(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107211027 B
(45)授权公告日 2020.09.15

(21)申请号 201680008715.9

(22)申请日 2016.02.03

(65)同一申请的已公布的文献号
申请公布号 CN 107211027 A

(43)申请公布日 2017.09.26

(30)优先权数据
62/128,631 2015.03.05 US

(66)本国优先权数据
PCT/CN2015/072161 2015.02.03 CN

(85)PCT国际申请进入国家阶段日
2017.08.03

(86)PCT国际申请的申请数据
PCT/US2016/016359 2016.02.03

(87)PCT国际申请的公布数据
WO2016/126816 EN 2016.08.11

(73)专利权人 杜比实验室特许公司
地址 美国加利福尼亚

(72)发明人 R·J·卡特莱特 G·N·迪金斯

(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
代理人 宿小猛

(51)Int.Cl.
H04L 29/06(2006.01)
H04M 3/56(2006.01)
H04M 15/00(2006.01)

(56)对比文件
CN 104010265 A,2014.08.27
US 2005240656 A1,2005.10.27

审查员 齐丽静

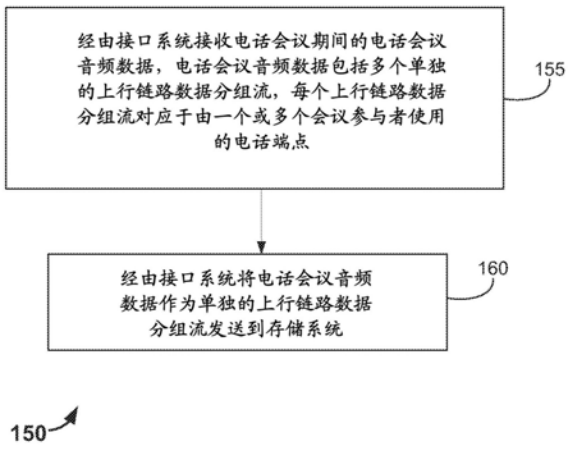
权利要求书3页 说明书102页 附图60页

(54)发明名称

感知质量比会议中原始听到的更高的后会议回放系统

(57)摘要

本公开的一些方面涉及与诸如电话会议的会议对应的音频数据的记录、处理和回放。在一些电话会议实现中,在会议记录被回放时所听到的音频体验可显著不同于在原始电话会议期间的单个会议参与者的音频体验。在一些实现中,记录的音频数据可包括在电话会议期间不可用的至少一些音频数据。在一些示例中,被回放的音频数据的空间特性可不同于电话会议的参与者收听到的音频的空间特性。



CN 107211027 B

1.一种用于处理音频数据的装置,所述装置包括:

接口系统;以及

控制系统,所述控制系统能够:

经由接口系统在电话会议期间接收电话会议音频数据,电话会议音频数据包括多个单独的上行链路数据分组流,每个上行链路数据分组流对应于一个或多个电话会议参与者使用的电话端点;以及

经由接口系统将电话会议音频数据作为单独的用于回放的上行链路数据分组流发送到存储系统,

其中,上行链路数据分组流中的至少一个对应于多个电话会议参与者,并且包括关于多个参与者中的每一个的空间信息,所述空间信息被用于在虚拟声学空间中渲染所述电话会议音频数据,使得每一个会议参与者在虚拟声学空间中具有各自的虚拟会议参与位置,

其中,上行链路数据分组流中的至少一个包括在电话会议的口到耳延迟时间阈值之后接收到并且在电话会议期间不用于再现音频数据的至少一个数据分组。

2.根据权利要求1所述的装置,其中所述控制系统能够:

确定不完整的上行链路数据分组流的迟到数据分组已经在迟到分组时间阈值之后被从电话端点接收到,所述迟到分组时间阈值大于或等于电话会议的口到耳延迟时间阈值;以及

将迟到数据分组添加到不完整的上行链路数据分组流。

3.根据权利要求2所述的装置,其中所述控制系统能够:

确定在大于迟到分组时间阈值的丢失分组时间阈值内没有从电话端点接收到不完整的上行链路数据分组流的丢失的数据分组;

经由接口系统向电话端点发送关于重新发送丢失的数据分组的请求;

接收所述丢失的数据分组;以及

将丢失的数据分组添加到不完整的上行链路数据分组流中。

4.根据权利要求2所述的装置,其中迟到分组时间阈值大于或等于1秒。

5.根据权利要求1所述的装置,其中单独的上行链路数据分组流是单独的编码的上行链路数据分组流。

6.根据权利要求5所述的装置,其中所述发送包括将电话会议音频数据作为单独的编码的上行链路数据分组流发送到存储系统。

7.根据权利要求1-6中任一项所述的装置,其中口到耳延迟时间阈值大于或等于100毫秒。

8.根据权利要求1-6中任一项所述的装置,其中控制系统能够提供电话会议服务器功能。

9.根据权利要求1-6中任一项所述的装置,其中接口系统包括网络接口,并且控制系统能够经由网络接口将电话会议音频数据发送到存储系统。

10.根据权利要求1-6中任一项所述的装置,其中,所述装置包括存储系统的至少一部分,并且其中所述接口系统包括所述存储系统的至少一部分和所述控制系统之间的接口。

11.一种用于处理音频数据的装置,所述装置包括:

接口系统;以及

控制部件,用于:

经由接口系统在电话会议期间接收电话会议音频数据,电话会议音频数据包括多个单独的上行链路数据分组流,每个上行链路数据分组流对应于一个或多个电话会议参与者使用的电话端点;以及

经由接口系统将电话会议音频数据作为单独的用于回放的上行链路数据分组流发送到存储系统,

其中,上行链路数据分组流中的至少一个对应于多个电话会议参与者,并且包括关于多个参与者中的每一个的空间信息,所述空间信息被用于在虚拟声学空间中渲染所述电话会议音频数据,使得每一个会议参与者在虚拟声学空间中具有各自的虚拟会议参与位置,

其中,上行链路数据分组流中的至少一个包括在电话会议的口到耳延迟时间阈值之后接收到并且在电话会议期间不用于再现音频数据的至少一个数据分组。

12.根据权利要求11所述的装置,其中单独的上行链路数据分组流是单独的编码的上行链路数据分组流。

13.一种用于处理音频数据的方法,所述方法包括:

经由接口系统在电话会议期间接收电话会议音频数据,电话会议音频数据包括多个单独的上行链路数据分组流,每个上行链路数据分组流对应于一个或多个电话会议参与者使用的电话端点;以及

经由接口系统将电话会议音频数据作为单独的用于回放的上行链路数据分组流发送到存储系统,

其中,上行链路数据分组流中的至少一个对应于多个电话会议参与者,并且包括关于多个参与者中的每一个的空间信息,所述空间信息被用于在虚拟声学空间中渲染所述电话会议音频数据,使得每一个会议参与者在虚拟声学空间中具有各自的虚拟会议参与位置,

其中,上行链路数据分组流中的至少一个包括在电话会议的口到耳延迟时间阈值之后接收到并且在电话会议期间不用于再现音频数据的至少一个数据分组。

14.根据权利要求13所述的方法,其中单独的上行链路数据分组流是单独的编码的上行链路数据分组流。

15.一种非暂态介质,在所述非暂态介质上存储有软件,所述软件包括用于通过控制至少一个设备进行以下操作以便处理音频数据的指令:

在电话会议期间接收电话会议音频数据,电话会议音频数据包括多个单独的上行链路数据分组流,每个上行链路数据分组流对应于一个或多个电话会议参与者使用的电话端点;以及

将电话会议音频数据作为单独的用于回放的上行链路数据分组流发送到存储系统,

其中,上行链路数据分组流中的至少一个对应于多个电话会议参与者,并且包括关于多个参与者中的每一个的空间信息,所述空间信息被用于在虚拟声学空间中渲染所述电话会议音频数据,使得每一个会议参与者在虚拟声学空间中具有各自的虚拟会议参与位置,

其中,上行链路数据分组流中的至少一个包括在电话会议的口到耳延迟时间阈值之后接收到并且在电话会议期间不用于再现音频数据的至少一个数据分组。

16.根据权利要求15所述的非暂态介质,其中单独的上行链路数据分组流是单独的编码的上行链路数据分组流。

17.一种用于处理音频数据的设备,包括:

一个或多个处理器,

一个或多个存储介质,存储有指令,所述指令在被所述一个或多个处理器执行时使得执行根据权利要求13-14中任一项所述的方法。

18.一种包括用于执行根据权利要求13-14中任一项所述的方法的部件的装置。

感知质量比会议中原始听到的更高的后会议回放系统

[0001] 相关申请的交叉引用

[0002] 本申请要求2015年2月3日提交的申请号为PCT/CN2015/072161的PCT专利申请;以及2015年3月5日提交的美国临时专利申请第62/128,631号的优先权,它们中的每一个的全文通过引用并入本文。

技术领域

[0003] 本公开涉及音频信号的处理。特别地,本公开涉及处理与会议相关的音频信号,包括但不限于处理用于电话会议或视频会议的音频信号。

背景技术

[0004] 在电话会议领域,通常提供设施以允许记录电话会议以供在电话会议结束之后进行回放。这可以使得那些无法参加会议的人听到会议中发生了什么。它还可以让那些在场的人刷新他们对电话会议期间所发生的事情的记忆。记录设施有时用于确保某些行业(如银行业)的法规遵从。

[0005] 典型的电话会议记录是包含所有各方到记录介质上的混合的单个单声道流。这通常通过将“虚拟”客户端或电话连接到电话会议桥或如下服务器来实现,该服务器对于桥看上去像普通客户端或电话,但实际上这可以是简单地记录其下行链路的机器。在这种系统中,聆听录音的回放的体验与原始电话会议期间在电话或客户端上被动地进行聆听的体验相同或基本相同。

发明内容

[0006] 根据本文公开的一些实现方式,一种方法可以涉及处理音频数据。一些这样的方法可以涉及接收对应于涉及多个会议参与者的会议的记录的音频数据。在一些例子中,会议可能是电话会议。然而,在一些例子中,会议可能是面对面会议(in-person conference)。

[0007] 根据一些示例,音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地,音频数据中的至少一些可以来自对应于多个会议参与者的单个端点。音频数据可以包括多个会议参与者中的每个会议参与者的空间信息。

[0008] 在一些实现中,该方法可以涉及分析音频数据以确定会话动态数据。在一些示例中,会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示至少两个会议参与者在其期间同时发言的会议参与者双讲话(doubletalk)的实例的数据、和/或指示会议参与者会话的实例的数据。

[0009] 一些公开的方法可以涉及将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量,该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。一些这样的方法可以涉及将优化技术应用于空间优化成本函数以确定局部最优解,

并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0010] 在一些实现中，虚拟声学空间可以相对于虚拟听众的头部在虚拟声学空间中的位置来确定。根据一些这样的实施方式，空间优化成本函数可以应用对于将参与会议参与者双讲话的会议参与者布置于如下虚拟会议参与者位置处的惩罚，该虚拟会议参与者位置位于相对于虚拟听众头部的位置被定义的“混淆锥(cone of confusion)”上或者与该“混淆锥”相距在预定的角距离内。通过混淆锥的圆锥切片可能具有相同的耳间时间差异。在一些示例中，空间优化成本函数可以应用对于将参与会议参与者相互会话的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚。

[0011] 根据一些示例，分析音频数据可以涉及确定哪些会议参与者(如果有的话)具有感知相似的语音。在一些这样的示例中，空间优化成本函数可以应用对于将具有感知相似的语音的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚。

[0012] 在一些示例中，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。在某些实例中，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于与较不频繁发言的会议参与者的虚拟会议参与者位置相比距虚拟听众头部的位置更远的虚拟会议参与者位置处的惩罚。在一些实现中，空间优化成本函数可以应用对于将很少发言的会议参与者布置于不在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。

[0013] 根据一些示例，优化技术可能涉及梯度下降技术、共轭梯度技术、牛顿法、Broyden-Fletcher-Goldfarb-Shanno算法、遗传算法、模拟退火算法、蚁群优化方法和/或蒙特卡罗方法。在一些示例中，分配虚拟会议参与者位置可以包括从一组预定的虚拟会议参与者位置中选择虚拟会议参与者位置。

[0014] 在一些实例中，音频数据可包括语音活动检测处理的输出。根据一些示例，分析音频数据可涉及识别对应于个体会议参与者的语音。

[0015] 在一些示例中，音频数据可对应于完整或基本上完整的会议的记录。一些示例可能涉及接收和处理来自多个会议的音频数据。

[0016] 一些公开的方法可涉及在电话会议期间接收(例如，经由接口系统)电话会议音频数据。在一些示例中，电话会议音频数据可以包括多个单独的上行链路数据分组流。每个上行链路数据分组流可以对应于一个或多个电话会议参与者使用的电话端点。该方法可以涉及(例如，经由接口系统)将电话会议音频数据作为单独的上行链路数据分组流发送到存储系统。

[0017] 一些方法可涉及确定不完整的上行链路数据分组流的迟到(late)数据分组已经在迟到分组时间阈值之后被从电话端点接收到。迟到分组时间阈值可以大于或等于电话会议的口到耳延迟时间阈值。在一些示例中，口到耳延迟时间阈值可以大于或等于100毫秒(ms)。在一些实例中，口到耳延迟时间阈值可以是150ms或更短。在一些示例中，迟到分组时间阈值可以是200ms，400ms，500ms或更大。在某些实现中，迟到分组时间阈值可以大于或等于1秒。一些这样的方法可涉及将迟到数据分组添加到不完整的上行链路数据分组流。

[0018] 一些方法可以涉及确定在大于迟到分组时间阈值的丢失分组时间阈值内没有从电话端点接收到不完整上行链路数据分组流的丢失数据分组。一些这样的方法可以涉及向

电话端点(例如,经由接口系统)发送关于重新发送丢失的数据分组的请求。如果电话端点重新发送丢失的数据分组,这样的方法可能涉及接收丢失的数据分组,并且将丢失的数据分组添加到不完整的上行数据分组流中。

[0019] 在一些示例中,单独的上行链路数据分组流可以是单独的编码的上行链路数据分组流。上行链路数据分组流中的至少一个可以包括在电话会议的口到耳延迟时间阈值之后接收到、因此在电话会议期间不用于再现音频数据的至少一个数据分组。在一些实例中,至少一个上行链路数据分组流可以对应于多个电话会议参与者,并且可以包括关于多个参与者中的每一个的空间信息。

[0020] 一些公开的方法可以涉及接收(例如,经由接口系统)所记录的电话会议的音频数据。记录的音频数据可以包括对应于由一个或多个电话会议参与者使用的电话端点的单独的上行链路数据分组流。一些这样的方法可以涉及分析单独的上行链路数据分组流中的数据分组的序列号数据。分析过程可以包括确定单独的上行链路数据分组流是否包括至少一个无序数据分组。如果上行链路数据分组流包括至少一个无序数据分组,则这样的方法可以包括根据序列号数据对单独的上行链路数据分组流重新排序。在一些实例中,单独的上行链路数据分组流的至少一个数据分组可能已经在电话会议的口到耳延迟时间阈值之后接收到。

[0021] 一些这样的方法可以包括接收(例如,经由接口系统)电话会议元数据并且至少部分地基于电话会议元数据来对单独的上行链路数据分组流进行索引。在一些实例中,记录的音频数据可以包括多个单独的编码的上行链路数据分组流。每个单独的编码的上行链路数据分组流可以对应于由一个或多个电话会议参与者使用的电话端点。这样的方法可以包括解码多个单独的编码的上行链路数据分组流,并分析该多个单独的上行链路数据分组流。

[0022] 一些方法可能涉及在一个或多个单独的解码的上行链路数据分组流中识别语音,并且产生语音识别结果数据。一些这样的方法可能涉及标识语音识别结果数据中的关键词并且对关键词位置进行索引。

[0023] 一些公开的方法可以涉及在单独的解码的上行链路数据分组流中识别多个电话会议参与者中的每一个的语音。一些这样的方法可能涉及生成发言者日志,该发言者日志指示多个电话会议参与者中的每一个发言时的时间。

[0024] 根据一些示例,分析多个单独的上行链路数据分组流可以涉及确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示至少两个会议参与者在其期间同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0025] 一些方法可能涉及接收对应于涉及多个会议参与者的会议的记录的音频数据。在一些例子中,会议可能是电话会议。然而,在一些例子中,会议可能是面对面会议。

[0026] 根据一些示例,音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地,音频数据中的至少一些可以来自对应于多个会议参与者的单个端点。音频数据可以包括多个会议参与者的每个会议参与者的空间信息。

[0027] 一些这样的方法可以涉及在虚拟声学空间中渲染会议参与者语音数据,使得各会

议参与者具有各自不同的虚拟会议参与者位置。这样的方法可以包括调度会议参与者语音回放，使得在会议参与者语音的至少两个输出讲话突发(talkspurt)之间的回放重叠量不同于(例如，大于)会议记录的两个对应的输入讲话突发(talkspurt)之间的原始重叠量。原始重叠量可以为零或非零。

[0028] 在一些示例中，调度可以至少部分地根据感知激发(motivated)规则的集合来执行。本文公开了各种感知激发规则。在一些实现中，感知激发规则集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则两个输出讲话突发在时间上不应该重叠。

[0029] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前开始。感知激发规则集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。

[0030] 根据一些实现，感知激发规则集合可以包括允许来自不同会议参与者的全部陈述(presentation)的并发回放的规则。在一些实现中，陈述可以对应于会议参与者语音的时间间隔，在该时间间隔期间，语音密度度量大于或等于静默阈值，双讲话比率小于或等于讨论阈值，并且主导度量大于陈述阈值。双讲话比率可以指示在该时间间隔中的在其期间至少两个会议参与者同时发言的语音时间的占比。语音密度度量可以指示在该时间间隔中的存在任何会议参与者语音的占比。主导度量可以指示在该时间间隔期间的由主导会议参与者发出的总语音的占比。主导会议参与者可以是在时间间隔内发言最多的会议参与者。

[0031] 在一些示例中，会议参与者语音中的至少一些可被调度为以比记录会议参与者语音的速率更快的速率来回放。根据一些这样的示例，可以通过使用WSOLA(基于波形相似性的重叠相加)技术来实现调度更快速率的语音回放。

[0032] 一些公开的方法可以涉及分析音频数据以确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据，指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。一些这样的方法可以涉及应用会话动态数据作为描述虚拟声学空间中的每个会议参与者的虚拟会议参与者位置的向量的空间优化成本函数的一个或多个变量。这样的方法可以包括将优化技术应用于空间优化成本函数以确定局部最优解，并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0033] 在一些示例中，音频数据可以包括语音活动检测处理的输出。一些实现可以涉及识别对应于各个会议参与者的语音。在一些实现中，音频数据对应于至少一个完整或基本上完整的会议的记录。

[0034] 一些方法可以涉及接收(例如，通过会话动态分析模块)对应于涉及多个会议参与者的会议的记录的音频数据。在一些例子中，会议可能是电话会议。然而，在一些例子中，会议可能是面对面会议。

[0035] 根据一些示例，音频数据可以包括来自多个端点的音频数据。多个端点中的每一

个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据中的至少一些可以来自对应于多个会议参与者的单个端点。音频数据可以包括用于标识多个会议参与者中的每个会议参与者的会议参与者语音的信息。

[0036] 一些这样的方法可能涉及分析会议记录的会话动态以确定会话动态数据。一些方法可能涉及搜索会议记录以确定多个段分类中的每一个的实例。每个段分类可以至少部分地基于会话动态数据。一些实现可以涉及将会议记录分成多个段。每个段可以对应于时间间隔和至少一个段分类。

[0037] 在一些示例中，分析、搜索和分段过程可以由会话动态分析模块执行。在一些实现中，搜索和分段过程可以是递归过程。在一些实现中，搜索和分段过程可以在不同的时间尺度上多次执行。

[0038] 根据一些实施方式，搜索和分段过程可以至少部分地基于段分类的层级结构。在一些示例中，段分类的层级结构可以基于特定段分类的段可被标识的置信水平、段的开始时间可被确定的置信水平、段的结束时间可被确定的置信水平和/或特定段分类包括对应于会议主题的会议参与者语音的可能性。

[0039] 在一些实现中，段分类的实例可以根据一组规则来确定。规则可以例如基于一个或多个会话动态数据类型，例如指示在时间间隔中的在其期间至少两个会议参与者同时发言的语音时间的占比的双讲话比率、指示在该时间间隔中的存在任何会议参与者语音的占比的语音密度度量、和/或指示在该时间间隔期间的由主导会议参与者发出的总语音的占比的主导度量。主导会议参与者可以是在时间间隔期间发言最多的会议参与者。

[0040] 在一些示例中，该组规则可以包括如果语音密度度量小于相互静默阈值则将段分类为相互静默段的规则。根据一些示例，该组规则可以包括如下规则，即如果语音密度度量大于或等于相互静默阈值并且双讲话比率比大于混串音阈值，则将段分类为混串音(Babble)段。在一些实现中，该组规则可以包括如下规则，即如果语音密度度量大于或等于静默阈值，并且如果双讲话比率小于或等于混串音阈值但是大于讨论阈值，则将段分类为讨论段。

[0041] 根据一些实现，该组规则可以包括如下规则，即如果语音密度度量大于或等于静默阈值，如果双讲话比率小于或等于讨论阈值，以及如果主导度量大于陈述阈值，则将段分类为陈述段。在一些示例中，该组规则可以包括如下规则，即如果语音密度度量大于或等于静默阈值，如果双讲话比率小于或等于讨论阈值，以及如果主导度量小于或等于陈述阈值但大于问答阈值，则将段分类为问答段。

[0042] 如上所述，在一些实现中，搜索和分段过程至少部分地基于段分类的层级结构。根据一些这样的实现，搜索过程的第一层级可以涉及搜索会议记录以确定混串音段的实例。在一些示例中，搜索过程的第二层级可以涉及搜索会议记录以确定陈述段的实例。

[0043] 根据一些示例，搜索过程的第三层级可以涉及搜索会议记录以确定问答段的实例。根据一些实施方式，搜索过程的第四层级可以包括搜索会议记录以确定讨论段的实例。

[0044] 然而，在一些替代实现中，段分类的实例可以根据机器学习分类器来确定。在一些示例中，机器学习分类器可以是自适应增强技术、支持向量机技术、贝叶斯网络模型技术、神经网络技术、隐式马尔可夫模型技术、或条件随机场技术。

[0045] 一些公开的方法可以包括接收(例如，通过主题分析模块)关于涉及多个会议参与

者的会议的记录的至少一部分的语音识别结果数据。语音识别结果数据可以包括多个语音识别格、以及语音识别格的多个假设词中的每一个的词语识别置信度分数。词语识别置信度分数可以对应于假设词与在会议期间由会议参与者说出的实际词正确对应的可能性。在一些示例中，接收语音识别结果数据可以涉及从两个或更多个自动语音识别过程接收语音识别结果数据。

[0046] 一些这样的方法可以涉及对于语音识别格中的多个假设词中的每一个确定主词候选(primary word candidate)和一个或多个替代词假设(alternative word hypotheses)。与一个或多个替代词假设中的任一个的词语识别置信度分数相比，主词候选的词语识别置信度分数指示更高的与在会议期间由会议参与者说出的实际词正确对应的可能性。

[0047] 一些方法可能包括计算主词候选和替代词假设的术语(term)频率度量。术语频率度量可以至少部分地基于语音识别格中的假设词的出现次数以及词语识别置信度分数。根据一些实现，计算术语频率度量可以至少部分地基于多个词含义。一些这样的方法可以包括根据术语频率度量来对主词候选和替代词假设，包括替代假设列表中的替代词假设，进行排序，并且根据替代假设列表对语音识别格的至少一些假设词进行重新评分。

[0048] 一些实现可以涉及形成词列表。词列表可以包括主词候选和每个主词候选词的术语频率度量。在一些示例中，术语频率度量可以与文档频率度量成反比。文档频率度量可以对应于主要候选词将在会议中出现的预期频率。根据一些示例，预期频率可以对应于主词候选在两个或更多个先前会议中出现的频率、或主词候选在语言模型中出现的频率。

[0049] 根据一些示例，词列表还可以包括关于每个主词候选的一个或多个替代词假设。在某些实例中，可以根据多种语言模型生成替代词假设。

[0050] 一些方法可以包括至少部分地基于词列表来生成会话主题的主题列表。在一些示例中，生成主题列表可以涉及确定词列表中的至少一个词的上位词。根据一些这样的示例，生成主题列表可以涉及确定主题分数。在一些示例中，主题分数可以包括上位词分数。根据一些这样的示例，包括过程可以涉及至少部分地基于主题分数将替代词假设包含在替代假设列表中。

[0051] 在一些实现中，可以执行至少确定、计算、排序、包括和重新评分过程的两次或多次迭代。根据一些示例，迭代可以涉及生成主题列表并确定主题分数。在一些示例中，替代假设列表可以在每次迭代之后被保留。

[0052] 一些实现可以涉及将语音识别格的至少一些假设词缩减到规范的基本形式。例如，缩减可以包括将语音识别格的名词缩减到规范的基本形式。规范的基本形式可以是名词的单数形式。作为替代地或者附加地，缩减可以包括将语音识别格的动词缩减到规范的基本形式。规范的基本形式可能是动词的不定式形式。

[0053] 根据一些示例，会议记录可以包括被分别记录的来自多个端点的会议参与者语音数据。作为替代地或者附加地，会议记录可以包括来自对应于多个会议参与者的单个端点的会议参与者语音数据，其可以包括用于标识多个会议参与者的每个会议参与者的会议参与者语音的信息。

[0054] 一些公开的方法可以涉及接收对应于涉及多个会议参与者的至少一个会议的记录的音频数据。音频数据可以包括被分别记录的来自多个端点的会议参与者语音数据、和/

或来自对应于多个会议参与者的单个端点的会议参与者语音数据，其可以包括多个会议参与者的每个会议参与者的空间信息。

[0055] 这样的方法可以涉及基于对音频数据的搜索来确定搜索结果。搜索可以是或可能已经基于一个或多个搜索参数。搜索结果可以对应于音频数据中的会议参与者语音的至少两个实例。会议参与者语音的实例可以包括讲话突发和/或讲话突发的部分。会议参与者语音的实例可以包括由第一会议参与者发出的第一语音实例和由第二会议参与者发出的第二语音实例。

[0056] 一些这样的方法可以包括将会议参与者语音的实例渲染到虚拟声学空间的至少两个不同的虚拟会议参与者位置，使得第一语音实例被渲染到第一虚拟会议参与者位置，并且第二语音实例被渲染到第二虚拟会议参与者位置。这样的方法可以包括调度会议参与者语音的实例的至少一部分进行同时回放，以产生回放音频数据。

[0057] 根据一些实现方式，确定搜索结果可能涉及接收搜索结果。例如，确定搜索结果可能涉及接收从通过另一装置(例如通过服务器)执行的搜索得到的搜索结果。

[0058] 然而，在一些实现中，确定搜索结果可能涉及执行搜索。根据一些示例，确定搜索结果可以包括执行音频数据的关于多个特征的并发搜索。根据一些实施方式，多个特征可以包括从一组特征中选择的两个或更多个特征。该组特征可以包括词语、会议段、时间、会议参与者情绪、端点位置和/或端点类型。在一些实现中，确定搜索结果可以涉及执行对应于多个会议的记录的音频数据的搜索。在一些示例中，调度过程可以包括至少部分地基于搜索相关性度量来调度会议参与者语音的实例进行回放。

[0059] 一些实现可以涉及修改会议参与者语音的至少一个实例的开始时间或结束时间。在一些示例中，修改过程可以涉及扩展对应于会议参与者语音的实例的时间间隔。根据一些示例，修改过程可以涉及合并对应于单个会议端点的、扩展后在时间上重叠的会议参与者语音的两个或更多个实例。

[0060] 在一些示例中，调度过程可以包括调度先前在时间上不重叠的会议参与者语音的实例以在时间上重叠地回放。作为替代地或者附加地，一些方法可以涉及调度先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地回放。

[0061] 根据一些实施方式，调度可以根据感知激发规则的集合来执行。在一些实现中，感知激发规则的集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则的集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则这两个输出讲话突发在时间上不应该重叠。

[0062] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。

[0063] 一些公开的方法可以涉及分析音频数据以确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示至少两个会议参与者在其期间同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。一些

这样的方法可以涉及将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。这样的方法可以涉及将优化技术应用于空间优化成本函数以确定局部最优解，并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0064] 一些实现可以涉及提供用于控制显示器以提供图形用户界面的指令。根据一些实现，用于控制显示器的指令可以包括用于进行会议参与者的展示的指令。用于执行搜索的一个或多个特征可以例如包括会议参与者的指示。

[0065] 在一些示例中，用于控制显示器的指令可以包括用于进行会议段的展示的指令。用于执行搜索的一个或多个特征可以例如包括会议段的指示。

[0066] 在一些实例中，用于控制显示器的指令可以包括用于进行用于搜索特征的显示区域的展示的指令。用于执行搜索的一个或多个特征可以例如包括词语、时间、会议参与者情绪、端点位置和/或端点类型。

[0067] 一些这样的实现可以涉及接收对应于用户与图形用户界面的交互的输入，并且至少部分地基于该输入来处理音频数据。在一些示例中，输入可以对应于用于执行音频数据的搜索的一个或多个特征。一些这样的方法可以包括将回放音频数据提供给扬声器系统。

