质，这是语言所无法传达的）。所以，好的肢体语言能帮助演讲者更好的诠释主题。通常，提高演讲水平最有效方法之一是通过专家的反馈和建议来练习，但雇用专家是昂贵的，因此大部分情况下不太实际。

随着大数据和人工智能技术的兴起，人们对多媒体数据的处理能力飞速增长。在演讲领域，人们对演讲的深度分析的需求越来越大，但始终没能得到很好的解决。在此之前，人们进行了很多诸如演讲打分，语音行为分析等一系列研究。但这些研究结果一方面需要人工标注，只能在小批量数据上进行。另一方面，没有给出具体的建议，无论是分数还是一系列的统计信息，都不能直观的给用户明了的建议。

发明内容

为了解决现有技术的不足，本发明提供了基于深度学习的人际交互肢体语言自动生成方法及系统；

本发明第一方面，提供了基于深度学习的人际交互肢体语言自动生成方法；

基于深度学习的人际交互肢体语言自动生成方法，包括：

步骤（1）：从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征；同时，从演讲视频中提取每一帧图像内人物的肢体训练特征，建立音频训练特征与肢体训练特征之间的的一一对应关系；

步骤（2）：构建双层循环神经网络模型；将音频训练特征作为双层循环神经网络模型的输入值，将肢体训练特征作为双层循环神经网络模型的输出值，对已构建的双层循环神经网络模型进行训练，得到训练好的双层循环神经网络模型；

步骤（3）：从待推荐肢体动作的音频中提取音频测试特征；

步骤(4): 将音频测试特征输入到训练好的双层循环神经网络模型中, 输出推荐的肢体测试特征。

作为本发明的进一步改进, 所述从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征; 是指:

将演讲视频的音频信号均匀划分为长度为 $\frac{1}{\text{framerate}}$ 秒的语音区间, 其中framerate为每秒的帧数; 对每个语音区间提取声音强度特征, 通过音频信号转为文本, 对文本进行分类得到感情类别特征; 将声音强度特征和感情类别特征进行特征融合作为音频训练特征。

作为本发明的进一步改进, 所述从演讲视频中提取每一帧图像内人物的肢体训练特征; 是指:

将演讲视频的视频信号以帧为单位划分为若干帧图像, 对每一帧图像使用肢体语言识别系统 OpenPose 检测出图像中人物各关节点坐标位置, 对各关节点坐标位置进行归一化后, 得到每一帧图像的人物关节与关节坐标位置关系矩阵; 根据每一帧图像的人物关节与关节坐标位置关系矩阵计算关节夹角的角度, 根据关节夹角的角度将当前帧图像映射到预先设定活动状态中最接近的一个状态作为肢体训练特征。

所述关节夹角是指: 在头部位置预先设定一个点为头部节点, 在颈部位置预先设定一个点为颈部节点, 在右肩关节位置预先设定一个点为右肩关节点, 在左肩关节位置预先设定一个点为左肩关节点, 在右胳膊肘位置预先设定一个点为右肘关节点, 在左胳膊肘位置预先设定一个点为左肘关节点, 在右手位置预先设定一个点为右手关节点, 在左手位置预先设定一个点为左手关节点;

第一个夹角是指: 头部节点与颈部节点连线和颈部节点与右肩关节点连线之间的夹角 α^H ;

第二个夹角是指: 颈部节点与右肩关节点连线和右肩关节点与右肘关节点连线之间的身体内侧夹角 α^{RS} ;

第三个夹角是指: 右肩关节点与右肘关节点连线和右肘关节点与右手关节点连线之间的身体内侧夹角 α^{RE} ;

第四个夹角是指: 颈部节点与左肩关节点连线和左肩关节点与左肘关节点连线之间的身体内侧夹角 α^{LS} ;

第五个夹角是指: 左肩关节点与左肘关节点连线和左肘关节点与左手关节点连线之间的身体内侧夹角 α^{LE} ;

预先设定活动状态, 包括: 第一个夹角设置 m 个状态, 第二个夹角设置 n 个状态, 第三

个夹角设置 o 个状态，第四个夹角设置 p 个状态，第五个夹角设置 q 个状态；根据各个夹角的状态的组合，将人物姿态划分为 $m*n*o*p*q$ 个状态。

例如：第一个夹角设置三个状态：

$$\alpha^H \in [60^\circ, 90^\circ, 120^\circ];$$

例如：第二个夹角设置四个状态：

$$\alpha^{RS} \in [60^\circ, 112.5^\circ, 157.5^\circ, 210^\circ];$$

例如：第三个夹角设置五个状态：

$$\alpha^{RE} \in [90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ];$$

例如：第四个夹角设置四个状态：

$$\alpha^{LS} \in [60^\circ, 112.5^\circ, 157.5^\circ, 210^\circ];$$

例如：第五个夹角设置五个状态：

$$\alpha^{LE} \in [90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ].$$

这里将人物状态划分为 1200 种状态。

所述肢体包括上肢和头部。

作为本发明的进一步改进，所述双层循环神经网络模型包括两个嵌套组合循环神经网络：第一循环神经网络和第二循环神经网络；第二循环神经网络在第一循环神经网络内部；第一循环神经网络包括第一编码器和第一解码器；第二循环神经网络包括第二编码器和第二解码器；第一编码器、第二编码器、第二解码器和第一解码器依次串联；第一编码器的输入值为音频训练特征，第一编码器的输出值输入到第二编码器中，第二编码器的输出值输入到第二解码器中，第二解码器的输出值输入到第一解码器中，第一解码器的输出值为肢体训练特征；

作为本发明的进一步改进，所述步骤（2）中，将音频训练特征作为双层循环神经网络模型的输入值，将肢体训练特征作为双层循环神经网络模型的输出值，对已构建的双层循环神经网络模型进行训练，得到训练好的双层循环神经网络模型，步骤为：

步骤（201）：根据声音强度将音频信号划分为若干个声音行为，划分过程中将声音强度超过设定阈值的连续音频信号视为一个声音行为，将声音强度未超过设定阈值的连续音频信号视为一个声音行为，每个声音行为包括若干个语音区间，每个语音区间都有对应的音频训练特征；将每个声音行为对应的音频训练特征输入到第一编码器中，第一编码器输出每个音频训练特征的隐藏层状态；

第一编码器输出每个音频训练特征的隐藏层状态：

$$h_{i,u} = GRU_{enc}^{unit}(h_{i,u-1}, x_{i,u}), u = 1, \dots, N_i,$$

其中, $x_{i,u}$ 代表第 i 个声音行为的第 u 个语音区间, $h_{i,u}$ 指对 $x_{i,u}$ 编码所得的隐藏层状态, 特别的, 记最终的隐藏层状态 h_{i,N_i} 为 h_i 。 GRU_{enc}^{unit} 是第一编码器; N_i 为第 i 个声音行为所具有的语音区间个数。

步骤 (202): 对于第一编码器输出的每个音频训练特征的隐藏层状态, 采用第二编码器进行编码, 得到第二编码器的编码结果:

$$l_v = GRU_{enc}^{seq}(l_{v-1}, h_v), v = 1, \dots, N_v,$$

其中 $l_v \in \mathbb{R}^{d_h}$, 表示每个声音行为的隐藏层状态, $l_0 = 0$, d_h 为隐藏层状态的向量空间的维度。第二编码器的最终编码结果 l_{N_v} 记为 c 。

