角色分离方法、会议纪要的记录方法、角色展示方法、装置、电子设备及计算机存储介质

本申请要求2020年06月28日递交的申请号为202010596049.3、发明名称为“角色分离方法、会议纪要的记录方法、角色展示方法、装置、电子设备及计算机存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明实施例涉及语音处理领域，尤其涉及一种角色分离方法、会议纪要的记录方法、角色展示方法、装置、电子设备及计算机存储介质。

背景技术

随着信息技术的不断发展，人们对于高精度信息化分析的需求越来越高。基于电子设备实现的通话或者会议是人们生活中不可或缺的一部分，对应于此，通话内容或者会议内容的记录、分析已成为相关技术领域人员研究的热点，例如，在公众报警电话、各类热线、公司会议等领域，可以对通话内容或者会议内容进行记录、分析，用以实现后期信息总结、检索等工作。

说话人角色分离作为会议内容分析的一个重要步骤，其分离的实时性直接影响用户体验。目前，大多基于声纹识别实现说话人角色分离。由于声纹识别需要积累一定时长的语音数据，才能保障较高的识别准确率。因此，市面上大多数基于声纹识别的角色分离系统，都是基于离线的语音数据完成角色的分离，很难实时地实现角色的分离。由此可见，如何实时地分离角色，从而提高用户体验成为当前亟待解决的技术问题。

发明内容

有鉴于此，本发明实施例提供一种角色分离方案，以至少部分解决上述技术问题。

根据本发明实施例的第一方面，提供了一种角色分离方法。所述方法包括：获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；基于所述待分离的角色的第一身份识别结果分离所述角色。

根据本发明实施例的第二方面，提供了一种角色分离方法。所述方法包括：向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色；接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

根据本发明实施例的第三方面，提供了一种角色分离方法。所述方法包括：接收语

音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求；基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果；基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

根据本发明实施例的第四方面，提供了一种会议纪要的记录方法。所述方法包括：获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果；基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

根据本发明实施例的第五方面，提供了一种角色展示方法。所述方法包括： 获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果；基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

根据本发明实施例的第六方面，提供了一种角色分离装置。所述装置包括：第一获取模块，用于获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；第一身份识别模块，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；分离模块，用于基于所述待分离的角色的第一身份识别结果分离所述角色。

根据本发明实施例的第七方面，提供了一种角色分离装置。所述装置包括：第一发送模块，用于向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色；第一接收模块，用于接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

根据本发明实施例的第八方面，提供了一种角色分离装置。所述装置包括：第二接收模块，用于接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求；第三获取模块，用于基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据；第二身份识别模块，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果；第二发送模块，用于基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

根据本发明实施例的第九方面，提供了一种会议纪要的记录装置。所述装置包括：第四获取模块，用于获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据；第三身份识别模块，用于基于所述声源角度数据，对所述会议角色

进行身份识别，以获得所述会议角色的身份识别结果；记录模块，用于基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

根据本发明实施例的第十方面，提供了一种角色展示装置。所述装置包括：第五获取模块，用于获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据；第四身份识别模块，用于基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果；第一展示模块，用于基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

根据本发明实施例的第十一方面，提供了一种电子设备，包括：处理器、存储器、通信接口和通信总线，所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信；所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行如第一方面、第二方面或者第三方面所述的角色分离方法对应的操作，或者执行如第四方面所述的会议纪要的记录方法对应的操作，或者执行如第五方面所述的角色展示方法对应的操作。

根据本发明实施例的第十二方面，提供了一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现如第一方面、第二方面或者第三方面所述的角色分离方法，或者实现如第四方面所述的会议纪要的记录方法，或者实现如第五方面所述的角色展示方法。

根据本发明实施例提供的角色分离方案，获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对待分离的角色进行身份识别，以获得待分离的角色的第一身份识别结果；再基于待分离的角色的第一身份识别结果分离角色，与现有的其它方式相比，基于语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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发明实施例中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

图 1A 为本申请实施例一中角色分离方法的步骤流程图；

图 1B 为根据本申请实施例一提供的近场模型下的声音传播示意图；

图 1C 为根据本申请实施例一提供的说话人分离方法的场景示意图；

图 2A 为本申请实施例二中角色分离方法的步骤流程图；

图 2B 为根据本申请实施例二提供的角色分离方法的场景示意图；

图 3A 为本申请实施例三中角色分离方法的步骤流程图；

图 3B 为根据本申请实施例三提供的角色分离方法的场景示意图；

图 4A 为本申请实施例四中角色分离方法的步骤流程图；

图 4B 为根据本申请实施例四提供的角色分离方法的场景示意图；

图 5 为本申请实施例五中会议纪要的记录方法的步骤流程图；

图 6 为本申请实施例六中角色展示方法的步骤流程图；

图 7 为本申请实施例七中角色分离装置的结构示意图；

图 8 为本申请实施例八中角色分离装置的结构示意图；

图 9 为本申请实施例九中角色分离装置的结构示意图；

图 10 为本申请实施例十中角色分离装置的结构示意图；

图 11 为本申请实施例十一中会议纪要的记录装置的结构示意图；

图 12 为本申请实施例十二中角色展示装置的结构示意图；

图 13 为本申请实施例十三中电子设备的结构示意图。

具体实施方式

为了使本领域的人员更好地理解本发明实施例中的技术方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本发明实施例一部分实施例，而不是全部的实施例。基于本发明实施例中的实施例，本领域普通技术人员所获得的所有其他实施例，都应当属于本发明实施例保护的范围。

下面结合本发明实施例附图进一步说明本发明实施例具体实现。

参照图 1A，示出了本申请实施例一中角色分离方法的步骤流程图。

具体地，本实施例提供的角色分离方法包括以下步骤：

在步骤 S101 中，获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据。

在本申请实施例中，所述语音采集设备可包括拾音器。所述待分离的角色可为待分离的会议讲话人、待分离的通话人等。所述语音数据帧可理解为时长为 20 毫秒至 30 毫秒的语音片段。所述声源角度数据可理解为所述待分离的角色在说话时与所述语音采集设备形成的角度。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，所述语音采集设备包括麦克风阵列。在获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据时，获取所述麦克风阵列中至少部分麦克风接收到的所述语音数据帧的协方差矩阵；对所述协方差矩阵进行特征值分解，以得到多个特征值；从所述多个特征值中选取第一数量个最大的特征值，并基于选取的

特征值对应的特征向量构成语音信号子空间，其中，所述第一数量与声源估计数量相当；基于所述语音信号子空间，确定所述声源角度数据。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，可以在支持语音交互功能的设备(如拾音器)上设置麦克风阵列，麦克风阵列用于接收附近的声音输入。麦克风阵列是一组位于空间不同位置的全向麦克风按一定的形状规则布置形成的阵列，是对空间传播声音输入进行空间采样的一种装置，采集到的信号包含了其空间位置信息。根据麦克风阵列的拓扑结构，可分为线性阵列、平面阵列、体阵列等。根据声源和麦克风阵列之间距离的远近，则可将阵列分为近场模型和远场模型。近场模型将声波看成球面波，它考虑麦克风阵元接收信号间的幅度差；远场模型则将声波看成平面波，它忽略各阵元接收信号间的幅度差，近似认为各接收信号之间是简单的时延关系。可以根据麦克风阵列中至少部分麦克风接收到的信号，进行声源定位，以确定角色的位置信息。所确定的位置信息可以是角色的二维位置坐标，也可以是角色相对于所述至少部分麦克风的方位角和距离。其中，方位角为角色在所述至少部分麦克风所在的坐标系中的方位角，也即是声源角度数据，距离为角色与所述至少部分麦克风的中心位置之间的距离。作为一个示例，根据麦克风阵列中的部分麦克风或全部麦克风接收到的信号，可以利用 MUSIC 算法(Multiple Signal classification，多信号分类算法)进行声源定位。MUSIC 算法的基本思想为将任意阵列输出数据的协方差矩阵进行特征值分解，从而得到与信号分量相对应的信号子空间和信号分量相正交的噪声子空间，然后利用这两个子空间的正交性来估计信号的参数(入射方向、极化信息和信号强度)。例如，可以利用这两个子空间的正交性构成空间扫描谱，进行全域搜索谱峰，从而实现信号的参数估计。

