

音乐推荐方法和装置

5 技术领域

本申请涉及人工智能领域，并且更具体的，涉及一种音乐推荐方法和装置。

背景技术

10 个性化音乐推荐技术能提升用户的音乐体验。传统的方法是基于用户的历史音乐播放信息，通过数据挖掘技术来推荐音乐。这种方法无法考虑用户当前的状态信息。目前的一些方法可以通过不同的传感器来收集用户的当前状态信息，例如通过感知环境信息，包括位置、天气、时间、季节、环境声音和环境画面等信息来推荐相关的音乐；或者通过测量用户当前的状态信息，例如通过采集脑电波来分析用户当前的心理状态，或者采集用户看到的画面，或者获取用户的心率等来推荐相关的音乐。

15 目前的方法中，根据拍摄采集用户看到的图像后进行音乐推荐，涉及音乐与图像的匹配过程。而在现实场景中，环境中可能包含许多景物，如果只根据图像整体来推荐音乐，降低了音乐匹配度。

发明内容

20 本申请提供一种音乐推荐方法和装置，通过用户的视点信息来确定用户在复杂环境中的注意模式，更加精准地匹配音乐。

第一方面，提供了一种音乐推荐方法，该方法包括：接收用户的视觉数据；根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长；根据所述至少一个注意单元的注意时长确定所述用户的注意模式；根据所述注意模式确定推荐的音乐信息。

25 本申请实施例的音乐推荐方法，根据用户的视觉信息来判断用户的注意模式，可以更加精确判断用户的注意内容，从而推荐更加适合的音乐，使得推荐的音乐符合用户真正感兴趣的事物，符合用户的真正的行为状态，提升用户的使用感受。

结合第一方面，在第一方面的一种可能的实施方式中，视觉数据包括所述用户的视点信息和所述用户所视的画面信息，所述视点信息包括视点的位置和所述视点的注意时长。

30 结合第一方面，在第一方面的一种可能的实施方式中，根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长，包括：根据所述画面信息获取所述至少一个注意单元；获取所述至少一个注意单元中的所述视点的注意时长之和，以作为所述至少一个注意单元的注意时长。

35 本申请实施例的音乐推荐方法，根据获取的画面信息确定初始的注意单元，根据用户的视点信息确定每个注意单元的时长，相比于现有技术中的仅根据用户所视的整个画面来推荐音乐，视点信息可以精确表示用户感兴趣的注意内容的所在，从而可以实现推荐的音乐更加符合用户的所需。

结合第一方面，在第一方面的一种可能的实施方式中，根据所述视觉数据获取至少一

个注意单元和所述至少一个注意单元的注意时长,还包括:判断所述至少一个注意单元中的第一注意单元和第二注意单元的相似度,所述第一注意单元和所述第二注意单元为不同时刻的注意单元;如果所述相似度大于或等于第一阈值,所述第二注意单元的注意时长等于所述第一注意单元的注意时长和所述第二注意单元的注意时长之和。

- 5 本申请实施例的音乐推荐方法中,第一注意单元和第二注意单元可以是一段预设时间内的不同时刻帧图像中的注意单元,也可以分别是历史库和新获取的注意单元。

结合第一方面,在第一方面的一种可能的实施方式中,根据所述至少一个注意单元的注意时长确定所述用户的注意模式,包括:如果所述至少一个注意单元的注意时长的标准差大于或等于第二阈值,确定所述用户的注意模式为盯着看;如果所述至少一个注意单元的注意时长的标准差小于第二阈值,确定所述用户的注意模式为扫描看。

结合第一方面,在第一方面的一种可能的实施方式中,根据所述注意模式确定音乐信息,包括:如果所述注意模式为扫描看,根据所述画面信息确定音乐信息;如果所述注意模式为盯着看,根据所述注意单元中关注度最高的注意单元确定音乐信息。

- 15 本申请实施例的音乐推荐方法中,在确定了用户的注意模式后,即可根据用户在这一段预设时间内的注意模式确定在这段预设时间内适合推荐给用户的音乐信息。当用户的注意模式为扫描看时,则认为用户在这段预设时间主要在感知环境,则可以根据画面信息(环境)来推荐音乐;当用户的注意模式为盯着看时,则认为用户在这段预设时间主要在感知感兴趣的事物,则可以根据关注度最高的注意单元(感兴趣的事物)来推荐音乐。

结合第一方面,在第一方面的一种可能的实施方式中,根据所述注意模式确定音乐信息,还包括:根据所述注意模式确定所述用户在第一时间段内的每个时刻的行为状态;根据所述每个时刻的状态确定所述用户在第一时间段内的行为状态;根据所述第一时间段内的行为状态确定音乐信息。

- 20 本申请实施例的音乐推荐方法中,在根据用户在一段预设时间内的注意模式确定了注意内容后,也可以先不确定音乐信息,而是确定用户在这段预设时间内的行为状态,然后根据多个预设时间段的行为状态确定用户在第一时间段内的总的行为状态,可以更加精确判断用户实际的行为状态,根据总的行为状态来推荐音乐,使得推荐的音乐更加符合用户的实际行为状态。

第二方面,提供了一种音乐推荐的装置,该装置包括:收发模块,用于接收用户的视觉数据;确定模块,用于根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长;所述确定模块还用于根据所述至少一个注意单元的注意时长确定所述用户的注意模式;所述确定模块还用于根据所述注意模式确定推荐的音乐信息。