步骤 (203): 将每个声音行为对应的肢体训练特征集合视为肢体动作行为; 将第二编码器每次输出的隐藏层状态作为与声音行为对应的肢体动作行为的隐藏层状态;

步骤 (203) 第二解码器解码输出隐表达 o_i 的条件概率 $P(o_i|f_{i-1}, c)$ 表示为:

$$P(o_i|f_{i-1}, c) = \prod_{i \in [0, N_x]} (o_i|o_1, \dots, o_{i-1}, f_{i-1}, c),$$

其中 f_{i-1} 为第二解码器的最后一个隐藏层状态输出, 当 $i = 1$ 时, f_{i-1} 设置为 0, $o_1 \dots o_{i-1}$ 为之前所有第二解码器解码得到的肢体动作行为的隐表达, c 为所有声音行为的隐藏层状态。

步骤 (204): 将每一帧图像的肢体训练特征视为肢体动作单元, 在得到每个肢体动作行为的隐表达之后, 构建肢体动作行为与肢体动作行为包含的每一帧图像的关系, 并通过第一解码器将肢体动作单元的隐表达解码为每个语音区间对应的帧的肢体动作特征的预测, 选取概率最大的肢体动作作为预测的肢体动作;

步骤 (204) 预测的概率 $P(f_{i,j}|g_{i,j}, o_i)$ 计算如下,

$$P(f_{i,j}|g_{i,j}, o_i) = \prod_{i \in [0, N_i]} (f_{i,j}|f_{i,1}, \dots, f_{i,j-1}, g_{i,j}, o_i),$$

其中 $f_{i,j}$ 表示 $x_{i,j}$ 所对应的预测值, $g_{i,j}$ 表示 $x_{i,j}$ 所对应的真实值, o_i 为解码出的肢体动作行为的隐表达;

步骤 (205): 利用损失函数计算预测的肢体动作特征与肢体训练特征之间的损失值, 如果损失值在设定范围内, 则表示双层循环神经网络模型训练合格, 进入步骤 (3); 否则表示双层循环神经网络模型训练不合格, 调整双层循环神经网络模型的参数后, 返回步骤 (201) 继续训练, 直至损失值在设定范围内, 结束。

调整双层循环神经网络模型的参数，例如：GRU 函数中的 $W_z, W_r, W_h, U_z, U_r, U_h$ ，以及 b_z, b_r, b_h 。

作为本发明的进一步改进，所述步骤（3）中，从待推荐肢体动作的音频中提取音频测试特征，步骤为：

对待测试音频信号划分为长度为 $\frac{1}{\text{framerate}}$ 秒的语音区间，其中 framerate 为每秒的帧数；对每个语音区间提取声音强度特征；将待测试音频信号转为文本，对文本进行分类得到感情类别特征；将声音强度特征和感情类别特征进行特征融合作为音频测试特征。

作为本发明的进一步改进，所述步骤（4）中，根据声音强度将音频信号划分为若干个声音行为，划分过程中将声音强度超过设定阈值的连续音频信号视为一个声音行为，将声音强度未超过设定阈值的连续音频信号视为一个声音行为，每个声音行为包括若干个语音区间，每个语音区间都有对应的音频测试特征；将每个声音行为对应的音频测试特征输入到训练好的双层循环神经网络模型中，输出推荐的肢体测试特征，对肢体测试特征进行展示。

本发明的第二方面，提供了基于深度学习的人际交互肢体语言自动生成系统；

基于深度学习的人际交互肢体语言自动生成系统，包括：

存储器、处理器以及存储在存储器上并在处理器上运行的计算机指令，所述计算机指令被处理器运行时，完成上述任一方法所述的步骤。

本发明的第三方面，提供了一种计算机可读存储介质；

一种计算机可读存储介质，其上存储有计算机指令，所述计算机指令被处理器运行时，完成上述任一方法所述的步骤。

与现有技术相比，本发明的有益效果是：

1、本发明能够利用数据分析方法以及人工智能技术为演讲者提供更为准确的肢体动作指导，使其能够与演讲内容深度结合，更好的表达交流内容，提高人际沟通的效率。

2、相比现有技术给出评分和各种统计信息的方式，本发明能给出更直观的建议。

3、设计了一个分层网络结构。一方面，网络结构中的单元层通过对声音行为和肢体动作内部连续的语音区间同时进行编码，保证了动作流的连续性。另一方面，网络结构中的序列层对不同的声音行为所产生的肢体动为以及不同肢体动作之间的概率转移关系进行建模，保证了肢体动作的完整性和丰富性。

附图说明

构成本申请的一部分的说明书附图用来提供对本申请的进一步理解，本申请的示意性实

施例及其说明用于解释本申请，并不构成对本申请的不当限定。

图 1 为本发明的流程图；

图 2 (a) 和图 2 (b) 为肢体动作提取效果图；

图 3 为模型流程图；

图 4 为编码器部分示意图；

图 5 为解码器部分示意图；

图 6 为声音行为检测示意图；

图 7 (a) ~图 7 (d) 为关节角度示意图。

具体实施方式

应该指出，以下详细说明都是例示性的，旨在对本申请提供进一步的说明。除非另有指明，本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解相同含义。

需要注意的是，这里所使用的术语仅是为了描述具体实施方式，而非意图限制根据本申请的示例性实施方式。如在这里所使用的，除非上下文另外明确指出，否则单数形式也意图包括复数形式，此外，还应当理解的是，当在本说明书中使用术语“包含”和/或“包括”时，其指明存在特征、步骤、操作、器件、组件和/或它们的组合。

本发明第一个实施例，提供了基于深度学习的人际交互肢体语言自动生成方法；

如图 1 所示，基于深度学习的人际交互肢体语言自动生成方法，包括：

步骤(1)：从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征；同时，从演讲视频中提取每一帧图像内人物的肢体训练特征，建立音频训练特征与肢体训练特征之间的一一对应关系；

作为本发明的进一步改进，所述从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征；是指：

将演讲视频的音频信号均匀划分为长度为 $\frac{1}{\text{framerate}}$ 秒的语音区间，其中framerate为每秒的帧数；对每个语音区间提取声音强度特征，通过音频信号转为文本，对文本进行分类得到感情类别特征；将声音强度特征和感情类别特征进行特征融合作为音频训练特征。

作为本发明的进一步改进，所述从演讲视频中提取每一帧图像内人物的肢体训练特征；是指：

将演讲视频的视频信号以帧为单位划分为若干帧图像，对每一帧图像使用肢体语言识别

系统 OpenPose 检测出图像中人物各关节坐标位置，对各关节坐标位置进行归一化后，得到每一帧图像的人物关节与关节坐标位置关系矩阵；根据每一帧图像的人物关节与关节坐标位置关系矩阵计算关节夹角的角度，根据关节夹角的角度将当前帧图像映射到预先设定活动状态中最接近的一个状态作为肢体训练特征。