[0068] 根据一些实现方式，确定搜索结果可能涉及搜索关键词检索索引。在一些示例中，关键词检索索引可以具有包括指向上下文信息的指针的数据结构。根据一些这样的示例，指针可以是或可以包括矢量量化索引。

[0069] 在一些示例中，确定搜索结果可以涉及例如根据一个或多个时间参数确定用于搜索的一个或多个会议的第一阶段。一些这样的方法可以涉及根据其他搜索参数来检索搜索结果的第二阶段。

[0070] 一些公开的方法可以涉及接收对应于会议记录的音频数据。音频数据可以包括对应于多个会议参与者中的每一个的会议参与者语音的数据。这样的方法可以包括仅将会议参与者语音的一部分选择为回放音频数据。

[0071] 根据一些实现，选择过程可以涉及根据所估计的会议参与者语音与一个或多个会议主题的相关性来选择用于回放的会议参与者语音的主题选择过程。在一些实现中，选择过程可以涉及根据所估计的会议参与者语音与会议段的一个或多个主题的相关性来选择用于回放的会议参与者语音的主题选择过程。

[0072] 在一些实例中，选择过程可以涉及去除具有低于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发。根据一些示例，选择过程可以包括讲话突发过滤过程，其去除具有等于或高于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发的一部分。

[0073] 作为替代地或者附加地，选择过程可以包括根据至少一个声学特征来选择用于回放的会议参与者语音的声学特征选择过程。在一些示例中，选择可以涉及迭代过程。一些这样的实现可以涉及将回放音频数据提供给扬声器系统以供回放。

[0074] 一些方法可以涉及接收目标回放持续时间的指示。根据一些这样的示例，选择过程可以包括使回放音频数据的持续时间在目标回放持续时间的阈值时间百分比和/或阈值时间差之内。在一些示例中，回放音频数据的持续时间可以至少部分地通过将会议参与者语音的至少一个选定部分的持续时间乘以加速系数来确定。

[0075] 根据一些示例，会议记录可以包括被分别记录的来自多个端点的会议参与者语音数据，或者来自对应于多个会议参与者的单个端点的会议参与者语音数据，其可以包括多个会议参与者的每个会议参与者的空间信息。一些这样的方法可以涉及在虚拟声学空间中渲染回放音频数据，使得其语音被包括在回放音频数据中的各会议参与者具有各自不同的虚拟会议参与者位置。

[0076] 根据一些实现，选择过程可以涉及主题选择过程。根据一些这样的示例，主题选择过程可以涉及接收会议主题的主题列表并且确定所选择的会议主题的列表。所选择的会议主题的列表可能是会议主题的子集。

[0077] 一些方法可以涉及接收主题排名(ranking)数据，其可以指示主题列表上的每个会议主题的估计的相关性。确定所选择的会议主题的列表可以至少部分地基于主题排名数据。

[0078] 根据一些实现，选择过程可以涉及讲话突发过滤过程。讲话突发过滤过程例如可以涉及去除输入讲话突发的初始部分。初始部分可以是从输入讲话突发开始时间到输出讲话突发开始时间的时间间隔。一些方法可以包括至少部分地基于输入讲话突发持续时间来计算输出讲话突发持续时间。

[0079] 一些这样的方法可以涉及确定输出讲话突发持续时间是否超过输出讲话突发时间阈值。如果确定输出讲话突发持续时间超过输出讲话突发时间阈值，讲话突发过滤过程可能涉及对于单个输入讲话突发生成会议参与者语音的多个实例。根据一些这样的示例，会议参与者语音的多个实例中的至少一个可以具有与输入讲话突发结束时间相对应的结束时间。

[0080] 根据一些实现，选择过程可以涉及声学特征选择过程。在一些示例中，声学特征选择过程可以涉及确定至少一个声学特征，例如音调变化、语速和/或响度。

[0081] 一些实现可以涉及修改会议参与者语音的至少一个实例的开始时间或结束时间。在一些示例中，修改过程可以涉及扩展对应于会议参与者语音的实例的时间间隔。根据一些示例，修改过程可以涉及将扩展后在时间上重叠的与单个会议端点对应的会议参与者语音的两个或更多个实例合并。

[0082] 在一些示例中，调度过程可以包括调度先前在时间上不重叠的会议参与者语音的实例以在时间上重叠地回放。作为替代地或者附加地，一些方法可以涉及调度先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地回放。

[0083] 根据一些实施方式，调度可以根据感知激发规则的集合来执行。在一些实现中，感知激发规则的集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则的集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则这两个输出讲话突发在时间上不应该重叠。

[0084] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。一些实现可以涉及至少部分地基于搜索相关性度量来调度会议参与者语音的实例以

供回放。

[0085] 一些公开的方法可以涉及分析音频数据以确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示至少两个会议参与者在其期间同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。一些这样的方法可以涉及将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。这样的方法可以涉及将优化技术应用于空间优化成本函数以确定局部最优解，并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0086] 一些实现可以涉及提供用于控制显示器以提供图形用户界面的指令。根据一些实现，用于控制显示器的指令可以包括用于进行会议参与者的展示的指令。在一些示例中，用于控制显示器的指令可以包括用于进行会议段的展示的指令。

[0087] 一些这样的实现可以涉及接收对应于用户与图形用户界面的交互的输入，并且至少部分地基于该输入来处理音频数据。在一些示例中，输入可以对应于目标回放持续时间的指示。一些这样的方法可以包括将回放音频数据提供给扬声器系统。

[0088] 本公开的至少一些方面可以经由装置来实现。例如，一个或多个设备可能能够至少部分地执行本文公开的方法。在一些实现中，装置可以包括接口系统和控制系统。接口系统可以包括网络接口、控制系统和存储系统之间的接口、控制系统与另一设备之间的接口和/或外部设备接口。控制系统可以包括通用单芯片或多芯片处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其他可编程逻辑器件、离散门或晶体管逻辑、或离散硬件组件中的至少一个。

[0089] 该控制系统可能能够至少部分地执行本文公开的方法。在一些实现中，控制系统可能够经由接口系统在电话会议期间接收电话会议音频数据。电话会议音频数据可以包括多个单独的上行链路数据分组流。每个上行链路数据分组流可以对应于一个或多个电话会议参与者使用的电话端点。在一些实现中，控制系统可能够经由接口系统将电话会议音频数据作为单独的上行链路数据分组流发送到存储系统。

[0090] 根据一些示例，控制系统可能够确定不完整的上行链路数据分组流的迟到数据分组已经在迟到分组时间阈值之后被从电话端点接收到。迟到分组时间阈值可以大于或等于电话会议的口到耳延迟时间阈值。控制系统可能够将迟到数据分组添加到不完整的上行链路数据分组流。

[0091] 在一些示例中，控制系统可能够确定在丢失分组时间阈值内没有从电话端点接收到不完整上行链路数据分组流的丢失数据分组。在一些示例中，该丢失分组时间阈值可大于迟到分组时间阈值。控制系统可能够经由接口系统向电话端点发送关于重新发送丢失的数据分组的请求，接收丢失的数据分组，并且将丢失的数据分组添加到不完整的上行数据分组流中。

[0092] 在一些实现中，单独的上行链路数据分组流可以是单独的编码的上行链路数据分组流。一些这样的实现可以涉及将电话会议音频数据作为单独的编码的上行链路数据分组流发送到存储系统。

[0093] 接口系统可以包括控制系统和存储系统的至少一部分之间的接口。根据一些实现，存储系统的至少一部分可以被包括在一个或多个或其他设备中，例如本地或远程存储

设备。在一些实现中，接口系统可以包括网络接口，并且控制系统可以能够经由网络接口将电话会议音频数据发送到存储系统。然而，根据一些示例，该装置可以包括存储系统的至少一部分。

[0094] 在一些示例中，上行链路数据分组流中的至少一个可包括已经在电话会议的口到耳延迟时间阈值之后接收到、因此不用于在电话会议期间再现音频数据的至少一个数据分组。根据一些示例，上行链路数据分组流中的至少一个可以对应于多个电话会议参与者，并且可以包括关于多个参与者中的每一个的空间信息。根据一些实施方式，控制系统可能能够提供电话会议服务器功能。

[0095] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可以能够经由接口系统接收电话会议的记录的音频数据。记录的音频数据可以包括对应于由一个或多个电话会议参与者使用的电话端点的单独的上行链路数据分组流。

[0096] 根据一些示例，控制系统可以能够分析单独的上行链路数据分组流中的数据分组的序列号数据。根据一些这样的示例，分析过程可以包括确定单独的上行链路数据分组流是否包括至少一个无序数据分组。如果上行链路数据分组流包括至少一个无序数据分组，则控制系统可以能够根据序列号数据重新排序单独的上行链路数据分组流。

[0097] 在一些实例中，控制系统可确定单独的上行链路数据分组流的至少一个数据分组已经在电话会议的口到耳延迟时间阈值之后接收到。根据一些这样的示例，控制系统可以能够接收(例如，经由接口系统)电话会议元数据并且至少部分地基于电话会议元数据来对单独的上行链路数据分组流进行索引。

[0098] 在一些示例中，记录的音频数据可以包括多个单独的编码的上行链路数据分组流。每个单独的编码的上行链路数据分组流可以对应于由一个或多个电话会议参与者使用的电话端点。根据一些实现，控制系统可以包括能够分析多个单独的上行链路数据分组流的联合分析模块。根据一些这样的示例，控制系统可以能够解码多个单独的编码的上行链路数据分组流，并且向联合分析模块提供多个单独的解码的上行链路数据分组流。

[0099] 在一些实现中，控制系统可以包括能够识别语音的语音识别模块。语音识别模块能够产生语音识别结果数据。根据一些示例，控制系统可以能够向语音识别模块提供一个或多个单独的解码的上行链路数据分组流。根据一些这样的示例，语音识别模块可以能够将语音识别结果数据提供给联合分析模块。

[0100] 根据一些实现，联合分析模块可以能够识别语音识别结果数据中的关键词。在一些示例中，联合分析模块可能能够对关键词位置进行索引。

[0101] 根据一些示例，控制系统可以包括发言者日志模块。在一些实例中，控制系统可能能够向发言者日志模块提供单独的解码的上行链路数据分组流。发言者日志模块可以例如能够识别单独的解码的上行链路数据分组流中的多个电话会议参与者中的每一个的语音。在一些示例中，发言者日志模块可以能够产生发言者日志，其指示多个电话会议参与者中的每一个讲话时的时间。发言者日志模块可以能够将发言者日志提供给联合分析模块。

[0102] 在一些实现中，联合分析模块可以能够确定会话动态数据。例如，会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0103] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可以能够经由接口系统接收对应于涉及多个会议参与者的会议的记录的音频数据。音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据。音频数据可以包括多个会议参与者的每个会议参与者的空间信息。

[0104] 在一些实现中，控制系统可以能够分析音频数据以确定会话动态数据。在一些示例中，会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0105] 根据一些示例，控制系统可以能够将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。例如，控制系统可以能够将优化技术应用于空间优化成本函数以确定局部最优解。控制系统可以能够至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0106] 根据一些实现，虚拟声学空间可以相对于虚拟听众的头部在虚拟声学空间中的位置来确定。在一些这样的实现中，空间优化成本函数可以应用对于将参与会议参与者双讲话的会议参与者布置在如下虚拟会议参与者位置处的惩罚，该虚拟会议参与者位置位于混淆锥上或者与该混淆锥相距在预定的角距离内。该混淆锥相对于虚拟听众头部的位置被定义。通过混淆锥的圆锥切片可能具有相同的耳间时间差异。

[0107] 在一些示例中，空间优化成本函数可以应用对于将参与会议参与者相互会话的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚。根据一些示例，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。在某些实例中，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于与较不频繁发言的会议参与者的虚拟会议参与者位置相比距虚拟听众头部的位置更远的虚拟会议参与者位置处的惩罚。然而，根据一些实现，分配虚拟会议参与者位置可涉及从一组预定的虚拟会议参与者位置中选择虚拟会议参与者位置。

[0108] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可以能够经由接口系统接收对应于涉及多个会议参与者的会议的记录的音频数据。音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据。音频数据可以包括多个会议参与者的每个会议参与者的空间信息。

[0109] 根据一些实施方式，控制系统可以能够将每个会议参与者的会议参与者语音数据渲染到虚拟声学空间中的单独的虚拟会议参与者位置。在一些实现中，控制系统可以能够调度会议参与者语音回放，使得在会议参与者语音的至少两个输出讲话突发之间的回放重叠量大于会议记录的两个对应的输入讲话突发之间的原始重叠量。

[0110] 在一些示例中，调度可以至少部分地根据感知激发规则的集合来执行。在一些实

现中，感知激发规则的集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则的集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则两个输出讲话突发在时间上不应该重叠。

[0111] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。

[0112] 根据一些示例，控制系统可能能够分析音频数据以确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0113] 在一些示例中，控制系统可以能够将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。在一些实现中，控制系统可以能够将优化技术应用于空间优化成本函数以确定局部最优解。根据一些实现，控制系统可以能够至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0114] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可以能够经由接口系统接收对应于涉及多个会议参与者的会议的记录的音频数据。音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据。音频数据可以包括用于识别多个会议参与者中的每个会议参与者的会议参与者语音的信息。

[0115] 根据一些实现，控制系统可以能够分析会议记录的会话动态以确定会话动态数据。在一些示例中，控制系统可以能够搜索会议记录来确定多个段分类中的每一个的实例。每个段分类可以至少部分地基于会话动态数据。

[0116] 根据一些这样的示例，控制系统可以能够将会议记录分成多个段。每个段可以对应于时间间隔和至少一个段分类。在一些示例中，控制系统可以能够在不同的时间尺度上多次执行搜索和分段处理。

[0117] 在一些实现中，搜索和分段过程可以至少部分地基于段分类的层级结构。根据一些这样的实现，段分类的层级结构可以基于一个或多个准则，例如特定段分类的段可被标识的置信水平、段的开始时间可被确定的置信水平、段的结束时间可被确定的置信水平、和/或特定分段分类包括对应于会议主题的会议参与者语音的可能性。

[0118] 在一些示例中，控制系统可能能够根据一组规则来确定段分类的实例。根据一些这样的示例，规则可以基于一个或多个会话动态数据类型，例如指示在时间间隔中的在其期间至少两个会议参与者同时发言的语音时间的占比的双讲话比率、指示在该时间间隔中的存在任何会议参与者语音的占比的语音密度度量、和/或指示在该时间间隔期间的由主导会议参与者发出的总语音的占比的主导度量。主导会议参与者可以是在时间间隔期间发

言最多的会议参与者。

[0119] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可以能够接收(例如，经由接口系统)关于涉及多个会议参与者的会议的记录的至少一部分的语音识别结果数据。在一些示例中，语音识别结果数据可以包括多个语音识别格、以及语音识别格的多个假设词中的每一个的词语识别置信度分数。词语识别置信度分数可以对应于假设词与在会议期间由会议参与者说出的实际词正确对应的可能性。

[0120] 在一些实现中，控制系统可以能够对于语音识别格中的多个假设词中的每一个确定主词候选和一个或多个替代词假设。与一个或多个替代词假设中的任一个的词语识别置信度分数相比，主词候选的词语识别置信度分数指示更高的与在会议期间由会议参与者说出的实际词正确对应的可能性。

[0121] 根据一些示例，控制系统可以能够计算主词候选和替代词假设的术语频率度量。在一些实例中，术语频率度量可以至少部分地基于语音识别格中假设词的出现次数。作为替代地或者附加地，术语频率度量可以至少部分地基于词语识别置信度分数。

[0122] 根据一些实现，控制系统可能能够根据术语频率度量来对主词候选和替代词假设进行排序。根据一些示例，控制系统可能能够在替代假设列表中包括替代词假设。根据一些这样的示例，控制系统可能能够根据替代假设列表对语音识别格中的至少一些假设词重新评分。

[0123] 在一些示例中，控制系统可能能够形成词列表。词列表可以包括主词候选和每个主词候选词的术语频率度量。根据一些示例，控制系统可以能够至少部分地基于词列表来生成会话主题的主题列表。在一些实现中，生成主题列表可以涉及确定词列表中的至少一个词的上位词。生成主题列表可以涉及确定包括上位词分数的主题分数。

[0124] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可能能够接收(例如，经由接口系统)与对应于涉及多个会议参与者的至少一个会议的记录相对应的音频数据。音频数据可以包括分别记录的来自多个端点的会议参与者语音数据，和/或来自对应于多个会议参与者的单个端点的会议参与者语音数据，其可以包括多个会议参与者中的每个会议参与者的空间信息。

[0125] 根据一些实现，控制系统可以能够确定与基于一个或多个搜索参数对音频数据的搜索相对应的搜索结果。搜索结果可以对应于音频数据中的会议参与者语音的至少两个实例。会议参与者语音的至少两个实例可以包括由第一会议参与者发出的至少第一语音实例和由第二会议参与者发出的至少第二语音实例。

[0126] 在一些示例中，控制系统可以能够将会议参与者语音的实例渲染到虚拟声学空间的至少两个不同的虚拟会议参与者位置，使得第一语音实例被渲染到第一虚拟会议参与者位置，并且第二语音实例被渲染给第二虚拟会议参与者位置。根据一些这样的示例，控制系统可以能够调度会议参与者语音的实例的至少一部分同时回放，以产生回放音频数据。

[0127] 在一些替代实现中，装置还可以包括诸如上述那些的接口系统。该装置还可以包括诸如上述那些的控制系统。根据一些这样的实现，控制系统可以能够接收(例如，经由接口系统)与会议记录对应的音频数据。音频数据可以包括对应于多个会议参与者中的每一

个的会议参与者语音的数据。

[0128] 根据一些示例，控制系统可以能够仅选择会议参与者语音的一部分作为回放音频数据。根据一些这样的示例，控制系统可以能够(例如，经由接口系统)将回放音频数据提供给用于回放的扬声器系统。

[0129] 根据一些实现，选择过程可以涉及根据所估计的会议参与者语音与一个或多个会议主题的相关性来选择用于回放的会议参与者语音的主题选择过程。在一些实现中，选择过程可以涉及根据所估计的会议参与者语音与会议段的一个或多个主题的相关性来选择用于回放的会议参与者语音的主题选择过程。

[0130] 在一些实例中，选择过程可以涉及去除具有低于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发。根据一些示例，选择过程可以包括讲话突发过滤过程，其去除具有等于或高于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发的一部分。

[0131] 作为替代地或者附加地，选择过程可以包括根据至少一个声学特征来选择用于回放的会议参与者语音的声学特征选择过程。在一些示例中，选择可以涉及迭代过程。

[0132] 根据一些示例，控制系统可以能够(例如，经由接口系统)接收目标回放持续时间的指示。根据一些这样的示例，选择过程可以包括使回放音频数据的持续时间在目标回放持续时间的阈值时间百分比和/或阈值时间差之内。在一些示例中，回放音频数据的持续时间可以至少部分地通过将会议参与者语音的至少一个选定部分的持续时间乘以加速系数来确定。

[0133] 本文所描述的方法中的一些或全部可以由一个或多个设备根据存储在非暂态介质上的指令(例如软件)执行。这种非暂态介质可以包括诸如本文所描述的那些的存储设备，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。因此，本公开中描述的主旨的各种创新方面可以在存储有软件的非暂态介质中实现。该软件可以例如包括用于控制至少一个设备来处理音频数据的指令。例如，该软件可由诸如本文公开的那些的控制系统的一个或多个组件执行。

[0134] 根据一些示例，软件可以包括用于在电话会议期间接收电话会议音频数据的指令。电话会议音频数据可以包括多个单独的上行链路数据分组流。每个上行链路数据分组流可以对应于一个或多个电话会议参与者使用的电话端点。在一些实现中，软件可以包括用于将电话会议音频数据作为单独的上行链路数据分组流发送到存储系统的指令。

[0135] 在一些示例中，单独的上行链路数据分组流可以是单独的编码的上行链路数据分组流。根据一些示例，上行链路数据分组流中的至少一个可包括已经在电话会议的口到耳延迟时间阈值之后接收到、因此不用于在电话会议期间再现音频数据的至少一个数据分组。根据一些示例，上行链路数据分组流中的至少一个可以对应于多个电话会议参与者，并且可以包括关于多个参与者中的每一个的空间信息。

[0136] 在一些实现中，软件可以包括用于接收对应于涉及多个会议参与者的会议的记录的音频数据的指令。根据一些示例，音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据，并且可以包括多个会议参与者的每个会议参与者的空间信息。

[0137] 根据一些实现，软件可以包括用于分析音频数据以确定会话动态数据的指令。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示至少两个会议参与者在其期间同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0138] 在一些实例中，软件可以包括用于将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量的指令，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。根据一些示例，软件可包括用于将优化技术应用于空间优化成本函数以确定局部最优解的指令。根据一些示例，软件可包括用于至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置的指令。

[0139] 在一些实现中，虚拟声学空间可以相对于虚拟听众的头部在虚拟声学空间中的位置来确定。根据一些这样的实现，空间优化成本函数可以应用对于将参与会议参与者双讲话的会议参与者布置在如下虚拟会议参与者位置处的惩罚，该虚拟会议参与者位置位于相对于虚拟听众头部的位置被定义的混淆锥上或者与该混淆锥相距在预定的角距离内。通过混淆锥的圆锥切片可能具有相同的耳间时间差异。在一些示例中，空间优化成本函数可以应用对于将参与会议参与者相互会话的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚。

[0140] 根据一些示例，分析音频数据可以涉及确定哪些会议参与者(如果有的话)具有感知相似的语音。在一些这样的示例中，空间优化成本函数可以应用对于将具有感知相似的语音的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚。

[0141] 在一些示例中，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。在某些实例中，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于与较不频繁发言的会议参与者的虚拟会议参与者位置相比距虚拟听众头部的位置更远的虚拟会议参与者位置处的惩罚。在一些实现中，空间优化成本函数可以应用对于将很少发言的会议参与者布置于不在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。

[0142] 根据一些示例，优化技术可能涉及梯度下降技术、共轭梯度技术、牛顿法、Broyden-Fletcher-Goldfarb-Shanno算法、遗传算法、模拟退火算法、蚁群优化方法和/或蒙特卡罗方法。在一些示例中，分配虚拟会议参与者位置可以包括从一组预定的虚拟会议参与者位置中选择虚拟会议参与者位置。

[0143] 在一些实现中，软件可以包括用于接收对应于涉及多个会议参与者的会议的记录的音频数据的指令。根据一些示例，音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据，并且可以包括多个会议参与者的每个会议参与者的空间信息。

[0144] 根据一些实现，软件可以包括用于在虚拟声学空间中渲染会议参与者语音数据以使得各会议参与者具有各自不同的虚拟会议参与者位置的指令。在一些示例中，软件可以包括如下指令，该指令用于调度会议参与者语音回放，使得在会议参与者语音的至少两个输出讲话突发之间的回放重叠量不同于(例如，大于)会议记录的两个对应的输入讲话突发

之间的原始重叠量。

[0145] 根据一些示例，软件可以包括用于至少部分地根据感知激发规则的集合执行调度过程的指令。在一些实现中，感知激发规则的集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则两个输出讲话突发在时间上不应该重叠。

[0146] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。感知激发规则集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。

[0147] 根据一些实现，感知激发规则集合可以包括允许来自不同会议参与者的全部陈述的并发回放的规则。在一些实现中，陈述可以对应于会议参与者语音的时间间隔，在该时间间隔期间，语音密度度量大于或等于静默阈值，双讲话比率小于或等于讨论阈值，并且主导度量大于陈述阈值。双讲话比率可以指示在该时间间隔中的在其期间至少两个会议参与者同时发言的语音时间的占比。语音密度度量可以指示在该时间间隔中的存在任何会议参与者语音的占比。主导度量可以指示在该时间间隔期间的由主导会议参与者发出的总语音的占比。主导会议参与者可以是在时间间隔内发言最多的会议参与者。

[0148] 在一些示例中，会议参与者语音中的至少一些可被调度为以比记录会议参与者语音的速率更快的速率来回放。根据一些这样的示例，可以通过使用WSOLA(基于波形相似性的重叠相加)技术来实现调度更快速率的语音回放。

[0149] 根据一些实现，软件可以包括用于分析音频数据以确定会话动态数据的指令。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据、指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。在一些示例中，软件可以包括用于应用会话动态数据作为描述虚拟声学空间中的每个会议参与者的虚拟会议参与者位置的向量的空间优化成本函数的一个或多个变量的指令。在一些实现中，软件可以包括如下指令，该指令用于将优化技术应用于空间优化成本函数以确定局部最优解，并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0150] 在一些实现中，软件可以包括用于接收对应于涉及多个会议参与者的会议的记录的音频数据的指令。根据一些示例，音频数据可以包括来自多个端点的音频数据。多个端点中的每一个的音频数据可能已被单独记录。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据，并且可以包括用于识别多个会议参与者中的每个会议参与者的会议参与者语音的信息。

[0151] 根据一些示例，软件可以包括用于分析会议记录的会话动态以确定会话动态数据的指令。在一些示例中，软件可以包括用于搜索会议记录以确定多个段分类中的每一个的实例的指令。每个段分类可以至少部分地基于会话动态数据。根据一些这样的示例，软件可以包括用于将会议记录分成多个段的指令。每个段可以对应于时间间隔和至少一个段分类。根据一些实现，软件可以包括用于在不同时间尺度上多次执行搜索和分段过程的指令。

[0152] 在一些示例中，搜索和分段过程可以至少部分地基于段分类的层级结构。根据一些这样的示例，段分类的层级结构可以至少部分地基于特定段分类的段可被标识的置信水平、段的开始时间可被确定的置信水平、段的结束时间可被确定的置信水平和/或特定段分类包括对应于会议主题的会议参与者语音的可能性。

[0153] 根据一些实现，软件可以包括根据一组规则来确定段分类的实例的指令。在一些这样的实现中，规则可以例如基于一个或多个会话动态数据类型，例如指示在时间间隔中的在其期间至少两个会议参与者同时发言的语音时间的占比的双讲话比率、指示在该时间间隔中的存在任何会议参与者语音的占比的语音密度度量、和/或指示在该时间间隔期间的由主导会议参与者发出的总语音的占比的主导度量。主导会议参与者可以是在时间间隔期间发言最多的会议参与者。

[0154] 在一些实现中，软件可以包括用于接收涉及多个会议参与者的会议的会议记录的至少一部分的语音识别结果数据的指令。在一些示例中，语音识别结果数据可以包括多个语音识别格。语音识别结果数据可以包括语音识别格的多个假设词中的每一个的词语识别置信度分数。根据一些这样的示例，词语识别置信度分数可以对应于假设词与在会议期间由会议参与者说出的实际词正确对应的可能性。

[0155] 根据一些示例，软件可以包括用于对于语音识别格中的多个假设词中的每一个确定主词候选和一个或多个替代词假设的指令。与一个或多个替代词假设中的任一个的词语识别置信度分数相比，主词候选的词语识别置信度分数指示更高的与在会议期间由会议参与者说出的实际词正确对应的可能性。

[0156] 根据一些实施方式，软件可以包括用于计算主词候选和替代词假设的术语频率度量的指令。在一些这样的实现中，术语频率度量可以至少部分地基于语音识别格中的假设词的出现次数和词语识别置信度分数。

[0157] 在一些示例中，软件可以包括根据术语频率度量将主词候选和替代词假设排序的指令。根据一些这样的示例，软件可以包括用于将替代词假设包括在替代假设列表中的指令。在一些这样的实现中，软件可以包括用于根据替代假设列表对语音识别格的至少一些假设词重新评分的指令。

[0158] 根据一些示例，软件可以包括用于形成词语列表的指令。词语列表可以例如包括主词候选和每个主词候选词的术语频率度量。根据一些这样的示例，软件可以包括用于至少部分地基于词语列表来生成会议主题的主题列表的指令。

[0159] 在一些实现中，生成主题列表可以涉及确定词语列表中的至少一个词语的上位词。根据一些这样的实现，生成主题列表可以涉及确定包括上位词分数的主题分数。

[0160] 在一些实现中，软件可以包括用于接收对应于涉及多个会议参与者的至少一个会议的记录的音频数据的指令。音频数据可以包括分别记录的来自多个端点的会议参与者语音数据、和/或来自对应于多个会议参与者的单个端点的会议参与者语音数据，其可以包括多个会议参与者的每个会议参与者的空间信息。

[0161] 根据一些示例，软件可以包括用于基于音频数据的搜索来确定搜索结果的指令。搜索可以是或可能已经基于一个或多个搜索参数。搜索结果可以对应于音频数据中的会议参与者语音的至少两个实例。会议参与者语音的实例可以例如包括讲话突发和/或讲话突发的部分。会议参与者语音的实例可以包括由第一会议参与者发出的第一语音实例和由第

二会议参与者发出的第二语音实例。

[0162] 在一些示例中，软件可以包括如下指令，该指令用于将会议参与者语音的实例渲染到虚拟声学空间的至少两个不同的虚拟会议参与者位置，使得第一语音实例被渲染到第一虚拟会议参与者位置，并且第二语音实例被渲染到第二虚拟会议参与者位置。根据一些示例，软件可以包括用于调度会议参与者语音的实例的至少一部分进行同时回放以产生回放音频数据的指令。