本申请实施例提供一种音乐推荐的装置,用于实现第一方面中的音乐推荐的方法。

结合第二方面,在第二方面的一种可能的实施方式中,视觉数据包括所述用户的视点信息和所述用户所视的画面信息,所述视点信息包括视点的位置和所述视点的注意时长。

- 35 结合第二方面,在第二方面的一种可能的实施方式中,确定模块根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长,包括:根据所述画面信息获取所述至少一个注意单元;获取所述至少一个注意单元中的所述视点的注意时长之和,以作为所述至少一个注意单元的注意时长。

结合第二方面,在第二方面的一种可能的实施方式中,确定模块根据所述视觉数据获

取至少一个注意单元和所述至少一个注意单元的注意时长，还包括：判断所述至少一个注意单元中的第一注意单元和第二注意单元的相似度，所述第一注意单元和所述第二注意单元为不同时刻的注意单元；如果所述相似度大于或等于第一阈值，所述第二注意单元的注意时长等于所述第一注意单元的注意时长和所述第二注意单元的注意时长之和。

5 结合第二方面，在第二方面的一种可能的实施方式中，确定模块根据所述至少一个注意单元的注意时长确定所述用户的注意模式，包括：如果所述至少一个注意单元的注意时长的标准差大于或等于第二阈值，确定所述用户的注意模式为盯着看；如果所述至少一个注意单元的注意时长的标准差小于第二阈值，确定所述用户的注意模式为扫描看。

10 结合第二方面，在第二方面的一种可能的实施方式中，确定模块用于根据所述注意模式确定音乐信息，包括：如果所述注意模式为扫描看，根据所述画面信息确定音乐信息；如果所述注意模式为盯着看，根据所述注意单元中关注度最高的注意单元确定音乐信息。

15 结合第二方面，在第二方面的一种可能的实施方式中，确定模块根据所述注意模式确定音乐信息，还包括：根据所述注意模式确定所述用户在第一时间段内的每个时刻的行为状态；根据所述每个时刻的状态确定所述用户在第一时间段内的行为状态；根据所述第一时间段内的行为状态确定音乐信息。

第三方面，提供了一种计算机可读存储介质，该计算机可读存储介质中存储有程序指令，当所述程序指令由处理器运行时，以实现上述第一方面和第一方面任一种实现方式的方法。

20 第四方面，提供了一种计算机程序产品，该计算机程序产品包括计算机程序代码，当所述计算机程序代码在计算机上运行时，以实现上述第一方面和第一方面任一种实现方式的方法。

25 第五方面，提供了一种音乐推荐系统，该系统包括数据采集设备和终端设备，所述终端设备包括处理器和存储器，所述存储器中存储有一个或多个程序，所述一个或多个计算机程序包括指令，其中，所述数据采集设备，用于采集用户的视觉数据；当所述指令被所述一个或多个处理器执行时，使得所述终端设备执行上述第一方面和第一方面任一种实现方式的方法。

附图说明

- 图 1 是本申请实施例的音乐推荐方法应用的系统架构；
- 30 图 2 是本申请实施例的音乐推荐方法应用的系统架构中的第一可穿戴设备的示意性框图；
- 图 3 是本申请实施例的音乐推荐方法应用的系统架构中的终端设备的示意性框图；
- 图 4 是本申请实施例的音乐推荐方法应用的系统架构中的第二可穿戴设备的示意性框图；
- 35 图 5 是本申请实施例的音乐推荐方法的示意性流程图；
- 图 6 是本申请实施例的音乐推荐方法的示意性框图；
- 图 7 是本申请实施例的音乐推荐装置的示意性框图；
- 图 8 是本申请实施例的音乐推荐设备的示意性框图。

具体实施方式

以下实施例中所使用的术语只是为了描述特定实施例的目的，而并非旨在作为对本申请的限制。如在本申请的说明书和所附权利要求书中所使用的那样，单数表达形式“一个”、“一种”、“所述”、“上述”、“该”和“这一”旨在也包括例如“一个或多个”这种表达形式，除非其上下文中明确地有相反指示。还应当理解，在本申请以下各实施例中，“至少一个”、“一个或多个”是指一个、两个或两个以上。术语“和/或”，用于描述关联对象的关联关系，表示可以存在三种关系；例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 的情况，其中 A、B 可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。

- 10 在本说明书中描述的参考“一个实施例”或“一些实施例”等意味着在本申请的一个或多个实施例中包括结合该实施例描述的特定特征、结构或特点。由此，在本说明书中的不同之处出现的语句“在一个实施例中”、“在一些实施例中”、“在其他一些实施例中”、“在另外一些实施例中”等不是必然都参考相同的实施例，而是意味着“一个或多个但不是所有的实施例”，除非是以其他方式另外特别强调。术语“包括”、“包含”、“具有”及它们的变形都意味着“包括但不限于”，除非是以其他方式另外特别强调。

下面将结合附图，对本申请中的技术方案进行描述。

- 20 现有的图像和应用匹配方法主要包括：一是提取音乐和图像两个模态的传统底层特征，再通过关系模型建立两者的联系，这种方法推荐的音乐与图像匹配度不高；二是先采集音乐与图像的匹配对数据，基于深度神经网络自动学习音乐与图像的匹配模型，这种方法在简单的场景下可以推荐合适的音乐。