所述关节夹角是指：在头部位置预先设定一个点为头部节点，在颈部位置预先设定一个点为颈部节点，在右肩关节位置预先设定一个点为右肩关节节点，在左肩关节位置预先设定一个点为左肩关节节点，在右胳膊肘位置预先设定一个点为右肘关节节点，在左胳膊肘位置预先设定一个点为左肘关节节点，在右手位置预先设定一个点为右手关节节点，在左手位置预先设定一个点为左手关节节点；

第一个夹角是指：头部节点与颈部节点连线和颈部节点与右肩关节节点连线之间的夹角 α^H ；

第二个夹角是指：颈部节点与右肩关节节点连线和右肩关节节点与右肘关节节点连线之间的身体内侧夹角 α^{RS} ；

第三个夹角是指：右肩关节节点与右肘关节节点连线和右肘关节节点与右手关节节点连线之间的身体内侧夹角 α^{RE} ；

第四个夹角是指：颈部节点与左肩关节节点连线和左肩关节节点与左肘关节节点连线之间的身体内侧夹角 α^{LS} ；

第五个夹角是指：左肩关节节点与左肘关节节点连线和左肘关节节点与左手关节节点连线之间的身体内侧夹角 α^{LE} ；

预先设定活动状态，包括：第一个夹角设置 m 个状态，第二个夹角设置 n 个状态，第三个夹角设置 o 个状态，第四个夹角设置 p 个状态，第五个夹角设置 q 个状态；根据各个夹角的状态的组合，将人物姿态划分为 $m*n*o*p*q$ 个状态。

如图 7 (a) ~图 7 (d) 所示。例如：第一个夹角设置三个状态：

$$\alpha^H \in [60^\circ, 90^\circ, 120^\circ];$$

例如：第二个夹角设置四个状态：

$$\alpha^{RS} \in [60^\circ, 112.5^\circ, 157.5^\circ, 210^\circ];$$

例如：第三个夹角设置五个状态：

$$\alpha^{RE} \in [90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ];$$

例如：第四个夹角设置四个状态：

$$\alpha^{LS} \in [60^\circ, 112.5^\circ, 157.5^\circ, 210^\circ];$$

例如：第五个夹角设置五个状态：

$$\alpha^{LE} \in [90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ].$$

这里将人物状态划分为 1200 种状态。

所述肢体包括上肢和头部。

步骤 (2): 构建双层循环神经网络模型; 将音频训练特征作为双层循环神经网络模型的输入值, 将肢体训练特征作为双层循环神经网络模型的输出值, 对已构建的双层循环神经网络模型进行训练, 得到训练好的双层循环神经网络模型;

作为本发明的进一步改进, 所述双层循环神经网络模型包括两个嵌套组合循环神经网络: 第一循环神经网络和第二循环神经网络; 第二循环神经网络在第一循环神经网络内部; 第一循环神经网络包括第一编码器和第一解码器; 第二循环神经网络包括第二编码器和第二解码器; 第一编码器、第二编码器、第二解码器和第一解码器依次串联; 第一编码器的输入值为音频训练特征, 第一编码器的输出值输入到第二编码器中, 第二编码器的输出值输入到第二解码器中, 第二解码器的输出值输入到第一解码器中, 第一解码器的输出值为肢体训练特征;

作为本发明的进一步改进, 所述步骤 (2) 中, 将音频训练特征作为双层循环神经网络模型的输入值, 将肢体训练特征作为双层循环神经网络模型的输出值, 对已构建的双层循环神经网络模型进行训练, 得到训练好的双层循环神经网络模型, 步骤为:

步骤 (201): 根据声音强度将音频信号划分为若干个声音行为, 划分过程中将声音强度超过设定阈值的连续音频信号视为一个声音行为, 将声音强度未超过设定阈值的连续音频信号视为一个声音行为, 每个声音行为包括若干个语音区间, 每个语音区间都有对应的音频训练特征; 将每个声音行为对应的音频训练特征输入到第一编码器中, 第一编码器输出每个音频训练特征的隐藏层状态;

第一编码器输出每个音频训练特征的隐藏层状态:

$$h_{i,u} = GRU_{enc}^{unit}(h_{i,u-1}, x_{i,u}), u = 1, \dots, N_i,$$

其中, $x_{i,u}$ 代表第 i 个声音行为的第 u 个语音区间, $h_{i,u}$ 指对 $x_{i,u}$ 编码所得的隐藏层状态, 特别的, 记最终的隐藏层状态 h_{i,N_i} 为 h_i 。 GRU_{enc}^{unit} 是第一编码器; N_i 为第 i 个声音行为所具有的语音区间个数。

步骤 (202): 对于第一编码器输出的每个音频训练特征的隐藏层状态, 采用第二编码器进行编码, 得到第二编码器的编码结果:

$$l_v = GRU_{enc}^{seq}(l_{v-1}, h_v), v = 1, \dots, N_v,$$

其中 $l_v \in \mathbb{R}^{d_h}$, 表示每个声音行为的隐藏层状态, $l_0 = 0$, d_h 为隐藏层状态的向量空间的

维度。第二编码器的最终编码结果 l_{N_p} 记为 c 。

步骤（203）：将每个声音行为对应的肢体训练特征集合视为肢体动作行为；将第二编码器每次输出的隐藏层状态作为与声音行为对应的肢体动作行为的隐藏层状态，第二解码器解码输出隐表达 o_i 的条件概率 $P(o_i|f_{i-1}, c)$ 表示为：

$$P(o_i|f_{i-1}, c) = \prod_{i \in [0, N_x]} (o_i|o_1, \dots, o_{i-1}, f_{i-1}, c),$$

其中 f_{i-1} 为第二解码器的最后一个隐藏层状态输出，当 $i = 1$ 时， f_{i-1} 设置为0， $o_1 \dots o_{i-1}$ 为之前所有第二解码器解码得到的肢体动作行为的隐表达， c 为所有声音行为的隐藏层状态。

步骤（204）：将每一帧图像的肢体训练特征视为肢体动作单元，在得到每个肢体动作行为的隐表达之后，构建肢体动作行为与肢体动作行为包含的每一帧图像的关系，并通过第一解码器将肢体动作单元的隐表达解码为每个语音区间对应的帧的肢体动作特征的预测，选取概率最大的肢体动作作为预测的肢体动作；

预测的概率 $P(f_{i,j}|g_{i,j}, o_i)$ 计算如下，

$$P(f_{i,j}|g_{i,j}, o_i) = \prod_{i \in [0, N_i]} (f_{i,j}|f_{i,1}, \dots, f_{i,j-1}, g_{i,j}, o_i),$$

其中 $f_{i,j}$ 表示 $x_{i,j}$ 所对应的预测值， $g_{i,j}$ 表示 $x_{i,j}$ 所对应的真实值， o_i 为解码出的肢体动作行为的隐表达；步骤（205）：利用损失函数计算预测的肢体动作特征与肢体训练特征之间的损失值，如果损失值在设定范围内，则表示双层循环神经网络模型训练合格，进入步骤（3）；否则表示双层循环神经网络模型训练不合格，调整双层循环神经网络模型的参数后，返回步骤（201）继续训练，直至损失值在设定范围内结束。