[0163] 根据一些实现方式，确定搜索结果可能涉及接收搜索结果。例如，确定搜索结果可能涉及接收通过另一设备(例如，通过服务器)执行的搜索而得到的搜索结果。

[0164] 然而，在一些实现中，确定搜索结果可能涉及执行搜索。根据一些示例，确定搜索结果可以包括执行音频数据的关于多个特征的并发搜索。根据一些实施方式，多个特征可以包括从一组特征中选择的两个或更多个特征。该组特征可以包括词语、会议段、时间、会议参与者情绪、端点位置和/或端点类型。在一些实现中，确定搜索结果可以涉及执行对应于多个会议的记录的音频数据的搜索。在一些示例中，调度过程可以包括至少部分地基于搜索相关性度量来调度会议参与者语音的实例进行回放。

[0165] 根据一些示例，软件可以包括用于修改会议参与者语音的至少一个实例的开始时间或结束时间的指令。在一些示例中，修改过程可以涉及扩展对应于会议参与者语音的实例的时间间隔。根据一些示例，修改过程可以涉及合并对应于单个会议端点的、扩展后在时间上重叠的会议参与者语音的两个或更多个实例。

[0166] 在一些示例中，软件可以包括用于调度先前在时间上不重叠的会议参与者语音的实例以在时间上重叠地回放。作为替代地或者附加地，软件可以包括用于调度先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地回放的指令。

[0167] 根据一些实施方式，调度可以根据感知激发规则的集合来执行。在一些实现中，感知激发规则的集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则的集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则这两个输出讲话突发在时间上不应该重叠。。

[0168] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。

[0169] 在一些实现中，软件可以包括用于接收对应于会议记录的音频数据的指令。音频数据可以包括对应于多个会议参与者中的每一个的会议参与者语音的数据。在一些示例中，软件可以包括用于仅选择会议参与者语音的一部分作为回放音频数据的指令。

[0170] 根据一些实现，选择过程可以涉及根据所估计的会议参与者语音与一个或多个会议主题的相关性来选择用于回放的会议参与者语音的主题选择过程。在一些实现中，选择过程可以涉及根据所估计的会议参与者语音与会议段的一个或多个主题的相关性来选择用于回放的会议参与者语音的主题选择过程。

[0171] 在一些实例中，选择过程可以涉及去除具有低于阈值输入讲话突发持续时间的输

入讲话突发持续时间的输入讲话突发。根据一些示例，选择过程可以包括讲话突发过滤过程，其去除具有等于或高于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发的一部分。

[0172] 作为替代地或者附加地，选择过程可以包括根据至少一个声学特征来选择用于回放的会议参与者语音的声学特征选择过程。在一些示例中，选择可以涉及迭代过程。一些这样的实现可涉及将回放音频数据提供给扬声器系统以供回放。

[0173] 根据一些实现，软件可以包括用于接收目标回放持续时间的指示的指令。根据一些这样的示例，选择过程可以包括使回放音频数据的持续时间在目标回放持续时间的阈值时间百分比和/或阈值时间差之内。在一些示例中，回放音频数据的持续时间可以至少部分地通过将会议参与者语音的至少一个选定部分的持续时间乘以加速系数来确定。

[0174] 根据一些示例，会议记录可以包括被分别记录的来自多个端点的会议参与者语音数据，或者来自对应于多个会议参与者的单个端点的会议参与者语音数据，其可以包括多个会议参与者的每个会议参与者的空间信息。根据一些这样的示例，软件可以包括如下指令，该指令用于在虚拟声学空间中渲染回放音频数据，使得其语音被包括在回放音频数据中的各会议参与者具有各自不同的虚拟会议参与者位置。

[0175] 根据一些实现，选择过程可以涉及主题选择过程。根据一些这样的示例，主题选择过程可以涉及接收会议主题的主题列表并且确定所选择的会议主题的列表。所选择的会议主题的列表可能是会议主题的子集。

[0176] 在一些示例中，软件可以包括用于接收主题排名数据的指令，该主题排名数据可以指示主题列表上的每个会议主题的估计的相关性。确定所选择的会议主题的列表可以至少部分地基于主题排名数据。

[0177] 根据一些实现，选择过程可以涉及讲话突发过滤过程。讲话突发过滤过程例如可以涉及去除输入讲话突发的初始部分。初始部分可以是从输入讲话突发开始时间到输出讲话突发开始时间的时间间隔。在一些实例中，软件可以包括用于至少部分地基于输入讲话突发持续时间来计算输出讲话突发持续时间的指令。

[0178] 根据一些这样的示例，软件可以包括用于确定输出讲话突发持续时间是否超过输出讲话突发时间阈值的指令。如果确定输出讲话突发持续时间超过输出讲话突发时间阈值，讲话突发过滤过程可能涉及对于单个输入讲话突发生成会议参与者语音的多个实例。根据一些这样的示例，会议参与者语音的多个实例中的至少一个可以具有与输入讲话突发结束时间相对应的结束时间。

[0179] 根据一些实现，选择过程可以涉及声学特征选择过程。在一些示例中，声学特征选择过程可以涉及确定至少一个声学特征，例如音调变化、语速和/或响度。

[0180] 在一些实现中，软件可以包括用于修改会议参与者语音的至少一个实例的开始时间或结束时间的指令。在一些示例中，修改过程可以涉及扩展对应于会议参与者语音的实例的时间间隔。根据一些示例，修改过程可以涉及合并对应于单个会议端点的、扩展后在时间上重叠的会议参与者语音的两个或更多个实例。

[0181] 在一些示例中，软件可以包括用于调度先前在时间上不重叠的会议参与者语音的实例以在时间上重叠地回放的指令。作为替代地或者附加地，软件可以包括用于调度先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地回放的指令。

[0182] 根据一些示例，调度可以根据感知激发规则的集合来执行。在一些实现中，感知激发规则的集合可以包括指示单个会议参与者的两个输出讲话突发不应该在时间上重叠的规则。感知激发规则的集合可以包括如下规则，该规则指示如果两个输出讲话突发对应于单个端点，则这两个输出讲话突发在时间上不应该重叠。

[0183] 根据一些实现，给定两个连续的输入讲话突发A和B，A已经在B之前发生，该感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。在一些这样的示例中，T可以大于零。一些实现可以涉及至少部分地基于搜索相关性度量来调度会议参与者语音的实例以供回放。

[0184] 根据一些实现，软件可以包括用于分析音频数据以确定会话动态数据的指令。会话动态数据可以例如包括指示会议参与者语音的频率和持续时间的数据、指示至少两个会议参与者在其期间同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0185] 在一些实例中，软件可以包括用于将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量的指令，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。根据一些示例，软件可以包括用于将优化技术应用于空间优化成本函数以确定局部最优解的指令。根据一些这样的示例，软件可以包括用于至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置的指令。

[0186] 在一些实现中，软件可包括用于控制显示器以提供图形用户界面的指令。根据一些实现，用于控制显示器的指令可以包括用于进行会议参与者的展示的指令。在一些示例中，用于控制显示器的指令可以包括用于进行会议段的展示的指令。

[0187] 在一些示例中，软件可包括如下指令，该指令用于接收对应于用户与图形用户界面的交互的输入，并且至少部分地基于该输入来处理音频数据。在一些示例中，输入可以对应于目标回放持续时间的指示。根据一些实现，软件可以包括用于将回放音频数据提供给扬声器系统的指令。

[0188] 本说明书中描述的主旨的一个或多个实现的细节在附图和下面的描述中被阐述。其他特征、方面和优点将从描述、附图和权利要求中变得显而易见。请注意，以下图形的相对尺寸可能未按比例绘制。

附图说明

[0189] 图1A示出了电话会议系统的组件的示例。

[0190] 图1B是示出能够实现本公开的各个方面的装置的组件的示例的框图。

[0191] 图1C是简述可由图1B的装置执行的方法的一个示例的流程图。

[0192] 图2A示出了电话会议系统的组件的附加示例。

[0193] 图2B示出了分组跟踪文件和会议元数据的示例。

[0194] 图3A是示出能够实现本公开的各个方面的装置的组件的示例的框图。

[0195] 图3B是简述可以由图3A的装置执行的方法的一个示例的流程图。

[0196] 图3C显示了电话会议系统的组件的附加示例。

[0197] 图4示出了上行链路分析模块的组件的示例。

[0198] 图5示出了联合分析模块的组件的示例。

[0199] 图6示出了回放系统和相关设备的组件的示例。

[0200] 图7示出了面对面会议实现的示例。

[0201] 图8是简述根据本公开的一些实现的方法的一个示例的流程图。

[0202] 图9示出了虚拟听众的头部和虚拟声学空间中的混淆锥的示例。

[0203] 图10示出了虚拟声学空间中的初始虚拟会议参与者位置的示例。

[0204] 图11示出了虚拟声学空间中的最终虚拟会议参与者位置的示例。

[0205] 图12是简述根据本公开的一些实施方式的方法的一个示例的流程图。

[0206] 图13是示出调度会议记录在小于输入时间间隔的输出时间间隔期间回放的示例的框图。

[0207] 图14示出了维持重叠的输入讲话突发和重叠的输出talkspur之间的类似的时间关系的示例。

[0208] 图15示出了确定不重叠的输入讲话突发的重叠量的示例。

[0209] 图16是示出应用感知激发规则以避免来自同一端点的输出讲话突发重叠的示例的框图。

[0210] 图17是示出能够调度来自不同会议参与者的全部陈述的并发回放的系统的示例的框图。

[0211] 图18A是简述会议分段方法的一个示例的流程图。

[0212] 图18B示出了用于至少部分地执行本文所述的会议分段方法和相关方法中的一些的系统的示例。

[0213] 图19简述了根据本文公开的一些实现的分段过程的初始阶段。

[0214] 图20简述了根据本文公开的一些实现的分段过程的后续阶段。

[0215] 图21简述了根据本文公开的一些实现的分段过程的后续阶段。

[0216] 图22简述了根据本文公开的一些实现的可由段分类器执行的操作。

[0217] 图23示出了根据本文公开的一些实现的最长段搜索处理的示例。

[0218] 图24是简述本文公开的某些主题分析方法的块的流程图。

[0219] 图25示出了主题分析模块元件的示例。

[0220] 图26示出了输入语音识别格的示例。

[0221] 包括图27A和27B的图27示出了修剪后的小语音识别格的一部分的示例。

[0222] 包括图28A和28B的图28示出了包括用于整个会议记录的词语云的用户界面的示例。

[0223] 包括图29A和29B的图29示出了包括用于多个会议段中的每一个的词语云的用户界面的示例。

[0224] 图30是简述本文公开的一些回放控制方法的块的流程图。

[0225] 图31示出了从词语云选择主题的示例。

[0226] 图32示出了从词语云选择主题和从会议参与者的列表中选择会议参与者两者的示例。

[0227] 图33是简述本文公开的某些主题分析方法的块的流程图。

[0228] 图34是示出搜索系统元件的示例的框图。

[0229] 图35示出了示例回放调度单元,合并单元和回放调度单元功能。

[0230] 图36示出了可以用于实现本公开的一些方面的图形用户界面的示例。

[0231] 图37示出了用于多维会议搜索的图形用户界面的示例。

[0232] 图38A示出了上下文增强语音识别格的示例部分。

[0233] 图38B和38C示出了可以通过使用如图38A所示的上下文增强语音识别格作为输入来生成的关键词检索索引数据结构的示例。

[0234] 图39显示了聚集的上下文特征的示例。

[0235] 图40是示出基于时间的分层索引的示例的框图。

[0236] 图41是示出上下文关键词搜索的示例的框图。

[0237] 图42示出了自上而下的基于时间戳的散列搜索的示例。

[0238] 图43是简述仅选择一部分会议参与者语音以供回放的一些方法的块的框图。

[0239] 图44示出了选择性摘要模块的示例。

[0240] 图45示出了选择性摘要模块的元件的示例。

[0241] 图46示出了用于将选择性摘要方法应用于分段会议的系统的示例。

[0242] 图47示出了根据一些实现的选择器模块的块的示例。

[0243] 图48A和48B示出了根据一些替代实现的选择器模块的块的示例。

[0244] 图49示出了根据其他替代实现的选择器模块的块的示例。

[0245] 各种附图中相同的附图标记和标号表示相同的元件。

具体实施方式

[0246] 以下描述针对出于描述本公开的一些创新方面的目的的某些实现、以及这些创新方面可以在其中实现的上下文的示例。然而,这里的教导可被以各种不同的方式应用。例如,虽然依照电话会议上下文中的音频数据处理的具体示例描述了各种实现,但是本文的教导可广泛地应用于其他已知的音频数据处理上下文,例如处理对应于面对面会议的音频数据。例如,这样的会议可以包括学术和/或专业会议、股票经纪人通话、医生/客户访问、个人日志(例如通过便携式记录设备,例如可穿戴式记录设备)等。

[0247] 此外,所描述的实施例可以在各种硬件,软件,固件等中实现。例如,本申请的各方面可至少部分地体现于装置(电话会议桥和/或服务器,分析系统,回放系统,诸如台式机,膝上型计算机或平板电脑计算机的个人计算机,诸如台式电话,智能电话或其他蜂窝电话的电话,电视机顶盒,数字媒体播放器等)、方法、计算机程序产品、包括多于一个的装置的系统(包括但不限于电话会议系统)等中。因此,本申请的各方面可以采取硬件实施例、软件实施例(包括固件,驻留软件,微代码等)和/或组合软件和硬件方面的实施例的形式。这样的实施例在本文中可以被称为“电路”、“模块”或“引擎”。本申请的一些方面可以采用体现在一个或多个非暂态介质中的计算机程序产品的形式,在该非暂态介质上包含有计算机可读程序代码。这种非暂态介质可以例如包括硬盘,随机存取存储器(RAM),只读存储器(ROM),可擦除可编程只读存储器(EPROM或闪存),便携式光盘只读存储器(CD-ROM),光存储设备,磁存储设备或上述的任何合适的组合。因此,本公开的教导并不预期被局限于图中所

示的和/或在此描述的实施方式，而是具有广泛的适用性。

[0248] 本公开的一些方面涉及对应于诸如电话会议的会议的音频数据的记录、处理和回放。在一些电话会议实现中，当会议的记录被回放时听到的音频体验可能与原始电话会议期间各会议参与者的音频体验显着不同。在一些实现中，所记录的音频数据可包括在电话会议期间不可用的至少一些音频数据。在一些示例中，回放的音频数据的空间和/或时间特性可以与电话会议的参与者听到的音频的空间和/或时间特性不同。

[0249] 图1A示出了电话会议系统的组件的示例。电话会议系统100的组件可以经由硬件、经由存储在非暂态介质上的软件、经由固件和/或通过他们的组合来实现。图1A中所示的组件的类型和数量仅作为示例示出。替代实现可以包括更多、更少和/或不同的组件。

[0250] 在该示例中，电话会议系统100包括电话会议装置200，其能够根据基于分组的协议提供电话会议服务器的功能，该协议在本实现中为VoIP(因特网协议语音)。电话端点1中的至少一些可以包括如下特征，该特征允许会议参与者使用在台式或膝上型计算机、智能电话，专用VoIP电话设备或另一个此类设备上运行的软件应用，以充当通过互联网连接到电话会议服务器的电话客户端。

[0251] 然而，一些电话端点1可以不包括这样的特征。因此，电话会议系统100可以经由PSTN(公共交换电话网络)提供接入，例如以将传统电话流从PSTN转换成VoIP数据分组流的桥的形式。

[0252] 在一些实施方式中，在电话会议期间，电话会议装置200接收来自多个电话端点1的多个单独的上行链路数据分组流7，以及发送去往多个电话端点1的多个单独的下行链路数据分组流8。电话端点1可以包括电话，个人计算机，移动电子设备(例如，蜂窝电话，智能电话，平板电脑等)或其他合适的设备。一些电话端点1可以包括耳机，例如立体声耳机。其他电话端点1可以包括传统的电话耳机。其他电话端点1可以包括可能由多个会议参与者使用的电话会议扬声器电话。因此，从一些这样的电话端点1接收的单独的上行链路数据分组流7可以包括来自多个会议参与者的电话会议音频数据。

[0253] 在该示例中，电话端点之一包括电话会议记录模块2。因此，电话会议记录模块2接收下行链路数据分组流8，但不发送上行链路数据分组流7。尽管在图1A中示出为单独的装置，电话会议记录模块2可以被实现为硬件，软件和/或固件。在一些示例中，电话会议记录模块2可以经由电话会议服务器的硬件，软件和/或固件来实现。然而，电话会议记录模块2仅仅是可选的。电话会议系统100的其他实现不包括电话会议记录模块2。

[0254] 分组网络上的语音传输受到通常称为抖动(jitter)的延时变化的影响。抖动可以例如依照到达时间间隔(IAT)变化或分组延时变化(PDV)来测量。IAT变化可以根据相邻分组的接收时间差来测量。PDV可以例如通过参考相对于数据或“锚”分组接收时间的时间间隔来测量。在基于互联网协议(IP)的网络中，固定延时可归因于由于材料和/或距离而导致的传播延时、处理延时和算法延时，而可变延时可能由IP网络流量的波动、互联网上不同的传输路径等导致的。

[0255] 电话会议服务器通常依赖于“抖动缓冲”来抵消抖动的负面影响。通过引入在接收到音频数据分组的时间与再现分组的时间之间的额外的延时，抖动缓冲器可以将到达分组的不均匀流动转换成分组的更规则流动，使得延时变化不会对于最终用户造成感知音质劣化。然而，语音通信是高度延时敏感的。根据ITU建议G.114，例如，对于正常会话，单向延时

(有时在文中被称为“口到耳延迟时间阈值”)对于正常会话应保持在150毫秒(ms)以下,高于400毫秒被认为是不可接受的。电话会议的典型延迟目标低于150ms,例如100ms或更低。

[0256] 低延迟要求可以对于在不打搅会议参与者的情况下电话会议装置200可以等待预期的上行链路数据分组到达的时间设置上限。对于在电话会议期间再现而言太晚到达的上行链路数据分组将不会被提供给电话端点1或电话会议记录模块2。相反,相应的下行链路数据分组流8将被提供给电话端点1和电话会议记录模块2,其中丢失或迟到的数据分组被丢弃。在本公开的上下文中,“迟到”数据分组是在电话会议期间到达太晚而不被提供给电话端点1或电话会议记录模块2的数据分组。

[0257] 然而,在本文公开的各种实现中,电话会议装置200可以能够记录更完整的上行链路数据分组流7。在一些实现中,电话会议装置200可能能够将迟到数据分组包含在所记录的上行链路数据分组流7中,该迟到数据分组在电话会议的口到耳延迟时间阈值之后被接收到,因此不用于在电话会议期间将音频数据再现给会议参与者。在一些这样的实现中,电话会议装置200能够确定不完整的上行数据分组流的迟到数据分组在迟到分组时间阈值内没有从电话端点接收到。迟到分组时间阈值可以大于或等于电话会议的口到耳延迟时间阈值。例如,在一些实现中,迟到分组时间阈值可以大于或等于200ms,400ms,500ms,1秒或更长。

[0258] 在一些示例中,远程会议装置200可能能够确定不完整上行链路数据分组流的数据分组在丢失分组时间阈值内没有从电话端点接收到,该丢失分组时间阈值大于迟到分组时间阈值。在一些这样的示例中,电话会议装置200可以能够向电话端点发送关于重新发送丢失的数据分组的请求。像迟到数据分组那样,丢失的数据分组也不会被电话会议记录模块2记录。在一些实现中,丢失分组时间阈值可以是数百毫秒甚至几秒,例如5秒,10秒,20秒,30秒等。在一些实现中,丢失分组时间阈值可以是1分钟或更长,例如2分钟,3分钟,4分钟,5分钟等。

[0259] 在该示例中,电话会议装置200能够记录各个上行链路数据分组流7,并将其作为各个上行链路数据分组流提供给会议记录数据库3。会议记录数据库3可以存储在一个或多个存储系统中,取决于特定的实现,该一个或多个存储系统可以与或可以不与电话会议装置200处于相同的位置。因此,在一些实现中,由电话会议装置200记录并存储在会议记录数据库3中的各个上行链路数据分组流可以比在电话会议期间可用的数据分组流更完整。

[0260] 在图1A所示的实现中,分析引擎307能够分析和处理所记录的上行链路数据分组流,以为回放进行准备。在该示例中,来自分析引擎307的分析结果存储在分析结果数据库5中,准备好由回放系统609进行回放。在一些示例中,回放系统609可以包括能够通过网络12(例如,因特网)流送分析结果的回放服务器。在图1A中,回放系统609被示出为将分析结果流送给多个收听站11(每个收听站11可以包括在本地设备上运行的一个或多个回放软件应用程序,例如计算机)。这里,其中一个收听台11包括头戴式受话器607,另一个收听台11包括扬声器阵列608。

[0261] 如上所述,由于延迟问题,回放系统609可以具有比在电话会议期间可用的数据分组更完整的可用于再现的数据分组。在一些实施方式中,在回放系统609再现的电话会议音频数据和可用于电话会议期间再现的电话会议音频数据之间可能存在其他差别和/或额外的差别。例如,电话会议系统通常将上行链路和下行链路数据分组的数据速率限制为可被

网络可靠维护的速率。此外，往往有经济动机来保持数据速率降低，这是因为如果系统的组合数据速率太高，则电话会议服务提供商可能需要提供更昂贵的网络资源。

[0262] 除了数据速率限制之外，还可能对于每秒可由网络组件(例如交换机和路由器)处理的以及还可由软件组件(诸如电话会议服务器的主机操作系统的内核中的TCP/IP栈)可靠地处理的IP分组的数量存在实际约束。这样的约束可能具有对于如何将对应于电话会议音频数据的数据分组流编码并分成IP分组的暗示。

[0263] 电话会议服务器需要足够快速地处理数据分组并执行混合操作等以避免会议参与者的感知质量劣化，并且这通常必须在计算资源的上限下进行。服务于单个会议参与者所需的计算开销越小，则单个服务器设备可以实时处理的会议参与者的数量越大。因此，保持计算开销相对较小为电话会议服务提供商提供了经济利益。

[0264] 大多数电话会议系统是所谓的“无预约”系统。这意味着电话会议服务器不提前“知道”预计会有多少个电话会议同时主持，或者有多少个会议参与者将连接到任何给定的电话会议。在电话会议期间的任何时间，服务器既没有指示有多少额外的会议参与者可能随后加入电话会议，也没有指示当前的会议参与者有多少可能提前离开电话会议。

[0265] 此外，电话会议服务器将通常在电话会议之前不会有会议动态信息，其是关于在电话会议期间预计会发生什么样的人际交互的。例如，将预先不知道一个或多个会议参与者是否会主导会话，以及如果是的话，哪个会议参与者将主导会话。在任何时刻，电话会议服务器必须仅基于在电话会议中直至该时刻为止发生的事情来决定在每个下行链路数据分组流中提供什么音频。

[0266] 然而，当分析引擎307处理存储在会议记录数据库3中的各个上行链路数据分组流时，上述约束通常将不适用。类似地，当回放系统609正在处理和再现已从分析引擎307输出的、来自分析结果数据库5的数据时，上述约束通常将不适用。

[0267] 例如，假设在电话会议完成之后进行分析和回放，则回放系统609和/或分析引擎307可以使用来自整个电话会议记录的信息，以便确定如何最好地处理，混合和/或渲染电话会议的任何时刻以供回放期间进行再现。即使电话会议记录仅对应于电话会议的一部分，对应于该整个部分的数据将可用于确定如何最佳地混合、渲染和以其他方式处理所记录的电话会议音频数据(以及可能的其他数据，例如电话会议元数据)以用于回放期间再现。

[0268] 在许多实现中，回放系统609可以向不试图与电话会议中的那些人进行交互的听众提供音频数据等。因此，回放系统609和/或分析引擎307可以具有在其中分析和/或处理记录的电话会议音频数据并使得电话会议可用于回放的数秒、数分、数小时、数天或甚至更长的时间段。这意味着分析引擎307和/或回放系统609可以使用计算量大且/或数据大的算法，该算法在可用硬件上仅可能执行得比实时慢。由于这些轻松的时间约束，一些实施可能涉及将用于分析的电话会议记录进行排队，以便在资源允许时对它们进行分析(例如，当先前记录的电话会议的分析完成时，或在电力或云计算资源更便宜或更容易获得的“非高峰”时段)。

[0269] 假设在电话会议完成之后进行分析和回放，分析引擎307和回放系统609可以访问一组完整的电话会议参与信息，例如关于哪些会议参与者参与电话会议以及每个会议参加者加入和离开电话会议的时间的信息。类似地，假设在电话会议完成之后进行分析和回放，

分析引擎307和回放系统609可以访问一组完整的电话会议音频数据和任何相关联的元数据，从该元数据确定(或至少估计)每个参加者何时发言。这个任务在这里可以被称为“发言者日志”。基于发言者日志信息，分析引擎307可以确定会话动态数据，诸如哪个(哪些)会议参与者发言最多，谁与谁交谈、谁打断谁、在电话会议期间发生多少双讲话(在其期间至少两个会议参与者同时发言的时间)，以及分析引擎307和/或回放系统609可以用于确定如何最佳地在回放期间混合和渲染会议的潜在的其他有用信息。即使电话会议记录仅对应于电话会议的一部分，对应于该整个部分的数据将仍可用于确定电话会议参与信息、会话动态数据等。

[0270] 本公开包括如下方法和设备，其用于记录，分析和回放电话会议音频数据，使得在回放期间渲染的电话会议音频数据可能与在原始电话会议期间会议参与者听到的和/或在原始电话会议期间由记录设备(诸如图1A所示的电话会议记录设备2)记录的明显不同。本文公开的各种实现利用了实时电话会议和回放使用情况之间的上述约束差异中的一个或多个，以便在回放期间产生更好的用户体验。在不丧失一般性的情况下，现在讨论用于记录、分析和回放电话会议音频数据以使得回放可以有利地与原始电话会议体验不同的数个具体实现和特定方法。

[0271] 图1B是示出能够实现本公开的各个方面的装置的组件的示例的框图。图1B中所示的组件的类型和数量仅仅是作为示例被示出的。替代性实现可以包括更多、更少和/或不同的组件。装置10可以例如是电话会议装置200的实例。在一些示例中，设备10可以是另一设备的组件。例如，在一些实现中，装置10可以是电话会议装置200的组件，例如线卡。

[0272] 在该示例中，装置10包括接口系统105和控制系统110。接口系统105可以包括一个或多个网络接口，控制系统110和存储系统之间的一个或多个接口、和/或一个或多个外部设备接口(诸如，一个或多个通用串行总线(USB)接口))。控制系统110可以例如包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑，和/或离散硬件组件。在一些实现中，控制系统110可能能够提供电话会议服务器功能。

[0273] 图1C是简述可由图1B的装置执行的方法的一个示例的流程图。如本文所述的其他方法那样，方法150的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0274] 在该实现中，块155涉及经由接口系统接收电话会议期间的电话会议音频数据。例如，在块155中，电话会议音频数据可以经由接口系统105被控制系统110接收。在该示例中，电话会议音频数据包括多个单独的上行链路数据分组流，诸如图1A所示的上行链路数据分组流7。因此，每个上行链路数据分组流对应于由一个或多个会议参与者使用的电话端点。

[0275] 在该示例中，块160涉及经由接口系统将电话会议音频数据作为单独的上行链路数据分组流发送到存储系统。因此，不是被记录为作为如图1A所示的下行链路数据分组流8之一被接收的混合音频数据，诸如由电话会议记录设备2记录的下行链路数据分组流8，经由每个上行链路数据分组流7接收的分组被记录和存储为单独的上行链路数据分组流。

[0276] 然而，在一些示例中，上行链路数据分组流中的至少一个可以对应于多个会议参与者。例如，块155可以涉及从由多个会议参与者使用的空间扬声器电话接收这样的上行链路数据分组流。因此，在某些实例中，对应的上行链路数据分组流可以包括关于多个参与者

中的每一个的空间信息。

[0277] 在一些实现中，在块155中接收的单独的上行链路数据分组流可以是单独的编码的上行链路数据分组流。在这样的实现中，块160可以包括将电话会议音频数据作为单独的编码的上行链路数据分组流发送到存储系统。

[0278] 如上所述，在一些示例中，接口系统105可以包括网络接口。在一些这样的示例中，块160可以包括经由网络接口将电话会议音频数据发送到另一设备的存储系统。然而，在一些实现中，装置10可以包括存储系统的至少一部分。接口系统105可以包括存储系统的至少一部分与控制系统之间的接口。在一些这样的实现中，块160可以涉及将电话会议音频数据发送到装置10的存储系统。

[0279] 至少部分地由于上述的电话会议延迟问题，上行链路数据分组流中的至少一个可以包括在电话会议的口到耳延迟时间阈值之后被接收到、因此不被用于在电话会议期间再现音频数据的至少一个数据分组。口到耳延迟时间阈值可能因实现而不同，但是在许多实现中，口到耳延迟时间阈值可以是150ms或更短。在一些示例中，口到耳延迟时间阈值可以大于或等于100ms。

[0280] 在一些实现中，控制系统110可以能够确定不完整的上行链路数据分组流的迟到数据分组在迟到分组时间阈值内没有从电话端点接收到。在一些实现中，迟到分组时间阈值可以大于或等于电话会议的口到耳延迟时间阈值。例如，在一些实现中，迟到分组时间阈值可以大于或等于200ms，400ms，500ms，1秒或更长。在一些示例中，控制系统110可能能够确定不完整上行链路数据分组流的数据分组在大于迟到分组时间阈值的丢失分组时间阈值内没有从电话端点接收到。在一些实现中，控制系统110可以能够经由接口系统105向电话端点发送请求以请求重新发送丢失的数据分组。控制系统110可以能够接收丢失的数据分组，并且将丢失的数据分组添加到不完整的上行链路数据分组流。