然而在现实场景中，环境可能包含许多景物和不同的风格元素，以上现有的方法没有考虑用户在当前环境中的兴趣所在，降低了音乐匹配度。例如，当用户关注场景中的云朵，与关注场景中的动物时，匹配的音乐应当是不同的。

- 25 因此本申请实施例提供了一种音乐推荐方法，通过获取用户的视点信息来得到用户在复杂环境中的注意区域，从而知道用户在当前环境中的真正兴趣所在，提高音乐匹配度。

- 30 图 1 示出了本申请实施例的音乐推荐方法应用的系统架构，如图 1 所示，包括第一可穿戴设备、第二可穿戴设备、移动终端设备。其中，第一可穿戴设备为可以采集用户视觉数据并记录用户头部运动数据的可穿戴设备，例如智能眼镜等，其上安装有先进摄影系统（advanced photo system, APS）摄像头、动态视觉传感器（dynamic vision sensor, DVS）摄像头、眼动仪和惯性测量单元（inertial measurement unit, IMU）传感器。第二可穿戴设备为可以播放音乐的可穿戴设备，例如耳机等。移动终端设备可以是手机、平板电脑、可穿戴设备（例如，智能手表）、车载设备、增强现实（augmented reality, AR）设备、虚拟现实（virtual reality, VR）设备、笔记本电脑、超级移动个人计算机（ultra-mobile personal computer, UMPC）、上网本、个人数字助理（personal digital assistant, PDA）等设备。
- 35 本申请实施例的终端设备可以包括触摸屏，用于向用户展示服务内容。本申请实施例中对终端设备的具体类型并不作任何限定。

应理解，上述只是对本申请实施例图 1 中的设备的举例，并不构成对本申请实施例的限定，除了上述举例的设备，本申请实施例图 1 中的设备还可以是其他可以实现相同功能的设备。

在应用本申请实施例的音乐推荐方法时，移动终端设备给第一可穿戴设备发送数据采集指令。第一可穿戴设备接收指令后，按一定频率采集全帧数据、记录画面变化数据，同时记录用户视点数据和局部画面数据，以及头部转动的加速度和角度数据，并不断发送到移动终端设备。移动终端设备接收数据后，判断用户的注意区域和注意模式，根据用户的注意模式和注意区域提取对应特征，匹配音乐。移动终端设备向第二可穿戴设备发送音频数据，第二可穿戴设备播放音乐。

图 2 示出了应用本申请实施例的音乐推荐方法时，第一可穿戴设备中包括的模块。

无线模块，用于建立无线链路，与其他节点通信，其中无线通信可以采用 wifi、蓝牙和蜂窝网络等通信方式。

视频帧采集模块，用于驱动第一可穿戴设备上的 APS 摄像头，采集描述环境的视频帧。

视点采集模块，用于驱动眼镜上的眼动仪，采集视点数据。其中，视点数据包括视点位置、获取时间、注视时间、瞳孔直径；

头部运动采集模块，用于驱动眼镜上的 IMU 模块，采集头部转动的速度和加速度。

画面变化捕捉模块，用于驱动眼镜上的 DVS 摄像头，采集画面的变化数据。

数据接收模块，用于接收移动终端设备发送的数据。

数据发送模块，用于将采集到的数据发送给移动终端设备。

图 3 示出了应用本申请实施例的音乐推荐方法时，移动终端设备中包括的模块。

无线模块，用于建立无线链路，与其他节点通信，其中无线通信可以采用 wifi、蓝牙和蜂窝网络等通信方式。

注意模式判别模块，用于根据眼镜采集到的数据，计算注意区域和注意模式。

特征提取和音乐匹配模块，用于根据注意模式类别，提取特征，并匹配音乐。

数据接收模块，用于接收来自第一可穿戴设备发送的数据。

数据发送模块，用于将音乐的音频数据和播放指令发送给第二可穿戴设备。

图 4 示出了应用本申请实施例的音乐推荐方法时，第二可穿戴设备中包括的模块。

无线模块，用于建立无线链路，与其他节点通信，其中无线通信可以采用 wifi、蓝牙和蜂窝网络等通信方式。

数据接收模块，用于接收移动终端设备发送的音频数据和播放指令。

音频播放模块，用于根据移动终端设备发送的音频数据和播放指令播放音乐。

图 5 示出了本申请实施例的一种音乐推荐方法的示意性流程图，包括步骤 501 至步骤 504，以下分别对这些步骤进行详细介绍。其中，图 5 中的音乐推荐方法可以由图 1 中的终端设备执行。

S501，接收用户的视觉数据。

具体的，可以是终端设备接收第一穿戴设备发送的用户的视觉数据，第一穿戴设备采集用户在一段预设时间内（例如 1 秒）的视觉数据，其中，用户的视觉数据包括用户的视点信息和用户所视的画面信息，视点信息包括视点的位置坐标（x,y）以及该视点的注意时长；画面信息包括 APS 摄像头采集到的视频帧图像和 DVS 摄像头采集到的画面变化数据。

S502，根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时

长。

具体的,根据画面信息获取至少一个注意单元,例如,将视频帧图像中的宏块作为注意单元,其中宏块可以是重叠的,也可以是不重叠的;或者根据用量化一个区域内是否存在一个物体的算法(例如 objectness 算法)来提取一个或多个对象矩形框作为注意单元;
5 还可以根据画面变化数据获取不同时刻的运动矩形框,然后将运动矩形框作为注意单元。其中,每个注意单元可以以最近时刻的帧图像中与该注意单元同一位置的图像数据作为该注意单元的内容。