调整双层循环神经网络模型的参数，例如：GRU 函数中的 $W_z, W_r, W_h, U_z, U_r, U_h$ ，以及 b_z, b_r, b_h 。

步骤（3）：从待推荐肢体动作的音频中提取音频测试特征；

作为本发明的进一步改进，所述步骤（3）中，从待推荐肢体动作的音频中提取音频测试特征，步骤为：

对待测试音频信号划分为长度为 $\frac{1}{\text{framerate}}$ 秒的语音区间，其中framerate为每秒的帧数；对每个语音区间提取声音强度特征或声音频率特征；将待测试音频信号转为文本，对文本进行分类得到感情类别特征；将声音强度特征和感情类别特征进行特征融合作为音频测试特征；

步骤（4）：将音频测试特征输入到训练好的双层循环神经网络模型中，输出推荐的肢体测试特征。

根据声音强度将音频信号划分为若干个声音行为，划分过程中将声音强度超过设定阈值的连续音频信号视为一个声音行为，将声音强度未超过设定阈值的连续音频信号视为一个声音行为，每个声音行为包括若干个语音区间，每个语音区间都有对应的音频测试特征；将每个声音行为对应的音频测试特征输入到训练好的双层循环神经网络模型中，输出推荐的肢体测试特征，对肢体测试特征进行展示。

本发明的第二个实施例，提供了基于深度学习的人际交互肢体语言自动生成系统；

基于深度学习的人际交互肢体语言自动生成系统，包括：

存储器、处理器以及存储在存储器上并在处理器上运行的计算机指令，所述计算机指令被处理器运行时，完成上述任一方法所述的步骤。

本发明的第三个实施例，提供了一种计算机可读存储介质；

一种计算机可读存储介质，其上存储有计算机指令，所述计算机指令被处理器运行时，完成上述任一方法所述的步骤。

本发明提出了一个双层循环编码器-解码器网络来编码语音和手势序列之间的关系以及手势层次结构。最后，当给定语音输入时，用学习出来的网络模型将生成适当的手势，并进行可视化以获得更有效的肢体语言推荐。

本发明一方面构建了大规模的数据库，且不需要人工标注，能在大量的数据中学习到更为准确的概率模型。另一方面，提出了用于分析语音和肢体语言之间关系的分层模型，能更好的模拟人在演讲时的状态。最后，我们将由音频生成的肢体动作可视化，直接展示给用户生成的结果。

在语言交流中，语言被认为是离散单元的有序组织，肢体语言也是如此。我们首先将模型的输入和输出，即音频和肢体动作离散化。通过对语音本身以及各种特征离散化为等长的单元作为语音的语音区间，对视频则将每一帧的肢体特征（见图 2（a）和图 2（b））作为一个动作语音区间。语音和视频的语音区间在时间轴上是一一对应的。在语义层面，我们将声音（动作）语音区间聚合为完整的声音（肢体动作），并分析完整的动作之间的概率转移关系。我们通过构建双层循环神经网络，对语音和动作在分别底层和语义层面分析建模，并建立输入和输出之间的映射关系，整个流程如图 3 所示。一方面我们保证了整个肢体动作的流畅性，另一方面我们确保了手势和手势之间根据输入音频的合理变换。

1. 基于语言结构模型的双层循环编码-解码网络包括：单元层面的编码、解码器，以及序列层面的编码、解码器。

我们定义 x 为视频中提取的语音特征集合， y 为视频中检测到的肢体的特征集合。

第一编码器：

为了将数量不同的语音区间分组并统一为等长的声音行为的表示，例如某一个声音行为包含 5 个语音区间，而另一段较长的声音行为由 10 个语音区间组成，我们要对这些语音区间进行编码，使用统一长度的表示方法。具体的，我们使用第一编码器进行编码：

$$h_{i,u} = GRU_{enc}^{unit}(h_{i,u-1}, x_{i,u}), u = 1, \dots, N_i,$$

其中 $x_{i,u}$ 代表第 i 个声音行为的第 u 个语音区间， N_i 为第 i 个声音行为所具有的语音区间个数。 $h_{i,u}$ 指对 $x_{i,u}$ 编码所得的隐藏状态， GRU_{enc}^{unit} 是编码函数。对于 GRU 函数，具体的，有：

$$\begin{aligned} z_t &= \sigma_g(W_z x_t + U_z h_{t-1} + b_z), \\ r_t &= \sigma_g(W_r x_t + U_r h_{t-1} + b_r), \\ \bar{h}_t &= \sigma_h(W_h x_t + U_h (r_t \cdot h_{t-1}) + b_h), \\ h_t &= (1 - z_t) \cdot h_{t-1} + z_t \cdot \bar{h}_t, \end{aligned}$$

其中 σ_g 为 sigmoid 函数， σ_h 为 tanh 函数。且有 $W_z, W_r, W_h \in \mathbb{R}^{d_u \times |x|}$ ， $U_z, U_r, U_h \in \mathbb{R}^{d_u \times d_u}$ ，以及 $b_z, b_r, b_h \in \mathbb{R}^{d_u}$ 。我们使用矩阵 W 来编码输入 x ，并以 h_t 作为输出的隐藏层状态。

最终，我们取每个声音行为输出的最后一个隐藏层状态作为整个声音行为的表示。此隐藏层状态包含了之前所有输入的信息，可以作为整个声音行为的所有语音区间的语义编码。

第二编码器：

当我们使用离散化的基本音频特征单元将不等长的声音行为表达为等长的特征表示 q_i 之后，我们得到了语义层面上的声音行为（如升高音调，语速加快等）的特征表示序列。对于这个序列，我们要建立不同的声音行为在大量数据下的概率转移关系。于是我们建立了第二层循环神经网络模型，用于对声音行为序列编码并得到整个序列的特定表示，

$$l_v = GRU_{enc}^{seq}(l_{v-1}, h_v), v = 1, \dots, N_v,$$

其中 $l_v \in \mathbb{R}^{d_h}$ ，表示每个声音行为的隐层表达，特殊的， $l_0 = 0$ ， d_h 为隐表达的向量空间的维度。整个编码过程如图 4 所示。

第二解码器：

当我们得到包含整个声音序列的编码信息的中间表示之后，便要建立声音行为和肢体语言的生成关系。对于整个声音序列的中间隐藏层状态 c ，我们按照与声音行为的时间区间相同的时间划分方式，将输出 y 分组，表示为与声音行为相对应的肢体动作行为，构建与第二编码器对应的解码器，并使解码器每次的输出隐藏层状态作为与此声音行为对应的肢体动作行为的隐表达，其概率表示为：

$$P(o_i|f_{i-1}, c) = \prod_{i \in [0, N_x]} (o_i|o_1, \dots, o_{i-1}, f_{i-1}, c),$$

其中 f_{i-1} 为前一个第一解码器的最后一个隐藏层状态输出，特别的，当 $i = 1$ 时， f_{i-1} 设置为 $GO_ID=0$ ， o_i 为之前所有解码得到的肢体动作行为的隐表达， c 为所有语音行为的中间隐变量。根据这一概率公式，我们便可以使用第二解码器解码出我们想要得到的肢体动作行为的隐表达。