以麦克风阵列应用于拾音器为例，麦克风阵列可以是线性阵列，声场模型可以视为近场模型。近场情况下，声源信号到达各阵列麦克风的时间差为 τ，和远场相比，不仅随着角度变化，也会随着距离变化。如图 1B 所示，设待分离的角色到麦克风阵列中各个麦克风的距离分别为 R_1，R_2，...，R_{N-1}，R_N，声波在空气中传播速度为 C，则声波到达第 i 个麦克风相对于到达第 1 个麦克风的时间差为 τ_i，其中，

$$\tau_i = \frac{R_i - R_1}{C}$$

近场模型下的声源定位过程描述如下。

首先可以获取麦克风阵列中至少部分麦克风接收到的信号的协方差矩阵。例如，协方差矩阵可以表示为 R(f)，$R(f)=E[X(f)X(f)^H]$，其中，X(f)是麦克风阵列中至少部分麦克风接收到的信号经过傅里叶变换(如短时傅里叶变换)后在不同频点 f 的数据，为频域数据。X(f)可以视为一个向量，向量中每一个元素代表一个麦克风接收到的信号经过傅里叶变换后在不同频点 f 的数据。例如，X(f)可以表示为

$$X(f)=\{X_1(f),\ X_2(f)\ldots X_M(f)\}$$

其中，$X_1(f)$、$X_2(f)$、$X_M(f)$表示不同麦克风接收到的信号经过傅里叶变换(如短时傅里叶变换)后在不同频点 f 的数据，M 为麦克风的个数。X(f)的表达式中实际上隐含了一个时间变量 t，完整表示应该为 X(f，t)，表示一个时间段 t 所含的数据。E 表示数学期望，求数学期望或者均值，实际是对时间 t 而言，可以理解为 $E[X(f,\ t)\ X(f,\ t)^H]$，或者

$$\frac{1}{N2-N1}\sum_{t=N1}^{N2}X(f,\ t)$$

其中，N2-N1 表示 X(f，t)对应的时间段，N1 表示起始时间，N2 表示结束时间。

然后对协方差矩阵进行特征值分解，可以得到多个特征值。从这多个特征值中可以选取第一数量个最大的特征值，所选取的特征值对应的特征向量就可以构成信号子空间。其中，剩余的特征值对应的特征向量可以构成噪声子空间，其中，第一数量与声源估计数量相当，如在认为有 3 个声源信号时，可以取最大的三个特征值对应的特征向量构成信号子空间。声源估计数量可以通过经验或其他估计方式计算得到，此处不再赘述。例如，对 R(f)进行特征值分解后为，$R(f)=U_s(f)\Sigma_sU_s(f)^H+U_N(f)\Sigma_NU_N(f)^H$，其中，$U_s(f)$是大特征值对应的特征向量构成的信号子空间，$U_N(f)$是小特征对应的特征向量构成的噪声子空间，S、N 表示对信号 U 的不同划分，S 代表信号，N 代表噪声，所划分的 U_s 代表信号子空间，U_N 代表噪声子空间。Σ 代表对角矩阵，表示由特征值组成的矩阵。实际上，对 R(f)进行特征值分解，有 $R(f)=U(f)\Sigma U(f)^H$，其中，Σ 是一个仅有主对角元素的矩阵，Σ 中的主对角元素即是分解得到的特征值，将 U 和 Σ 按 Σ 的主对角元素(特征值)大小进行分类，可以分为较大的一类 S(即由大特征值对应的特征向量构成的信号子空间)，和较小的一类 N(由剩余的小特征值对应的特征向量构成的噪声子空间)，就有 $R(f)=U_s(f)\Sigma_sU_s(f)^H+U_N(f)\Sigma_NU_N(f)^H$。

基于信号子空间，可以确定声源位置。例如，可以基于信号子空间，确定信号在二维空间的最大响应，基于最大响应所对应的波达方向(DOA)，可以确定声源位置，即角色的位置信息。

作为示例，目标信号在二维空间的响应计算公式为

$$S_{R,\theta}=\sum_f a^H(R,\theta,f)U_s(f)U_s(f)^Ha(R,\theta,f)$$

f 是一个取值范围，a(R,θ,f)可由相对时间差 τ 求得。其中，a(R,θ,f)表示麦克风阵列的导向矢量。R 是声源和麦克风阵列中心的距离，θ 是声源在阵列坐标系中的方位角。设声源在(R,θ)位置时，相对时间差 τ 定义为：声源到达各个麦克风所需时间相对于到达第一个麦克风所需时间的差 $\tau=(\tau_1,\tau_2,\ldots,\tau_M)$，$\tau_1=0$，然后可以求得对应位置(R ,θ)在频率 f 的导向矢量 $a(R,\theta,f)=(a_1,a_2,\ldots,a_M)$，其中 $\alpha_i=e^{-2\pi f\tau_i}$。待分离的角色二维坐标为 $(R_{target},\ \theta_{target})=\text{argmax}(R,\theta)S_{R,\theta}$。也就是说，响应 $S_{R,\theta}$ 最大时的(R，θ)即为待分离的角色的位置。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，在获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据之后，所述方法还包括：对所述待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧；基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；基于所述过滤平滑后的语音数据帧，对所述声源角度数据进行更新，以获得更新后的声源角度数据。籍此，通过对待分离的角色的语音数据帧进行语音端点检测和过滤平滑，能够获得更加稳定的声源角度数据。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，语音端点检测(Voice Activity Detection，VAD)，也可以称为语音活动检测，是从包含语音的一段信号中确定出语音的起点及终点，进而提取出相应的非静音语音信号，从而排除静音段和非语音信号的干扰，使得处理质量得到保证。此外，有效的端点检测还能使处理时间减到最少。在对所述待分离的角色的语音数据帧进行语音端点检测时，可以基于空间熵进行语音端点检测，还可以基于神经网络模型进行语音端点检测。其中，基于空间熵进行语音端点检测的过程如下：麦克风阵列接收到的声音信号中可能存在待分离的角色的声音以及周围的环境噪声。因此，可以根据麦克风阵列中至少部分麦克风接收到的声音信号的信号空间的混乱程度，来进行语音端点检测。在本实施例中，可以用空间熵来表征信号空间的混乱程度。可以在空间熵较小的情况下，认为存在语音活动，在空间熵较大的情况下，认为不存在语音活动。作为示例，可以首先获取麦克风阵列中至少部分麦克风接收到的信号的协方差矩阵，对协方差矩阵进行特征值分解，以得到多个特征值。如上文所述，大特征值构成的信号子空间可以视为是语音子空间，小特征值构成的信号子空间可以视为噪声子空间，因此可以通过对多个特征值进行分析，确定是否存在语音活动。如可以将每个特征值视为一个信号子空间(也即信号源)，计算这多个特征值的熵(即空间熵)，根据计算得到的空间熵的大小，可以判断是否存在语音活动。例如，可以对多个特征值进行归一化处理，并计算经过归一化处理后得到的多个值的空间熵，将空间熵与预定阈值进行比较，基于空间熵与预定阈值的比较结果，判断是否存在语音活动。如可以在空间熵小于预定阈值的情况下，判定存在语音活动，大于或等于预定阈值的情况下，判定不存在语音活动。其中，预定阈值的取值可以根据实际情况设定，如可以与选取的定位频带有关，例如在定位频带取 500-5000(HZ)时，预定阈值可以取 1，空间熵小于 1 时可以判定存在语音活动，反之则可以判断为噪音，不存在语音活动。其中，空间熵为 ES，

$$ES = -\sum_{i=1}^{N} p_i \log(p_i)$$

p_i 为特征值经过归一化处理后得到的值，N 为对协方差矩阵进行特征值分解后得到的特征值的个数，log 的底数为大于 1 的数值，如可以是 2、10、e，对此本公开不做限定。其中，基于神经网络模型进行语音端点检测的过程如下：可以使用语音活动检测模