当用户所视画面静止或者在一帧图像中,此时 DVS 摄像头没有采集到画面变化数据,获取注意单元的注意时长可以根据所有的视点对注意单元投票,即当一个视点位于一个注意单元内时,则将该视点的注意时长累加到该注意单元的注视时长中。
10

可选的,根据视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长,还包括:当用户所视画面是变化的,此时 DVS 摄像头采集到画面变化数据,则在一帧图像中时依然按照上述方法对该帧图像中的注意单元投票,由此可以获得每帧图像中的注意单元的注意时长。对于任意两个相邻时刻的图像中的注意单元,以后一时刻图像中的一个注意单元为例,命名为第二注意单元,在前一时刻图像中找到与第二注意单元距离小于预设值的 N 个注意单元,其中,注意单元之间的距离为两个注意单元中心坐标的欧氏距离, N 可以是人为规定的值,也可以是满足条件的注意单元的个数的最大值。以 N 个注意单元中的一个注意单元为例,命名为第一注意单元,判断第一注意单元和第二注意单元的相似度,即匹配第一注意单元和第二注意单元的特征,其中,第一注意单元和第二注意单元的特征匹配方法可以是现有的任一种图像特征匹配方法,本申请实施例在此不作具体限定,如果判定第一注意单元和第二注意单元的特征相似,即第一注意单元和第二注意单元的相似度大于或等于第一阈值,则认为第一注意单元和第二注意单元是同一事物在不同时刻的呈现,则令第二注意单元的注意时长等于第一注意单元的注意时长和第二注意单元的注意时长之和,然后令第一注意单元的注意时长为零;如果判定第一单元和第二单元的特征不相似,即第一注意单元和第二注意单元的相似度小于第一阈值,则保留第一单元和第二单元的注意时长。对于任意两个相邻时刻图像中的注意单元均按照上述方法进行判定。
15
20
25

可选的,根据视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长,还包括:建立注意单元的历史库,历史库的大小固定,例如只能存放 10 个注意单元。判断新获取的注意单元与历史库中的注意单元的相似度,例如判断新获取的第二注意单元与历史库中的第一注意单元的相似度,可以分别提取第一注意单元和第二注意单元的视觉特征,然后计算视觉特征之间的相似度。如果判定第一注意单元和第二注意单元的特征相似,即第一注意单元和第二注意单元的相似度大于或等于第三阈值,则第二注意单元的注意时长等于第一注意单元的注意时长和第二注意单元的注意时长之和,然后用第二注意单元替换第一注意单元存放在历史库中;如果判定第一单元和第二单元的特征不相似,即第一注意单元和第二注意单元的相似度小于第三阈值,则保留历史库中的第一注意单元。如此,可以在一段预设时间内,例如 1 秒中得到历史库中的注意单元和每个注意单元的注意时间,然后按照 S503 中的方法确定用户在这 1 秒内的注意模式。然后删除历史库中存在时间超过 1 秒且注意时长少于 600 毫秒的注意单元,并补充新获取的注意单元。
30
35

S503,根据所述至少一个注意单元的注意时长确定用户的注意模式。

在一段预设时间内，如果所有注意单元的注意时长的标准差大于或等于第二阈值，则确定所述用户的注意模式为盯着看；如果所有注意单元的注意时长的标准差小于第二阈值，则确定所述用户的注意模式为扫描看。

S504，根据所述注意模式确定推荐的音乐信息。

- 5 如果用户的注意模式为扫描看，则将 APS 摄像头采集的帧图像直接作为用户的注意内容；如果用户的注意模式为盯着看，则将一段预设时间内的所有注意单元中关注度最高的注意单元作为用户的注意内容。其中，关注度可以根据注意时长来判定，例如将注意时长最长的注意单元作为关注度最高的注意单元；或者可以根据用户瞳孔缩放程度来判定，例如将用户瞳孔放大程度最大的注意单元作为关注度最高的注意单元；或者可以根据用户复看次数来判定，例如用户在注视一个注意单元后，又多次复看该注意单元，如果复看次数大于预设值，则将该注意单元作为关注度最高的注意单元；或者同时考虑这三者来估计注意单元的关注度，例如关注度是用户瞳孔放大程度乘以注意时长乘以复看次数的积。

- 10 然后根据注意内容确定音乐信息，根据注意内容确定音乐信息可以是现有的根据图像匹配音乐的方法，例如将注意内容（帧图像或关注度最高的注意单元）作为神经网络模型的输入，将神经网络模型输出的概率值最大的音乐类别作为判断结果，例如当概率值大于 0.8 时，即认为图像和音乐的匹配度足够高。

- 15 可选的，在根据用户在一段预设时间内的注意模式确定了注意内容后，也可以先不确定音乐信息，而是确定用户在这一段预设时间内的行为状态。根据注意内容确定行为状态的方法可以采用现有的分类的机器学习方法，例如把注意内容作为神经网络模型的输入，然后将神经网络模型的输出概率值最大的行为状态类别作为判断结果。其中，行为状态包括开车、学习、旅行、运动等。由此，可以确定用户在第一时间段内的多个预设时间的行为状态，例如第一时间段为 10 秒，一段预设时间为 1 秒，则可以确定用户在 10 秒内的 10 个行为状态。对该 10 个行为状态进行投票，例如，10 个行为状态中有 7 个被判定为学习，2 个被判定为运动，1 个被判定为旅行，则认为用户在这 10 秒中的行为状态为学习。
- 20 最后根据用户在第一时间段内的行为状态来匹配音乐，根据行为状态来匹配音乐的方法可以是现有的方法，例如可以根据行为状态的标签信息来匹配音乐，本申请实施例对此不作具体限定。