第一解码器：

类似的，得到每个肢体动作行为的隐表达之后，我们便可以构建肢体动作行为与其包含的每一帧的生成关系，并通过单元层解码将肢体动作单元的隐表达解码为每个语音区间对应的帧，具体对应的概率计算如下，

$$P(f_{i,j}|g_{i,j}, o_i) = \prod_{i \in [0, N_i]} (f_{i,j}|f_{i,1}, \dots, f_{i,j-1}, g_{i,j}, o_i),$$

其中 $f_{i,j}$ ， $g_{i,j}$ 分别表示 $x_{i,j}$ 所对应的预测值和真实值， o_i 为输入的肢体动作行为的隐表达，整个解码过程如图5所示。

2. 模型训练

我们使用交叉熵来作为网络训练的损失函数，并通过梯度下降和反向传播来更新四个GRU函数，包括 GRU_{enc}^{unit} ， GRU_{enc}^{seq} ， GRU_{dec}^{seq} ， GRU_{dec}^{unit} 内的参数。其损失函数定义如下：

$$L = -\frac{1}{N_f} \sum_{i,j} g_{i,j} \log P(f_{i,j}) + (1 - g_{i,j}) \log P(1 - f_{i,j})$$

其中 N_f 为语音（肢体动作）语音区间的个数。

如图6所示，阴影的边界将声音序列划分为若干个声音行为。

以上所述仅为本申请的优选实施例而已，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

1、基于深度学习的人际交互肢体语言自动生成方法，其特征是，包括：

步骤（1）：从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征；同时，从演讲视频中提取每一帧图像内人物的肢体训练特征，建立音频训练特征与肢体训练特征之间的的一一对应关系；

步骤（2）：构建双层循环神经网络模型；将音频训练特征作为双层循环神经网络模型的输入值，将肢体训练特征作为双层循环神经网络模型的输出值，对已构建的双层循环神经网络模型进行训练，得到训练好的双层循环神经网络模型；

步骤（3）：从待推荐肢体动作的音频中提取音频测试特征；

步骤（4）：将音频测试特征输入到训练好的双层循环神经网络模型中，输出推荐的肢体测试特征。

2、如权利要求 1 所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，所述从演讲视频中提取每一帧图像所对应时间范围内人物的音频训练特征；是指：

将演讲视频的音频信号均匀划分为长度为 $\frac{1}{\text{framerate}}$ 秒的语音区间，其中framerate为每秒的帧数；对每个语音区间提取声音强度特征，通过音频信号转为文本，对文本进行分类得到感情类别特征；将声音强度特征和感情类别特征进行特征融合作为音频训练特征。

3、如权利要求 1 所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，所述从演讲视频中提取每一帧图像内人物的肢体训练特征；是指：

将演讲视频的视频信号以帧为单位划分为若干帧图像，对每一帧图像使用肢体语言识别系统 OpenPose 检测出图像中人物各关节点坐标位置，对各关节点坐标位置进行归一化后，得到每一帧图像的人物关节与关节坐标位置关系矩阵；根据每一帧图像的人物关节与关节坐标位置关系矩阵计算关节夹角的角度，根据关节夹角的角度将当前帧图像映射到预先设定活动状态中最接近的一个状态作为肢体训练特征。

4、如权利要求 3 所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，

所述关节夹角是指：在头部位置预先设定一个点为头部节点，在颈部位置预先设定一个点为颈部节点，在右肩关节位置预先设定一个点为右肩关节点，在左肩关节位置预先设定一个点为左肩关节点，在右胳膊肘位置预先设定一个点为右肘关节点，在左胳膊肘位置预先设定一个点为左肘关节点，在右手位置预先设定一个点为右手关节点，在左手位置预先设定一个点为左手关节点；

第一个夹角是指：头部节点与颈部节点连线和颈部节点与右肩关节点连线之间的夹角 α^H ；

第二个夹角是指：颈部节点与右肩关节点连线和右肩关节点与右肘关节点连线之间的身体内侧夹角 α^{RS} ；

第三个夹角是指：右肩关节点与右肘关节点连线和右肘关节点与右手关节点连线之间的身体内侧夹角 α^{RE} ；

第四个夹角是指：颈部节点与左肩关节点连线和左肩关节点与左肘关节点连线之间的身体内侧夹角 α^{LS} ；

第五个夹角是指：左肩关节点与左肘关节点连线和左肘关节点与左手关节点连线之间的身体内侧夹角 α^{LE} ；

预先设定活动状态，包括：第一个夹角设置 m 个状态，第二个夹角设置 n 个状态，第三个夹角设置 o 个状态，第四个夹角设置 p 个状态，第五个夹角设置 q 个状态；根据各个夹角的状态的组合，将人物姿态划分为 $m*n*o*p*q$ 个状态。

5、如权利要求 1 所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，

所述双层循环神经网络模型包括两个嵌套组合循环神经网络：第一循环神经网络和第二循环神经网络；第二循环神经网络在第一循环神经网络内部；第一循环神经网络包括第一编码器和第一解码器；第二循环神经网络包括第二编码器和第二解码器；第一编码器、第二编码器、第二解码器和第一解码器依次串联；第一编码器的输入值为音频训练特征，第一编码器的输出值输入到第二编码器中，第二编码器的输出值输入到第二解码器中，第二解码器的输出值输入到第一解码器中，第一解码器的输出值为肢体训练特征。

6、如权利要求 1 所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，

所述步骤（2）中，将音频训练特征作为双层循环神经网络模型的输入值，将肢体训练特征作为双层循环神经网络模型的输出值，对已构建的双层循环神经网络模型进行训练，得到训练好的双层循环神经网络模型，步骤为：

步骤（201）：根据声音强度将音频信号划分为若干个声音行为，划分过程中将声音强度超过设定阈值的连续音频信号视为一个声音行为，将声音强度未超过设定阈值的连续音频信号视为一个声音行为，每个声音行为包括若干个语音区间，每个语音区间都有对应的音频训练特征；将每个声音行为对应的音频训练特征输入到第一编码器中，第一编码器输出每个音频训练特征的隐藏层状态；

步骤（203）：将每个声音行为对应的肢体训练特征集合视为肢体动作行为；将第二编码器每次输出的隐藏层状态作为与声音行为对应的肢体动作行为的隐藏层状态；

步骤（204）：将每一帧图像的肢体训练特征视为肢体动作单元，在得到每个肢体动作行

为的隐表达之后，构建肢体动作行为与肢体动作行为包含的每一帧图像的关系，并通过第一解码器将肢体动作单元的隐表达解码为每个语音区间对应的帧的肢体动作特征的预测，选取概率最大的肢体动作作为预测的肢体动作；

步骤（205）：利用损失函数计算预测的肢体动作特征与肢体训练特征之间的损失值，如果损失值在设定范围内，则表示双层循环神经网络模型训练合格，进入步骤（3）；否则表示双层循环神经网络模型训练不合格，调整双层循环神经网络模型的参数后，返回步骤（201）继续训练，直至损失值在设定范围内，结束。

7、如权利要求1所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，所述步骤（3）中，从待推荐肢体动作的音频中提取音频测试特征，步骤为：