型对基于麦克风阵列中至少部分麦克风获取的语音数据帧进行预测，以判断是否存在语音活动。其中，语音活动检测模型用于对输入的语音数据帧的语音活动状态进行预测，语音活动检测模型可以是基于神经网络的模型，可以通过有监督的机器学习的方式得到预测模型。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，在基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧时，通过中值滤波器，基于所述待分离的角色的语音数据帧的能量频谱的谱平度，对所述具有语音端点的语音数据帧进行过滤平滑，以获得所述过滤平滑后的语音数据帧。籍此，通过中值滤波器，基于待分离的角色的语音数据帧的能量频谱的谱平度，对具有语音端点的语音数据帧进行过滤平滑，能够有效提升针对具有语音端点的语音数据帧的过滤平滑效果。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，中值滤波是一种非线性数字滤波器技术，经常用于去除图像或者其它信号中的噪声。中值滤波器的设计思想就是检查输入信号中的采样并判断它是否代表了信号，使用奇数个采样组成的观察窗实现这项功能。观察窗口中的数值进行排序，位于观察窗中间的中值作为输出。然后，丢弃最早的值，取得新的采样，重复上面的计算过程。所述待分离的角色的语音数据帧的能量频谱的谱平度可理解为所述待分离的角色的语音数据帧的能量频谱的平坦度，是能量频谱的特征参数，可通过对所述待分离的角色的语音数据帧的能量频谱进行计算，获得所述待分离的角色的语音数据帧的能量频谱的平坦度。在通过中值滤波器，基于所述待分离的角色的语音数据帧的能量频谱的谱平度，对所述具有语音端点的语音数据帧进行过滤平滑时，通过中值滤波器，基于所述待分离的角色的语音数据帧的能量频谱的谱平度，对所述具有语音端点的语音数据帧的能量频谱进行过滤平滑，以获得过滤平滑后的能量频谱；基于所述过滤平滑后的能量频谱，确定所述过滤平滑后的语音数据帧。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，在基于所述过滤平滑后的语音数据帧，对所述声源角度数据进行更新时，获取所述过滤平滑后的语音数据帧所对应的声源角度数据；使用所述过滤平滑后的语音数据帧所对应的声源角度数据更新所述待分离的角色的语音数据帧所对应的声源角度数据。其中，所述获取所述过滤平滑后的语音数据帧所对应的声源角度数据的具体实施方式与上述获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S102 中，基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果。

在一些可选实施例中，在基于所述声源角度数据，对所述待分离的角色进行身份识

别，以获得所述待分离的角色的第一身份识别结果时，对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的第一身份识别结果。其中，所述角色身份标识可为角色的姓名、昵称或者身份编码等。籍此，通过对声源角度数据进行顺序聚类，能够准确地获得待分离的角色的第一身份识别结果。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，顺序聚类通常是用于对一定数量的同质数据集合(或集群)中的数据进行分类的非监督方法。在这种情况下，集群的数量没有预先确定，而是根据给定的标准按顺序逐渐增加(一个接一个)，直到满足适当的停止条件。顺序聚类算法的优点是双重的。首先，避免了不必要的大量集群的冗余计算。其次，集群通常按有序顺序提取，从最重要的集群(具有最大容量的集群)到最不重要(最小容量)的集群。在对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果时，确定所述声源角度数据与声源角度顺序聚类中心的距离；基于所述声源角度数据与所述声源角度顺序聚类中心的距离，确定所述声源角度数据的顺序聚类结果。其中，所述声源角度顺序聚类中心可理解为声源角度每一顺序聚类的中心角度，例如，假设有三个声源角度的顺序聚类，那么三个声源角度的顺序聚类的中心角度可为 30 度、60 度、90 度。在确定所述声源角度数据与声源角度顺序聚类中心的距离时，确定所述声源角度数据与所述声源角度顺序聚类中心的差的绝对值为所述声源角度数据与声源角度顺序聚类中心的距离。在基于所述声源角度数据与所述声源角度顺序聚类中心的距离，确定所述声源角度数据的顺序聚类结果时，将所述声源角度数据与所述声源角度顺序聚类中心的距离与预设的距离阈值进行比较，若所述声源角度数据与所述声源角度顺序聚类中心的距离小于预设的距离阈值，则确定所述声源角度数据的顺序聚类结果为所述声源角度顺序聚类中心所在的顺序聚类，若所述声源角度数据与所述声源角度顺序聚类中心的距离等于或大于预设的距离阈值，则确定所述声源角度数据的顺序聚类结果不为所述声源角度顺序聚类中心所在的顺序聚类。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S103 中，基于所述待分离的角色的第一身份识别结果分离所述角色。

在本申请实施例中，在获得所述待分离的角色的第一身份识别结果之后，便可使用所述待分离的角色的第一身份识别结果区分所述待分离的角色，从而实现了所述待分离的角色的分离。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，如图 1C 所示，待分离的角色的语音数据帧由语音采集设备采集。当语音采集设备采集到待分离的角色的语音数据帧之后，对待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧，并基于所述待分离的角

色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；再对所述过滤平滑后的语音数据帧进行声源定位，以确定声源角度数据；再对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；再确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的身份识别结果，最后基于所述待分离的角色的身份识别结果分离所述角色。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

通过本申请实施例提供的角色分离方法，获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对待分离的角色进行身份识别，以获得待分离的角色的第一身份识别结果；再基于待分离的角色的第一身份识别结果分离角色，与现有的其它方式相比，基于语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

本实施例提供的角色分离方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：摄像头、终端、移动终端、PC 机、服务器、车载设备、娱乐设备、广告设备、个人数码助理（PDA）、平板电脑、笔记本电脑、掌上游戏机、智能眼镜、智能手表、可穿戴设备、虚拟显示设备或显示增强设备等。

参照图 2A，示出了本申请实施例二的角色分离方法的步骤流程图。

具体地，本实施例提供的角色分离方法包括以下步骤：

在步骤 S201 中，获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据。

由于该步骤 S201 的具体实施方式与上述步骤 S101 的具体实施方式类似，在此不再赘述。

在步骤 S202 中，基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果。

由于该步骤 S202 的具体实施方式与上述步骤 S102 的具体实施方式类似，在此不再赘述。

在步骤 S203 中，对所述待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得所述待分离的角色的第二身份识别结果。

在本申请实施例中，声纹(Voiceprint)指的是人类语音中携带言语信息的声波频谱，具备独特的生物学特征，具有身份识别的作用。声纹识别(Voiceprint Identification)，又称角色识别(Speaker Identification)，该技术是从角色发出的语音信号中提取语音特征，并据此对角色进行身份验证的生物识别技术。声纹识别的过程通常是，预先存储某个或某些用户的声纹信息(存储了声纹信息的用户为注册用户)，将从角色语音信号中提取出来的语音特征与预先存储的声纹进行比对，得到一个相似度分值，然后将该分值与阈值进行