[0281] 图2示出了电话会议系统的组件的附加示例。图2所示的组件的类型和数量仅仅是作为示例被示出的。替代性实现可以包括更多、更少和/或不同的组件。在该示例中，电话会议装置200包括VoIP电话会议桥。在该示例中，会议参与者正在使用五个电话端点，包括两个耳机端点206，空间扬声器电话端点207和两个PSTN端点208。空间扬声器端点207可能能够提供对应于多个会议参与者中的每一个的位置的空间信息。这里，PSTN桥209在IP网络和PSTN端点208之间形成网关，将PSTN信号转换成IP数据分组流，反之亦然。

[0282] 图2A示出了电话会议系统的组件的附加示例。图2A所示的组件的类型和数量仅仅是作为示例被示出的。替代性实现可以包括更多、更少和/或不同的组件。在该示例中，电话会议装置200包括VoIP电话会议桥。在该示例中，会议参与者正在使用五个电话端点，包括两个耳机端点206，空间扬声器电话端点207和两个PSTN端点208。空间扬声器端点207可能能够提供对应于多个会议参与者中的每一个的位置的空间信息。这里，PSTN桥209在IP网络和PSTN端点208之间形成网关，将PSTN信号转换成IP数据分组流，反之亦然。

[0283] 在图2A中，电话会议装置200正在接收各对应于五个电话端点之一的上行链路数据分组流201A-205A。在一些实例中，可能有多个会议参与者通过空间扬声器终端207参与电话会议。如果是这样的话，则上行链路数据分组流203A可以包括多个会议参与者中的每一个的音频数据和空间信息。

[0284] 在一些实现中，上行链路数据分组流201A-205A中的每一个可以包括每个数据分

组的序列号、以及数据分组有效载荷。在一些示例中，上行链路数据分组流201A-205A中的每一个可以包括与包括在上行链路数据分组流中的每个讲话突发相对应的讲话突发数。例如，每个电话终端(或与电话端点相关联的设备，诸如PSTN桥209)可以包括能够检测语音和非语音的实例的语音活动性检测器。电话端点或相关联设备可以将讲话突发数包含在与这种语音实例相对应的上行链路数据分组流的一个或多个数据分组中，并且每当语音活动检测器确定语音已经在非语音时段之后重新开始时，可增加讲话突发数。在一些实现中，讲话突发数可以是在每个讲话突发开始时在1和0之间切换的单个比特。

[0285] 在该示例中，电话会议装置200为每个接收到的上行链路数据分组分配“接收”时间戳。这里，电话会议装置200向会议记录数据库3发送分组跟踪文件201B-205B，分组跟踪文件201B-205B之一与上行数据分组流201A-205A之一对应。在此实现中，分组跟踪文件201B-205B包括针对每个接收到的上行链路数据分组的接收时间戳，以及所接收的序列号，讲话突发数和数据分组有效载荷。

[0286] 在该示例中，电话会议装置200还向会议记录数据库3发送会议元数据210。会议元数据210可以例如包括关于各个会议参与者的数据，例如会议参与者姓名，会议参与者位置等。会议元数据210可以指示各个会议参与者与分组跟踪文件201B-205B之一之间的关联。在一些实现中，分组跟踪文件201B-205B和会议元数据210可以共同在会议记录数据库3中形成电话会议记录。

[0287] 图2B示出了分组跟踪文件和会议元数据的示例。在该示例中，会议元数据210和分组跟踪文件201B-204B具有被表示为包括四列(这里也被称为字段)的表的数据结构。图2B中所示的特定数据结构仅仅是作为示例；其他示例可以包括更多或更少的字段。如本文其他地方所描述的，在一些实现中，会议元数据210可以包括在图2B中未示出的其他类型的信息。

[0288] 在该示例中，会议元数据210数据结构包括会议参与者姓名字段212、连接时间字段214(指示相应的会议参与者何时加入会议)、断开时间字段216(指示相应的会议参与者何时离开会议)、和分组跟踪文件字段218。在该示例中可以看出，在会议元数据210数据结构中可以多次列出同一会议参与者，每次他或她加入或重新加入会议就列出一次。分组跟踪文件字段218包括用于识别对应的分组跟踪文件的信息。

[0289] 因此，会议元数据210提供了会议的一些事件的总结，包括谁参与、多长时间等等。在一些实施方式中，会议元数据210可以包括诸如端点类型(例如耳机，移动设备，扬声器电话等)的其他信息。

[0290] 在该示例中，分组跟踪文件201B-204B中的每一个还包括四个字段，每个字段对应于不同类型的信息。这里，分组跟踪文件201B-204B中的每一个包括接收时间字段222，序列号字段224，讲话突发标识字段226和有效载荷数据字段228。可以包括在分组有效载荷中的序列号和讲话突发数使得能够以正确的顺序排列有效载荷。在该示例中，由有效载荷数据字段228指示的有效载荷数据的每个实例对应于已经去除了序列号和讲话突发数之后的分组的有效载荷的剩余部分，包括对应于相应的会议参与者的音频数据。例如，分组跟踪文件201B-204B中的每一个可以包含源自诸如图2A所示的那些的端点的分组的有效载荷数据。一个分组跟踪文件可以包括来自大量分组的有效载荷数据。

[0291] 尽管图2B中未示出，会议元数据210对应于特定的会议。因此，用于会议的元数据

和分组跟踪文件201B-204B(包括有效载荷数据)可以被存储,以用于根据例如会议代码进行的后续检索。

[0292] 随着更多信息的添加,分组跟踪文件201B-204B和会议元数据210可以在会议的持续时间内改变。根据一些实现,这种改变可能在本地发生,并且最终的分组跟踪文件和会议元数据210在会议结束之后被发送到会议记录数据库3。作为替代地或者附加地,分组跟踪文件201B-204B和/或会议元数据210可以在会议记录数据库3上被创建然后被更新。

[0293] 图3A是示出能够实现本公开的各个方面的装置的组件的示例的框图。图3A中所示的组件的类型和数量仅仅是作为示例被示出的。替代性实现可以包括更多、更少和/或不同的组件。装置300可以例如是分析引擎307的实例。在一些示例中,装置300可以是另一设备的组件。例如,在一些实现中,装置300可以是分析引擎307的组件,例如本文别处描述的上行链路分析模块。

[0294] 在该示例中,装置300包括接口系统325和控制系统330。接口系统325可以包括一个或多个网络接口,控制系统330和存储系统之间的一个或多个接口、和/或一个或多个外部设备接口(诸如,一个或多个通用串行总线(USB)接口))。控制系统330可以例如包括通用单芯片或多芯片处理器,数字信号处理器(DSP),专用集成电路(ASIC),现场可编程门阵列(FPGA)或其他可编程逻辑器件,离散门或晶体管逻辑,和/或离散硬件组件。

[0295] 图3B是简述可以由图3A的装置执行的方法的一个示例的流程图。如本文所述的其它方法那样,方法350的块不一定按照所示的顺序执行。此外,这样的方法可以包括比所示的和/或描述的块更多或更少的块。

[0296] 在该实现中,块355涉及经由接口系统接收电话会议的先前存储的音频数据(这里也称为记录的音频数据)。例如,在块355中,记录的音频数据可以由控制系统330经由接口系统325接收。在该示例中,记录的音频数据包括与由一个或多个会议参与者使用的电话端点对应的至少一个单独的上行链路数据分组流。

[0297] 这里,所接收的单独的上行链路数据分组流包括对应于该单独的上行数据分组流的数据分组的时间戳数据。如上所述,在一些实现中,电话会议装置200可以向每个接收的上行链路数据分组分配接收时间戳。电话会议装置200可以以电话会议服务器200接收的顺序存储被加时间戳的数据分组,或者使得数据分组被这样存储。因此,在一些实现中,块355可以涉及从会议记录数据库3(诸如上述图1A所示的)接收记录的音频数据,包括包含时间戳数据的单独的上行链路数据分组流。

[0298] 在该示例中,块360涉及分析单独的上行链路数据分组流中的数据分组的时间戳数据。这里,块360的分析处理涉及确定该单独的上行链路数据分组流是否包括至少一个无序数据分组。在该实现中,如果该单独的上行链路数据分组流包括至少一个无序数据分组,则在块365中将根据时间戳数据将该单独的上行链路数据分组流重新排序。

[0299] 在一些实现中,单独的上行链路数据分组流的至少一个数据分组可能在电话会议的口到耳延迟时间阈值之后被接收到。如果是这样,则单独的上行链路数据分组流包括不可用于包含在用于再现给会议参与者或用于在电话端点处记录的下行链路数据分组流中的数据分组。根据具体情况,口到耳延迟时间阈值之后接收到的数据分组可能被无序地接收或可能不被无序地接收。

[0300] 图3A的控制系统330可以具有各种其他功能。例如,控制系统330可能能够经由接

口系统325接收电话会议元数据，并至少部分地基于电话会议元数据来索引单独的上行链路数据分组流。

[0301] 由控制系统330接收的记录的音频数据可以包括多个单独的编码的上行链路数据分组流，每个单独的编码的上行链路数据分组流对应于由一个或多个会议参与者使用的电话端点。在一些实现中，如下面更详细描述的，控制系统330可以包括能够分析多个单独的上行链路数据分组流的联合分析模块。联合分析模块可以能够确定会话动态数据，诸如指示会议参与者语音的频率和持续时间的数据、指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据、和/或指示会议参与者会话的实例的数据。

[0302] 控制系统330可以能够解码多个单独的编码的上行链路数据分组流中的每一个。在一些实现中，控制系统330可以能够向语音识别模块提供一个或多个解码的上行链路数据分组流，该语音识别模块能够识别语音并产生语音识别结果数据。语音识别模块可以能够将语音识别结果数据提供给联合分析模块。在一些实现中，联合分析模块可以能够标识语音识别结果数据中的关键词并且对关键词位置进行索引。

[0303] 在一些实现中，控制系统330可以能够向发言者日志模块提供一个或多个解码的上行链路数据分组流。发言者日志模块可以能够标识单独的解码的上行链路数据分组流中的多个会议参与者中的每一个的语音。发言者日志模块可以能够生成指示多个会议参与者中的每一个正在发言的时间的发言者日志，并且向联合分析模块提供发言者日志。在一些实现中，控制系统330可以能够向联合分析模块提供多个单独的解码的上行链路数据分组流。

[0304] 图3C显示了电话会议系统的组件的附加示例。图3C所示的组件的类型和数量仅仅是示例性的。替代实现可以包括更多，更少和/或不同的组件。在该实现中，分析引擎307接收来自会议记录数据库3的各种文件以及来自会议数据库308的信息。分析引擎307及其组件可以经由硬件、经由存储在非暂态介质上的软件、经由固件和/或它们的组合来实现。来自会议数据库308的信息可以例如包括关于哪些会议记录存在的信息、关于谁有权听取和/或修改每个会议记录的信息、关于哪些会议被调度的信息、和/或关于谁被各会议邀请的信息等等。

[0305] 在该示例中，分析引擎307接收来自会议记录数据库3的分组跟踪文件201B-205B，它们中的每一个对应于先前已被电话会议装置200接收的上行链路数据分组流201A-205A之一。分组跟踪文件201B-205B可以例如包括用于每个接收到的上行链路数据分组的接收时间戳、以及接收的序列号，讲话突发数和数据分组有效载荷。在该示例中，将分组跟踪文件201B-205B中的每一个提供给上行链路分析模块301-305中的单独一个以供处理。在一些实现中，上行链路分析模块301-305可以能够将分组跟踪文件的数据分组进行重新排序，例如，如上文参考图3B所述。下面参考图4描述上行链路分析模块功能的一些附加示例。

[0306] 在该示例中，上行链路分析模块301-305中的每一个输出各上行链路分析结果301C-305C中的相应一个。在一些实现中，各上行链路分析结果301C-305C可被回放系统609用于回放和可视化。下面参考图6描述一些示例。

[0307] 这里，上行链路分析模块301-305中的每个还向联合分析模块306提供输出。联合分析模块306可以能够分析与多个单独的上行链路数据分组流相对应的数据。

[0308] 在一些示例中，联合分析模块306可以能够分析会话动态，并确定会话动态数据。

以下参考图5更详细地描述联合分析模块功能的这些和其它示例。

[0309] 在该示例中，联合分析模块306输出会议概述信息311，其可以包括会议的时间，参与者的姓名等。在一些实现中，会议概述信息311可以包括会话动态数据。这里，联合分析模块306还输出分段和词语云数据309和搜索索引310，这两者都在下文参照图5被描述。

[0310] 这里，分析引擎307也接收会议元数据210.如本文其他地方所述，会议元数据210可以包括关于各会议参与者的数据，例如会议参与者姓名和/或会议参与者位置、各会议参与者和分组跟踪文件201B-205B之一之间的关联等。在该示例中，会议元数据210被提供给联合分析模块306。

[0311] 图4示出了上行链路分析模块的组件的示例。上行链路分析模块301及其组件可以经由硬件、经由存储在非暂态介质上的软件、经由固件和/或它们的组合来实现。图4中所示的组件的类型和数量仅仅是通过示例的方式示出的。替代实现可以包括更多、更少和/或不同的组件。

[0312] 在该实现中，上行链路分析模块301被示出为接收分组跟踪文件201B。这里，对应于单独的上行数据分组流的分组跟踪文件201B被分组流归一化模块402接收和处理。在该示例中，分组流归一化模块402能够分析分组跟踪文件201B中的数据分组的序列号数据，并且确定该单独的上行链路数据分组流是否包括至少一个无序数据分组。如果分组流归一化模块402确定单独的上行链路数据分组流包括至少一个无序数据分组，则在该示例中，分组流归一化模块402将根据序列号将该单独的上行链路数据分组进行重新排序。

[0313] 在该实现中，分组流归一化模块402输出有序回放流40B作为由上行链路分析模块301输出的上行链路分析结果301C的一个分量。在一些实现中，分组流归一化模块402可以包括对应于有序回放流401B的每个数据分组的回放时间戳和数据分组有效载荷。这里，有序回放流40B包括编码数据，但是在替换实现中，有序回放流40B可以包括解码数据或代码转换数据。在该示例中，分组流索引模块403输出的回放流索引401A是上行分析结果301C的另一分量。回放流索引401A可以有助于回放系统609的随机访问回放。

[0314] 例如，分组流索引模块403可以确定会议参与者的讲话突发的实例(例如，根据输入上行链路分组跟踪的讲话突发数)，并且在回放流索引401A中包括相应的索引信息，以便于有助于回放系统609对会议参与者讲话突发的随机访问回放。在一些实现中，分组流索引模块403可能能够根据时间进行索引。例如，在一些示例中，分组流索引模块403可能能够形成分组流索引，该分组流索引指示对于对应回放时间的编码音频的回放流内的字节偏移量。在一些这样的实现中，在回放期间，回放系统609可以查找分组流索引中的特定时间(例如，根据时间粒度，例如10秒粒度)，并且分组流索引可以指示对于该回放时间的编码音频的回放流内的字节偏移量。这可能是有用的，因为编码音频可能具有可变的比特率，或者因为当静默时可能不存在分组(所谓的“DTX”或“不连续传输”)。在任一情况中，分组流索引可以有助于回放过程期间的快速寻找，至少部分是因为在时间与回放流内的字节偏移量之间常常可能存在非线性关系。

[0315] 在图4所示的示例中，解码模块404还从分组流归一化模块402接收有序回放流40B1。在该实现中，解码模块404对编码的有序回放流401B进行解码，并且将解码的回放流提供给自动语音识别模块405，可视化分析模块406和发言者日志模块407。在一些示例中，解码的回放流可以是脉冲编码调制(PCM)流。

[0316] 根据一些实现，解码模块404和/或回放系统609可以应用与在原始电话会议期间使用的解码过程不同的解码过程。由于时间，计算和/或带宽约束，相同的音频分组可能在电话会议期间以具有最小计算需求的低保真度解码，而由解码模块404以具有更高计算需求的更高保真度解码。解码模块404的更高保真度的解码例如可能涉及解码到较高采样率，开启谱带宽复制(SBR)以获得更好的感知结果，运行迭代解码过程的更多迭代等。

[0317] 在图4所示的示例中，自动语音识别模块405分析由解码模块404提供的解码回放流中的音频数据，以确定与解码回放流对应的电话会议部分中说出的词语。自动语音识别模块405将语音识别结果401F输出到联合分析模块306。

[0318] 在该示例中，可视化分析模块406分析解码回放流中的音频数据，以确定讲话突发的发生，讲话突发的幅度和/或讲话突发的频率内容等，并输出可视化数据401D。例如，可视化数据401D可以提供关于当电话会议被回放时回放系统609可以显示的波形的信息。

[0319] 在该实现中，发言者日志模块407根据单个会议参与者还是多个会议参与者正使用与输入上行链路分组跟踪201B对应的同一电话端点，分析解码回放流中的音频数据，以标识并记录来自一个或多个会议参与者的语音的出现。发言者日志模块407输出发言者日志401E，其与可视化数据401D一起被包括作为由分析引擎307输出的上行链路分析结果301C(参见图3C)的一部分。实质上，发言者日志401E指示哪个(哪些)会议参与者发言，会议参与者何时讲话。

[0320] 上行链路分析结果301C以及语音识别结果401F一起被包括在提供给联合分析模块306的可用于联合分析的上行链路分析结果401中。多个上行链路分析模块中的每一个可以将可用于联合分析的上行链路分析结果的实例输出到联合分析模块306。

[0321] 图5示出了联合分析模块的组件的示例。联合分析模块306及其组件可以经由硬件、经由存储在非暂态介质上的软件、经由固件和/或它们的组合来实现。图5所示的组件的类型和数量仅仅是作为示例被示出的。替代性实现可以包括更多，更少和/或不同的组件。

[0322] 在该示例中，图3C所示的上行链路分析模块301-305中的每一个输出可用于联合分析的上行链路分析结果401-405中的相应一个，所有这些在图5中被示出为由联合分析模块306接收到。在该实现中，其中语音识别结果401F-405F被提供给关键词检索和索引模块505以及主题分析模块525，语音识别结果401F-405F分别来自可用于联合分析的上行链路分析结果401-405中的每一个。在该示例中，语音识别结果401F-405F对应于特定电话会议的所有会议参与者。语音识别结果401F-405F可以例如是文本文件。

[0323] 在该示例中，关键词检索和索引模块505能够分析语音识别结果401F-405F，标识在电话会议期间由所有会议参与者说出的频繁出现的词语，以及并且对频繁出现的词语的出现建立索引。在一些实现中，关键词检索和索引模块505可以确定并记录每个关键词的实例的数量。在该示例中，关键词检索和索引模块505输出搜索索引310。

[0324] 在图5所示的示例中，会话动态分析模块510接收发言者日志401E-405E，发言者日志401E-405E分别来自可用于联合分析的上行链路分析结果401-405中的每一个。会话动态分析模块510可以能够确定会话动态数据，诸如指示会议参与者语音的频率和持续时间的数据、指示在其期间至少两个会议参与者同时发言的会议参与者“双讲话”的实例的数据、指示会议参与者会话的实例的数据、和/或指示一个会议参与者打断一个或多个其他会议参与者的实例的数据等。

[0325] 在该示例中，会话动态分析模块510输出会话动态数据文件515a-515d，每个会话动态数据文件对应于不同的时间尺度。例如，会话动态数据文件515a可以对应于会议段(陈述，讨论等)为大约1分钟长的时间尺度，会话动态数据文件515b可以对应于会议段为大约3分钟长的时间尺度，会话动态数据文件515c可以对应于会议段为大约5分钟长的时间尺度，以及会话动态数据文件515d可以对应于会议段为大约7分钟长或者更长时间的时间尺度。在其他实现中，会话动态分析模块510可以输出更多或更少的会话动态数据文件515。在该示例中，会话动态数据文件515a-515d仅输出到主题分析模块525，但是在其他实现中，会话动态数据文件515a-515d可以被输出到一个或多个其他模块，和/或从整个分析引擎307输出。因此，在一些实施方式中，会话动态数据文件515a-515d可供回放系统609使用。

[0326] 在一些实现中，主题分析模块525可以能够分析语音识别结果401F-405F并且标识可能的会议主题。在一些示例中，如这里，主题分析模块525可以接收和处理会议元数据210。下面详细描述主题分析模块525的各种实施方式。在该示例中，主题分析模块525输出段和词语云数据309，其可以包括用于多个会话段中的每一个的主题信息和/或用于多个时间间隔中的每一个的主题信息。

[0327] 在图5所示的示例中，联合分析模块包括概述(overview)模块520。在该实现中，概述模块520接收会议元数据210以及来自会议数据库308的数据。会议元数据210可以包括关于各会议参与者的数据，例如会议参与者姓名和会议参与者位置，指示会议的时间和日期的数据等。会议元数据210可以指示各会议参与者和电话端点之间的关联。例如，会议元数据210可以指示各会议参与者和由分析引擎输出的分析结果301C-305C之一(参见图3C)之间的关联。会议数据库308可以向概述模块520提供关于哪些会议被调度的数据，关于会议主题的数据和/或谁被邀请参加每个会议的数据等。在该示例中，概述模块520输出会议概述信息311，其可以包括会议元数据210的总结和来自会议数据库308的数据的总结。

[0328] 在一些实现中，分析引擎307和/或电话会议系统100的其他组件可以能够具有其他功能。例如，在一些实现中，分析引擎307，回放系统609或电话会议系统100的另一组件可能能够至少部分地基于会话动态数据来在虚拟声学空间中分配虚拟会议参与者位置。在一些示例中，会话动态数据可以基于整个会议。

[0329] 图6示出了回放系统的组件和相关设备的示例。回放系统609及其组件可以经由硬件、经由存储在非暂态介质上的软件、经由固件和/或它们的组合来实现。图6中所示的组件的类型和数量仅仅作为示例被示出。替代性实现可以包括更多，更少和/或不同的组件。

[0330] 在该示例中，回放系统609正在接收与包括三个电话端点的电话会议相对应的数据，而不是如上所述包括五个电话端点的电话会议。因此，回放系统609被示出为接收分析结果301C-303C、以及段和词语云数据309、搜索索引310和会议概述信息311。

[0331] 在该实现中，回放系统609包括多个解码单元601A-603A。这里，解码单元601A-603A接收分别来自分析结果301C-303C中的每一个的有序回放流401B-403B。在一些示例中，回放系统609可以每个回放流调用一个解码单元，因此解码单元的数量可以根据所接收的回放流的数量而改变。

[0332] 根据一些实现，解码单元601A-603A可以应用与在原始电话会议期间使用的解码过程不同的解码过程。如本文其他地方所述，在原始电话会议期间，由于时间，计算和/或带宽约束，音频数据可能以具有最小计算需求的低保真度解码。但是，有序回放流401B-430B

可由解码模块601A-603A以具有更高计算需求的更高保真度解码。解码模块601A-603A的更高保真度的解码例如可能涉及解码到较高采样率，开启谱带宽复制(SBR)以获得更好的感知结果，运行迭代解码过程的更多迭代等。

[0333] 在该示例中，解码单元601A-603A中的每一个将解码的回放流提供给后处理模块601B-603B中的相应的一个。如下面更详细地讨论的，在一些实现中，后处理模块601B-603B可以能够进行一种或多种类型的处理，以加速有序回放流401B-403B的回放。在一些这样的示例中，后处理模块601B-603B可能能够从有序回放流401B-403B去除静默部分，使有序回放流401B-403B的先前未重叠的部分重叠，改变有序回放流401B-403B的先前重叠部分的重叠量，和/或用于加速有序回放流401B-403B的回放的其他处理。

[0334] 在该实现中，混合和渲染模块604接收来自后处理模块601B-603B的输出。这里，混合和渲染模块604能够混合从后处理模块601B-603B接收到的各个回放流，并且渲染所得到的回放音频数据以供通过诸如耳机607和/或扬声器阵列608的扬声器系统再现。在一些示例中，混合和渲染模块604可将回放音频数据直接提供给扬声器系统，而在其它实施方式中，混合和渲染模块604可以将回放音频数据提供给可能能够与扬声器系统通信的另一设备(诸如显示设备610)。在一些实现中，混合和渲染模块604可以根据由分析引擎307确定的空间信息来渲染混合音频数据。例如，混合和渲染模块604可以能够基于这样的空间信息将每个会议参与者的混合音频数据渲染至虚拟声学空间中被分配的虚拟会议参与者位置。在一些替代实现中，混合和渲染模块604还可能能够确定这样的空间信息。在一些实例中，混合和渲染模块604可以根据与在原始电话会议期间渲染所使用的空间参数不同的空间参数来渲染电话会议音频数据。

[0335] 在一些实现中，回放系统609的一些功能可以至少部分地根据“基于云的”系统来提供。例如，在一些实现中，回放系统609可能能够经由网络与一个或多个其他设备(诸如一个或多个服务器)进行通信。在图6所示的示例中，回放系统609被示出为经由一个或多个网络接口(未示出)与可选的回放控制服务器650和可选的再现服务器660进行通信。根据一些这样的实现，在其他实现中可以由混合和渲染模块604执行的功能中的至少一些可以由渲染服务器660执行。类似地，在一些实现中，在其他实现中可由回放控制模块605执行的功能中的至少一些可以由回放控制服务器650执行。在一些实现中，解码单元601A-603A和/或后处理模块601B-603B的功能可以由一个或多个服务器执行。根据一些示例，整个回放系统609的功能可以由一个或多个服务器来实现。结果可以被提供给诸如显示设备610的客户端设备以用于回放。

[0336] 在该示例中，回放控制模块605接收分别来自分析结果301C-303C中的每一个的回放流索引401A-403A。尽管在图6中未示出，但是回放控制模块605还可以接收来自分析结果301C-303C的其他信息，以及段和词语云数据309，搜索索引310和会议概述信息311。至少部分地，回放控制模块605可以至少部分地基于用户输入(在本例中可以经由显示设备610被接收)、基于分析结果301C-303C、基于段和词语云数据309、搜索索引310和/或基于会议概述信息311，控制回放过程(包括来自混合和渲染模块604的音频数据的再现)。

[0337] 在该示例中，显示设备610被示出为提供图形用户界面606，其可以用于与回放控制模块605进行交互以控制音频数据的回放。显示设备610可以例如是膝上型计算机，平板计算机，智能电话或其他类型的设备。在一些实现中，用户可能能够经由显示设备610的用

户界面系统与图形用户界面606交互，例如通过触摸覆盖触摸屏，经由通过相关联的键盘和/或鼠标进行交互，通过经由麦克风和显示设备610的相关软件的语音命令等。

[0338] 在图6所示的示例中，图形用户界面606的每一行615对应于特定的会议参与者。在该实现中，图形用户界面606指示会议参与者信息620，其可以包括会议参与者姓名，会议参与者位置，会议参与者照片等。在该示例中，对应于每个会议参与者的语音的实例的波形625也被在图形用户界面606示出。显示设备610可以例如根据来自回放控制模块605的指令来显示波形625。这样的指令可以例如基于包括在分析结果301C-303C中的可视化数据410D-403D。在一些示例中，用户可能能够根据要展示的会议的期望时间间隔来改变图形用户界面606的比例。例如，用户可能能够“放大”或扩大图形用户界面606的至少一部分以显示较小的时间间隔，或者“缩小”图形用户界面606的至少一部分以显示较大的时间间隔。根据一些这样的示例，回放控制模块605可以访问对应于改变的时间间隔的会话动态数据文件515的不同实例。

[0339] 在一些实现中，用户可能不仅能够根据诸如暂停，播放等的典型命令来控制音频数据的再现，而且还可以根据基于更丰富的关联数据和元数据集合的附加能力来控制音频数据的再现。例如，在一些实现中，用户可能能够选择仅回放所选会议参与者的语音。在一些示例中，用户可以选择仅回放会议的正在讨论特定关键词和/或特定主题的那些部分。

[0340] 在一些实现中，图形用户界面606可以至少部分地基于段和词语云数据309来显示一个或多个词语云(word cloud)。在一些实现中，所显示的词云可以至少部分地基于用户输入和/或基于在特定时间正在回放的会议的特定部分。本文公开了各种示例。

[0341] 虽然以上主要在电话会议上下文中描述了音频数据处理的各种示例，但是本公开可更广泛地应用于其他已知的音频数据处理上下文，例如处理对应于面对面会议的音频数据。这样的面对面会议可以例如包括学术和/或专业会议、医生/客户访问、个人日志(例如通过便携式记录设备，例如可穿戴式记录设备)等。

[0342] 图7示出了面对面会议实现的示例。图7所示的组件的类型和数量仅作为示例示出。替代性实现可以包括更多，更少和/或不同的组件。在该示例中，会议地点700包括会议参与者桌子705和听众座位区域710。在该实现中，麦克风715a-715d位于会议参与者桌子705上。因此，会议参与者桌子705被设置为使得四个会议参与者中的每一个将具有他或她的单独的麦克风。

[0343] 在该实现中，线缆712a-712d中的每一个将单独的音频数据流从麦克风715a-715d中的相应的一个传送到在此实例中位于会议参与者桌子705下方的记录设备720。在替代示例中，麦克风715a-715d可以经由无线接口与记录设备720通信，使得不需要线缆712a-712d。会议地点700的一些实现可以包括用于听众座位区域710和/或用于在听众座位区域710和会议参与者桌子705之间的区域的额外的麦克风715，麦克风715可以是也可以不是无线麦克风。