对待测试音频信号划分为长度为 $\frac{1}{\text{framerate}}$ 秒的语音区间，其中framerate为每秒的帧数；对每个语音区间提取声音强度特征；将待测试音频信号转为文本，对文本进行分类得到感情类别特征；将声音强度特征和感情类别特征进行特征融合作为音频测试特征。

8、如权利要求1所述的基于深度学习的人际交互肢体语言自动生成方法，其特征是，所述步骤（4）中，根据声音强度将音频信号划分为若干个声音行为，划分过程中将声音强度超过设定阈值的连续音频信号视为一个声音行为，将声音强度未超过设定阈值的连续音频信号视为一个声音行为，每个声音行为包括若干个语音区间，每个语音区间都有对应的音频测试特征；将每个声音行为对应的音频测试特征输入到训练好的双层循环神经网络模型中，输出推荐的肢体测试特征，对肢体测试特征进行展示。

9、基于深度学习的人际交互肢体语言自动生成系统，其特征是，包括：

存储器、处理器以及存储在存储器上并在处理器上运行的计算机指令，所述计算机指令被处理器运行时，完成上述权利要求1-8任一方法所述的步骤。

10、一种计算机可读存储介质，其特征是，其上存储有计算机指令，所述计算机指令被处理器运行时，完成上述权利要求1-8任一方法所述的步骤。

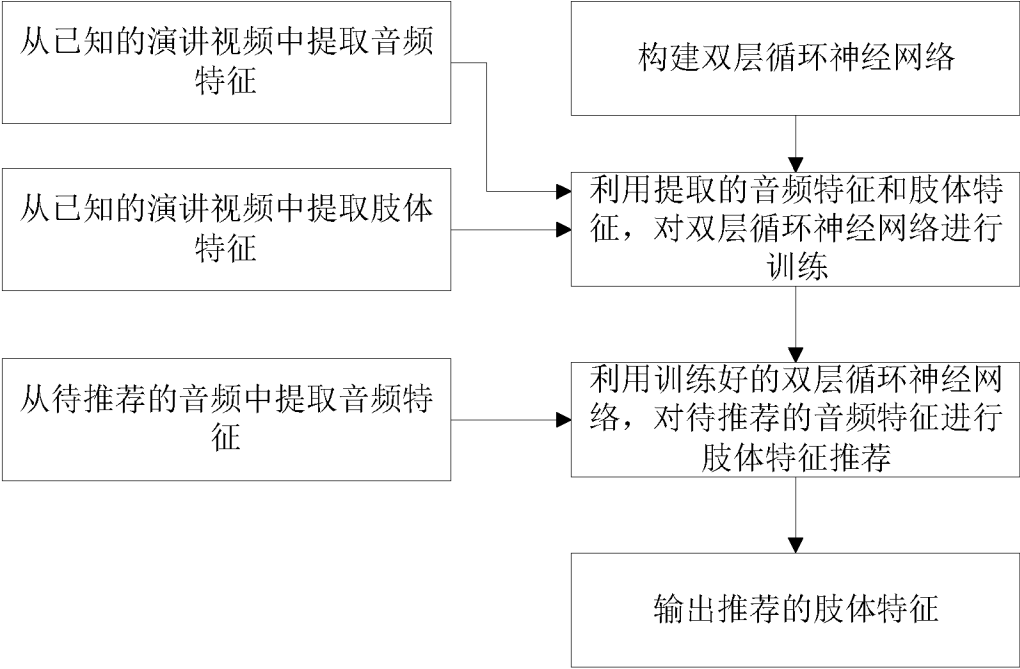


图 1



图 2 (a)

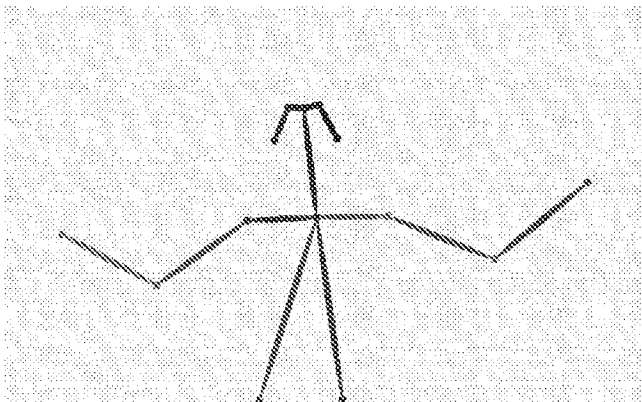


图 2 (b)