确定了音乐信息后，终端设备可以根据音乐信息向第二可穿戴设备发送音乐播放指令，第二可穿戴设备则播放指定音乐。或者终端设备也可以根据音乐信息播放音乐。

- 30 本申请实施例的音乐推荐方法，根据用户的视觉信息来判断用户的注意模式，可以更加精确判断用户的注意内容，从而推荐更加适合的音乐，使得推荐的音乐符合用户真正感兴趣的事物，符合用户的真正的行为状态，提升用户的使用感受。

以下根据具体示例对本申请实施例的音乐推荐方法进行详细介绍。其中，第一可穿戴设备以智能眼镜为例，第二可穿戴设备以耳机为例，移动终端设备以手机为例。

- 35 图 6 示出了本申请实施例提供的音乐推荐方法的示意性框图，如图 6 所示，包括如下步骤。

1、数据采集

手机给智能眼镜发送数据采集指令。智能眼镜接收到手机发送的数据采集指令后，开始采集数据，并将采集到的数据不断传输给手机端。采集的数据包括：

(1) 帧数据: 按一定频率 (例如 30Hz) 采集用户通过智能眼镜可以看到的整幅图像的帧数据;

(2) 视点数据: 记录用户的视点位置坐标 (x,y)、瞳孔直径、获取时间和注视时间;

(3) 头部运动数据: 头部转动的角度和加速度;

5 (4) 画面变化数据: DVS 摄像头采集到的事件数。

2、基于采集到的数据进行分析, 并提取特征匹配音乐

I、判定用户在一段时间内的一个或多个注意单元和每个注意单元对应的注意时间。

具体的, 上述一段时间可以是 1 秒。在 1 秒的开始时刻拍摄一次 APS 帧, 开始记录画面变化和眼动数据, 这段时刻的末尾分析数据, 并提取特征匹配音乐。如果在这 1 秒中情况发生了变化, 例如在 500 毫秒时, 用户的头部发生了较大转动, 则也可以只分析这 10 500 毫秒的数据, 但是如果上述一段时间小于 100 毫秒, 不足以生成一个注视点, 则丢掉数据。

其中, 注意单元可以是宏块、对象矩形框或运动矩形框。当注意单元是宏块时, 宏块可以是重叠的, 也可以是不重叠的; 当注意单元是对象矩形框时, 初始时刻的注意单元可以用量化一个区域内是否存在一个物体的算法 (例如 objectness 算法) 来提取一个或多个 15 对象矩形框作为注意单元; 当注意单元是运动矩形框时, 可以基于 DVS 摄像头采集到的事件数据来获取每个时刻的运动矩形框, 具体的, 在每个时刻, 先将 DVS 摄像头采集到的事件数据表示为帧数据, 即事件的像素位置的灰度值为 255, 其余像素位置的灰度值为 0, 然后在帧数据上先腐蚀后膨胀得到运动区域, 最后将能覆盖整个连通运动区域的最小 20 矩形框作为注意单元。

当用户头部不动 (头部转动角度小于等于 5 度时), 且 DVS 摄像头在这 1 秒内无局部输出, 即用户所见画面静止时:

(1) 当一个注视点位于一个注意单元内, 累积该注视点的注意时长到当前注意单元的注意时长。

25 (2) 去掉注意时长为 0 的注意单元, 根据非极大值抑制 (non maximum suppression, NMS) 方法去掉面积高度重合的注意单元。

当用户头部不动 (头部转动角度小于等于 5 度时), 且 DVS 摄像头在这 1 秒内有局部输出, 即用户所见画面有变化, 可能发生追视行为时:

30 (1) 在同一时刻, 当一个注视点位于一个注意单元内, 累积该注视点的注意时长到当前注意单元的注意时长。

(2) 去掉每个时刻中注意时长为 0 的注意单元, 根据 NMS 方法去掉面积高度重合的注意单元。

35 (3) 在相邻两个时刻, 对于后一时刻的一个注意单元 A, 在前一时刻找到与该注意单元距离最近的 N 个注意单元, N 为大于或等于 1 的正整数, 其中两个注意单元之间的距离为这两个注意单元中心坐标的欧式距离。将 N 个注意单元中每个注意单元分别与该注意单元 A 进行特征匹配, 如果前一时刻的注意单元 B 与注意单元 A 的特征相似, 则认为这两个注意单元是同一个物体在不同时刻的呈现, 则删掉前一时刻的注意单元 B, 并将前一时刻的注意单元 B 的注意时间累积到注意单元 A; 如果前一时刻的注意单元 B 与注意单元 A 的特征不相似, 则保留这两个注意单元。

本申请实施例的音乐推荐方法适用于用户头部不动时,如果用户头部在动时,则此时不进行音乐匹配,等到用户头部不动时,再执行本申请实施例的音乐推荐方法。

II、判定注意模式和注意内容。

注意模式:

5 根据上述判定方法:

(1) 如果注意单元的个数为 0 时,则判定注意模式为“扫描看”;

(2) 如果注意单元的个数不为 0 时,且不同注意单元的注意时间的均方差大于预设值,例如 100ms,则判定注意模式为“盯着看”,否则为“扫描看”。

注意内容:

10 (1) 当注意模式为“扫描看”时,此时认为用户主要在感知环境,因此将 APS 帧图像作为注意内容;

(2) 当注意模式为“盯着看”时,此时认为用户在感知兴趣物体,并将关注度最高的注意单元作为注意内容。

III、根据注意模式和注意内容提取特征、匹配音乐。

15 本申请实施例提供了两种根据注意模式和注意内容提取特征、匹配音乐的方法策略。

(1) 短时 (short term) 策略

直接匹配当前时段的注意内容的视觉特征和音乐的音频特征。例如,采用分类的机器学习方法,将注意内容作为深度卷积神经网络的输入,并将神经网络的输出中概率值最大的类别作为判断结果,例如,当概率值大于 0.8 时,判断该视觉特征和该音乐的匹配度高,20 则该音乐是符合用户的当前感知的。根据图像匹配音乐的过程可以根据现有的任一种图像匹配音乐的方法,本申请实施例在此不做具体限定。

(2) 长时 (long term) 策略

判定每一个时刻用户注意区域内容属于的状态类别,关联不同时刻状态类别信息,得出一段时间内用户的状态,并根据状态的标签信息来匹配音乐。其中,状态类别可以是“开车”、“学习”、“旅行”、“运动”等听歌高频场景。根据某一时刻用户注意区域内容来判断用户状态类别的过程,可以采用分类的机器学习方法,例如将注意内容作为深度卷积神经网络的输入,并将网络的输出中概率值最大的类别作为判断结果。关联不同时刻状态类别信息,可以采用与时间无关的投票取最高的方法,也可以采用与时间相关的时间加权方法,例如在一段时间内分为十个时刻,其中用户有 8 个时刻被判定为学习,有 2 个时刻被判定为运动,则可以得出用户在这段时间内的状态为学习。

30 3、耳机端播放音乐

耳机端接收手机端发送的音频数据和播放指令后,播放相应的音乐。

可选的,在图 6 所示的音乐推荐方法中,本申请实施例还提供另外一种基于采集到的数据进行分析,并提取特征匹配音乐的方法。以下对该另一种方法进行介绍。

35 I、判断用户在一段时间内的一个或多个注意单元和每个注意单元对应的注意时长。

建立注意单元的历史库,其中历史库的大小固定,例如设定为该历史库可以储存 10 个注意单元。历史库刚建立时空,将用户产生的注意单元放入历史库直至历史库满,其中注意单元的注意时长可以根据上述方法中的视点投票确定。历史库满后,将每一个新产生的注意单元与历史库中的每一个注意单元进行匹配,其中,新产生的注意单元的注意时

- 长也可以根据上述方法中的视点投票确定,如果历史库中的注意单元 A 与新产生的注意单元 B 的相似度最高,则将注意单元 A 对应的注意时间累积到注意单元 B 对应的注意时间,然后删除注意单元 A,将注意单元 B 放入历史库中。其中,匹配不同注意单元的相似性的过程为,对不同单元分别提取视觉特征,根据快速鲁棒性特征 (speeded up robust features, SURF) 的算法计算不同单元特征之间的相似度。如果历史库中有存在时间超过 1 秒且注意时间低于 600 毫秒的注意单元,则删除该注意单元,并随机填入一个新产生的注意单元。

II、根据历史库内的注意单元和注意时间,判断注意模式和注意内容。

每隔 1 秒,根据历史库中的注意单元和注意时间,量化不同注意单元的注意分配均衡度。

- 10 当用户的转头角度大于 90 度且小于 270 度时,即用户的视角变化大,则清空注意单元的历史库,当用户头部不动后,再重新累积历史库,1 秒后再次量化注意单元的均衡度。

注意模式:

根据上述判定方法:

- (1) 如果历史库中注意单元的个数为 0 时,则判定注意模式为“扫描看”;
- 15 (2) 如果历史库中注意单元的个数不为 0 时,且不同注意单元的关注时间的均方差大于预设值,例如 100ms,则判定注意模式为“盯着看”,否则为“扫描看”。

注意内容:

(1) 当注意模式为“扫描看”时,此时认为用户主要在感知环境,因此将 APS 帧图像作为注意内容;

- 20 (2) 当注意模式为“盯着看”时,此时认为用户在感知兴趣物体,并将关注度最高的注意单元作为注意内容。

III、根据注意模式和注意内容提取特征、匹配音乐。

本申请实施例提供了两种根据注意模式和注意内容提取特征、匹配音乐的方法策略。

(1) 短时 (short term) 策略

- 25 直接匹配当前时段的注意内容的视觉特征和音乐的音频特征。例如,采用分类的机器学习方法,将注意内容作为深度卷积神经网络的输入,并将神经网络的输出中概率值最大的类别作为判断结果,例如,当概率值大于 0.8 时,判断该视觉特征和该音乐的匹配度高,则该音乐是符合用户的当前感知的。根据图像匹配音乐的过程可以根据现有的任一种图像匹配音乐的方法,本申请实施例在此不做具体限定。