[0344] 在该示例中，记录设备720不混合各个音频数据流，而是分别记录每个单独的音频数据流。在一些实现中，麦克风715a-715d中的每一个或记录设备720可以包括模数转换器，使得来自麦克风715a-715d的音频数据流可以由记录设备720记录为单独的数字音频数据流。

[0345] 麦克风715a-715d有时可能被称为“端点”的示例，因为它们类似于上文在电话会

议上下文中讨论的电话端点。因此，图7所示的实现提供了另一个示例，其中在本示例中由麦克风715a-715d表示的多个端点中的每个端点的音频数据将被单独记录。

[0346] 在替代实现中，会议参与者桌子705可以包括诸如声场麦克风的麦克风阵列。声场麦克风可以例如能够产生A格式或B格式的高保真环绕声信号(例如Core Sound TetraMic™)，Zoom H4n™，MH Acoustics Eigenmike™，或诸如杜比会议电话(Dolby Conference Phone™)的空间扬声器。麦克风阵列在本文中可以被称为单个端点。然而，来自这样的单个端点的音频数据可以对应于多个会议参与者。在一些实现中，麦克风阵列可能能够检测每个会议参与者的空间信息，并且将每个会议参与者的空间信息包含在提供给记录设备720的音频数据中。

[0347] 鉴于上述内容，本公开涵盖了可以记录涉及多个会议参与者的会议的音频数据的各种实施方式。在一些实现中，会议可以是电话会议，而在其他实现中，会议可以是面对面会议。在各种示例中，可以分别记录多个端点中的每一个的音频数据。作为替代地或附加地，来自单个端点的记录的音频数据可以对应于多个会议参与者，并且可以包括每个会议参与者的空间信息。

[0348] 各种公开的实现涉及按前述方式之一或两者记录的数据的处理和/或回放。一些这样的实现涉及确定虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。虚拟声学空间内的位置可以相对于虚拟收听者的头部来确定。在一些示例中，在给定会议的会话动态的情况下，虚拟会议参与者位置可以至少部分地根据人类声音定位的心理物理学，根据影响语音可懂度的空间参数，和/或根据揭示听众已经发现什么样的讲话者位置会相对或多或少地令人反感的经验数据。

[0349] 在一些实现中，对应于整个会议或至少电话会议的显著部分的音频数据可用于确定虚拟会议参与者位置。因此，可以确定会议的完整或基本上完整的会话动态数据的集合。在一些示例中，虚拟会议参与者位置可以至少部分地根据会议的完整或基本上完整的会话动态数据的集合被确定。

[0350] 例如，会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据。在听力练习中已经发现，许多人反对在会议中的主导发言者被渲染至在听众后面或旁边的虚拟位置。当听一个讲话者的长篇演讲(例如在商务演讲中)时，很多听众都报告说，他们希望与该讲话者对应的音源被定位在听众面前，就好像听众在讲座或研讨会中那样。对于一个讲话者的长篇演讲，定位在后面或旁边常常会引起这似乎不自然的评论，或者在某些情况下听众的个人空间受到侵犯的评论。因此，会议参与者语音的频率和持续时间可以是针对相关联的会议记录的回放来分配和/或渲染虚拟会议参与者位置的过程的有用输入。

[0351] 在一些实现中，会话动态数据可以包括指示会议参与者会话的实例的数据。已经发现，将参加会话的会议参与者渲染到大不相同的虚拟会议参与者位置可以提高听众在任何给定时间辨别哪个会议参与者正在讲话的能力，并且可以提高听众理解每个会议参与者正在说什么的能力。

[0352] 会话动态数据可以包括所谓的“双讲话”的实例，在该“双讲话”期间至少有两名会议参与者同时发言。已经发现，与将参加双讲话的参与者渲染到相同虚拟位置相比，将进行双讲话的参与者渲染到大不相同的虚拟会议参与者位置有利于听众。这种差异化的定位为听众提供了关于选择性地倾听参加双讲话的会议参与者之一和/或了解每个会议参与者在

说什么的更好的线索。

[0353] 在一些实现中，会话动态数据可以被应用为空间优化成本函数的一个或多个变量。成本函数可以是描述虚拟声学空间中的多个会议参与者中的每一个的虚拟会议参与者位置的向量的函数。

[0354] 图8是简述根据本公开的一些实现的方法的一个示例的流程图。在一些示例中，方法800可以由诸如图3A的装置的装置执行。与本文所述的其它方法一样，方法800的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0355] 在此实现中，块805涉及接收对应于涉及多个会议参与者的会议的记录的音频数据。根据一些示例，音频数据可以对应于完整的或基本上完整的会议的记录。在一些实现中，在块805中，诸如图3A的控制系统330的控制系统可经由接口系统325接收音频数据。

[0356] 在一些实现中，会议可以是电话会议，而在其他实现中，会议可以是面对面会议。在该示例中，音频数据可以包括被分别记录的来自多个端点的音频数据。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的、并且包括多个会议参与者的每个会议参与者的空间信息的音频数据。例如，单个端点可以是空间扬声器电话端点。

[0357] 在一些实现中，在块805中接收的音频数据可以包括语音活动检测过程的输出。在一些替代实现中，方法800可以包括语音活动检测过程。例如，方法800可以包括标识对应于各个会议参与者的语音。

[0358] 在该示例中，块810涉及分析音频数据以确定会话动态数据。在这种实例中，会话动态数据包括以下中的一个或多个：指示会议参与者语音的频率和持续时间的数据；指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据；以及指示会议参与者会话的实例的数据。

[0359] 在此实现中，块815涉及将会话动态数据应用作为空间优化成本函数的一个或多个变量。这里，空间优化成本函数是描述虚拟声学空间中每个会议参与者的虚拟会议参与者位置的向量的函数。虚拟声学空间内的位置可以相对于虚拟听众头部的位置来定义。下面描述合适的成本函数的一些示例。在回放期间，虚拟听众头部的位置可以与实际听众头部的位置相对应，在实际听众佩戴耳机的情况下尤其如此。在下面的讨论中，术语“虚拟听众的头部”和“听众的头部”有时可以互换使用。同样地，术语“虚拟听众”和“听众”有时可以互换使用。

[0360] 在该示例中，块820涉及将优化技术应用于空间优化成本函数以求解。在这种实现中，该解是局部最优解。块820可以例如包括应用梯度下降技术、共轭梯度技术、牛顿法、Broyden-Fletcher-Goldfarb-Shanno算法、遗传算法、模拟退火算法、蚁群优化方法和/或蒙特卡罗方法。在该实现中，块825包括至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0361] 例如，成本函数的变量可以至少部分地基于指示会议参与者语音的频率和持续时间的会话动态数据。如上所述，当收听一个会话参与者的长篇讲话(例如，在商业演讲中)时，许多听众已经表示他们喜欢将会话参与者定位在他们前面，就好像他们在讲座或研讨会中那样。因此，在一些实现中，空间优化成本函数可以包括倾向于将频繁发言的会话参与

者布置在听众前面的加权因子，惩罚函数，成本或另一个这样的术语(它们中的任一个或者全部在本文中可以被称为“惩罚”)。例如，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。

[0362] 作为替代地或者附加地，成本函数的变量可以至少部分地基于指示参与会议参与者双讲话的会议参与者的会话动态数据。之前已经注意到，与将参与双讲话的会议参与者渲染到相同的虚拟位置相比，将参加双讲话的会议参与者渲染到大不相同的虚拟会议参与者位置可以为听众带来便利。

[0363] 为了量化这种差异化定位，空间优化成本函数的一些实现可能涉及应用对于将参与会议参与者双讲话的会议参与者布置在如下虚拟会议参与者位置处的惩罚，该虚拟会议参与者位置位于相对于虚拟听众头部被定义的所谓“混淆锥”上或者接近于位于该混淆锥上。

[0364] 图9示出了虚拟声学空间中的虚拟听众的头部和混淆锥的示例。在该示例中，在虚拟声学空间900中，坐标系905相对于虚拟听众头部910的位置来定义。在该示例中，坐标系905的y轴与在虚拟听众头部910的耳朵915之间通过的耳间轴线重合。这里，z轴是穿过虚拟听众头部910的中心的垂直轴线，并且x轴在虚拟听众头部910所面向的方向上是正的。在这个例子中，原点是在耳朵915之间的中点。

[0365] 图9还示出了在该示例中相对于耳间轴和声源925被定义的混淆锥920的示例。这里，声源925被定位在与耳间轴相距半径R处，并且被示出为发射声波930。在该示例中，半径R平行于x轴和z轴，并且限定圆锥切片935。因此，沿圆锥切片935的所有点与虚拟听众头部910的每个耳朵915等距。因此，来自位于圆锥切片935或者通过混淆锥920的任何其他圆锥切片上的任何地方的声源的声音将产生相同的耳间时间差。这种声音也将产生非常相似(尽管不一定相同)的耳间水平差异。

[0366] 由于相同的耳间时间差，听众区辨别在混淆锥上或附近的声源的位置会非常有挑战性。虚拟声学空间中的声源位置对应于会议参与者的语音将被渲染到的位置。因此，由于虚拟声学空间中的源位置对应于虚拟会议参与者位置，所以术语“源”和“虚拟会议参与者位置”在文中可以互换使用。如果将两个不同的会议参与者的声音渲染到位于混淆锥上或接近混淆锥的虚拟会议参与者位置，则虚拟会议参与者位置可能似乎是相同或基本相同的。

[0367] 为了充分区分至少一些会议参与者(例如参与双讲话的参与者)的虚拟会议参与者位置，可能有利地是定义相对于混淆锥的预定角距离，例如如图9中所示的相对于混淆锥920的角度α。角度α可以定义与混淆锥920具有相同的轴线(这里，y轴))的、在混淆锥920的内部和/或外部的锥环。因此，空间优化成本函数的一些实现可以涉及应用对于将参与会议参与者双讲话的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚，该混淆锥相对于虚拟听众头部被定义。在一些实现中，惩罚可能与源A和B位于其上的混淆锥之间的角距离成反比。换句话说，在某些这样的实现中，两个源越接近于位于共同的混淆锥，惩罚越大。为了避免突然变化和/或不连续性，惩罚可能会平滑地变化。

[0368] 作为替代地或者附加地，成本函数的变量可至少部分地基于指示会议参与者会话

的实例的会话动态数据。如上所述，将进行会话的会议参与者渲染至大不相同的虚拟会议参与者位置可以提高听众在任何给定时间辨别哪个会议参与者正在讲话的能力，并且可以提高听众理解每个会议参与者正在说什么的能力。因此，空间优化成本函数的一些实现可以涉及应用对于将参与会议参与者相互会话的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚，该混淆锥相对于虚拟听众头部被定义。例如，随着虚拟会议参与者位置越接近共同的混乱锥，惩罚可以平滑地增加。

[0369] 对于在会议期间仅短暂插话(short interjection)(或主要是短暂插话)的会议参与者，将相应的虚拟会议参与者位置渲染至听众后面或旁边是可以接受的，甚至是希望的。置于听众旁边或后面使得就好像观众同伴提问或评论。

[0370] 因此，在一些实施方式中，空间优化成本函数可以包括一个或多个项，其倾向于避免将与在会议期间仅短暂插话(或主要是短暂插话)的会议参与者对应的虚拟会议参与者位置渲染至听众前面的位置。根据一些这样的实现，空间优化成本函数可以应用对于将很少发言的会议参与者布置于不在虚拟听众的头部位置旁边、后面、上方或下方的虚拟会议参与者位置处的惩罚。

[0371] 在群组情况中进行交谈时，听众可能倾向于更靠近他或她想要收听的讲话者，而不是保持距离。这种行为可能有社会原因以及声学原因。本文中公开的一些实现方式可以通过将较频繁讲话的会议参与者的虚拟会议参与者位置渲染为与较不频繁讲话的会议参与者相比更接近虚拟听众来模拟这种行为。例如，在一些这样的实现中，空间优化成本函数可以应用对于将频繁发言的会议参与者布置于与较不频繁发言的会议参与者的虚拟会议参与者位置相比距虚拟听众头部更远的虚拟会议参与者位置处的惩罚。

[0372] 根据一些实现，成本函数可以表达如下：

[0373] $F(a)=F_{conv}(a)+F_{dt}(a)+F_{front}(a)+F_{dist}(a)+F_{int}(a)$ (式1)

[0374] 在式1中，F_{conv}表示违反如下准则的感知成本，即参与会话的会话参与者不应在位于混淆锥上或附近的虚拟会议参与者位置处被渲染。在式1中，F_{dt}表示违反如下准则的感知成本，即参与双讲话的会话参与者不应在位于混淆锥上或附近的虚拟会议参与者位置处被渲染。在式1中，F_{front}表示违反如下准则的感知成本，即频繁发言的会话参与者应该在处于听众前面的虚拟会议参与者位置处被渲染。在式1中，F_{dist}表示违反如下准则的感知成本，即频繁发言的会议参与者应该在与较不频繁发言的会议参与者相比更接近听众的虚拟会议参与者位置处被渲染。在式1中，F_{int}表示违反如下准则的感知成本，即仅短暂插话和/或很少发言的会话参与者不应在处于听众前面的虚拟会议参与者位置处被渲染。

[0375] 在替代实现中，成本函数可以包括更多、更少和/或不同的项。一些替代实现可以省略式1的F_{int}变量和/或一个或多个其他项。

[0376] 在式1中，a表示描述N个会议参与者中的每一个的在虚拟声学空间中的D维虚拟会议参与者位置的向量。例如，如果渲染器具有每个位置三个自由度(使得D=3)，并且这些是给定源i的方位角(θ_i)、仰角(Φ_i)和距离(d_i)(其中$1<i<N$)的极坐标(欧拉角坐标)，则向量a可以定义如下：

[0377] $$a = \begin{bmatrix} \theta_1 \\ \phi_1 \\ d_1 \\ \vdots \\ \theta_N \\ \phi_N \\ d_N \end{bmatrix}$$

（式2）

[0378] 然而，在许多情况中，可以通过改为在笛卡尔坐标中工作来获得更简单和更数值稳定的解。例如，可以定义一个(x,y,z)坐标系，如图9所示。在一个这样的例子中，可以将x_i定义为源i(诸如图9的声源925)到虚拟听众头部的中心的沿着在听众前面从听众的鼻子向外延伸的轴线的距离。可以将y_i定义为源i到听众头部的中心的沿着垂直于第一轴线延伸到听众左侧的轴线的距离。最后，可以将z_i定义为源i到听众头部的中心的沿着垂直于其它两个轴线向上延伸的轴线的距离。使用的距离单位可以是任意的。然而，在下面的描述中，将假设距离被归一化以适合于渲染系统，使得在距离收听者一个单位的虚拟距离处，收听者定位源的能力将被最大化。

[0379] 如果使用刚刚描述的笛卡尔坐标系，则矢量a可以定义如下：

[0380] $$a = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ \vdots \\ x_N \\ y_N \\ z_N \end{bmatrix}$$

（式3）

[0381] 上述段落提供了感知成本函数F(a)的示例，其根据各种类型的会话动态数据描述虚拟会议参与者位置的特定向量a的适应度(适合度)。现在可以找到导致最小感知成本(换句话说，最大适应度)的源位置a_{opt}的向量。鉴于上述新颖的成本函数，一些实现可能涉及应用已知的数值优化技术来求解，例如梯度下降技术、共轭梯度技术、牛顿法、Broyden-Fletcher-Goldfarb-Shanno算法、遗传算法、模拟退火算法、蚁群优化方法和/或蒙特卡罗方法。在一些实现中，解可以是局部最优解，已知上述示例技术是非常适合于该局部最优解。

[0382] 在一些实施例中，空间优化成本函数的输入可以是VAD(语音活动检测器)输出的矩阵V。例如，矩阵对于会议的每个离散时间分析帧可以具有一行，并且可以具有N列，每个会议参与者一列。在一个这样的示例中，分析帧大小可能是20ms，这意味着V包含VAD的对于每个源的每个20ms分析帧包含语音的概率的估计。在其他实现中，分析帧可以对应于不同的时间间隔。为了简单起见，进一步假设在下面描述的示例中，每个VAD输出可以是0或1。也就是说，VAD输出指示每个源在每个分析帧内包含或不包含语音。

[0383] 为了进一步简化讨论，可以假设在会议记录完成之后进行虚拟会议参与者位置的

优化布置，使得该过程可以随机访问会议的所有分析帧。然而，在替代示例中，可以为会议的任何部分(诸如会议的不完整记录)生成解，这考虑了为该会议的该部分生成的VAD信息。

[0384] 在该示例中，该过程可以涉及使得矩阵V通过聚合过程，以便生成会议的聚合特征。根据一些这样的实现，聚合特征可以对应于在会议期间的双讲话和话轮转换(turn-taking)的实例。根据一个这样的示例，聚合特征对应于双讲话矩阵C_{dt}和话轮转换矩阵C_{turn}。

[0385] 例如，C_{dt}可以是对称N×N矩阵，在行i，j中描述了会议期间的会议参与者i和j同时包含语音的分析帧的数量。因此，C_{dt}的对角线元素描述了每个会议参与者的语音帧数，矩阵的其他元素描述了在会议期间参与双讲话的特定会议参与者对的帧数。

[0386] 在一些实现中，计算C_{dt}的算法可如下进行。首先，C_{dt}可被初始化，使得所有元素都为零。然后，可以依次考虑V的每行v(换句话说，每个分析帧)。对于每个帧，可以向C_{dt}的每个元素c_{ij}加1，其中v的列i和j都为零。作为替代地，C_{dt}可以通过矩阵乘法来计算，例如如下：

[0387] $C_{dt}=V^{T}V$ (式4)

[0388] 在式4中，V^{T}表示应用于矩阵V的常规矩阵转置操作。

[0389] 然后，可以通过将C_{dt}除以会议中的讲话的总量(换句话说，矩阵C_{dt}的迹(trace))创建归一化的双讲话矩阵N_{dt}，例如如下：

[0390] $$N_{dt}=\frac{C_{dt}}{tr(C_{dt})} \quad (式5)$$

[0391] 在式5中，$tr(C_{dt})$表示矩阵C_{dt}的迹。

[0392] 为了计算C_{turn}，在初始化为零之后，一些实现涉及定位每个讲话突发的开始。例如，一些实现可以涉及考虑V中的每个会议参与者i，并且查找V中的每一行r，其中在列i中为零而在行r+1中为1。然后，对于每个讲话突发，一些这样的示例涉及确定哪个会议参与者j在该讲话突发之前最近发言。这将是涉及会议参与者i和j的“话轮转换”的例子，这也可以在这里被称为“转换”的例子。

[0393] 这样的例子可能涉及在时间上向后看(换句话说，查看行r和以上的行)，以便标识哪个会议参与者j在该讲话突发之前最近发言。在一些这样的例子中，对于发现的每个这样的话轮转换实例，可以向C_{turn}的行i、列j加“1”。一般来说，C_{turn}可能是非对称的，因为它保留了与时间顺序有关的信息。

[0394] 给定上述信息，可以例如通过将C_{turn}除以会议中的总转换数(换句话说，除以矩阵中的所有元素的总和)来创建归一化的话轮转换矩阵N_{turn}，例如如下：

[0395] $$N_{turn}=\frac{C_{turn}}{\sum_i \sum_j C_{turn,ij}} \quad (式6)$$

[0396] 在式6中，$\sum_i \sum_j C_{turn,ij}$代表$C_{turn}$矩阵中的所有元素的总和。在替代实现中，矩阵$C_{dt}$和$C_{turn}$以及归一化因子$tr(C_{dt})$和$\sum_i \sum_j C_{turn,ij}$可以通过分析VAD输出(每次一个分析帧)来计算。换句话说，不需要一次可使用整个矩阵V。除了C_{dt}，C_{turn}，$tr(C_{dt})$和$\sum_i \sum_j C_{turn,ij}$之外，一些这样的方法仅需要最新的讲话者的身份被保持为状态，因为该过程每次一个帧地迭代地分析VAD输出。

[0397] 在一些实现中，聚合特征N_{dt}和N_{turn}与位置向量a的初始条件一起可以形成空间优

化成本函数的输入。几乎任何一组初始的虚拟会议参与者位置都是合适的。然而，优选的是，例如为了确保成本函数的梯度被明确定义，任何两个源最初不位于同一位置。一些实现涉及使所有初始的虚拟会议参与者位置位于听众后面。在一些这样的实现中，成本函数可以不包括F_{int}项或倾向于将短暂插话/很少发言的说话者的虚拟会议参与者位置移动到听众后面的位置的对应项。换句话说，两个一般选项如下：(a)使所有初始虚拟会议参与者的位置位于听众后面，并省去F_{int}项或对应项；或(b)包括F_{int}项或对应项，并使初始虚拟会议参与者的位置位于任何方便的位置。对于短暂插话者来说，F_{front}可能很小，因为他们很少说话。因此，涉及选项(a)的实现可能没有将短暂插话者朝收听者前面移动的强烈倾向。

[0398] 图10示出了虚拟声学空间中的初始虚拟会议参与者位置的示例。类似于如图9所示，图10所示的虚拟声学空间的坐标系是基于虚拟听者头部910的位置的。在该示例中，示出了11个初始虚拟会议参与者位置，每个初始虚拟会议参与者位置已经根据以下被确定：

[0399] $x_i = -0.5$ (式7)

[0400] $$y_i = -1 + \frac{2i}{N-1} \quad (式8)$$

[0401] $$z_i = \left|-1 + \frac{2i}{N-1}\right| \quad (式9)$$

[0402] 在式7-9中，x_i，y_i和z_i表示会话参与者i的初始(x，y，z)坐标，N表示会话参与者的总数。在图10中，编号的点对应于虚拟会议参与者位置。点大小表示相应的会议参与者的语音的相对量，较大的点表示相对较多的语音。附连到点的垂直线表示水平面上方的距离，对应于每个虚拟会议参与者位置的z坐标。单元球1005(其表面距原点为一个单位的距离)被示出作为参考。

[0403] 在一个示例中，可以通过应用下式(在迭代k)来执行梯度下降优化，直到达到收敛标准：

[0404] $$a_{k+1} = a_k - \beta_k \nabla F(a_k) \quad (式10)$$

[0405] 在式10中，β_k表示适当的步长大小，这将在下面进一步详细讨论。在一个示例中，可以对其中以下条件成立的连续优化步骤的数量n计数：

[0406] $|F(a_{k+1}) - F(a_k)| < T$ (式11)

[0407] 在式11中，T表示常数，其可以被设为适当小的值。一些实现的常数T的合适的示例值是10^{-5}。在替代实现中，T可以被设为另一个值。然而，在这种替代实现中，可以比平均成本F(a)(例如在大量会议条件下被平均化)小数个数量级。在一些示例中，收敛标准可以是$n \geq 10$，表明在过去10个连续优化步骤中成本的变化非常小，并且现在非常接近局部最小值(或者至少在成本函数的非常“平坦的”区域，在该区域中任何进一步变化不太可能被听众感知到)。

[0408] 为了在下面的讨论中清楚起见，请注意，可以从式10以扩展形式写出梯度表达式如下：

[0409] $$\nabla F(a) = \begin{bmatrix} \frac{\partial F(a)}{\partial x_1} \\ \frac{\partial F(a)}{\partial y_1} \\ \frac{\partial F(a)}{\partial z_1} \\ \vdots \\ \frac{\partial F(a)}{\partial x_N} \\ \frac{\partial F(a)}{\partial y_N} \\ \frac{\partial F(a)}{\partial z_N} \end{bmatrix} \quad (式\ 12)$$

[0410] 图11示出了虚拟声学空间中最终的虚拟会议参与者位置的示例。图11示出了给定图10所示的初始虚拟会议参与者位置，对于11个会话参与者应用前述过程的示例。在该示例中，所有最终的虚拟会议参与者位置都位于单位球1005上或附近。在图11中，与最频繁发言的会话参与者对应的所有最大的点已被移动到虚拟听众的头部910之前。与会话参与者1和3相对应的小点是最小的，表示这些会话参与者最少发言，因此保留在虚拟听众的头部910之后。在该示例中，与会话参与者5和8相对应的点小，但略大于会话参与者1和3的点，表示这些会话参与者比会话参与者1和3更频繁，但不如其他对象参与者1和3那样多。因此，与会话参与者5和8相对应的点从他们的在虚拟听众头部910之后初始位置向前偏移，但不是非常强烈。由于F_{dist}的影响，与会话参与者5和8相对应的虚拟会议参与者位置保持在虚拟侦听器头部910的上方，F_{dist}在本实施例中倾向于将所有虚拟会议参与者位置保持在距原点一个单位的半径之处。

[0411] 以下是根据一些实现对式1的项的更详细的描述。在一些示例中，式1的对应于涉及会议参与者会话的会话动态数据的项可以如下确定

[0412] $$F_{conv}(a) = \sum_{i=1}^{N} \sum_{j=1}^{N} F_{conv,ij}(a) \quad (式\ 13)$$

[0413] 在式13中，$F_{conv,ij}(a)$代表由一对源i和j接近混淆锥而贡献的成本的分量。因为如果源的y坐标相等(假定它们位于单位球上)，则源会位于混淆锥上，因此在一些例子中，$F_{conv,ij}(a)$可被如下确定：

[0414] $$F_{conv,ij}(a) = \begin{cases} 0, & i = j \\ \frac{K_{conv} N_{turn,ij}}{(y_i - y_j)^2 + \varepsilon}, & 其它 \end{cases} \quad (式\ 14)$$

[0415] 在式14中，K_{conv}以及ε表示常数。在一些示例中，两个常数可以被设为相对较小的值，例如0.001。在这个例子中，当数据源正好位于混淆锥上时，ε会阻止成本达到无穷大的值。K_{conv}可以关于其他参数被调整，以实现良好的分离，同时还允许多源在前面。如果K_{conv}被设得太高，F_{conv}将倾向于主宰所有其他成本函数元素，并将源散布到球体周围。因此，尽管可以在各种实现中使用K_{conv}以及ε的替代值，但是这些和其它参数是相互关联的，并且可以被联合调整以产生期望的结果。

[0416] 式14的基本假设是源位于单位球上，这是因为在一些实现中，$F_{dist}(a)$(其一个例子在下面被更具体地定义)将可靠地将源保持在单位球附近。如果作为替代地，$F_{dist}(a)$被

定义为使得它不会可靠地将源保持在单位球附近，则可能需要在计算$F_{conv,ij}(a)$之前对y坐标进行归一化，例如如下：

[0417] $$\hat{y}_i = \frac{y_i}{\sqrt{x_i^2+y_i^2+z_i^2}} \quad (式\ 15)$$

[0418] $$F_{conv,ij}(a) = \begin{cases} 0, & i=j \\ \frac{K_{conv}N_{turn,ij}}{(\hat{y}_i-\hat{y}_j)^2+\varepsilon}, & 其它 \end{cases} \quad (式\ 16)$$

[0419] 一些替代示例可以涉及直接计算与耳间时间差的倒数成正比的成本。

[0420] 在一些实现中，$F_{dt}(a)$可以如下计算：

[0421] $$F_{dt}(a) = \sum_{i=1}^{N}\sum_{j=1}^{N}F_{dt,ij}(a) \quad (式\ 17)$$

[0422] 在一些示例中，式17的项$F_{dt,ij}(a)$可以确定如下：

[0423] $$F_{dt,ij}(a) = \begin{cases} 0, & i=j \\ \frac{K_{dt}N_{dt,ij}}{(y_i-y_j)^2+\varepsilon}, & 其它 \end{cases} \quad (式\ 18)$$

[0424] $$\frac{\partial F_{dt,ij}}{\partial y_i} = \frac{-2K_{dt}N_{dt,ij}}{[(y_i-y_j)^2+\varepsilon]^2} \quad (式\ 19)$$

[0425] $$\frac{\partial F_{dt,ij}}{\partial y_j} = \frac{2K_{dt}N_{dt,ij}}{[(y_i-y_j)^2+\varepsilon]^2} \quad (式\ 20)$$

[0426] 在式18-20中，K_{dt}和ε表示常数。在一些实例中，K_{dt}可以是为0.002，ε可以为0.001。尽管在替代实现中可以使用K_{dt}和ε的各种其他值，但是这些和其它参数是相互关联的，并且可以被联合调整以产生期望的结果。

[0427] 在一些实现中，式(1)中的$F_{front}(a)$的变量对于不位于听众前面施加惩罚，该惩罚与已参与会议的会话参与者的数量的平方成正比。结果，相对讲话更多的会话参与者的虚拟会议参与者位置终止于与虚拟声学空间中的虚拟听众相比更靠近前方中心位置。在一些这样的示例中，$F_{front}(a)$可以如下确定：

[0428] $$F_{front}(a) = \sum_{i=1}^{N}F_{front,i}(a) \quad (式\ 21)$$

[0429] $F_{front,i}(a) = K_{front}N_{dt,ii}{}^2[(x_i-1)^2+y_i^2+z_i^2]^2$ (式22)

[0430] 在式22中，K_{front}表示常数，在一些示例中，K_{front}可能是5。尽管在替代实现中可以使用K_{front}的各种其他值，但是此参数可能与其他参数相互关联。例如，K_{front}应该足够大，以便可以将讲话最多的会话参与者的虚拟会议参与者位置拉倒前面，但不会太大而使得F_{front}始终超越F_{conv}和F_{dt}的贡献。在一些示例中，由于$F_{front}(a)$导致的对梯度的贡献可以如下确定：