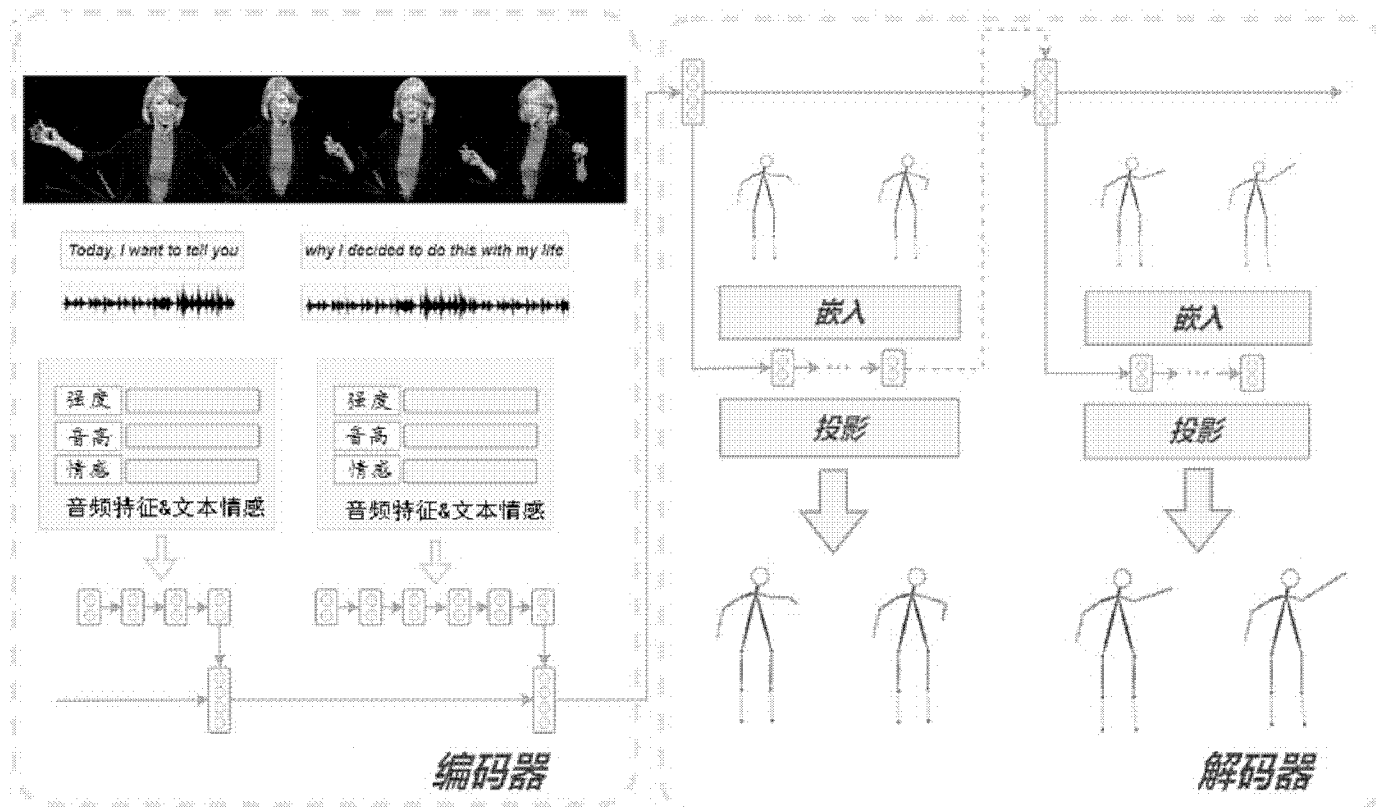


图 3

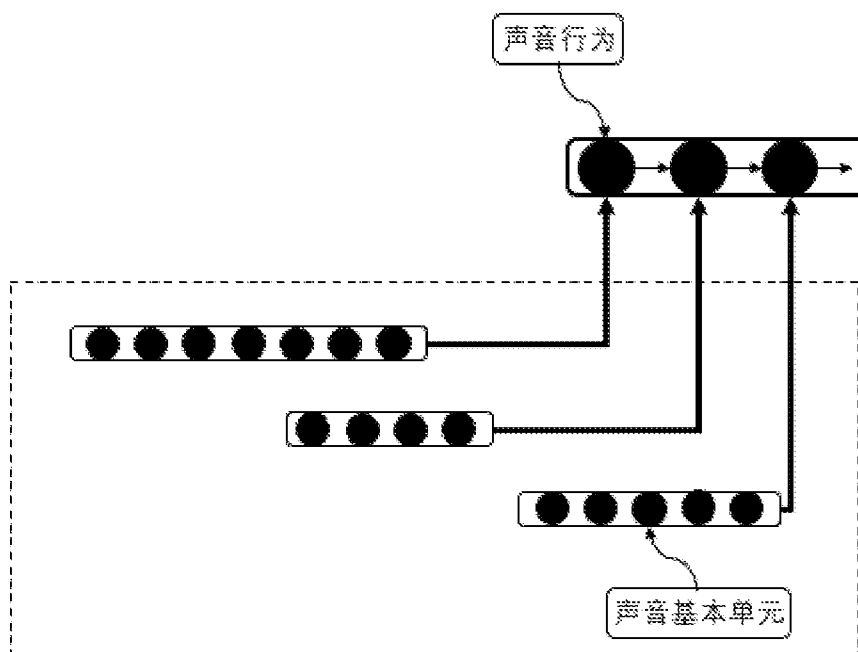


图 4

替换页 (细则第26条)