比较，若分值大于阈值，则认为角色就是该声纹所对应的注册用户；若分值小于等于阈值，则认为角色不是该声纹所对应的注册用户。所述预设时间段可由本领域技术人员根据实际需求进行设定，本申请实施例对此不做任何限定。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，所述待分离的角色在预设时间段内的语音数据帧可以在声纹识别之前经受不同层次的预处理。这种预处理可以促进更加高效的声纹识别。在各种实施方式中，预处理可以包括：采样；量化；去除非语音的音频数据和静默的音频数据；对包括语音的音频数据进行分帧、加窗，以供后续处理等等。经过预处理之后，可以提取待分离的角色在预设时间段内的语音数据帧的语音特征，并基于语音数据帧的语音特征将语音数据帧与用户的声纹进行匹配。语音特征可以是滤波器组 FBank(Filter Bank)、梅尔频率倒谱系数 MFCC(Mel Frequency Cepstral Coefficents)、感知线性预测系数 PLP、深度特征 Deep Feature、以及能量规整谱系数 PNCC 等特征中的一种或者多种的组合。在一种实施例中，还可以对提取得到的语音特征进行归一化处理。而后，基于语音数据帧的语音特征，将语音数据帧与用户的声纹进行匹配，以得到语音数据帧与用户的声纹之间的相似评分，并根据该相似评分来确定与语音数据帧相匹配的用户。具体地，在一些实施方式中，用户的声纹以声纹模型来描述，例如隐马尔可夫模型(HMM 模型)、高斯混合模型(GMM 模型)等等。用户的声纹模型以语音特征为特征，利用包括用户语音的音频数据(后文简称为用户的音频数据)训练得到。可以采用匹配运算函数来计算语音数据帧的与用户的声纹之间的相似度。例如可以计算语音数据帧的语音特征与用户的声纹模型相匹配的后验概率来作为相似评分，也可以计算语音数据帧的语音特征与用户的声纹模型之间的似然度来作为相似评分。但由于训练好用户的声纹模型需要大量该用户的音频数据，因此在一些实施方式中，用户的声纹模型可以基于与用户无关的通用背景模型，利用少量用户的音频数据训练得到(同样以语音特征为特征)。例如，可以先使用与用户无关的、多个角色的音频数据，通过期望最大化算法 EM 训练得到通用背景模型(Universal Background Model，UBM)，以表征用户无关的特征分布。再基于该 UBM 模型，利用少量的用户的音频数据通过自适应算法(如最大后验概率 MAP，最大似然线性回归 MLLR 等)训练得到 GMM 模型(这样得到的 GMM 模型称之为 GMM-UBM 模型)，以表征用户的特征分布。该 GMM-UBM 模型即为用户的声纹模型。此时，可以基于语音数据帧的语音特征，分别将语音数据帧与用户的声纹模型和通用背景模型进行匹配，以得到语音数据帧与用户的声纹之间的相似评分。例如，分别计算语音数据帧的语音特征与上述 UBM 模型和 GMM-UBM 模型之间的似然度，然后将这两个似然度相除后取对数，将得到的值作为语音数据帧与用户的声纹之间的相似评分。

在另一些实施方式中，用户的声纹以声纹向量来描述，例如 i-vector、d-vector、x-vector 和 j-vector 等等。可以至少基于语音数据帧的语音特征，提取语音数据帧的声纹向量。根

据一种实施例，可以先利用语音数据帧的语音特征训练待分离的角色的声纹模型。如前文类似地，可以基于预先训练好的与用户无关的上述通用背景模型，利用语音数据帧的语音特征训练得到待分离的角色的声纹模型。在得到待分离的角色的声纹模型之后，可以根据该声纹模型提取语音数据帧的均值超矢量。例如，可以将待分离的角色的GMM-UBM 模型的各个 GMM 分量的均值进行拼接，得到待分离的角色的 GMM-UBM 模型的均值超矢量，即语音数据帧的均值超矢量。之后，可以采用联合因子分析法(JFA)或者简化的联合因子分析法，从语音数据帧的均值超矢量中提取得到低维的声纹向量。以 i-vector 为例，在训练得到与用户无关的上述通用背景模型(UBM 模型)之后，可以提取该通用背景模型的均值超矢量，并估计全局差异空间(Total Variability Space，T)矩阵。而后基于语音数据帧的均值超矢量、T 矩阵、通用背景模型的均值超矢量来计算语音数据帧的 i-vector。具体地，i-vector 可以根据以下公式计算得到：

$$M_{s,h}=m_u+T\omega_{s,h}$$

其中，$M_{s,h}$ 是从角色 s 的语音 h 中得到的均值超矢量，m_u 是通用背景模型的均值超矢量，T 是全局差异空间矩阵，$\omega_{s,h}$ 是全局差异因子，也就是 i-vector。

根据另一种实施例，还可以利用训练好的深度神经网络(Deep Neural Network，DNN)来得到语音数据帧的声纹向量。以 d-vector 为例，DNN 可以包括输入层、隐层和输出层。可以先将语音数据帧的 FBank 特征输入到 DNN 输入层，DNN 最后一个隐层的输出即为 d-vector。

在得到语音数据帧的声纹向量之后，可以基于语音数据帧的声纹向量和用户的声纹向量，来计算语音数据帧与用户的声纹之间的相似评分。其中，可以采用支持向量机(SVM)、LDA(Linear Discriminant Analysis，线性判别分析)、PLDA(Probabilistic Linear Discriminant Analysis，概率线性判别分析)、似然度和余弦距离(Cosine Distance)等算法来计算语音数据帧与用户的声纹之间的相似评分。

以 PLDA 算法为例，假设语音由 I 个角色的语音组成，其中每个角色有 J 段不一样的语音，并且定义第 i 个角色的第 j 段语音为 Y_{ij}。那么，定义 Y_{ij} 的生成模型为：

$$Y_{ij}=\mu+Fh_i+Gw_{ij}+\varepsilon_{ij}$$

其中，μ 是声纹向量的均值，F、G 是空间特征矩阵，各自代表角色类间特征空间和类内特征空间。F 的每一列，相当于类间特征空间的特征向量，G 的每一列，相当于类内特征空间的特征向量。向量 h_i 和 w_{ij} 可以看作是该语音分别在各自空间的特征表示，ε_{ij} 则是噪声协方差。如果两条语音的 h_i 特征相同的似然度越大，即相似评分越高，那么它们来自同一个角色的可能性就越大。

PLDA 的模型参数包括 4 个，即 μ、F、G 和 ε_{ij}，是采用 EM 算法迭代训练而成。通常地，可以采用简化版的 PLDA 模型，忽略类内特征空间矩阵 G 的训练，只训练类间特征空间矩阵 F，即：

$$Y_{ij}=\mu+Fh_i+\varepsilon_{ij}$$

可以基于语音数据帧的声纹向量，参照上述公式得到语音数据帧的 h_i 特征。同样地，基于用户的声纹向量，参照上述公式得到用户语音的 h_i 特征。而后，可以计算两个 h_i 特征的对数似然比或余弦距离来作为语音数据帧与用户的声纹之间的相似评分。

应当注意的是，声纹并不限于上述声纹向量(i-vector、d-vector 和 x-vector 等等)和上述声纹模型(HMM 模型和 GMM 模型等等)，相应的相似评分算法也可依据所选定的声纹来任意选取，本发明对此不做限制。

在各种实施方式中，如果得到的相似评分超过相似阈值，则确定语音数据帧与该用户的声纹相匹配，也就是确定语音数据帧与该声纹对应的用户相匹配。否则确定语音数据帧不与该用户的声纹相匹配。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S204 中，若所述第一身份识别结果与所述第二身份识别结果不相同，则使用所述第二身份识别结果更正所述第一身份识别结果，以获得所述待分离的角色的最终身份识别结果。

在本申请实施例中，若所述第一身份识别结果与所述第二身份识别结果相同，则无需使用所述第二身份识别结果更正所述第一身份识别结果，并将所述第一身份识别结果确定为所述待分离的角色的最终身份识别结果。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S205 中，基于所述待分离的角色的最终身份识别结果分离所述角色。