[0431] $$\frac{\partial F_{front,i}}{\partial x_i} = 2K_{front}N_{dt,ii}{}^2(x_i-1) \quad (式\ 23)$$

[0432] $$\frac{\partial F_{front,i}}{\partial y_i} = 2K_{front}N_{dt,ii}{}^2 y_i \quad (式\ 24)$$

[0433] $$\frac{\partial F_{front,i}}{\partial z_i} = 2K_{front}N_{dt,ii}{}^2 z_i \quad (式\ 25)$$

[0434] 在一些实现中，式1的$F_{dist}(a)$分量可对不将虚拟会议参与者位置布置在单位球上而施加惩罚。在一些这样的例子中，讲话更多的会议参与者的惩罚可能更高。在某些实例中，$F_{dist}(a)$可以如下确定：

[0435] $$F_{dist}(a) = \sum_{i=1}^{N} F_{dist,i}(a) \quad (式\ 26)$$

[0436] $F_{dist,i}(a) = K_{dist}N_{dt,ii}[x_i{}^2+y_i{}^2+z_i{}^2-1]^2$ (式27)

[0437] 在式27中，K_{dist}表示常数，在一些示例中，K_{dist}可能是1。尽管在替代实现中可以使用K_{dist}的各种其他值，但此参数可能与其他参数相互关联。例如，如果K_{dist}太小，则F_{dist}的效果可能太弱，而源将趋向于偏离单位球。在一些示例中，由于$F_{dist}(a)$导致的对梯度的贡献可以如下确定：

[0438] $$\frac{\partial F_{dist,i}}{\partial x_i} = 4K_{dist}N_{dt,ii}x_i[x_i{}^2 + y_i{}^2 + z_i{}^2 - 1] \quad (式\ 28)$$

[0439] $$\frac{\partial F_{dist,i}}{\partial y_i} = 4K_{dist}N_{dt,ii}y_i[x_i{}^2 + y_i{}^2 + z_i{}^2 - 1] \quad (式\ 29)$$

[0440] $$\frac{\partial F_{dist,i}}{\partial z_i} = 4K_{dist}N_{dt,ii}z_i[x_i{}^2 + y_i{}^2 + z_i{}^2 - 1] \quad (式\ 30)$$

[0441] 在一些实施例中，式1的项$F_{int}(a)$可以被设置为零。这例如在初始条件将源布置在虚拟听众头部后面的实现中可能是可以接受的。因为$F_{front}(a)$的各种实现仅对于位于听众后面的极少讲话的源施加了弱惩罚，所以它们将停留在虚拟听众头部的后面，除非收敛标准非常严厉。在一些替代实施例中，小的惩罚可以与不在虚拟听众头部后面的任何源相关联。在许多实现中，这个小的惩罚趋向于被$F_{front,i}(a)$主导，除了在极少讲话的会话参与者的实例中之外。

[0442] 现在将描述收敛标准和过程的一些更详细的例子。再次参考式10，一些实施方式包括随着优化进行通过使用所谓的线搜索来调整步长大小β_k。在一些这样的实现中，可以将β_{-1}的值初始化为0.1。根据一些这样的示例，在每一步中，β_k可以根据以下过程来调整：

[0443] 1.假设$\hat{\beta}_k = \beta_{k-1}$。

[0444] 2.计算$F_1 = F(a_k - \hat{\beta}_k \nabla F(a_k))$，在步长大小为$\hat{\beta}_k$时的新成本。

[0445] 3.如果$F_1 > F(a_k)$则以$\hat{\beta}_k$步进将超过最小值，因此将$\hat{\beta}_k$减半，并返回到步骤2。

[0446] 4.计算$F_2 = F(a_k - 2\hat{\beta}_k \nabla F(a_k))$，在步长大小为$2\hat{\beta}_k$时的新成本。

[0447] 5.如果$F_1 > F_2$，以$2\hat{\beta}_k$步进将低于最小值，所以将$\hat{\beta}_k$加倍，并返回到步骤2。

[0448] 6.$\hat{\beta}_k$和$2\hat{\beta}_k$之间某处的步长大小应该会导致最小值附近的值。一些示例在成本函数的形状可由在$\hat{\beta}_k$的通过点$(0, F(a_k)), (\hat{\beta}_k, F_1), (2\hat{\beta}_k, F_2)$的二次方来近似的假设下操作，并找到最小值如下：

[0449] $$\beta_k = \hat{\beta}_k + \frac{F_2 - F(a_k)}{2F_1 - 3F(a_k) - F_2} \quad (式\ 31)$$

[0450] 7.然后，箝位$\hat{\beta}_k$确保它在$[\hat{\beta}_k, 2\hat{\beta}_k]$中。

[0451] 在一些实施例中，空间优化成本函数可以考虑会话参与者的感知区别性。有充分的证据表明，当同时讲话者在他们的声音被感知为区别明显时可被更好地理解。这在引起声音区别性的特质被描述为分类(例如，讲话者被认为是男性还是女性，活着声音被感知为是“干净”还是“嘈杂”)或连续(例如，音调，声道长度等)

[0452] 因此，一些实现可以涉及确定哪些会议参与者(如果有的话)具有感知类似的语音。在一些这样的实施方式中，空间优化成本函数可以应用对于将具有感知相似的语音的会议参与者布置于位于混淆锥上或者与混淆锥相距在预定的角距离内的虚拟会议参与者位置处的惩罚，该混淆锥相对于虚拟听众头部被定义。一些这样的实现可以涉及向式1添加另一个变量。

[0453] 然而，替代实现可以涉及修改式1的变量之一。例如，虽然$F_{conv}(a)$和$F_{dt}(a)$的有些实现旨在惩罚将分别交谈和双讲话的会议参与者定位在可混淆空间布置中，但是一些替代实现涉及在所关注的会议参与者在感知上是相似的情况下，修改$F_{conv}(a)$和/或$F_{dt}(a)$以进一步惩罚这样的布置。

[0454] 一些这样的示例可以涉及第三N×N聚合矩阵N_{dsim}，其量化了参与会议的每对会议参与者的不相似性。为了计算N_{dsim}，一些实现首先确定由会议记录中每个会议参与者的B特性特征组成的“特性特征向量”s，其中每个特性特征$s[k]_i$是讲话者i的感知相关度量。其中B＝2的一个例子如下：

[0455] $$s_i = \begin{bmatrix} s[1]_i \\ s[2]_i \end{bmatrix} \quad (式\ 32)$$

[0456] 在式32中，$s[1]_i$表示中间音调，$s[2]_i$表示会议参与者i的估计的声道长度。可以通过聚合来自会议参与者在会议期间发出的许多(可能所有)的讲话话语的信息来估计特性特征。在其他实现中，可以使用其他特性特征，例如口音和语速，来量化一对会议参与者的不相似性。还有其他实现可能涉及量化一对会议参与者的相似性而不是相似性。

[0457] 在一些实现中，特性特征向量可以由一组B个时域滤波器产生，每个时域滤波器之后可以是具有适当时间常数的包络检测器。可以通过应用离散傅里叶变换(DFT)来产生特性特征向量，在该离散傅里叶变换之前可以是适当的加窗，而之后可以是适当的带化(banding)过程。带化过程可以将DFT箱分组成具有大致相等的感知尺寸的带。在一些示例中，可以在DFT和带化过程之后计算Mel频率倒谱系数。如果会议以使用频域编码(例如，根据修改的离散余弦变换(MDCT)过程)的编码格式被存储，则一些实现可以使用编码域系数，然后进行适当的带化。

[0458] 在一些实现中，特性特征向量可以由线性预测系数(诸如在线性预测编码(LPC)方案中使用的那些)来产生。一些示例可以涉及感知线性预测(PLP)方法，例如用于语音识别的那些。

[0459] 根据一些实现，在计算特性特征向量之后，可以在每对特性特征矢量s_i，s_j之间应用合适的距离度量来计算N_{dsim}中的每个元素。这样的距离度量的示例是均方差，其可以如下地计算：

[0460] $$N_{dsim,ij} = \frac{1}{B}\sum_{k=1}^{B}(s_i(k) - s_j(k))^2 \quad (式\ 33)$$

[0461] 在式33中，k表示B个特性特征中的一个在s中的索引(在该示例中，s是B维或B特征向量)。根据式33，考虑每个特征，确定每两个特征之间的差，该差被平方并在所有维度上求和。例如，对于式32给出的二维示例，B是2，并且变量k上的总和采用对应于式32中所示的字面数字1和2的值＝1和k＝2。一些实现可以涉及基于跨多个会议的信息来计算用于特定会议参与者的特性特征向量s。一些这样的实现可以涉及基于多个会议的音频数据确定长期平均值。

[0462] 在某些实现中，可能有会议参与者的性别的先验知识。例如，作为注册或登记过程的一部分，可能需要或鼓励会议参与者指定他们是男性还是女性。当这种知识可用于回放系统时，用于计算$N_{dsim,ij}$的替代示例方法可以如下：

[0463] $$N_{dsim,ij} = \begin{cases} K_{homo}, & 讲话者\ i\ 和\ j\ 为相同性别 \\ K_{hetero}, & 讲话者\ i\ 和\ j\ 为不同性别 \end{cases} \quad (式\ 34)$$

[0464] 在式34中，K_{homo}和K_{hetero}表示常数。在一个例子中，K_{homo}和K_{hetero}等于1.0，并且K_{hetero}可以是例如在[0.1,0.9]*K_{homo}，或等于0.5。

[0465] 基于上述任何一个例子，可以重新定义$F_{conv,ij}(a)$和$F_{dt,ij}(a)$，并包括频谱相似性聚合$N_{dsim,ij}$，例如如下所示：

[0466] $$F_{conv,ij}(a) = \begin{cases} 0, & i = j \\ \dfrac{K_{conv} N_{turn,ij} N_{dsim,ij}}{(y_i - y_j)^2 + \varepsilon}, & 其它 \end{cases} \quad (式\ 35)$$

[0467] $$F_{dt,ij}(a) = \begin{cases} 0, & i = j \\ \dfrac{K_{dt} N_{dt,ij} N_{dsim,ij}}{(y_i - y_j)^2 + \varepsilon}, & 其它 \end{cases} \quad (式\ 36)$$

[0468] 根据一些实施例，分配虚拟会议参与者位置可以包括从一组预定的虚拟会议参与者位置中选择虚拟会议参与者位置。在一些这样的示例中，每个源可以仅被布置在大小为A的虚拟会议参与者位置的固定集合之一中。在这样的实现中，可以通过表查找来直接计算每个成本函数分量，而不通过基于位置坐标的计算。例如，每个成本函数分量可以如下计算：

[0469] $F_{conv,ij}(a) = K_{conv,ij} N_{turn,ij} N_{dsim,ij}$ (式37)

[0470] 在式37中，$K_{conv,ij}$表示固定矩阵(例如，查找表)，其描述了来自位置i的语音将在多大程度上在感知上掩盖来自位置j的语音。例如，$K_{conv,ij}$可能从大规模的主观测试得出。在该示例中，优化过程涉及将每个源分配给A个虚拟会议参与者位置之一。因为搜索空间不

再是连续的，因此在这样的示例中，离散优化技术(诸如模拟退火和遗传算法)可能比本文所提及的一些其它优化技术更适用。

[0471] 一些实现可以涉及一种混合解决方案，其中一些虚拟会议参与者位置被分配给预定的虚拟会议参与者位置，而其他虚拟会议参与者位置在不参考预定的虚拟会议参与者位置的情况下被确定。例如，当要确定的虚拟会议参与者位置的数量超过预定的虚拟会议参与者位置的数量时，可以使用这样的实现。在一些这样的示例中，如果存在A个预定的虚拟会议参与者位置，但是多于A个虚拟会议参与者位置待确定的，则可以将预定的虚拟会议参与者位置用于讲话最多的A个会议参与者，并且可以对于剩余的会议参与者计算动态位置，例如通过使用诸如式1的空间优化成本函数。

[0472] 这里公开的一些实现允许收听者快速地回放和/或扫描会议记录，同时保持有关注感兴趣的词语、主题和讲话者的能力。一些这样的实现通过利用空间渲染技术并且根据一组感知激发规则引入(或改变)会议参与者语音的实例之间的重叠来减少回放时间。作为替代地或者附加地，一些实现可以涉及加速被回放的会议参与者语音。

[0473] 图12是简述根据本公开的一些实现的方法的一个示例的流程图。在一些示例中，方法1200可以由诸如图3A的装置的装置和/或图6的回放系统609的一个或多个组件来执行。在一些实现中，方法1200可以由至少一个设备根据存储在一个或多个非暂态介质上的软件执行。类似于本文描述的其它方法，方法1200的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0474] 在该实现中，块1205涉及接收对应于涉及多个会议参与者的会议的记录的音频数据。在一些实现中，在块1205中，诸如图3A的控制系统330的控制系统可以经由接口系统325接收音频数据。

[0475] 在一些实现中，会议可以是电话会议，而在其他实现中，会议可以是面对面会议。在该示例中，音频数据可以包括被分别记录的来自多个端点的音频数据。作为替代地或者附加地，音频数据可以包括来自对应于多个会议参与者的单个端点的音频数据，并且包括多个会议参与者中的每个会议参与者的空间信息。例如，单个端点可以包括麦克风阵列，诸如声场麦克风或空间扬声器电话的阵列。根据一些示例，音频数据可以对应于完整的或基本上完整的会议的记录。

[0476] 在一些实现中，音频数据可以包括语音活动检测过程的输出。因此，在一些这样的实现中，音频数据可以包括语音和/或非语音分量的指示。然而，如果音频数据不包括语音活动检测过程的输出，则在一些示例中，方法1200可以涉及识别对应于各个会议参与者的语音。对于其中在块1205中接收到来自对应于多个会议参与者的单个端点的会议参与者语音数据的实现，方法1200可以包括根据“发言者日志”过程的输出来标识对应于各个会议参与者的语音，该“发言者日志”过程标识说出了语音的每一个实例的会议参与者。

[0477] 在该示例中，块1210涉及将每个会议参与者的会议参与者语音数据渲染到虚拟声学空间中的单独的虚拟会议参与者位置。在一些实现中，块1210可以涉及如本文别处所描述的虚拟会议参与者位置。

[0478] 因此，在一些这样的实现中，块1210可以涉及分析音频数据以确定会话动态数据。在某些实例中，会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据；指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据；以及指示会

议参与者会话的实例的数据。一些实现可以涉及分析音频数据以确定其他类型的会话动态数据和/或会议参与者语音的相似性。

[0479] 在一些这样的实现中，块1210可以涉及将会话动态数据应用作为空间优化成本函数的一个或多个变量。该空间优化成本函数可以是描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置的向量的函数。虚拟声学空间内的位置可以相对于虚拟听众头部的位置来定义。块1210可以包括将优化技术应用于空间优化成本函数，以确定局部最优解并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0480] 然而，在其他实现中，块1210可以不涉及空间优化成本函数。例如，在一些替代实现中，块1210可以涉及将各会议参与者的会议参与者语音数据渲染给多个预定的虚拟会议参与者位置中的单独的一个。块1210的一些替代实现可以涉及在不参考会话动态数据的情况下确定虚拟会议参与者位置。

[0481] 在各种实现中，方法1200可以包括根据感知激发规则的集合来回放会议参与者语音。在该示例中，块1215涉及回放会议参与者语音，使得根据感知激发规则的集合，先前在时间上不重叠的会议参与者语音中的至少一些以重叠方式被回放。

[0482] 根据诸如方法1200的方法，听众可以受益于通过对于多个会议参与者中的每一个从空间中的各种独特位置回放音频数据而提供的双耳优点。例如，听众可能能够容忍被渲染到不同的位置的来自会议参与者的语音的严重重叠，并且仍然保持关注(不失一般性)感兴趣的话语，主题，声音或讲话者的能力。在一些实现中，一旦已经识别出感兴趣的部分，收听者可以具有切换到非重叠回放模式的选项以更详细地收听该部分，例如，通过与回放系统(例如图6的回放系统609)的一个或多个元件的交互。

[0483] 方法1200中以及本文提供的其他方法中应用的规则被称为“感知激发”，因为它们基于现实世界的听觉体验。例如，在一些实现中，感知激发规则的集合可以包括指示单个会议参与者的两个语音部分不应该在时间上重叠的规则。这个规则是由于如下这样的观察而激发的，即尽管听到多位讲话者同时发言(例如在鸡尾酒会上)是人体验的自然部分，但听到同一个讲话者的两个副本同时发言并不是一个自然的体验。在现实世界中，人类每次只能发出单一的言语流，而且一般来说，每个人都有独特的可标识的说话语音。

[0484] 一些实现可以涉及上述规则的一个或多个变型。例如，在一些实现中，感知激发规则的集合可以包括指示如果两个语音段对应于单个端点，则这两个语音端不应该在时间上重叠的规则。在许多实例中，单个端点将仅对应于单个会议参与者。在这种实例中，这种变型是对于单个会议参与者的两个语音段在时间上重叠来表达上述规则的另一种方式。然而，在一些实施方式中，即使对应于多个会议参与者的单个端点也可以应用该变型。

[0485] 在一些实现中，一组感知激发的规则可以试图防止在多个会议参与者之间讨论和/或交互期间的发言顺序以不自然的方式变得无序。例如，在现实世界中，一个会议参与者可在另一个会议参与者结束阐述问题之前回答该问题。然而，通常不会期望在问题本身之前听到该问题的完整答案。

[0486] 考虑两个连续的输入讲话突发A和B，其中讲话突发A发生在讲话突发B之前。根据一些实现，感知激发规则的集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。

[0487] 在一些实现中，上限(在本文中有时称为T)可以施加于任何两个连续的输入讲话突发(例如A和B)之间引入的重叠量，以便防止在多个会议参与者之间进行讨论和/或交互期间回放具有显著的非因果关系。因此，在一些示例中，感知激发规则的集合可以包括可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前的时间T开始。

[0488] 在某些实例中，记录的音频数据可以包括先前在时间上(在原始会议期间)重叠的输入讲话突发。在某些实现中，感知激发规则的集合可以包括一个或多个规则，其指示对应于先前重叠的输入讲话突发的输出讲话突发在回放期间应该保持重叠。在一些示例中，感知激发规则的集合可以包括允许与先前重叠的输入讲话突发对应的输出讲话突发在时间上进一步重叠地回放的规则。这样的规则可能受制于掌控允许的重叠量的一个或多个其他规则，例如上述段落中指出那些。

[0489] 在一些实现中，会议参与者语音中的至少一些可以以比记录会议参与者语音的速率更快地速率被回放。根据一些这样的实现，可以通过使用WSOLA(基于波形相似性的重叠相加)技术来实现更快速率的语音的回放。在替代实现中，可以通过使用诸如间距同步重叠和相加(PSOLA)或相位声码器法的其他时间尺度修正(TSM)方法来实现更快速率的语音的回放。

[0490] 图13是示出在小于输入时间间隔的输出时间间隔期间调度用于回放的会议记录的示例的框图。图13所示的特征的类型和数量仅仅是作为示例被示出的。替代实现可以包括更多、更少和/或不同的特征。

[0491] 在图13所示的示例中，示出了回放调度器1306接收会议记录的输入会议段1301。在该示例中，输入时间间隔1310对应于输入的会议段1301的记录时间间隔。在图13中，输入时间间隔1310从输入时间t_{i0}开始，在输入时间t_{i1}结束。回放调度器1306输出对应的输出回放调度1311，其具有相对于输入时间间隔1310较小的输出时间间隔1320。这里，输出时间间隔1320从输出时间t_{o0}开始，在输出时间t_{o1}结束。

[0492] 回放调度器1306能够至少部分地执行本文公开的各种方法。例如，在一些实现中，回放调度器1306能够至少部分地执行图12的方法1200。回放调度器1306可以根据具体的实施方式在各种硬件，软件，固件等中实现。回放调度器1306可以例如是回放系统的元件的实例，诸如图6所示的回放系统609的回放控制模块605。在替代示例中，回放调度器1306可以至少部分地经由诸如回放控制服务器650或分析引擎307的另一设备和/或模块来实现，或者可以是诸如图3A的控制系统330的另一个设备的组件或者经由该另一设备实现的模块。

[0493] 因此，在一些示例中，回放调度器1306可以包括接口系统和控制系统，诸如图3A所示的那些。接口系统可以包括一个或多个网络接口、控制系统和存储系统之间的一个或多个接口、和/或一个或多个外部设备接口(诸如一个或多个通用串行总线(USB)接口)。控制系统可以例如包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑，和/或离散硬件组件。在一些示例中，回放调度器1306可以根据存储在非暂态介质上的指令(例如，软件)来实现。这种非暂态介质可以包括诸如本文所描述的那些的存储设备，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。

[0494] 在图13所示的示例中，输入会议段1301包括输入会议记录的来自每个端点1302-

1305的输入讲话突发。在一些实现中，端点1302-1305中的每一个可以对应于电话端点，诸如图1A所示的电话端点1。在其他实现中，端点1302-1305中的每一个可以对应于面对面会议端点，诸如图7所示的麦克风715a-715d。这里，输入会议段1301包括来自端点1302的输入讲话突发1302A-1302D，来自端点1303的输入讲话突发1303A-1303C，来自端点1304的输入讲话突发1304A和1304B以及来自端点1305的输入讲话突发1305A和1305B。

[0495] 输入会议段1301和输出回放时间表1311的水平轴表示时间。因此，图13所示的每个讲话突发的水平尺寸对应于讲话突发时间间隔的示例。每个输入讲话突发具有开始时间t_{start}和结束时间t_{end}。例如，输入讲话突发1302B的输入开始时间t_{start}和输入结束时间t_{end}如图13所示。因此，根据一些实现，输入会议段可以被描述为输入讲话突发的列表L_i，每个输入通话突峰T_i具有输入开始时间$t_{start}(Ti)$和输入结束时间$t_{end}(Ti)$，并且与端点相关联。

[0496] 在该示例中，输出回放调度1311指示多个空间端点回放位置1312-1315和相应的输出讲话突发。在一些实现中，空间端点回放位置中的每一个可以对应于虚拟声学空间中的每个会议参与者的虚拟会议参与者位置，例如，如本文别处所描述的。在该示例中，输出回放调度1311包括：输出讲话突发1312A-D，其与端点回放位置1312相关联并分别基于输入讲话突发1302A-D；输出讲话突发1313A-C，其与端点回放位置1313相关联，并且分别基于输入讲话突发1303A-C；输出讲话突发1314A和1314B，其与端点回放位置1314相关联，并且分别基于输入讲话突发1304A和1304B；以及输出讲话突发1315A和1315B，其与端点回放位置1315相关联并且分别基于输入讲话突发1305A和1305B。

[0497] 每个输出讲话突发具有开始时间t_{start}和结束时间t_{end}。例如，输出讲话突发1315A的输出开始时间t_{start}和输出结束时间t_{end}如图13所示。因此，根据一些实现，可以将输出回放调度描述为输出讲话突发的列表L_o，每个输出讲话突发T_o具有输出开始时间$t_{start}(T_o)$和输出结束时间$t_{end}(T_o)$，并且与端点和空间端点回放位置相关联。每个输出讲话突发也可以与相应的输入讲话突发input(T_i)相关联，并且可以被调度为在输出时间$t_{start}(T_o)$播放。

[0498] 取决于具体的实现方式，回放调度器1306可以根据各种方法使输出时间间隔1320小于输入时间间隔1310。例如，可以至少部分地通过删除与至少一些输入讲话突发之间的非语音间隔或“间隙”相对应的音频数据，使得输出时间间隔1320小于输入时间间隔1310。一些替代实现还可以涉及删除对应于至少一些会议参与者发声(例如笑声)的音频数据。通过将输入会议段1301与输出回放调度1311进行比较，可以看出，输入讲话突发1302A，1302B和1302C在它们之间具有间隙，但是回放调度器1306已经去除了对应的输出讲话突发1303A-1303C之间的间隙。

[0499] 此外，在图13所示的例子中，以前在时间上不重叠的会议参与者语音中的至少一些被调度以重叠的方式回放。例如，通过将输入会议段1301与输出回放调度1311进行比较，可以看出，输入讲话突发1302A和1303A以前在时间上没有重叠，但是回放调度器1306已经调度了相应的输出讲话突发1312A和1313A在回放过程中在时间上重叠。

[0500] 在该示例中，回放调度器1306根据感知激发规则的集合调度各种输出讲话突发在回放期间时间重叠。在该实现中，回放调度器1306调度输出讲话突发回放，使得对应于单个端点的两个语音段不应该在时间上重叠。例如，虽然回放调度器1306已经去除了全部对应于端点1302的对应的输出讲话突发1303A-1303C之间的间隙，但是回放调度器1306没有使输出通话单元1303A-1303C中任何一个重叠。

[0501] 此外，回放调度器1306调度输出讲话突发回放，使得给定两个连续的输入讲话突发A和B，讲话突发A发生在讲话突发B之前，对应于B的输出讲话突发的回放可以在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放开始之前。例如，连续的输入通话突峰1302C和1303B对应于重叠的输出讲话突发1312C和1313B。这里，回放调度器1306已经调度了输出讲话突发1313B在输出讲话突发1313C的回放完成之前开始，而不早于输出通话突发1313C的回放开始。

[0502] 在一些实现中，回放调度器1306可以调度输出讲话突发以为原始语音速率倍数的速度因子S来回放。例如，在图13中可以看出，输出讲话突发1312A-1312D被调度为在比对应的输入讲话突发1302A-1302D的间隔更短的时间间隔内回放。在一些实现中，回放调度器1306可以根据WSOLA方法或通过使用诸如PSOLA或相位声码器法的另一时间尺度修正(TSM)方法，使得以更快的速率回放语音。

[0503] 给定输入讲话突发的列表L_i，速度因子S，重叠时间t_{over}和输出开始时间t_{o0}，根据一些实现，回放调度器1306可以如下操作。回放调度器1306可以将最新输入时间t_{i1}初始化输入段的开始时间t_{i0}。回放调度器1306可以将对于每个端点的最新输出时间$t_{out,e}$初始化为t_{o0}。回放调度器1306可以将输出重叠时间t_{over}初始化为t_{o0}。回放调度器1306可以将输出结束时间t_{o1}初始化为t_{o0}。回放调度器1306可以将输出讲话突发的列表L_o初始化为空列表。

[0504] 每个输入讲话突发T_i可以按照输入开始时间的顺序被考虑。在一些示例中，对于每个输入会话突峰T_i，回放调度器1306可以如下确定用于回放的输出讲话突发T_o的暂定开始回放时间：

[0505] $$t'_{start}(T_o) = \min\left(t_{oover}, t_{o1} - \frac{\max(t_{i1} - t_{start}(T_i), 0)}{S}\right) \quad (式\ 38)$$

[0506] 在式38中，$t'_{start}(T_o)$表示输出讲话突发T_o的暂定开始回放时间，$t_{start}(T_i)$表示输入讲话突发T_i的开始时间，S代表速度因子，它可以表示为要回放输出讲话突发的原始语音速率的倍数。在式38的示例中，min()的第二自变量(argument)的影响是在输出回放调度1311中，根据以下感知激发规则来保持输入讲话突发T_i和最新结束的已考虑的输入讲话突发之间的时间关系，该激发规则为：(a)当考虑两个连续的输入讲话突发A和B进行重叠时，不允许与B对应的输出讲话突发开始回放，直到对应于A的输出讲话突发的回放开始之后的预定时间；和(b)当两个输入讲话突发在输入时间上重叠时，对应的输出讲话突发应保持重叠，在输出时间具有类似的时间关系。

[0507] 图14示出了保持重叠的输入讲话突发和重叠的输出讲话突发之间的类似时间关系的示例。在该示例中，回放调度器1306正在评估输入通话突峰1402A。因此，输入讲话突发1402A是输入讲话突发T_i的一个例子。在该示例中，与输入讲话突发1402A在时间上重叠的最新结束且已经考虑的输入讲话突发1401A在输入时间t_{i1}结束。这里，回放调度器1306已经调度与输入讲话突发1401A相对应的输出讲话突发1401B，以在输出时间t_{o1}结束。

[0508] 在图14中，输出讲话突发1402B是与输入讲话突发T_i对应的输出讲话突发T_o的示例。在该示例中，回放调度器1306根据式38调度输出讲话突发1402B的暂定开始回放时间。由于式38中的min()的第二自变量，输出讲话突发1402B已被调度为与1401B重叠($t_{o1}-t_{start}(T_o)$)，其等于按速度因子S缩放的输入讲话突发1402A与输入讲话突发1401A的重叠时间量(($t_{i1}-t_{start}(T_i)$)。

[0509] 回放调度器1306可以通过式38来实现其他感知激发规则。一种这样的感知激发规则可以是给定两个连续的输入讲话突发A和B，A在B之前发生，对应于B的输出讲话突发的回放可以不早于在对应于A的输出讲话突发的回放完成之前的预定时间开始。在一些示例中，即使输入讲话突发A和B最初没有重叠，也可以应用这种感知激发规则。

[0510] 图15示出了确定用于不重叠的输入讲话突发的重叠量的示例。在该实现中，回放调度器1306根据式38确定输出讲话突发T_o的输出时间。这里，输出讲话突发1501是最新结束的输出讲话突发。在该示例中，根据式38中的min()的第二自变量，块1502A对应于用于输出讲话突发T_o的暂定开始回放时间。然而，在该示例中，如块1502B所示，输出讲话突发T_o的开始回放时间被临时设置在时间t_{oover}，以便与输出讲话突发1501以重叠时间t_{over}重叠：在该示例中，由于式38中的min()的运算，$t'_{start}(T_o)=t_{oover}$。