- 30 (2) 长时 (long term) 策略

判定每一个时刻用户注意区域内容属于的状态类别,关联不同时刻状态类别信息,得出一段时间内用户的状态,并根据状态的标签信息来匹配音乐。其中,状态类别可以是“开车”、“学习”、“旅行”、“运动”等听歌高频场景。根据某一时刻用户注意区域内容来判断用户状态类别的过程,可以采用分类的机器学习方法,例如将注意内容作为深度卷积神经网络的输入,并将网络的输出中概率值最大的类别作为判断结果。关联不同时刻状态类别信息,可以采用与时间无关的投票取最高的方法,也可以采用与时间相关的时间加权方法,例如在一段时间内分为十个时刻,其中用户有 8 个时刻被判定为学习,有 2 个时刻被判定为运动,则可以得出用户在这段时间内的状态为学习。

其中,数据采集的方法和耳机端播放音乐的方法与前一音乐推荐方法中数据采集的方

法和耳机端播放音乐的方法一致，为了简洁，本申请实施例在此不再赘述。

本申请实施例的音乐推荐方法根据用户注意不同的内容时，推荐不同的音乐，从而提供更好的音乐体验。本申请实施例的音乐推荐方法通过获取用户的视点数据、头部运动数据

5 和环境数据，来判断用户当前的注意模式，根据判断结果选择全帧图像或者局部注意区域来作为匹配音乐的依据。

以上介绍了本申请实施例的音乐推荐方法，下面介绍本申请实施例的音乐推荐装置。

图 7 示出了本申请实施例的音乐推荐装置的示意性框图，如图 7 所示，包括收发模块 710 和确定模块 720，以下对收发模块 710 和确定模块 720 的作用分别进行介绍。

收发模块 710，用于接收用户的视觉数据。

10 确定模块 720，用于根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长。

所述确定模块 720 还用于根据所述至少一个注意单元的注意时长确定所述用户的注意模式。

所述确定模块 720 还用于根据所述注意模式确定推荐的音乐信息。

15 可选的，所述视觉数据包括所述用户的视点信息和所述用户所视的画面信息，所述视点信息包括视点的位置和所述视点的注意时长。

可选的，所述确定模块 720 根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长，包括：根据所述画面信息获取所述至少一个注意单元；获取所述至少一个注意单元中的所述视点的注意时长之和，以作为所述至少一个注意单元的注意时

20 长。

可选的，所述确定模块 720 根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长，还包括：判断所述至少一个注意单元中的第一注意单元和第二注意单元的相似度，所述第一注意单元和所述第二注意单元为不同时刻的注意单元；如果所述相似度大于或等于第一阈值，所述第二注意单元的注意时长等于所述第一注意单元的注意

25 时长和所述第二注意单元的注意时长之和。

可选的，所述确定模块 720 根据所述至少一个注意单元的注意时长确定所述用户的注意模式，包括：如果所述至少一个注意单元的注意时长的标准差大于或等于第二阈值，确定所述用户的注意模式为盯着看；如果所述至少一个注意单元的注意时长的标准差小于第二阈值，确定所述用户的注意模式为扫描看。

30 可选的，所述确定模块 720 用于根据所述注意模式确定音乐信息，包括：如果所述注意模式为扫描看，根据所述画面信息确定音乐信息；如果所述注意模式为盯着看，根据所述注意单元中关注度最高的注意单元确定音乐信息。

所述确定模块 720 根据所述注意模式确定音乐信息，还包括：根据所述注意模式确定所述用户在第一时间段内的每个时刻的行为状态；根据所述每个时刻的状态确定所述用户在第一时间段内的行为状态；根据所述第一时间段内的行为状态确定音乐信息。

35

应理解，本申请实施例的音乐推荐装置 700 中的收发模块 710 可以用于执行图 5 中 S501 的方法，确定模块 720 可以用于执行图 5 中 S502 至 S504 的方法，具体描述可以参照上述对于图 5 的介绍，为了简洁，本申请实施例在此不再赘述。

图 8 是本申请实施例的音乐推荐设备 800 的示意性框图。该音乐推荐设备 800 可以用

于执行上文实施例提供的音乐推荐的方法，为了简洁，此处不再赘述。音乐推荐设备 800 包括：处理器 810，处理器 810 与存储器 820 耦合，存储器 820 用于存储计算机程序或指令，处理器 810 用于执行存储器 820 存储的计算机程序或指令，使得上文方法实施例中的方法被执行。

- 5 本申请实施例还提供一种计算机可读存储介质，计算机可读存储介质中存储有程序指令，当所述程序指令由处理器运行时，以实现本申请实施例的音乐推荐的方法。

 本申请实施例还提供一种计算机程序产品，其特征在于，所述计算机程序产品包括计算机程序代码，当所述计算机程序代码在计算机上运行时，以实现本申请实施例的音乐推荐的方法。

- 10 本申请实施例还提供一种音乐推荐系统，其特征在于，所述系统包括数据采集设备和终端设备，所述终端设备包括处理器和存储器，所述存储器中存储有一个或多个程序，所述一个或多个计算机程序包括指令，其中，所述数据采集设备，用于采集用户的视觉数据；当所述指令被所述一个或多个处理器执行时，使得所述终端设备执行本申请实施例的音乐推荐的方法。

- 15 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

- 20 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

- 在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。
- 25 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

- 另外，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单元中，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单元中。
- 30 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机

- 软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（read-only memory，ROM）、随机存取存储器（random access memory，RAM）、磁碟或者光盘等各种可以存储程序代码的

- 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机
- 35 软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（read-only memory，ROM）、随机存取存储器（random access memory，RAM）、磁碟或者光盘等各种可以存储程序代码的