在本申请实施例中，在获得所述待分离的角色的最终身份识别结果之后，便可使用所述待分离的角色的最终身份识别结果区分所述角色，从而实现了所述待分离的角色的分离。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，在获得所述待分离的角色的最终身份识别结果之后，所述方法还包括：获取图像采集装置采集的所述待分离的角色的人脸图像数据；对所述人脸图像数据进行人脸识别，以获得所述待分离的角色的第三身份识别结果；若所述第三身份识别结果与所述第二身份识别结果不相同，则使用所述第三身份识别结果更正所述第二身份识别结果，以获得所述待分离的角色的最终身份识别结果。籍此，在对待分离的角色在预设时间段内的语音数据帧进行声纹识别的结果与对人脸图像数据进行人脸识别的结果不相同的情况下，使用对人脸图像数据进行人脸识别的结果更正对待分离的角色在预设时间段内的语音数据帧进行声纹识别的结果，能够准确地获得角色的身份识别结果，进而能够根据角色的身份识别结果准确地分离角色。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，所述图像采集装置可为摄像头。语音采集设备可从摄像头中获取摄像头采集的待分离的角色的人脸图像数据。在对所述人脸图像数据进行人脸识别

时，可通过人脸识别模型，对所述人脸图像数据进行人脸识别，以获得所述待分离的角色的第三身份识别结果。其中，所述人脸识别模型可为用于人脸识别的神经网络模型。若所述第三身份识别结果与所述第二身份识别结果相同，则无需使用所述第三身份识别结果更正所述第二身份识别结果，并将所述第二身份识别结果确定为所述待分离的说角色的最终身份识别结果。在获得所述待分离的角色的最终身份识别结果之后，便可使用所述待分离的角色的最终身份识别结果区分所述角色，从而实现了所述待分离的角色的分离。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，如图 2B 所示，待分离的角色的语音数据帧由语音采集设备采集。当语音采集设备采集到待分离的角色的语音数据帧之后，对待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧，并基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；再对所述过滤平滑后的语音数据帧进行声源定位，以确定声源角度数据；再对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；再确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的第一身份识别结果，再对所述待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得所述待分离的角色的第二身份识别结果，若所述第一身份识别结果与所述第二身份识别结果不相同，则使用所述第二身份识别结果更正所述第一身份识别结果，以获得所述待分离的角色的最终身份识别结果，最后基于所述待分离的角色的最终身份识别结果分离所述角色。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在上述实施例一的基础上，对待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得待分离的角色的第二身份识别结果，若第一身份识别结果与第二身份识别结果不相同，则使用第二身份识别结果更正第一身份识别结果，以获得待分离的角色的最终身份识别结果，并基于待分离的角色的最终身份识别结果分离待分离的角色，与现有的其它方式相比，在基于声源角度数据进行身份识别的结果与对角色在预设时间段内的语音数据帧进行声纹识别的结果不相同的情况下，使用对角色在预设时间段内的语音数据帧进行声纹识别的结果更正基于声源角度数据进行身份识别的结果，能够准确地获得角色的身份识别结果，进而能够根据角色的身份识别结果准确地分离角色。

本实施例提供的角色分离方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：摄像头、终端、移动终端、PC 机、服务器、车载设备、娱乐设备、广告设备、个人数码助理（PDA）、平板电脑、笔记本电脑、掌上游戏机、智能眼镜、智能手表、可穿戴设备、虚拟显示设备或显示增强设备等。

参照图 3A，示出了本申请实施例三的角色分离方法的步骤流程图。

具体地，本实施例提供的角色分离方法包括以下步骤：

在步骤 S301 中，向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色。

在本申请实施例中，语音采集设备向云端发送携带有待分离的角色的语音数据帧的角色分离请求，所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色。其中，所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据的具体实施方式与上述实施例一中获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据的具体实施方式类似，在此不再赘述。所述云端基于所述声源角度数据，对所述待分离的角色进行身份识别的具体实施方式与上述实施例一中基于所述声源角度数据，对所述待分离的角色进行身份识别的具体实施方式类似，在此不再赘述。所述云端基于所述待分离的角色的身份识别结果分离所述角色的具体实施方式与上述实施例一中基于所述待分离的角色的身份识别结果分离所述角色的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S302 中，接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

在本申请实施例中，语音采集设备接收所述云端基于所述角色分离请求发送的所述角色的分离结果。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，如图 3B 所示，待分离的角色的语音数据帧由语音采集设备采集。当语音采集设备采集到待分离的角色的语音数据帧之后，将待分离的角色的语音数据帧发送至云端进行角色分离。具体地，云端对待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧，并基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；再对所述过滤平滑后的语音数据帧进行声源定位，以确定声源角度数据；再对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；再确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的身份识别结果，最后基于所述待分离的角色的身份识别结果分离所述角色，并将所述角色的分离结果发送至语音采集设备。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

通过本申请实施例提供的角色分离方法，语音采集设备向云端发送携带有待分离的角色的语音数据帧的角色分离请求，云端基于角色分离请求，获取语音数据帧所对应的声源角度数据，并基于声源角度数据，对待分离的角色进行身份识别，再基于待分离的

角色的身份识别结果分离角色，语音采集设备接收云端基于角色分离请求发送的角色的分离结果，与现有的其它方式相比，基于角色分离请求携带的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

本实施例提供的角色分离方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：摄像头、终端、移动终端、PC 机、服务器、车载设备、娱乐设备、广告设备、个人数码助理（PDA）、平板电脑、笔记本电脑、掌上游戏机、智能眼镜、智能手表、可穿戴设备、虚拟显示设备或显示增强设备等。

参照图 4A，示出了本申请实施例四的角色分离方法的步骤流程图。

具体地，本实施例提供的角色分离方法包括以下步骤：

在步骤 S401 中，接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求。

在本申请实施例中，云端接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S402 中，基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据。

在本申请实施例中，所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据的具体实施方式与上述实施例一中获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S403 中，基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果。

在本申请实施例中，所述云端基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果的具体实施方式与上述实施例一中基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S404 中，基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

在本申请实施例中，所述云端基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，如图 4B 所示，待分离的角色的语音数据帧由语音采集设备

采集。当语音采集设备采集到待分离的角色的语音数据帧之后，将待分离的角色的语音数据帧发送至云端进行角色分离。具体地，云端对待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧，并基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；再对所述过滤平滑后的语音数据帧进行声源定位，以确定声源角度数据；再对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；再确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的第一身份识别结果，再对所述待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得所述待分离的角色的第二身份识别结果，若所述第一身份识别结果与所述第二身份识别结果不相同，则使用所述第二身份识别结果更正所述第一身份识别结果，以获得所述待分离的角色的最终身份识别结果，最后基于所述待分离的角色的最终身份识别结果分离所述角色，并将所述待分离的角色的分离结果发送至语音采集设备。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

通过本申请实施例提供的角色分离方法，云端接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求，并基于角色分离请求，获取语音数据帧所对应的声源角度数据，再基于声源角度数据，对待分离的角色进行身份识别，以获得待分离的角色的身份识别结果，再基于待分离的角色的身份识别结果分离角色，并向语音采集设备发送针对角色分离请求的角色分离结果，与现有的其它方式相比，基于角色分离请求携带的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

本实施例提供的角色分离方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：摄像头、终端、移动终端、PC 机、服务器、车载设备、娱乐设备、广告设备、个人数码助理（PDA）、平板电脑、笔记本电脑、掌上游戏机、智能眼镜、智能手表、可穿戴设备、虚拟显示设备或显示增强设备等。

参照图 5，示出了本申请实施例五的会议纪要的记录方法的步骤流程图。

具体地，本实施例提供的会议纪要的记录方法包括以下步骤：

在步骤 S501 中，获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据。

在本申请实施例中，所述位于会议室的语音采集设备可为位于会议室的拾音器。所述会议角色可理解为参加会议的人员。其中，所述获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据的具体实施方式与上述实施例一中获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做

任何限定。

在步骤 S502 中，基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果。

在本申请实施例中，所述基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果的具体实施方式与上述实施例一中基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S503 中，基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

在本申请实施例中，在获得所述会议角色的身份识别结果之后，便可使用所述会议角色的身份识别结果区分所述会议角色，进而可以实时记录会议角色的会议纪要。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，在基于所述会议角色的身份识别结果记录所述会议角色的会议纪要时，基于所述会议角色的身份识别结果，对所述会议角色的会议纪要语音数据进行语音识别，以获得所述会议角色的会议纪要文本数据，并记录所述会议角色的会议纪要文本数据。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