[0511] 回放调度器1306可以实现其他感知激发规则。图16是示出应用感知激发规则以避免来自同一端点的输出讲话突发重叠的示例的框图。在该示例中，回放调度器1306通过如下地确保输出讲话突发T_o不会与来自相同端点e的任何已经调度的输出讲话突发重叠来实现该规则：

[0512] $t_{start}(T_o)=\max(t'_{start}(T_o),t_{out,e})$ (式39)

[0513] 在图16所示的示例中，通过式38的运算，如块1602A的位置所示，用于输出讲话突发T_o的开始回放时间的初始候选被设置为$t'_{start}(T_o)$。然而，在该示例中，来自同一端点的输出讲话突发1601已经被调度为回放，直到在$t'_{start}(T_o)$之后的时间$t_{out,e}$。因此，通过式39的运算，输出讲话突发T_o被调度为在时间$t_{start}(T_o)$开始回放，如块1602B的位置所示。

[0514] 在一些示例中，输出讲话突发T_o的输出结束时间可以如下计算：

[0515] $$t_{end}(T_o)=t_{start}(T_o)+\frac{(t_{end}(T_i)-t_{start}(T_i))}{S} \quad (式40)$$

[0516] 在式40的示例中，$t_{end}(T_o)$表示输出讲话突发T_o的输出结束时间。在该示例中，通过将输入讲话突发时间间隔($t_{end}(T_i)-t_{start}(T_i)$)除以速度因子S来减小在其期间输出讲话突发$T_o$被调度回放的时间间隔。

[0517] 在一些实现中，然后将输出讲话突发T_o添加到输出讲话突发列表L_o。在一些示例中，讲话突发T_o的对于端点e的最新输出时间可以根据下式被更新：

[0518] $t_{out,e}=t_{end}(T_o)$ (式41)

[0519] 在一些示例中，输出重叠时间可以根据下式来更新：

[0520] $t_{oover}=\max(t_{oover},t_{end}(T_o)-t_{over})$ (式42)

[0521] 根据一些实现，最新的输入结束时间可以根据以下来更新：

[0522] $t_{i1}=\max(t_{i1},t_{start}(T_i))$ (式43)

[0523] 在某些实例中，最新的输出结束时间可根据以下更新：

[0524] $t_{o1}=\max(t_{o1},t_{end}(T_o))$ (式44)

[0525] 可以重复上述过程，直到已经处理了所有输入讲话突发。已排定的输出列表L_o可被返回。

[0526] 一些会议可能涉及多个会议参与者的陈述。如本文所使用的，“陈述”可以对应于延长时间间隔(其可以例如为几分钟或更长)，在其期间单个会议参与者是主要发言者，或在某些实例中是唯一的发言者。在一些实现中，感知激发规则的集合可以包括允许来自不

同会议参与者的全部陈述并发地回放的规则。根据一些这样的实施方式，会议参与者语音中的至少一些可被以比记录会议参与者语音的速率更快的速率回放。

[0527] 图17是示出能够调度来自不同会议参与者的全部陈述的并发回放的系统的示例的框图。图17中所示的特征的类型和数量仅作为示例示出。替代实现可以包括更多，更少和/或不同的特征。

[0528] 在图17所示的示例中，系统1700包括段调度器单元1710，其被示出为接收分段的会议记录1706A。在一些示例中，分段的会议记录1706A可以根据会话动态数据被分段，以允许讨论，陈述和/或其他类型的会议段被标识。下面提供了会话动态数据的会议分段的一些示例。在该示例中，分段的会议记录1706A包括讨论段1701A，其后是陈述段1702A-1704A，其后是讨论段1705A。

[0529] 段调度器单元1710和系统1700的其他元素能够至少部分地执行本文公开的各种方法。例如，在一些实现中，段调度器单元1710和系统1700的其他元件可以能够调度分段会议记录的段，以用于来自不同会议参与者的陈述的并发回放。取决于具体实现，段调度器单元1710和系统1700的其他元件可以以各种硬件，软件，固件等来实现。例如，段调度器单元1710和/或系统1700的其他元件可以经由通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑，和/或离散硬件组件来实现。在一些示例中，段调度器单元1710和/或系统1700的其他元件可以根据存储在非暂态介质上的指令(例如，软件)来实现。这种非暂态介质可以包括诸如本文所描述的那些的存储设备，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。段调度器单元1710和/或系统1700的其他元件可以例如是诸如图6所示的回放控制模块605的回放系统609的组件。在替代示例中，段调度器单元1710和/或系统1700的其他元件可以被实现在另一设备或模块中，诸如回放控制服务器650或分析引擎307，或者可以由另一设备或模块(例如图3A的控制系统330)实现。

[0530] 在图17所示的示例中，段调度器单元1710能够确定是否存在能够并行播放的分别由不同的陈述者陈述的连续陈述段。这里，该处理的结果是段调度1706B。在该实现中，段调度1706B包括基于讨论段1701A并将首先被回放的讨论段1701B。这里，段调度1706B分别包括基于陈述段1702A-1704A的陈述段1702B-1704B。在此实现中，陈述段1702B-1704B将与讨论段1701B同时地播放并且在其之后还播放。

[0531] 在该示例中，插话过滤模块1702C-1704C能够从陈述段1702B-1704B去除插话。在这里，插话是并非“陈述者”(正进行陈述的会议参与者)的语音的讲话突发。在一些实现中，可以不从陈述段去除插话，例如在陈述段不被调度为与另一个陈述段并行地回放的情况下。因此，插话过滤模块1702C-1704C可以确保来自同一端点的语音不被同时回放。

[0532] 在该实现中，系统1700包括回放调度器单元1306，诸如图13所示的。这里，回放调度器单元1306包括模块1701D-1705D，每个模块能够独立地调度会议段中的一个以供回放。模块1701D和1705D分别接收讨论段1701B和1705B，并输出相应的讨论回放安排(schedule)170IF和1705F。模块1702D-1704D接收对应于陈述段1702B-1704B的插话过滤模块1702C-1704C的输出，并输出相应的独立的陈述回放安排。在一些替代实现中，可以为每个段创建回放调度器单元1306的单独实例。在一些实现中，每个段可以依次传递给调度器功能，使得调度过程对于每个段都重新开始。

[0533] 在该示例中，系统1700还包括合并单元1702E。这里，合并单元1702E能够将要并发地回放的段的回放安排(输出时间)合并成单个回放安排。在该实现中，模块1702D-1704D向合并单元1702E提供对应于陈述段1702B-1704B的独立的陈述回放安排，其输出合并的渲染回放安排1702F。在该示例中，合并的陈述回放安排1702F具有等于任何输入安排的最大长度的长度。

[0534] 在图17所示的实现中，系统1700包括拼接单元1706G。在该示例中，拼接单元1706G能够拼接第一讨论回放安排1701F、合并的陈述回放安排1702F和第二讨论回放安排1705F，并输出单个输出回放安排1706H。

[0535] 根据段调度器单元1710的一些实现，输出安排1076H可以被初始化为空列表。调度器单元1710可以按顺序处理会议记录的每个段，依次考虑每个段。当所考虑的段不是陈述段时，其可被调度以产生段安排(例如，1701F)，然后以适当的输出时间偏移量拼接到输出回放安排1076H，使得该段被调度为在输出回放安排1076H中的当前最后一个讲话突发之后开始。段调度器单元1710然后可以继续下一个段。

[0536] 当所考虑的段是陈述安排时，段调度器单元1710也可以考虑之后的段，只要它们是来自不同陈述者的陈述即可。一旦已经发现可以并行回放的陈述段的运行，则可以针对每个陈述段进行插话过滤，然后使用回放调度器605分别进行调度。合并单元1702E然后可以通过将所有对应的输出讲话突发组合成按输出开始时间排序的单个列表来合并来自每个陈述段的安排。然后，拼接单元1706G可以将合并的陈述调度以适当的输出时间偏移量拼接到输出安排1076H，使得它们在输出安排中的当前最后一个讲话突发之后开始。段调度器单元1710然后可以继续下一个段。

[0537] 听众通常难以在不收听整个录音的情况下在会议记录中找到感兴趣的区域。在听众没有出席会议的情况下尤其如此。本公开引入了各种新颖的技术来帮助听众在会议记录中找到感兴趣的区域。

[0538] 本文描述的各种实现涉及将会议记录分类为不同的段，这是基于看起来在各段中主导地发生的人交互的类别的。这些段可以对应于与人交互的类别对应的时间间隔和至少一个段分类。例如，如果从时间T1到时间T2，会议参与者A似乎已经做出了陈述，则可以在从时间T1到时间T2的时间间隔中标识“陈述”段。陈述段可以与会议参与者A相关联。如果会议参与者A似乎已经从时间T2到时间T3回答他或她的观众的问题，则可以在从时间T2到时间T3的时间间隔中标识“问答”或“Q&A”段。Q&A段可能与会议参与者A相关联。如果在时间T3之后的会议记录的剩余时间期间，会议参与者A似乎已经参与了与其他会议参与者的讨论，则可以在时间T3之后的时间间隔中标识“讨论”。讨论段可能与参与讨论的会议参与者相关联。

[0539] 所得的会议记录的分段可能以各种方式潜在地有用。分段可以补充基于内容的搜索技术，例如关键词检索和/或主题确定。例如，不是在全长3小时的会议录音中搜索术语“helicopter(直升飞机)”，一些实现可允许听众在该记录内来自特定会议参与者的特定的30分钟陈述中搜索术语“helicopter”。以这种方式进一步改进搜索的能力可以减少在电话会议记录中找到感兴趣的特定区域和/或事件所花费的时间。

[0540] 本文公开的一些回放系统实现提供图形用户界面，其可以包括会议段的视觉描绘。在这种实现中，会议段的可视描述对于向回放系统的用户提供会议的事件的可视概述

可能是有用的。该可视概述可以帮助用户浏览会议内容。例如，一些实现可以允许听众浏览所有讨论段和/或涉及特定会议参与者的所有讨论段。

[0541] 此外，这种会议分段在下游注释和搜索技术中可能是有用的。例如，一旦会议已经基于会话动态被分解成段，则可以通过利用自动语音识别来向用户指示该分段期间涵盖什么主题的想法。例如，收听者可能希望浏览涉及特定主题的所有表现段或讨论段。

[0542] 图18A是简述会议分段方法的一个示例的流程图。在一些示例中，方法1800可以由装置(诸如图3A的装置和/或图1A或图3C的分析引擎307的一个或多个组件)执行。

[0543] 在一些实现中，方法1800可以由至少一个设备根据存储在一个或多个非暂态介质上的软件执行。与本文所述的其它方法一样，方法1800的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0544] 在该实现中，块1805涉及接收对应于涉及多个会议参与者的会议的记录的音频数据。在该示例中，音频数据包括：(a)分别记录的来自多个端点的会议参与者语音数据；和/或(b)来自对应于多个会议参与者的单个端点的会议参与者语音数据。

[0545] 在一些实现中，音频数据可以包括语音活动检测过程的输出。因此，在一些这样的实现中，音频数据包括语音和/或非语音分量的指示。然而，如果音频数据不包括语音活动检测过程的输出，则在一些示例中，方法1800可以包括语音活动检测过程。

[0546] 根据图18A所示的示例，来自对应于多个会议参与者的单个端点的会议参与者语音数据还包括用于标识多个会议参与者中的每个会议参与者的会议参与者语音的信息。这样的信息可以从发言者日志过程输出。然而，如果音频数据不包括来自发言者日志过程的输出，则在一些示例中，方法1800可以包括发言者日志过程。

[0547] 在一些实现中，在块1805中，诸如图3A的控制系统330的控制系统可经由接口系统325接收音频数据。在一些示例中，控制系统可以能够执行方法1800的块1805-1820。在一些实现中，控制系统可以能够执行本文公开的其他与分段有关的方法，诸如本文参考图18B-23所描述的那些。在一些示例中，方法1800可以至少部分地由联合分析模块306的一个或多个组件执行，诸如图5的会话动态分析模块510。根据一些这样的实施方式，块1805可以包括会话动态分析模块510接收音频数据。

[0548] 在一些实现中，会议可以是电话会议，而在其他实现中，会议可以是面对面会议。根据一些示例，音频数据可以对应于完整的或基本上完整的会议的记录。

[0549] 在该示例中，块1810涉及分析音频数据以确定会话动态数据。在一些实例中，会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据；指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的双讲话数据等等。在一些实现中，块1810可以涉及确定双讲话比率，其可指示在该时间间隔中的在其期间至少两个会议参与者同时发言的语音时间的部分。

[0550] 这里描述的一些实现涉及评估分析音频数据以确定其他类型的会话动态数据。例如，在一些实现中，在块1810中确定的会话动态数据可以包括语音密度度量，其指示该时间间隔的存在任何会议参与者语音的部分。在一些实现中，块1810可以涉及确定主导度量，其指示在该时间间隔期间的由主导会议参与者发出的总语音的部分。主导会议参与者可以是在时间间隔期间发言最多的会议参与者。

[0551] 在该实现中，块1815涉及搜索会议记录以确定多个段分类中的每一个的实例。在

该示例中，每个段分类至少部分地基于会话动态数据。以下描述各种示例。

[0552] 在一些实施方式中，块1815可以涉及确定混串音段的实例，混串音段是在其期间至少两个会议参与者正在并发地讲话的段。在一些示例中，混串音段可以根据双讲话数据的实例被标识，例如在阈值时间间隔期间连续的双讲话的实例和/或其中存在双讲话的时间间隔的一部分。在实质性讨论、陈述等之前，常常会在会议(特别是包含至少一个多方端点的会议)开始时找到混串音段。

[0553] 根据一些实现，块1815可以涉及确定相互静默段的实例，这些段是在其期间语音可忽略不计(例如，小于相互静默阈值量)的时间间隔。这例如可能发生在电话会议中当一个会议参与者在未被注意地情况下暂时离开他或她的端点而其他人等待他或她返回时和/或当一个会议参与者正在等待他人加入电话会议时。在某些实现中，相互静默段可以至少部分地基于可以在块1810中确定的语音密度度量。

[0554] 部分地由于它们独特的会话动态特性，混串音段的实例可被以高置信水平来标识，并且相互静默段的实例可以用非常高的置信水平来标识。此外，可以用相对较高的置信水平来标识混串音段和相互静默段的开始时间和结束时间。由于混串音段包括对应于感兴趣的会议主题的可理解语音的可能性较低，而相互静默分段包括与感兴趣的会议主题相对应的任何语音的可能性非常低，所以查看会议记录的人可以是有理由相信他或她可以安全地省略对这些会议部分的查看。因此，在回放会议记录期间，标识混串音段和相互静默段可能会导致听众节省时间。

[0555] 在一些实现中，块1815可以涉及确定陈述段的实例，该陈述段是在其中一个会议参与者正在进行绝大多数讲话、而其他会议参与者保持基本上静默的段。根据一些实现方式，确定陈述段的实例可以至少部分地基于语音密度度量和主导度量。陈述通常包含很小的双讲话。因此，在一些实现中，确定陈述段的实例可以至少部分地基于双讲话量度，例如双讲话比率。

[0556] 部分地由于它们独特的会话动态特征，陈述段的实例可以用较高的置信水平来标识。在一些实现中，可以以相当高的置信水平来标识陈述段的开始时间和结束时间，但是该置信水平通常比可标识混串音段和相互静默段的使用时间和结束时间的置信水平更低。因为陈述段包含与感兴趣的会议主题相对应的语音的可能性很高，所以查看者标识这样的会议段可能是有利的。在提供关于会议段的附加信息的实现(例如涉及关键词标识、主题确定等的实现)中，这样的潜在优点可被增强。例如，听众可能选择仅查看其中发出特定词语的陈述段，或者在其中讨论了特定主题的陈述段。因此，在回放会议记录期间，标识陈述段段可导致听众的时间节省。

[0557] 在一些实现中，块1815可以涉及确定讨论段的实例，该讨论段是在其期间多个会议参与者发言、但是没有单个会议参与者占据任何明显的主导的段。根据一些实现，确定讨论段的实例可以至少部分地基于语音密度度量和主导度量。一些讨论可能涉及大量的双讲话，但通常不如混串音段的双讲话那么多。因此，在一些实现中，确定讨论段的实例可以至少部分地基于双讲话量度，例如双讲话比率。

[0558] 在一些实现中，块1815可以涉及确定Q&A段的实例，该Q&A段是与如下时间间隔相对应的段，在该事件间隔期间多个会议参与者提出问题，而单个会议参与者进行回复或较小会议参与者子集中的一个参与者进行回复。例如，Q&A段常常可能在陈述段结束之后。在

陈述结束后，陈述会议参与者可以回答正在听陈述的其他会议参与者所提出的问题。在问答环节期间，单个会议参与者常常进行回复，因此会议参与者可能会比任何其他会议参与者做更多的发言。因此，主导度量可能小于关于陈述的主导度量，而大于关于讨论的主导度量。因此，根据一些实现，确定Q&A段的实例可以至少部分地基于语音密度度量和主导度量。有时在问答环节可能会有大量的双讲话(例如，比陈述期间更多的双讲话)，但在问答环节中可能会有比在讨论期间更少的双讲话。因此，在某些实现中，确定Q&A段的实例可以至少部分地基于双讲话量度，例如双讲话比率。

[0559] 在一些实现中，讨论段和Q&A段可能不被以与例如相互静默段，混串音段或甚至陈述段相同的置信水平标识。在一些实现中，讨论段和Q&A段的开始时间和结束时间可被以中等的置信水平来标识，但是通常以比可标识混串音段和相互静默段的开始时间和结束时间的置信水平更低的置信水平来标识。然而，由于讨论段或Q&A段可能包括与感兴趣的会议主题相对应的语音具有合理的可能性，所以查看者标识这样的会议段是有利的。在提供关于会议段的附加信息的实现(例如涉及关键词标识、主题确定等的实现)中，这样的潜在优点可被增强。例如，听众可以选择仅查看(review)其中发出特定词语或讨论了特定主题的陈述段，讨论段和/或Q&A段。因此，标识讨论段和/或Q&A段可以在会议记录回放期间节省收听者的时间。

[0560] 这里，块1820涉及将会议记录分段成多个段。在该示例中，每个段对应于段分类中的至少一个和时间间隔。段可以对应于附加信息，例如在该段期间发言的会议参与者(如果有的话)。

[0561] 根据一些实现，搜索和/或分段过程可以是递归的。在一些实现中，分析、搜索和分段过程可以都是递归的。以下提供了各种示例。

[0562] 在下面的描述中，可以观察到若干搜索过程可以涉及时间阈值(诸如t_{min}和t_{snap})，这将在下面描述。这些时间阈值具有将段的大小限制为不小于阈值时间的效果。根据一些实现，当向用户显示分段过程的结果时(例如，当图6的回放系统609使得在显示器上提供相应的图形用户界面时)，用户可能能够适时地缩放(例如，通过与触摸屏进行交互，通过使用鼠标或通过激活放大或缩小命令)。在这种实例中，可能期望以不同的时间尺度(可能涉及应用t_{min}和t_{snap}的不同值)多次执行分段过程。在回放期间，可能有利的是在不同时间尺度的分段结果之间动态切换，其结果可以基于当前缩放水平被显示给用户。根据一些示例，该过程可能涉及选择将不包含在当前缩放水平下在宽度上占据小于X个像素的段的分段时间尺度。X的值至少部分地基于显示器的分辨率和/或尺寸。在一个示例中，X可以等于100个像素。在替代示例中，X可以等于50个像素，150个像素，200个像素，250个像素，300个像素，350个像素，400个像素，450个像素，500个像素或某些其他数量的像素。如图5所示的会话动态数据文件515a-515e是在不同时间尺度的分段结果的示例，其可用于基于当前缩放水平来快速调整显示。

[0563] 然而，在其他实现中，块1810-1820可以不递归地执行，而是可以各自执行预定次数，例如仅一次，仅两次等。作为替代地或者附加地，在一些实现中，块1810-1820可以仅在一个时间尺度上执行。这样的实现的输出可能不像递归过程那样准确或那样方便于收听者。然而，一些这样的实现可以比递归实现和/或对于多个时间尺度执行的实现更快地执行。作为替代地或者附加地，这样的实现可能比递归实现和/或对于多个时间尺度执行的实

现更简单。

[0564] 在一些实现中，搜索和分段过程(并且在一些实现中，分析过程)可以至少部分地基于段分类的层级结构。根据一些实现，分析，搜索和分段过程全部可以至少部分地基于段分类的层级结构。如上所述，可以以变化的置信度来标识不同的段类型，以及不同段类型的开始和结束时间。因此，根据一些实现，段分类的层级结构至少部分地基于特定段分类的段可被标识的置信水平，段的开始时间可被确定的置信水平、和/或段的结束时间可被确定的置信水平。

[0565] 例如，段分类的层级结构的第一或最高级别可以与混串音段或相互静默段(可被用高(或非常高)置信水平来表示)相对应。混串音段和相互静默段的开始和结束时间也可以用高(或非常高)的置信水平确定。因此，在一些实现中，搜索和分段过程(以及在一些实现中，分析过程)的第一阶段可以涉及定位混串音段或相互静默段。

[0566] 此外，不同的段类型具有不同的包括感兴趣的主旨(例如对应于会议主题的会议参与者语音)，感兴趣的关键词等的可能性。标识哪些会议段可以被跳过以及哪些会议段可能包括感兴趣的主旨可能是有利的。例如，混串音段和相互静默段具有低的或者非常低的包括对应于会议主题的会议参与者语音、感兴趣的关键词等的可能性。陈述段可能具有高的包括对应于会议主题的会议参与者语音，感兴趣的关键词等的可能性。因此，根据一些实现，段分类的层级结构至少部分地基于特定段分类包括对应于会议主题的会议参与者语音的可能性。

[0567] 根据一些实现，搜索和分段过程(以及在一些实现中，分析过程)可以包括首先定位混串音段，然后是陈述段，然后是Q&A段，然后是其他段。这些过程可以是递归过程。其他实现可以涉及以一个或多个不同序列定位段。

[0568] 图18B示出了用于至少部分地执行本文所述的会议分段方法和相关方法中的一些的系统的示例。与本文提供的其它图一样，图18B所示的元件的数量和类型仅仅是作为示例被示出的。在该示例中，音频记录1801A-1803A正由发言者日志单元1801B-1803B接收。在一些实现中，音频记录1801A-1803A可以与上面参考图3C和图4描述的分组跟踪文件201B-205B相对应，每个分组跟踪文件可以对应于上行链路数据分组流201A-205A之一。在一些实现中，发言者日志单元1801B-1803B可以是图4所示的发言者日志模块407的实例。

[0569] 在该示例中，音频记录1801A-1803A中的每一个来自电话端点。这里，音频记录1801A是来自多方端点(例如扬声器电话)的记录，而音频记录1802A和1803A是单方端点(例如标准电话和/或耳机)的记录。

[0570] 在该示例中，发言者日志单元1801B-1803B能够确定每个会议参与者何时发出语音。当处理来自单方端点的音频数据(诸如音频记录1802B和1803B)时，发言者日志单元1802B和1803B可以用作语音活动检测器。当处理来自多方端点的音频数据(诸如音频记录1801A)时，发言者日志单元1801C可以估计存在多少个会议参与者(例如，在会议期间有多少个会议参与者在发言)，并且可以尝试标识哪个会议参与者发出了每个讲话突发。在一些实现中，发言者日志单元1801B-1803B可以使用本领域普通技术人员已知的方法。例如，在一些实现中，发言者日志单元1801B-1803B可以使用高斯混合模型来对每个讲话者进行建模，并且可以根据隐马尔可夫模型为每个讲话者分配相应的讲话突发。

[0571] 在图18B所示的实现中，发言者日志单元1801B-1803B输出发言者活动文档1801C-

1803C。这里，发言者活动文件1801C-1803C中的每一个指示在相应端点每个会议参与者何时发出了语音。在一些实施方式中，发言者活动文件1801C-1803C可以是可用于图5所示的联合分析401-405的上行链路分析结果的实例。

[0572] 在该示例中，发言者活动文档1801C-1803C由分段单元1804接收以用于进一步处理。分段单元1804产生至少部分地基于发言者活动文档1801C-1803C的分段记录1808。在一些实现中，分段单元1804可以是图5的会话动态分析模块510的实例。在一些这样的实现中，分段记录1808可以是被示出为由图5中的会话动态分析模块510输出的会话动态数据文件515a-515e之一的实例。

[0573] 根据具体示例，分段单元1804和发言者日志单元1801B-1803B可以通过硬件，软件和/或固件来实现，例如经由可以包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑、或离散硬件组件中的至少一个的控制系统的一部分来实现。在一些示例中，分段单元1804和发言者日志单元1801A-1803B可以根据存储在诸如随机存取存储器(RAM)设备，只读存储器(ROM)设备等非暂态介质上的指令(例如，软件)来实现。

[0574] 在该示例中，分段单元1804包括合并单元1806，其能够将多个发言者活动文档1801C-1803C组合成全局发言者活动图1809。图18B中示出了关于对应于该示例中的整个会议的从t_0到t_1的时间间隔的全局发言者活动图1809。全局发言者活动图1809指示在会议期间哪个会议参与者在哪个时间间隔期间以及在哪个端点处讲话。

[0575] 在该示例中，分段单元1804包括分段引擎1807，其能够执行诸如上面参考图18A所描述的那些的分析，搜索和分段过程。分析，搜索和分段过程有时在这里可以被统称为“分段过程”。在这种实现中，分段引擎1807能够执行分层级的和递归的分段过程，从定位混串音段的过程开始。在替代实现中，分段引擎1807可以从定位另一段分类(诸如相互静默段或陈述段)的过程开始。

[0576] 在该示例中，分段记录1808是在会议中发现的段1808A-1808F的列表。这里，段1808A-1808F中的每一个具有开始时间，结束时间和段分类标识符。在该示例中，段分类标识符将指示该段是相互静默段，混串音段，陈述段，讨论段还是问答(Q&A)段。其他实现可能涉及更多或更少的段分类。在该示例中，段1808A和1808F是混串音段，段1808B和1808D是陈述段，段1808C是Q&A段，段1808E是讨论段。

[0577] 图19简述了根据本文公开的一些实现的分段过程的初始阶段。根据一些这样的实现，分段处理的所有阶段可以至少部分地由图18B的分段引擎1807执行。在该示例中，分段引擎1807能够执行从“产生混串音(Make Babble)”过程1901开始的递归分段处理。在这个例子中，已经对包含用于产生混串音过程1901的指令的子例程进行了函数调用。这里，根据产生混串音过程1901的结果，产生混串音过程1901生成包含一个或多个混串音段的部分分段记录1903A或不包含混串音段的部分分段记录1903B。

[0578] 这里，因为这是分段过程的第一和最高级部分，输入到产生混串音过程1901的发言者活动图是全局发言者活动图1809，其指示对于整个会议的发言者活动。因此，在该示例中，时间t_0到t_i之间的时间间隔包括整个会议。然而，在其他示例中，产生混串音过程1901可以接收具有较小时间间隔的发言者活动图，以便生成对应于较小时间尺度的部分分段记录。

[0579] 在这个例子中，产生混串音过程1901包括最长混串音段搜索过程1904。在该示例中，最长混串音段搜索处理1904能够搜索全局发言者活动图1809以在时间t_0和t_1之间定位最长混串音段。如果不能找到合适的混串音段，则不包含混串音段的部分分段记录1903B被传送到下面参照图20描述的进行陈述过程2001。

[0580] 然而，在这个例子中，最长混串音段搜索过程1904定位最长混串音段1906B 1，其具有开始时间t_2和结束时间t_3，其输入到部分分段记录1903A。这里，在先发言者活动图1906A是输入的全局发言者活动图1809的在最长混串音段1906B1的时间间隔之前的时间间隔(从时间t_0到时间t_2)期间的剩余未分段部分。在该示例中，后续发言者活动图1906C是输入的全局发言器活动图1809的在最长混串音段1906B1的时间间隔之后的时间间隔(从时间t_3到时间t_1)期间的剩余未分段部分。在先发言者活动图1906A和后续发言者活动图1906C可以作为输入被提供给“产生混串音”过程1901的一个或多个后续递归。

[0581] 然而，根据一些实现，可以评估在先发言者活动图1906A和后续发言者活动图1906C的时间间隔以被评估以确定它们是否短于阈值t_{snap}。例如，如果确定在先发言者活动图1906A的时间间隔比阈值t_{snap}短，则最长混串音比特段1906B将“咬合(snap)”，以通过使$t_2=t_0$来跨越在先发言者活动图1906A的时间间隔。否则，在先发言者活动图1906A被输入到在先发言者活动递归1907A。根据一些这样的实施方式，如果后续发言者活动图1906C的时间间隔短于阈值t_{snap}，则最长混串音段1906B将“咬合”以通过让$t_3=t_1$来跨越随后的发言者活动图1906C的时间间隔。否则，后续发言者活动图1906C被输入到后续发言者活动递归1907C。

[0582] 在图19所示的示例中，在先发言者活动图1906A和后续发言者活动图1906C的时间间隔都比阈值t_{snap}长。这里，在先发言者活动递归1907A输出在先部分分段记录1908A，其包括附加的混串音段1906B2和1906B3，它们在图19中示出为具有与最长混串音段1906B1相同类型的填充。在该示例中，后续发言者活动递归1907C输出后续部分分段记录1908C，其包括附加的混串音段实例。这些混串音段也在图19中示出为具有与最长混串音段1906B 1相同类型的填充。在该示例中，在先部分分段记录1908A，最长混串音段1906B 1和后续部分分段记录1908C被拼接以形成部分分段记录1903A。