介质。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以所述权利要求的保护范围为准。

权 利 要 求 书

- 1、一种音乐推荐方法，其特征在于，包括：
接收用户的视觉数据；
- 5 根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长；
根据所述至少一个注意单元的注意时长确定所述用户的注意模式；
根据所述注意模式确定推荐的音乐信息。
- 2、根据权利要求 1 所述的方法，其特征在于，所述视觉数据包括所述用户的视点信息和所述用户所视的画面信息，所述视点信息包括视点的位置和所述视点的注意时长。
- 10 3、根据权利要求 2 所述的方法，其特征在于，所述根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长，包括：
根据所述画面信息获取所述至少一个注意单元；
获取所述至少一个注意单元中的所述视点的注意时长之和，以作为所述至少一个注意单元的注意时长。
- 15 4、根据权利要求 3 所述的方法，其特征在于，所述根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长，还包括：
判断所述至少一个注意单元中的第一注意单元和第二注意单元的相似度，所述第一注意单元和所述第二注意单元为不同时刻的注意单元；
如果所述相似度大于或等于第一阈值，所述第二注意单元的注意时长等于所述第一注意单元的注意时长和所述第二注意单元的注意时长之和。
- 20 5、根据权利要求 1 至 4 中任一项所述的方法，其特征在于，所述根据所述至少一个注意单元的注意时长确定所述用户的注意模式，包括：
如果所述至少一个注意单元的注意时长的标准差大于或等于第二阈值，确定所述用户的注意模式为盯着看；
如果所述至少一个注意单元的注意时长的标准差小于第二阈值，确定所述用户的注意模式为扫描看。
- 25 6、根据权利要求 5 所述的方法，其特征在于，所述根据所述注意模式确定音乐信息，包括：
如果所述注意模式为扫描看，根据所述画面信息确定音乐信息；
如果所述注意模式为盯着看，根据所述注意单元中关注度最高的注意单元确定音乐信息。
- 30 7、根据权利要求 1 至 5 中任一项所述的方法，其特征在于，所述根据所述注意模式确定音乐信息，还包括：
根据所述注意模式确定所述用户在第一时间段内的每个时刻的行为状态；
根据所述每个时刻的状态确定所述用户在第一时间段内的行为状态；
根据所述第一时间段内的行为状态确定音乐信息。
- 35 8、一种音乐推荐的装置，其特征在于，包括：
收发模块，用于接收用户的视觉数据；

确定模块,用于根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长;

所述确定模块还用于根据所述至少一个注意单元的注意时长确定所述用户的注意模式;

5 所述确定模块还用于根据所述注意模式确定推荐的音乐信息。

9、根据权利要求 8 所述的装置,其特征在于,所述视觉数据包括所述用户的视点信息和所述用户所视的画面信息,所述视点信息包括视点的位置和所述视点的注意时长。

10、根据权利要求 9 所述的装置,其特征在于,所述确定模块根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长,包括:

10 根据所述画面信息获取所述至少一个注意单元;

获取所述至少一个注意单元中的所述视点的注意时长之和,以作为所述至少一个注意单元的注意时长。

11、根据权利要求 10 所述的装置,其特征在于,所述确定模块根据所述视觉数据获取至少一个注意单元和所述至少一个注意单元的注意时长,还包括:

15 判断所述至少一个注意单元中的第一注意单元和第二注意单元的相似度,所述第一注意单元和所述第二注意单元为不同时刻的注意单元;

如果所述相似度大于或等于第一阈值,所述第二注意单元的注意时长等于所述第一注意单元的注意时长和所述第二注意单元的注意时长之和。

12、根据权利要求 8 至 11 中任一项所述的装置,其特征在于,所述确定模块根据所述至少一个注意单元的注意时长确定所述用户的注意模式,包括:

20 如果所述至少一个注意单元的注意时长的标准差大于或等于第二阈值,确定所述用户的注意模式为盯着看;

如果所述至少一个注意单元的注意时长的标准差小于第二阈值,确定所述用户的注意模式为扫描看。

25 13、根据权利要求 12 所述的装置,其特征在于,所述确定模块用于根据所述注意模式确定音乐信息,包括:

如果所述注意模式为扫描看,根据所述画面信息确定音乐信息;

如果所述注意模式为盯着看,根据所述注意单元中关注度最高的注意单元确定音乐信息。

30 14、根据权利要求 8 至 12 中任一项所述的装置,其特征在于,所述确定模块根据所述注意模式确定音乐信息,还包括:

根据所述注意模式确定所述用户在第一时间段内的每个时刻的行为状态;

根据所述每个时刻的状态确定所述用户在第一时间段内的行为状态;

根据所述第一时间段内的行为状态确定音乐信息。

35 15、一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有程序指令,当所述程序指令由处理器运行时,以实现权利要求 1 至 7 中任一项所述的方法。

16、一种计算机程序产品,其特征在于,所述计算机程序产品包括计算机程序代码,当所述计算机程序代码在计算机上运行时,以实现权利要求 1 至 7 中任一项所述的方法。

17、一种音乐推荐系统,其特征在于,所述系统包括数据采集设备和终端设备,所述

终端设备包括处理器和存储器，所述存储器中存储有一个或多个程序，所述一个或多个计算机程序包括指令，其中，

所述数据采集设备，用于采集用户的视觉数据；

当所述指令被所述一个或多个处理器执行时，使得所述终端设备执行权利要求 1 至 7 中任一项所述的方法。