通过本申请实施例提供的会议纪要的记录方法，获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对会议角色进行身份识别，以获得会议角色的身份识别结果，再基于会议角色的身份识别结果记录会议角色的会议纪要，与现有的其它方式相比，基于位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据，对会议角色进行身份识别，再基于会议角色的身份识别结果记录会议角色的会议纪要，能够实时地记录会议角色的会议纪要，从而有效提高会议角色的会议纪要的记录效率。

本实施例提供的会议纪要的记录方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：摄像头、终端、移动终端、PC 机、服务器、车载设备、娱乐设备、广告设备、个人数码助理（PDA）、平板电脑、笔记本电脑、掌上游戏机、智能眼镜、智能手表、可穿戴设备、虚拟显示设备或显示增强设备等。

参照图 6，示出了本申请实施例六的角色展示方法的步骤流程图。

具体地，本实施例提供的角色展示方法包括以下步骤：

在步骤 S601 中，获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据。

在本申请实施例中，所述获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据的具体实施方式与上述实施例一中获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据的具体实施方式类似，在此不再赘述。可以理解的是，

以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S602 中，基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果。

在本申请实施例中，所述基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果的具体实施方式与上述实施例一中基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果的具体实施方式类似，在此不再赘述。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，在获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据之后，所述方法还包括：在所述声源角度数据指示的声源方向上开启所述语音采集设备的灯具。籍此，通过在声源角度数据指示的声源方向上开启语音采集设备的灯具，能够有效指示声源方向。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，所述语音采集设备的灯具以阵列的方式排布在所述语音采集设备的各个方位，从而能够有效指示声源方向。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在步骤 S603 中，基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

在本申请实施例中，所述语音采集设备的交互界面可为语音采集设备的触控屏。所述角色的身份数据可为所述角色的人脸图像数据、身份标识数据等。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一些可选实施例中，所述方法还包括：在所述语音采集设备的交互界面上展示所述角色的说话动作图像或者语音波形图像。籍此，能够更加生动地展示角色说话时的形象。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

在一个具体的例子中，可以在所述语音采集设备的交互界面上动态展示所述角色的说话动作图像序列或者语音波形图像序列。可以理解的是，以上描述仅为示例性的，本申请实施例对此不做任何限定。

通过本申请实施例提供的角色展示方法，获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对角色进行身份识别，以获得角色的身份识别结果，再基于角色的身份识别结果，在语音采集设备的交互界面上展示角色的身份数据，与现有的其它方式相比，基于语音采集设备采集的角色的语音数据帧所对应的声源角度数据，对角色进行身份识别，再基于角色的身份识别结果，在语音采集设备的交互界面上展示角色的身份数据，能够实时地展示角色的身份数据，从而使得用户体验更加流畅。

本实施例提供的角色展示方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：摄像头、终端、移动终端、PC 机、服务器、车载设备、娱乐设备、广告设备、个人数码助理（PDA）、平板电脑、笔记本电脑、掌上游戏机、智能眼镜、智能手表、可穿戴设备、虚拟显示设备或显示增强设备等。

参照图 7，示出了本申请实施例七中角色分离装置的结构示意图。

本实施例提供的角色分离装置包括：第一获取模块 701，用于获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；第一身份识别模块 702，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；分离模块 703，用于基于所述待分离的角色的第一身份识别结果分离所述角色。

本实施例提供的角色分离装置用于实现前述多个方法实施例中相应的角色分离方法，并具有相应的方法实施例的有益效果，在此不再赘述。

参照图 8，示出了本申请实施例八中角色分离装置的结构示意图。

本实施例提供的角色分离装置包括：第一获取模块 801，用于获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；第一身份识别模块 805，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；分离模块 808，用于基于所述待分离的角色的第一身份识别结果分离所述角色。

可选地，所述第一获取模块 801 之后，所述装置还包括：检测模块 802，用于对所述待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧；过滤平滑模块 803，用于基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；更新模块 804，用于基于所述过滤平滑后的语音数据帧，对所述声源角度数据进行更新，以获得更新后的声源角度数据。

可选地，所述过滤平滑模块 803，具体用于：通过中值滤波器，基于所述待分离的角色的语音数据帧的能量频谱的谱平度，对所述具有语音端点的语音数据帧进行过滤平滑，以获得所述过滤平滑后的语音数据帧。

可选地，所述第一身份识别模块 805，包括：聚类子模块 8051，用于对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；确定子模块 8052，用于确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的第一身份识别结果。

可选地，所述聚类子模块 8051，具体用于：确定所述声源角度数据与声源角度顺序聚类中心的距离；基于所述声源角度数据与所述声源角度顺序聚类中心的距离，确定所述声源角度数据的顺序聚类结果。

可选地，所述第一身份识别模块 805 之后，所述装置还包括：声纹识别模块 806，用于对所述待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得所述待分离的角色的第二身份识别结果；第一更正模块 807，用于若所述第一身份识别结果与所述第二身份识别结果不相同，则使用所述第二身份识别结果更正所述第一身份识别结果，以获得所述待分离的角色的最终身份识别结果。

可选地，所述语音采集设备包括麦克风阵列，所述第一获取模块 801，具体用于：获取所述麦克风阵列中至少部分麦克风接收到的所述语音数据帧的协方差矩阵；对所述协方差矩阵进行特征值分解，以得到多个特征值；从所述多个特征值中选取第一数量个最大的特征值，并基于选取的特征值对应的特征向量构成语音信号子空间，其中，所述第一数量与声源估计数量相当；基于所述语音信号子空间，确定所述声源角度数据。

可选地，所述第一更正模块 807 之后，所述装置还包括：第二获取模块 809，用于获取图像采集装置采集的所述待分离的角色的人脸图像数据；人脸识别模块 810，用于对所述人脸图像数据进行人脸识别，以获得所述待分离的角色的第三身份识别结果；第二更正模块 811，用于若所述第三身份识别结果与所述第二身份识别结果不相同，则使用所述第三身份识别结果更正所述第二身份识别结果，以获得所述待分离的角色的最终身份识别结果。

本实施例提供的角色分离装置用于实现前述多个方法实施例中相应的角色分离方法，并具有相应的方法实施例的有益效果，在此不再赘述。

参照图 9，示出了本申请实施例九中角色分离装置的结构示意图。

本实施例提供的角色分离装置包括：第一发送模块 901，用于向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色；第一接收模块 902，用于接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

本实施例提供的角色分离装置用于实现前述多个方法实施例中相应的角色分离方法，并具有相应的方法实施例的有益效果，在此不再赘述。

参照图 10，示出了本申请实施例十中角色分离装置的结构示意图。

本实施例提供的角色分离装置包括：第二接收模块 1001，用于接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求；第三获取模块 1002，用于基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据；第二身份识别模块 1003，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果；第二发送模块 1004，用于基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

本实施例提供的角色分离装置用于实现前述多个方法实施例中相应的角色分离方法，并具有相应的方法实施例的有益效果，在此不再赘述。

参照图 11，示出了本申请实施例十一中会议纪要的记录装置的结构示意图。

本实施例提供的会议纪要的记录装置包括：第四获取模块 1101，用于获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据；第三身份识别模块 1102，用于基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果；记录模块 1103，用于基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

本实施例提供的会议纪要的记录装置用于实现前述多个方法实施例中相应的会议纪要的记录方法，并具有相应的方法实施例的有益效果，在此不再赘述。

参照图 12，示出了本申请实施例十二中角色展示装置的结构示意图。

本实施例提供的角色展示装置包括：第五获取模块 1201，用于获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据；第四身份识别模块 1203，用于基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果；第一展示模块 1204，用于基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