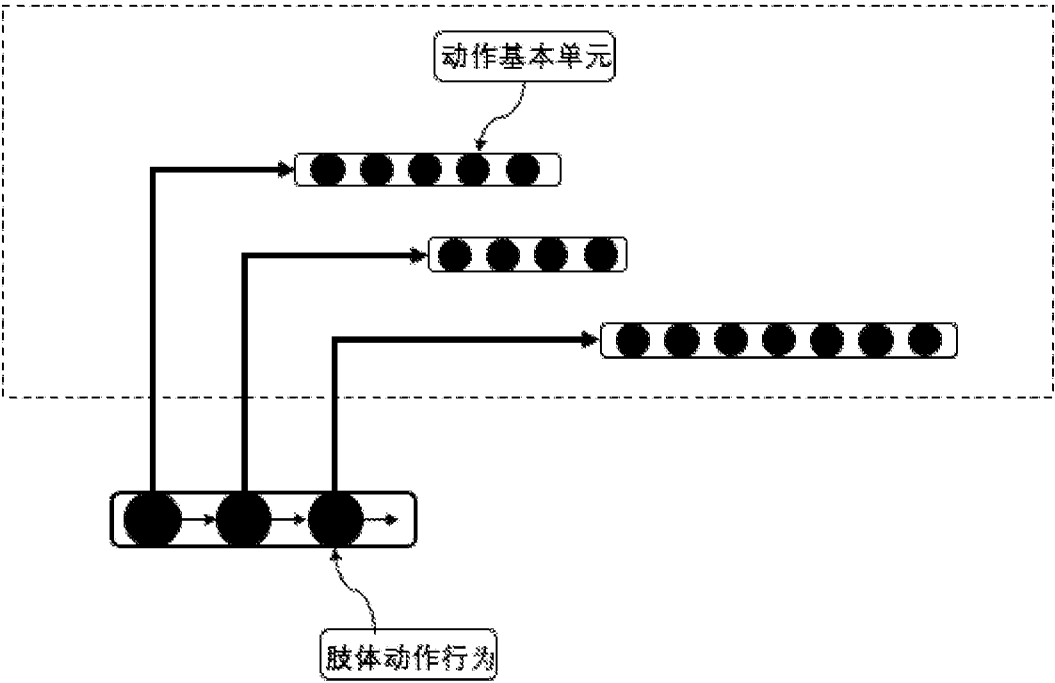


图 5

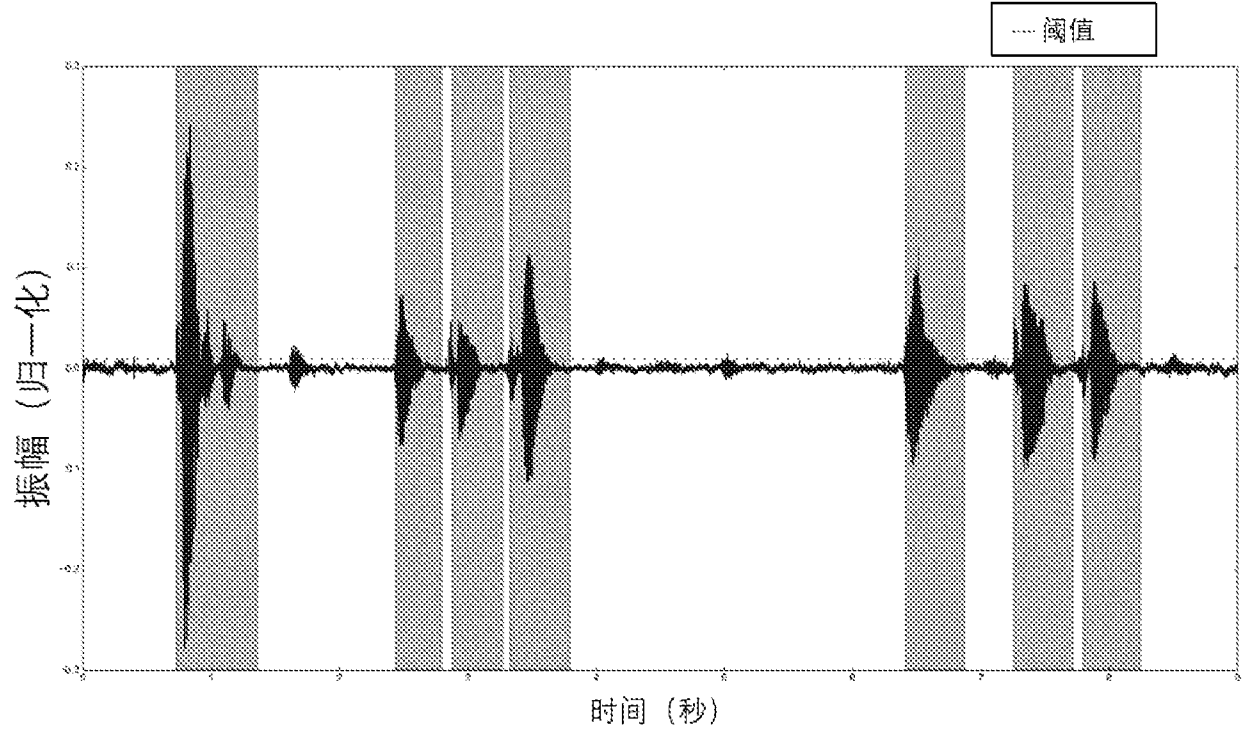


图 6

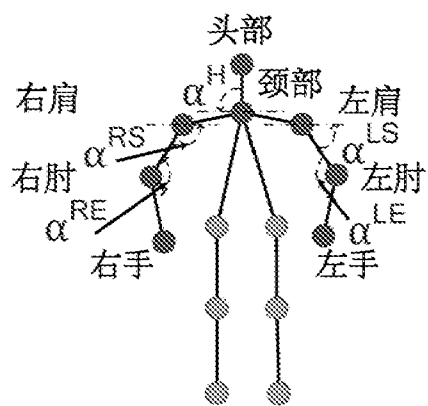


图 7 (a)

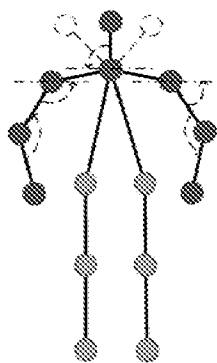


图 7 (b)

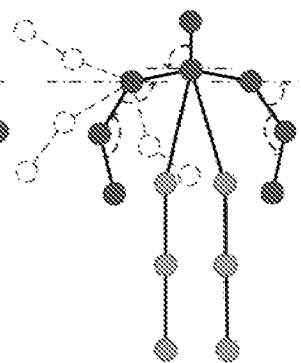


图 7 (c)

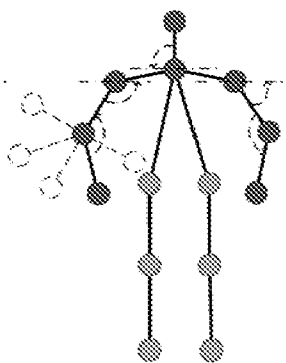


图 7 (d)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/122942

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/16(2006.01)i; G10L 15/16(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

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)

CNPAT, WPI, EPODOC, CNKI, GOOGLE: mn, 演讲, 动作, voice, 语音, recognition, gesture, 训练, neural, 肢体, 测试, 深度学习, network, 模型, hand, speech, 音频, 手势, recurrent, 视频, 山东大学, 甘甜, 马志鑫, 宋雪萌, 聂礼强

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 108921284 A (SHANDONG UNIVERSITY) 30 November 2018 (2018-11-30) claims 1-10, and description, paragraphs [0075]-[0160]	1-10
A	CN 106991372 A (BEIJING UNIVERSITY OF TECHNOLOGY) 28 July 2017 (2017-07-28) description, paragraphs [0056]-[0068], and figures 1 and 2	1-10
A	CN 104182772 A (DALIAN UNIVERSITY OF TECHNOLOGY) 03 December 2014 (2014-12-03) entire document	1-10
A	CN 105957531 A (SHANGHAI JIAO TONG UNIVERSITY) 21 September 2016 (2016-09-21) entire document	1-10
A	CN 104484656 A (ANHUI COSWIT INFORMATION TECHNOLOGY CO., LTD.) 01 April 2015 (2015-04-01) 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

11 March 2019

Date of mailing of the international search report

22 March 2019

Name and mailing address of the ISA/CN

**National Intellectual Property Administration, PRC (ISA/
CN)**
**No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088**
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/122942

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	108921284	A	30 November 2018	None	
CN	106991372	A	28 July 2017	None	
CN	104182772	A	03 December 2014	None	
CN	105957531	A	21 September 2016	None	
CN	104484656	A	01 April 2015	None	

国际检索报告

国际申请号

PCT/CN2018/122942

A. 主题的分类 G10L 15/16(2006.01)i; G10L 15/16(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 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)) CNPAT, WPI, EPDOC, CNKI, GOOGLE:rnn, 演讲, 动作, voice, 语音, recognition, gesture, 训练, neural, 肢体, 测试, 深度学习, network, 模型, hand, speech, 音频, 手势, recurrent, 视频, 山东大学, 甘甜, 马志鑫, 宋雪萌, 聂礼强																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 108921284 A (山东大学) 2018年 11月 30日 (2018 - 11 - 30) 权利要求1-10, 说明书[0075]-[0160]段</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 106991372 A (北京工业大学) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0056]-[0068]段, 附图1-2</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 104182772 A (大连理工大学) 2014年 12月 3日 (2014 - 12 - 03) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 105957531 A (上海交通大学) 2016年 9月 21日 (2016 - 09 - 21) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 104484656 A (安徽寰智信息科技股份有限公司) 2015年 4月 1日 (2015 - 04 - 01) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 108921284 A (山东大学) 2018年 11月 30日 (2018 - 11 - 30) 权利要求1-10, 说明书[0075]-[0160]段	1-10	A	CN 106991372 A (北京工业大学) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0056]-[0068]段, 附图1-2	1-10	A	CN 104182772 A (大连理工大学) 2014年 12月 3日 (2014 - 12 - 03) 全文	1-10	A	CN 105957531 A (上海交通大学) 2016年 9月 21日 (2016 - 09 - 21) 全文	1-10	A	CN 104484656 A (安徽寰智信息科技股份有限公司) 2015年 4月 1日 (2015 - 04 - 01) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 108921284 A (山东大学) 2018年 11月 30日 (2018 - 11 - 30) 权利要求1-10, 说明书[0075]-[0160]段	1-10																		
A	CN 106991372 A (北京工业大学) 2017年 7月 28日 (2017 - 07 - 28) 说明书第[0056]-[0068]段, 附图1-2	1-10																		
A	CN 104182772 A (大连理工大学) 2014年 12月 3日 (2014 - 12 - 03) 全文	1-10																		
A	CN 105957531 A (上海交通大学) 2016年 9月 21日 (2016 - 09 - 21) 全文	1-10																		
A	CN 104484656 A (安徽寰智信息科技股份有限公司) 2015年 4月 1日 (2015 - 04 - 01) 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 2019年 3月 11日		国际检索报告邮寄日期 2019年 3月 22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		受权官员 林芳 电话号码 86-(10)-53961337																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/122942

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	108921284	A	2018年 11月 30日	无	
CN	106991372	A	2017年 7月 28日	无	
CN	104182772	A	2014年 12月 3日	无	
CN	105957531	A	2016年 9月 21日	无	
CN	104484656	A	2015年 4月 1日	无	