可选地，所述第五获取模块 1201 之后，所述装置还包括：开启模块 1202，用于在所述声源角度数据指示的声源方向上开启所述语音采集设备的灯具。

可选地，所述装置还包括：第二展示模块 1205，用于在所述语音采集设备的交互界面上展示所述角色的说话动作图像或者语音波形图像。

本实施例提供的角色展示装置用于实现前述多个方法实施例中相应的角色展示方法，并具有相应的方法实施例的有益效果，在此不再赘述。

参照图 13，示出了根据本发明实施例十三的一种电子设备的结构示意图，本发明具体实施例并不对电子设备的具体实现做限定。

如图 13 所示，该电子设备可以包括：处理器(processor)1302、通信接口(Communications Interface)1304、存储器(memory)1306、以及通信总线 1308。

其中：

处理器 1302、通信接口 1304、以及存储器 1306 通过通信总线 1308 完成相互间的通信。

通信接口 1304，用于与其它电子设备或服务器进行通信。

处理器 1302，用于执行程序 1310，具体可以执行上述角色分离方法实施例中的相关步骤。

具体地，程序 1310 可以包括程序代码，该程序代码包括计算机操作指令。

处理器 1302 可能是中央处理器 CPU，或者是特定集成电路 ASIC（Application Specific

Integrated Circuit），或者是被配置成实施本发明实施例的一个或多个集成电路。智能设备包括的一个或多个处理器，可以是同一类型的处理器，如一个或多个 CPU；也可以是不同类型的处理器，如一个或多个 CPU 以及一个或多个 ASIC。

存储器 1306，用于存放程序 1310。存储器 1306 可能包含高速 RAM 存储器，也可能还包括非易失性存储器（non-volatile memory），例如至少一个磁盘存储器。

程序 1310 具体可以用于使得处理器 1302 执行以下操作：获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；基于所述待分离的角色的第一身份识别结果分离所述角色。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据之后，对所述待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧；基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；基于所述过滤平滑后的语音数据帧，对所述声源角度数据进行更新，以获得更新后的声源角度数据。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧时，通过中值滤波器，基于所述待分离的角色的语音数据帧的能量频谱的谱平度，对所述具有语音端点的语音数据帧进行过滤平滑，以获得所述过滤平滑后的语音数据帧。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果时，对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的第一身份识别结果。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果时，确定所述声源角度数据与声源角度顺序聚类中心的距离；基于所述声源角度数据与所述声源角度顺序聚类中心的距离，确定所述声源角度数据的顺序聚类结果。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在获得所述待分离的角色的第一身份识别结果之后，对所述待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得所述待分离的角色的第二身份识别结果；若所述第一身份识别结果与所述第二身份识别结果不相同，则使用所述第二身份识别结果更正所述第一身份识别结果，以获得所述待分离的角色的最终身份识别结果。

在一种可选的实施方式中，所述语音采集设备包括麦克风阵列，程序 1310 还用于使得处理器 1302 在获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据时，获取所述麦克风阵列中至少部分麦克风接收到的所述语音数据帧的协方差矩阵；对所述协方差矩阵进行特征值分解，以得到多个特征值；从所述多个特征值中选取第一数量个最大的特征值，并基于选取的特征值对应的特征向量构成语音信号子空间，其中，所述第一数量与声源估计数量相当；基于所述语音信号子空间，确定所述声源角度数据。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在获得所述待分离的角色的最终身份识别结果之后，获取图像采集装置采集的所述待分离的角色的人脸图像数据；对所述人脸图像数据进行人脸图像识别，以获得所述待分离的角色的第三身份识别结果；若所述第三身份识别结果与所述第二身份识别结果不相同，则使用所述第三身份识别结果更正所述第二身份识别结果，以获得所述待分离的角色的最终身份识别结果。

程序 1310 中各步骤的具体实现可以参见上述角色分离方法实施例中的相应步骤和单元中对应的描述，在此不赘述。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，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

通过本实施例的电子设备，获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对待分离的角色进行身份识别，以获得待分离的角色的第一身份识别结果；再基于待分离的角色的第一身份识别结果分离待分离的角色，与现有的其它方式相比，基于语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

程序 1310 具体可以用于使得处理器 1302 执行以下操作：向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色；接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

程序 1310 中各步骤的具体实现可以参见上述角色分离方法实施例中的相应步骤和单元中对应的描述，在此不赘述。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，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

通过本实施例的电子设备，语音采集设备向云端发送携带有待分离的角色的语音数据帧的角色分离请求，云端基于角色分离请求，获取语音数据帧所对应的声源角度数据，并基于声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别

结果分离角色，语音采集设备接收云端基于角色分离请求发送的角色的分离结果，与现有的其它方式相比，基于角色分离请求携带的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

程序 1310 具体可以用于使得处理器 1302 执行以下操作：接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求；基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果；基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

程序 1310 中各步骤的具体实现可以参见上述角色分离方法实施例中的相应步骤和单元中对应的描述，在此不赘述。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，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

通过本实施例的电子设备，云端接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求，并基于角色分离请求，获取语音数据帧所对应的声源角度数据，再基于声源角度数据，对待分离的角色进行身份识别，以获得待分离的角色的身份识别结果，再基于待分离的角色的身份识别结果分离角色，并向语音采集设备发送针对角色分离请求的角色分离结果，与现有的其它方式相比，基于角色分离请求携带的待分离的角色的语音数据帧所对应的声源角度数据，对待分离的角色进行身份识别，再基于待分离的角色的身份识别结果分离角色，能够实时地分离角色，进而使得用户体验更流畅。

程序 1310 具体可以用于使得处理器 1302 执行以下操作：获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果；基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

程序 1310 中各步骤的具体实现可以参见上述会议纪要的记录方法实施例中的相应步骤和单元中对应的描述，在此不赘述。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，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

通过本实施例的电子设备，获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对会议角色进行身份识别，以获得会议角色的身份识别结果，再基于会议角色的身份识别结果记录会议角色的会议纪要，与现有的其它方式相比，基于位于会议室的语音采集设备采集的会议角色的语音数据帧

所对应的声源角度数据，对会议角色进行身份识别，再基于会议角色的身份识别结果记录会议角色的会议纪要，能够实时地记录会议角色的会议纪要，从而有效提高会议角色的会议纪要的记录效率。

程序 1310 具体可以用于使得处理器 1302 执行以下操作：获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据；基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果；基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据之后，在所述声源角度数据指示的声源方向上开启所述语音采集设备的灯具。

在一种可选的实施方式中，程序 1310 还用于使得处理器 1302 在所述语音采集设备的交互界面上展示所述角色的说话动作图像或者语音波形图像。

程序 1310 中各步骤的具体实现可以参见上述角色展示方法实施例中的相应步骤和单元中对应的描述，在此不赘述。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，上述描述的设备和模块的具体工作过程，可以参考前述方法实施例中的对应过程描述，在此不再赘述。

通过本实施例的电子设备，获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据，并基于声源角度数据，对角色进行身份识别，以获得角色的身份识别结果，再基于角色的身份识别结果，在语音采集设备的交互界面上展示角色的身份数据，与现有的其它方式相比，基于语音采集设备采集的角色的语音数据帧所对应的声源角度数据，对角色进行身份识别，再基于角色的身份识别结果，在语音采集设备的交互界面上展示角色的身份数据，能够实时地展示角色的身份数据，从而使得用户体验更加流畅。

需要指出，根据实施的需要，可将本发明实施例中描述的各个部件/步骤拆分为更多部件/步骤，也可将两个或多个部件/步骤或者部件/步骤的部分操作组合成新的部件/步骤，以实现本发明实施例的目的。

上述根据本发明实施例的方法可在硬件、固件中实现，或者被实现为可存储在记录介质（诸如 CD ROM、RAM、软盘、硬盘或磁光盘）中的软件或计算机代码，或者被实现通过网络下载的原始存储在远程记录介质或非暂时机器可读介质中并将被存储在本地记录介质中的计算机代码，从而在此描述的方法可被存储在使用通用计算机、专用处理器或者可编程或专用硬件（诸如 ASIC 或 FPGA）的记录介质上的这样的软件处理。可以理解，计算机、处理器、微处理器控制器或可编程硬件包括可存储或接收软件或计算机代码的存储组件（例如，RAM、ROM、闪存等），当所述软件或计算机代码被计算机、处理器或硬件访问且执行时，实现在此描述的角色分离方法、会议纪要的记录方法，或者角色展示方法。此外，当通用计算机访问用于实现在此示出的角色分离方法、会议纪

要的记录方法，或者角色展示方法的代码时，代码的执行将通用计算机转换为用于执行在此示出的角色分离方法、会议纪要的记录方法，或者角色展示方法的专用计算机。

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方法步骤，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发明实施例的范围。

以上实施方式仅用于说明本发明实施例，而并非对本发明实施例的限制，有关技术领域的普通技术人员，在不脱离本发明实施例的精神和范围的情况下，还可以做出各种变化和变型，因此所有等同的技术方案也属于本发明实施例的范畴，本发明实施例的专利保护范围应由权利要求限定。

权 利 要 求 书

1. 一种角色分离方法，所述方法包括：

获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；

基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；

基于所述待分离的角色的第一身份识别结果分离所述角色。

2. 根据权利要求 1 所述的方法，其中，所述获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据之后，所述方法还包括：

对所述待分离的角色的语音数据帧进行语音端点检测，以获得具有语音端点的语音数据帧；

基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧；

基于所述过滤平滑后的语音数据帧，对所述声源角度数据进行更新，以获得更新后的声源角度数据。

3. 根据权利要求 2 所述的方法，其中，所述基于所述待分离的角色的语音数据帧的能量频谱，对所述具有语音端点的语音数据帧进行过滤平滑，以获得过滤平滑后的语音数据帧，包括：

通过中值滤波器，基于所述待分离的角色的语音数据帧的能量频谱的谱平度，对所述具有语音端点的语音数据帧进行过滤平滑，以获得所述过滤平滑后的语音数据帧。

4. 根据权利要求 1 所述的方法，其中，所述基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果，包括：

对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果；

确定所述声源角度数据的顺序聚类结果对应的角色身份标识为所述待分离的角色的第一身份识别结果。

5. 根据权利要求 4 所述的方法，其中，所述对所述声源角度数据进行顺序聚类，以获得所述声源角度数据的顺序聚类结果，包括：

确定所述声源角度数据与声源角度顺序聚类中心的距离；

基于所述声源角度数据与所述声源角度顺序聚类中心的距离，确定所述声源角度数据的顺序聚类结果。

6. 根据权利要求 1 所述的方法，其中，所述获得所述待分离的角色的第一身份识别结果之后，所述方法还包括：

对所述待分离的角色在预设时间段内的语音数据帧进行声纹识别，以获得所述待分离的角色的第二身份识别结果；

若所述第一身份识别结果与所述第二身份识别结果不相同，则使用所述第二身份识别结果更正所述第一身份识别结果，以获得所述待分离的角色的最终身份识别结果。

7. 根据权利要求 1 所述的方法，其中，所述语音采集设备包括麦克风阵列，所述获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据，包括：

获取所述麦克风阵列中至少部分麦克风接收到的所述语音数据帧的协方差矩阵；

对所述协方差矩阵进行特征值分解，以得到多个特征值；

从所述多个特征值中选取第一数量个最大的特征值，并基于选取的特征值对应的特征向量构成语音信号子空间，其中，所述第一数量与声源估计数量相当；

基于所述语音信号子空间，确定所述声源角度数据。

8. 根据权利要求 6 所述的方法，其中，所述获得所述待分离的角色的最终身份识别结果之后，所述方法还包括：

获取图像采集装置采集的所述待分离的角色的人脸图像数据；

对所述人脸图像数据进行人脸识别，以获得所述待分离的角色的第三身份识别结果；

若所述第三身份识别结果与所述第二身份识别结果不相同，则使用所述第三身份识别结果更正所述第二身份识别结果，以获得所述待分离的角色的最终身份识别结果。

9. 一种角色分离方法，所述方法包括：

向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色；

接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

10. 一种角色分离方法，所述方法包括：

接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求；

基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据；

基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果；

基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

11. 一种会议纪要的记录方法，所述方法包括：

获取位于会议室的语音采集设备采集的会议角色的语音数据帧所对应的声源角度数据；

基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果；

基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

12. 一种角色展示方法，所述方法包括：

获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据；

基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果；

基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

13. 根据权利要求 12 所述的方法，其中，所述获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据之后，所述方法还包括：

在所述声源角度数据指示的声源方向上开启所述语音采集设备的灯具。

14. 根据权利要求 12 所述的方法，其中，所述方法还包括：

在所述语音采集设备的交互界面上展示所述角色的说话动作图像或者语音波形图像。

15. 一种角色分离装置，所述装置包括：

第一获取模块，用于获取语音采集设备采集的待分离的角色的语音数据帧所对应的声源角度数据；

第一身份识别模块，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的第一身份识别结果；

分离模块，用于基于所述待分离的角色的第一身份识别结果分离所述待分离的角色。

16. 一种角色分离装置，所述装置包括：

第一发送模块，用于向云端发送携带有待分离的角色的语音数据帧的角色分离请求，使得所述云端基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据，并基于所述声源角度数据，对所述待分离的角色进行身份识别，再基于所述待分离的角色的身份识别结果分离所述角色；

第一接收模块，用于接收所述云端基于所述角色分离请求发送的所述角色的分离结果。

17. 一种角色分离装置，所述装置包括：

第二接收模块，用于接收语音采集设备发送的携带有待分离的角色的语音数据帧的角色分离请求；

第三获取模块，用于基于所述角色分离请求，获取所述语音数据帧所对应的声源角度数据；

第二身份识别模块，用于基于所述声源角度数据，对所述待分离的角色进行身份识别，以获得所述待分离的角色的身份识别结果；

第二发送模块，用于基于所述待分离的角色的身份识别结果分离所述角色，并向所述语音采集设备发送针对所述角色分离请求的角色分离结果。

18. 一种会议纪要的记录装置，所述装置包括：

第四获取模块，用于获取位于会议室的语音采集设备采集的会议角色的语音数据帧

所对应的声源角度数据；

第三身份识别模块，用于基于所述声源角度数据，对所述会议角色进行身份识别，以获得所述会议角色的身份识别结果；

记录模块，用于基于所述会议角色的身份识别结果记录所述会议角色的会议纪要。

19. 一种角色展示装置，所述装置包括：

第五获取模块，用于获取语音采集设备采集的角色的语音数据帧所对应的声源角度数据；

第四身份识别模块，用于基于所述声源角度数据，对所述角色进行身份识别，以获得所述角色的身份识别结果；

第一展示模块，用于基于所述角色的身份识别结果，在所述语音采集设备的交互界面上展示所述角色的身份数据。

20. 一种电子设备，包括：处理器、存储器、通信接口和通信总线，所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信；

所述存储器用于存放至少一可执行指令，所述可执行指令使所述处理器执行如权利要求 1-8 中任意一项权利要求所述的角色分离方法对应的操作，或者执行如权利要求 9 所述的角色分离方法对应的操作，或者执行如权利要求 10 所述的角色分离方法对应的操作，或者执行如权利要求 11 所述的会议纪要的记录方法对应的操作，或者执行如权利要求 12-14 中任意一项权利要求所述的角色展示方法对应的操作。

21. 一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时实现如权利要求 1-8 中任意一项权利要求所述的角色分离方法，或者实现如权利要求 9 所述的角色分离方法，或者实现如权利要求 10 所述的角色分离方法，或者实现如权利要求 11 所述的会议纪要的记录方法，或者实现如权利要求 12-14 中任意一项权利要求所述的角色展示方法。