[0583] 根据一些实现，为了启动最长混串音段搜索过程1904，可以产生双讲话段的列表。例如，双讲话段的列表可以按照双讲话段长度的降序来产生。双讲话段是包括双讲话的实例的会议的部分，在该实例期间至少有两个会议参与者正在同时讲话。可以依次(例如，按长度的降序)考虑这些双字节段中的每一个作为根候选混串音段，并且可以对于每一个进行最长混串音段搜索过程1904。从任何根候选开始发现的最长混串音段被返回。在替代实施例中，搜索可以从每个根候选依次前进，直到它们中的任何一个返回有效的混串音段。找到的第一个混串音段可能会被返回，搜索可能会终止。对于任一类型的实现，如果在搜索每个根候选之后没有发现混串音段，那么最长混串音搜索过程1904可以报告不能发现混串音段，例如，通过输出不包含混串音段的部分分段记录1903B。

[0584] 在一些实现中，为了被包括在候选混串音段中，讲话突发的持续时间必须至少为阈值候选段时间间隔(例如，600ms长，700ms长，800ms长，900ms长，1秒长等)，并且必须被分类为混串音(例如，根据图22中所示的分类器2301的确定)。根据一些示例，候选混串音段可以根据本文中称为“混串音率”的度量(其可以被定义为在其期间存在双讲话的候选段内的

时间占比)被分类为混串音。例如,对于从时间50开始并在时间54结束的候选混串音段(4秒长),关于被分类为混串音的从时间51到53的单个讲话突发(2秒长),该混串音率为50%。一些这样的示例可能要求候选混串音段至少具有阈值混串音率(例如,40%,45%,50%,55%,60%等),以便被分类为混串音段。

[0585] 本文公开的一些实现可以区分混串音率和"双讲话比率",这将在下面更详细地讨论。在一些这样的实现中,双讲话比率是与在其期间存在双讲话的候选段相对应的时间间隔的语音时间的占比(与时间间隔的总持续时间相反)。

[0586] 根据一些实现,可以将持续时间至少为阈值候选段时间的下一个混串音讲话突发添加到先前的候选混串音段,以形成一个新的候选混串音段。在一些示例中,该下一个混串音讲话突发必须在先前候选混串音段的阈值候选段时间间隔内,以便被添加到先前的候选混串音段。

[0587] 同样地,可以将持续时间至少为阈值候选段时间间隔的先前的混串音讲话突发添加到先前的候选混串音段以形成第二新的候选混串音段。在一些示例中,先前的混串音讲话突发必须在先前的候选混串音段的阈值候选段时间间隔内,以便被添加到先前的候选混串音段。因此,根据这样的实现,在每个步骤中可以生成零个、一个或两个候选混串音段。

[0588] 在诸如下面参考图23所描述的替代实现中,可以在一个步骤中评估下一个混串音讲话突发,然后可以在第二步骤中评估先前的混串音讲话突发。根据这样的实现,在每个步骤可以生成零个或一个候选混串音段。

[0589] 图20简述了根据本文公开的一些实现的分段过程的后续阶段。在该示例中,已经对包括用于产生陈述(Make Presentation)过程2001的指令的子例程进行了函数调用。根据一些实现,产生陈述过程2001可以类似于产生混串音过程1901。这里,根据产生陈述过程2001的结果,产生陈述过程2001产生包含一个或多个陈述段的部分分段记录2003A,或者不包含陈述段的部分分段记录2003B。

[0590] 产生陈述过程2001的输入发言者活动图2002可能取决于具体的实现。在一些实现中,输入发言者活动图2002可以是全局发言者活动图1809,其指示整个会议的发言者活动,或者对应于较小时间间隔的发言者活动图。然而,在一些实现中,产生陈述过程2001可以从产生混串音过程接收指示会议的哪些时间间隔(或部分或会议的哪个时间间隔)对应于混串音段的输入。根据一些这样的实现,输入发言者活动图2002可以对应于不对应于混串音段的时间间隔。

[0591] 在这个例子中,产生陈述过程2001包括最长陈述段搜索过程2004。在该示例中,最长陈述段搜索过程2004能够搜索输入的发言者活动图2002以定位在时间t_0到t_1之间的最长陈述段。如果没有找到合适的陈述段,则分段过程可以继续到后续过程,诸如下面参考图21所描述的产生其他(Make other)过程2101。

[0592] 然而,在该示例中,最长陈述段搜索过程2004定位了具有开始时间t_2和结束时间t_3的最长陈述段2006B 1,其进入到部分分段记录2003A中。这里,在先发言者活动图2006A是在最长陈述段2006B1之前的时间间隔(从时间t_0到时间t_2)期间的输入全局发言者活动图1809的剩余未分段部分。在该示例中,后续发言者活动图2006C是在最长陈述段2006B 1之后的时间间隔(从时间t_3到时间t_1)期间的输入全局发言者活动图1809的剩余未分段部分。在先发言者活动图2006A和后续发言者活动图2006C可以作为输入提供给产生陈述过程

2001的一个或多个后续递归。

[0593] 然而，根据一些实施方式，可以评估在先发言者活动图2006A和后续发言者活动图2006C的时间间隔，以确定它们是否短于阈值t_{snap}。例如，如果确定在先发言者活动图2006A的时间间隔比阈值t_{snap}短，则最长的陈述段2006B1将“咬合(snap)”，以通过使$t_2=t_0$来跨越在先发言者活动图1906A的时间间隔。否则，在先发言者活动图2006A被输入到在先发言者活动递归2007A。根据一些这样的实施方式，如果后续发言者活动图2006C的时间间隔短于阈值t_{snap}，则最长陈述段2006B1将“咬合”以通过让$t_3=t_1$来跨越随后的发言者活动图2006C的时间间隔。否则，后续发言者活动图2006C被输入到后续发言者活动递归2007C。

[0594] 在图20所示的示例中，在先发言者活动图2006A和后续发言者活动图2006C的时间间隔都比阈值t_{snap}长。这里，在先发言者活动递归2007A输出在先部分分段记录2008A，其包括附加的陈述段2006B2和2006B3，它们在图20中示出为具有与最长陈述音段2006B 1相同类型的填充。在该示例中，后续发言者活动递归2007C输出后续部分分段记录2008C，其包括附加的陈述段实例。这些陈述段也在图20中示出为具有与最长陈述段2006B 1相同类型的填充。在该示例中，在先部分分段记录2008A，最长陈述段2006B 1和后续部分分段记录2008C被拼接以形成部分分段记录2003A。

[0595] 在一些示例中，当搜索陈述段时，每个根候选段可以是对应于单个讲话突发的段。搜索可以依次(例如，按长度的降序)在每个根候选段处开始，直到搜索到所有根候选，并返回最长陈述。

[0596] 在替代实施例中，搜索可以从每个根候选依次前进，直到它们中的任何一个返回有效的陈述段。找到的第一个陈述段可能会被返回，搜索可能会终止。如果在搜索每个根候选之后没有发现陈述段，那么最长陈述段搜索过程2004可以报告不能发现陈述段(例如，通过输出不包含陈述段的部分分段记录2003B)。

[0597] 根据一些实现，在最长陈述段搜索过程2004中生成候选陈述段可以包括在每个步骤中产生多达两个新候选陈述段。在一些示例中，可以通过采用现有的候选陈述段并且使得结束时间稍晚以包括在被评估的时间间隔(在此也可以称为作为“感兴趣区域”)内由相同参与者发出的下一讲话突发，生成第一新候选陈述段。可以通过采用现有的候选陈述段，并将开始时间提前以包括在感兴趣区域内由同一个参与者发出的前一个讲话突发，生成第二新候选陈述段。如果在感兴趣区域内没有由同一参与者发出的下一个或前一个讲话突发，则可能不会生成新候选陈述段中的一个或两者。下面将参照图23描述生成候选陈述段的替代方法。

[0598] 在一些示例中，最长陈述段搜索过程2004可能涉及评估关于新候选陈述段的一个或多个接受准则。根据一些这样的实现，可以为每个新候选陈述段计算主导度量。在一些这样的实现中，主导度量可以指示在包括新候选陈述段的时间间隔期间由主导会议参与者发出的总语音的占比。主导会议参与者可能是在该时间间隔内发言最多的会议参与者。在一些示例中，具有大于主导阈值的主导度量的新候选陈述段将被添加到现有候选陈述段。在一些实现中，主导阈值可以是0.7，0.75，0.8，0.85等。否则，搜索可终止。

[0599] 在一些实现中，可以在产生陈述过程2001期间，例如在最长陈述段搜索过程2004期间，评估双对话比率和/或语音密度度量。下面将参考图22来描述一些示例。

[0600] 图21简述了根据本文公开的一些实现的分段过程的后续阶段。在该示例中，已经

对包括用于产生其他过程2101的指令的子例程进行了函数调用。

[0601] 产生其他过程2101的输入发言者活动图2102可以取决于具体的实现。在一些实现中，输入发言者活动图2102可以是全局发言者活动图1809，其指示整个会议的发言者活动，或者是对应于较小时间间隔的发言者活动图。然而，在一些实现中，产生其他过程2101可以从分段过程的一个或多个先前阶段(诸如产生混串音过程1901和/或产生陈述过程2001)接收输入，指示会议的哪些时间间隔(或部分或会议的哪些时间间隔)对应于先前标识的段(诸如先前标识的混串音段或陈述段)。根据一些这样的实现，输入发言者活动图2102可以对应于与先前标识的段的时间间隔不对应的时间间隔。

[0602] 在该示例中，进行过程2101包括最长段搜索处理2104，其可能能够定位包含来自一个会议参与者的语音的感兴趣区域中的最长段。这里，根据最长段搜索处理2104的结果，产生其他过程2101产生包含一个或多个被分类的段的部分分段记录2103A，或包含单个被分类的段的部分分段记录2103B。在一些示例中，如果进行过程2101产生部分分段记录2103B，则它将被输入到分类器，诸如下面参考图22描述的分类器2201。产生其他过程2101可以涉及对于在感兴趣区域中已经标识了其语音的每个会议参与者执行段搜索过程2104的迭代过程。

[0603] 在该示例中，可以基本上如上文参考最长陈述段搜索过程2004所述地生成根候选段。对于每个根候选讲话突发，一些实现涉及搜索由与根候选相同的会议参与者所发出的兴趣区域中的所有讲话突发。一些例子包括构建包括包含根候选的这种讲话突发的最长行程的候选段。

[0604] 一些这样的示例涉及应用一个或多个接受准则。在一些实现中，一个这样的准则是没有两个讲话突发可以被大于阈值候选段时间间隔t_{window}分隔开。t_{window}的示例性设置是$t_{min}/2$，其中t_{min}表示阈值候选段时间(候选段的最小持续时间)。其他实现可以应用不同的阈值候选段时间间隔和/或其他接受准则。一些实现可以涉及通过评估同一个会议参与者的下一个讲话突发和/或同一会议参与者的前一个讲话突发来构建候选段，如上文所述或如下文参照图23所述。

[0605] 搜索完成后，最长候选段(在分析所有根候选之后)可以被分类。在该示例中，最长候选段被传递到分类器2201，分类器2201返回分类的最长段2106B。在图21所示的示例中，将在先发言者活动图2106A输入到在先发言者活动递归2107A，其输出在先的部分分段记录2108A。这里，后续发言者活动图2106C被输入到后续发言者活动递归2107C，后续发言者活动递归2107C输出后续部分分段记录1908C。

[0606] 图22简述了根据本文公开的一些实现的段分类器可执行的操作。在该示例中，给定关于时间t_0到t_1的发言者活动图2202作为输入，分类器2201能够确定段分类2209A-2209E之一的实例。在该示例中，发言者活动图2202包括全局发言者活动图1809的一部分，并且被限制为包含仅在时间t_0到t_1之间的感兴趣时间区域中的信息。在一些实现中，分类器2201可以与本文别处描述的递归分段过程中的一个或多个结合使用。然而，在替代实现中，分类器2201可以用于非递归分段过程。根据一些这样的实现，分类器2201可以用于在会议记录或其一部分的多个时间间隔(例如，顺序时间间隔)中的每一个中标识段。

[0607] 在该实现中，分类器2201包括特征提取器2203，其能够分析发言者活动图2202的会话动态，并且标识会话动态数据类型DT，DEN和DOM，其在本示例中分别对应于双讲话比

率、语音密度度量和主导度量。这里，分类器2201能够根据一组规则来确定段分类的实例，该组规则在此示例中是基于由特征提取器2203标识的一个或多个会话动态数据类型的。

[0608] 在该示例中，该组规则包括以下规则：如果语音密度度量DEN小于相互静默阈值DEN_S，则将段分类为相互静默段2209A。这里，该规则由相互静默确定过程2204应用。在一些实现中，相互静默阈值DEN_S可以为0.1，0.2，0.3等。

[0609] 在该示例中，如果相互静默确定过程2204确定语音密度度量大于或等于相互静默阈值，则下一个过程是混串音确定过程2205。这里，该组规则包括以下规则：如果语音密度度量大于或等于相互静默阈值，并且双讲话比率DT大于混串音阈值DT_B，则将段分类为混串音段。在一些实现中，多路复用阈值DT_B可以为0.6，0.7，0.8等因此，如果混串音确定处理2205确定双讲话比率大于混串音阈值，则混串音确定过程2205将该段分类为混串音段2209B。

[0610] 这里，如果混串音确定过程2205确定双讲话比率小于或等于混串音阈值，则下一过程是讨论确定过程2206。这里，该组规则包括以下规则：如果语音密度度量大于或等于静默阈值，并且如果双讲话比率小于或等于混串音阈值但大于讨论阈值DT_D，则将段分类为讨论段。在一些实现中，讨论阈值DT_D可以是0.2，0.3，0.4等因此，如果讨论确定过程2206确定双讲话比率大于讨论阈值DT_D，则将段分类为讨论段2209C。

[0611] 在该实现中，如果讨论确定过程2206确定双讲话比率不大于讨论阈值DT_D，则下一个过程是陈述确定过程2207。这里，该组规则包括以下规则：如果语音密度度量大于或等于静默阈值，如果双讲话比率小于或等于讨论阈值，并且如果主导度量DOM大于陈述阈值DOM_P，则将段分类为陈述段。在一些实现中，陈述阈值DOM_P可以为0.7，0.8，0.9等因此，如果陈述确定过程2207确定主导度量DOM大于陈述阈值DOM_P，则陈述确定过程2207将该段分类为陈述段2209D。

[0612] 在该示例中，如果陈述确定过程2207确定主导度量DOM不大于陈述阈值DOM_P，则下一过程是问答确定过程2208。这里，该组规则包括如下规则：如果语音密度度量大于或等于静默阈值，如果双讲话比率小于或等于讨论阈值，以及如果主导度量小于或等于陈述阈值但是大于问答阈值，则将段分类为问答段。

[0613] 在一些实现中，问答阈值可以是全部会议参与者的数量N的函数或者在感兴趣区域中已经标识了其语音的会议参与者的数量N的函数。根据一些示例，问答阈值可以是DOM_Q/N，其中DOM_Q表示常数。在一些例子中，DOM_Q可以等于1.5，2.0，2.5等。

[0614] 因此，如果问答确定过程2208确定主导度量大于问答阈值，则在该示例中，该段将被分类为Q&A段2209E。如果没有，在这个例子中，段将被分类为讨论段2209C。

[0615] 图23示出了根据本文公开的一些实现的最长段搜索过程的示例。根据一些实现，例如上述那些，产生混串音，产生陈述和产生其他过程各自包含相应的最长段搜索过程。在一些这样的实现中，最长段搜索过程可以如下地进行。此示例将涉及最长陈述段搜索过程。

[0616] 这里，评估被包括在输入发言者活动图2301中的候选种子讲话突发2302A-2302F的列表。在一些示例中，如这里，即使候选种子讲话突发的列表在图23中根据开始和结束时间排列，候选种子讲话突发的列表仍可以按照长度的降序排序。接下来，可以依次考虑候选种子讲话突发中的每一个。在该示例中，首先考虑最长候选种子讲话突发(2302C)。对于每个候选种子讲话突发，可以指定候选段。这里，候选段2304A最初被指定用于候选种子讲话

突发2302C。

[0617] 在该实现中，第一迭代2303A涉及对候选段2304A(这里，通过分类器2201)进行分类，以确保其会话动态数据类型(例如，上述的DEN，DT和/或DOM会话动态数据类型)不排除候选段2304A属于在最长段搜索过程中寻找的特定段分类。在该示例中，候选段2304A仅包括被分类为陈述段(2305A)的候选讲话突发2302C。因为这是在最长段搜索过程中寻找的段分类，所以最长段搜索过程继续。

[0618] 在该示例中，最长段搜索过程的第二迭代2303B涉及将以下讲话突发2302D添加到候选段2304A，以创建候选段2304B，并对候选段2304B进行分类。在一些实现中，在先的和/或随后的讲话突发可能需要在候选段的阈值时间间隔内，以便有资格被添加到候选段。如果添加随后的讲话突发排除分类作为正在寻找的段分类，则随后的讲话突发可能不会被包含在候选分段中。然而，在该示例中，候选段2304B被分类为陈述段(2305B)，因此保留候选段2304B并继续迭代。

[0619] 在该实现中，最长段搜索过程的第三迭代2303C涉及将在先的讲话突发2302B添加到候选段2304B，以创建候选段2304C，并对候选段2304C进行分类。在该示例中，候选段2304C被分类为陈述段(2305C)，因此保留候选段2304C并继续迭代。

[0620] 在该示例中，最长段搜索过程的第四迭代2303D包括将随后的讲话突发2302E添加到候选段2304C，以创建候选段2304D，并对候选段2304D进行分类。在该示例中，候选段2304D被分类为陈述段(2305D)，因此保留候选段2304D并继续迭代。

[0621] 随后的和/或在先的讲话突发可以继续被添加到候选段，直到添加任何讲话突发意味着候选段不再是所寻找的类别。这里，例如，最长段搜索过程的第五迭代2303E涉及将在先的讲话突发2302A添加到候选段2304D，以创建候选段2304E，并对候选段2304E进行分类。在该示例中，候选段2304E被分类为Q&A段(2305E)，因此不保留候选段2304E。

[0622] 然而，在这个例子中，该过程将继续进行，以便评估随后的讲话突发。在图23所示的例子中，最长段搜索过程的第六迭代2303F涉及随后的讲话突发2302F添加到候选段2304D，以创建候选段2304E，并对候选段2304F进行分类。在该示例中，候选段2304F被分类为Q&A段(2305E)，因此不保留候选段2304C，并且迭代停止。

[0623] 如果所得的候选段不短于阈值候选段时间t_{min}，则候选段可以被指定为最长段。否则，最长段搜索过程可能会报告没有合适的段存在。如本文其他地方所述，阈值候选段时间t_{min}可根据可能对应于感兴趣区域的时间间隔的时间尺度而变化。在该示例中，候选段2304D长于阈值候选段时间t_{min}，因此最长段搜索过程输出陈述段2306。

[0624] 会议记录通常包括大量的音频数据，其可能包括相当大量的混串音和非实质性的讨论。通过音频回放找到相关会议主题可能非常耗时。自动语音识别(ASR)有时被用于将会议记录转换为文本，以使得能够基于文本搜索和浏览。

[0625] 不幸的是，基于自动语音识别的准确会议转录已被证明是一项具有挑战性的任务。例如，美国国家标准与技术研究所(NIST)的领先的基准已经示出尽管近几十年来各种语言的ASR的误码率(WER)大幅度下降，但会议语音的WER仍然保持显著高于其他类型语音的WER。根据2007年发布的NIST报告，会议语音的WER通常超过25％，而对于涉及多个会议参加者的会议，往往超过50％。(Fiscus，Jonathan G.等人，“The Rich Transcription 2007Meeting Recognition Evaluation”(NIST 2007)。)

[0626] 尽管已知会议语音的WER高，但是自动生成会议主题的先前尝试通常是基于会议记录的ASR结果产生会议参与者所说的词语的完美转录的假设的。本公开包括用于确定会议主题的各种新颖技术。一些实现涉及词语云生成，其可以在回放期间是交互式的。一些例子能够进行高效的主题挖掘，同时解决了ASR误差提供的挑战。

[0627] 根据一些实现，给定话语的许多假设(例如，如在语音识别格中所描述的)可能有助于词语云。在一些示例中，全会议(或多会议)上下文可以通过编译在整个会议中发现的和/或在多个会议中发现的许多词语的替代假设的列表而被引入。一些实现可以涉及在多次迭代上应用全会议(或多会议)上下文，以对语音识别格的假设词重新评分(例如，通过淡化较不频繁的替代词)，从而去除一些话语级别歧义。

[0628] 在一些示例中，可以使用“术语频率度量”来将主词候选和替代词假设排序。在一些这样的示例中，术语频率度量可以至少部分地基于语音识别格中的假设词的出现次数和语音识别器报告的词语识别置信度分数。在一些示例中，术语频率度量可以至少部分地基于底层语言中的词语的频率和/或词语可能具有的不同含义的数量。在一些实现中，可以使用可以包括上位词信息的本体来将词语概括为主题。

[0629] 图24是简述本文公开的某些主题分析方法的块的流程图。与本文所述的其它方法一样，方法2400的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0630] 在一些实施方式中，方法2400可以至少部分地经由存储在诸如本文所描述的那些的非暂态介质(包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等)上的指令(例如，软件)来实现。在一些实现中，方法2400可以至少部分地由装置(诸如图3A所示的装置)来实现。根据一些这样的实现，方法2400可以至少部分地由图3C和5所示的分析引擎307的一个或多个元件实现，例如由联合分析模块306实现。根据一些这样的示例，方法2400可以至少部分地由图5的主题分析模块525来实现。

[0631] 在该示例中，块2405包括接收对于涉及多个会议参与者的会议的会议记录的至少一部分的语音识别结果数据。在一些示例中，语音识别结果数据可以由主题分析模块在块2405中接收。这里，语音识别结果数据包括多个语音识别格、以及语音识别格中的多个假设词中的每一个的词语识别置信度分数。在该实现中，词语识别置信度分数对应于假设词与在会议期间会议参与者所说的实际词语正确对应的可能性。在一些实现中，在块2405中可以接收来自两个或更多个自动语音识别过程的语音识别结果数据。下面介绍一些例子。

[0632] 在一些实现中，会议记录可以包括被分别记录的来自多个端点的会议参与者语音数据。作为替代地或者附加地，会议记录可以包括来自对应于多个会议参与者的单个端点的会议参与者语音数据，并且包括用于标识多个会议参与者的每个会议参与者的会议参与者语音的信息。

[0633] 在图24所示的例子中，块2410涉及对于语音识别格中的多个假设词中的每一个确定主词候选和一个或多个替代词假设。这里，与任何替代词假设的词语识别置信度分数相比，主词候选的词语识别置信度分数指示更高的与在会议期间会议参与者所说出的实际词语正确对应的可能性。

[0634] 在该实现中，块2415涉及为主词候选和替代词假设计算“术语频率度量”。在该示例中，术语频率度量至少部分地基于语音识别格中的假设词的出现次数以及基于词语识别

置信度分数。

[0635] 根据一些示例，术语频率度量可以至少部分地基于“文档频率度量”。在一些这样的示例中，术语频率度量可以与文档频率度量成反比。例如，文档频率度量可以对应于主词候选将在会议中出现的预期频率。

[0636] 在一些实现中，文档频率度量可以对应于主词候选已经在两个或更多个先前会议中出现的频率。例如，先前的会议可以是同一类别的会议，例如商业会议，医疗会议，工程会议，法律会议等。在一些实现中，会议可以按子类进行分类，例如，工程会议的类别可以包括电气工程会议、机械工程会议、音频工程会议、材料科学会议、化学工程会议等的子类别。同样，商务会议的类别可能包括销售会议、财务会议、营销会议等的子类别。在一些示例中，会议可以至少部分地根据会议参与者进行分类。

[0637] 作为替代地或者附加地，文档频率度量可以对应于主词候选在至少一个语言模型中出现的频率，其可以估计不同词语和/或短语的相对似然性，例如通过根据概率分布将概率分配给词语序列。(一个或多个)语言模型可以提供上下文以区分听起来相似的词语和短语。语言模型可以例如是统计语言模型，例如词袋模型，N-gram模型，因子化语言模型等。在一些实现中，语言模型可以与会议类型相对应，例如与会议的预期主旨相对应。例如，与和非医学语言有关的语言模型相比，和医学术语相关的语言模型可能给“脾”和“梗塞”分配更高的概率。

[0638] 根据一些实现，在块2405中，会议类别，会议子类别和/或语言模型信息可以与语音识别结果数据一起被接收。在一些这样的实现中，这样的信息可以包括在由图5的主题分析模块525接收的会议元数据210中。

[0639] 本文公开了确定术语频率度量的各种替代示例。在一些实现中，术语频率度量可以至少部分地基于数个词语含义。在一些这样的实现中，术语频率度量可以至少部分地基于标准参考书(例如特定的词典或字典)中的对应词的定义的数量。

[0640] 在图24所示的示例中，块2420涉及根据术语频率度量对主词候选和替代词假设进行排序。在一些实现中，块2420可以包括以术语频率度量的降序对主词候选和替代词假设进行排序。

[0641] 在该实现中，块2425涉及将替代词假设包括在替代假设列表中。在一些实现中，方法2400的至少一些过程的迭代可以至少部分地基于替代假设列表。因此，一些实现可以涉及在一个或多个这样的迭代期间保留替代假设列表，例如在每次迭代之后。

[0642] 在该示例中，块2430涉及根据替代假设列表对语音识别格的至少一些假设词重新评分。换句话说，在确定、计算、排序，包括和/或重新评分的一个或多个这样的迭代期间，可以改变在框2405中针对语音识别格的一个或多个假设词语接收的词语识别置信度分数。以下提供进一步的细节和实例。

[0643] 在一些示例中，方法2400可以包括形成包括主词候选和用于每个主词候选的术语频率度量的词语列表。在一些示例中，词语列表还可以包括用于每个主词候选的一个或多个替代词假设。例如，替代词假设可以根据语言模型生成。

[0644] 一些实现可以涉及至少部分地基于词语列表来生成会话主题的主题列表。主题列表可以包括词语列表中的一个或多个词语。一些这样的实现可能涉及确定主题分数。例如，这样的实现可以至少部分地基于主题分数来确定是否在主题列表上包括词语。根据一些实

现，主题分数可以至少部分地基于术语频率度量。

[0645] 在一些示例中，主题分数可以至少部分地基于用于主题概括的本体。在语言学中，下位词是其语义场被包含在另一个词语(其已知为上位词)的语义场中的词语或短语。下位词与其上位词共享“类型”的关系。例如，“知更鸟”，“椋鸟”，“麻雀”，“乌鸦”和“鸽子”都是“鸟”(它们的上位词)的下位词;“鸟”又是“动物”的下位词。

[0646] 因此，在一些实现中，产生主题列表可以涉及确定词语列表中的一个或多个词语的至少一个上位词。这样的实现可以涉及至少部分地基于上位词分数来确定主题分数。在一些实现中，上位词不需要已被会议参与者说出以便成为主题分数确定过程的一部分。下面提供了一些例子。

[0647] 根据一些实施方式，方法2400的至少一些过程的多次迭代可以包括生成主题列表并确定主题分数的迭代。在一些这样的实现中，块2425可以包括至少部分地基于主题分数在替代假设列表中包括替代词假设。下面描述一些实现，然后是作为确定主题分数的过程的一部分的使用上位词的一些示例。

[0648] 在一些示例中，方法2400可以包括将语音识别格的至少一些假设词缩减为规范的基本形式。在一些这样的示例中，缩减过程可以包括将语音识别格点的名词缩减为规范的基本形式。规范的基本形式可以是名词的单数形式。作为替代地或者附加地，缩减过程可以包括将语音识别格点的动词缩减为规范的基本形式。规范的基本形式可能是动词的不定式形式。

[0649] 图25示出了主题分析模块元素的示例。与本文公开的其他实现一样，主题分析模块525的其他实现可以包括更多，更少和/或其他元素。主题分析模块525可以例如经由控制系统(诸如图3A所示的控制系统)来实现。控制系统可以包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑或离散硬件组件中的至少一个。在一些实现中，主题分析模块525可以通过存储在非暂态介质上的指令(例如，软件)来实现。这种非暂态介质为诸如本文所描述的那些介质，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。

[0650] 在该示例中，主题分析模块525被示出为接收语音识别格2501。语音识别格2501例如可以是语音识别结果(诸如上文参照图4和图5所描述的语音识别结果401F-405F)的实例。语音识别格的一些例子如下所述。

[0651] 主题分析模块525的该示例包括格子重评分单元2502。在一些实现中，格子重评分单元2502可以能够根据替代假设列表对语音识别格2501中的至少一些假设词重新评分。例如，格子重评分单元2502可能能够改变在替代假设列表2507中找到的假设词的词语识别置信度分数，使得这些假设词被淡化。该过程可能取决于用于词语识别置信度分数的特定度量。例如，在一些实现中，词语识别置信度分数可以用成本来表示，其的值可以是假设词语有多不可能是正确的度量。根据这样的实现，淡化这样的假设词可能涉及增加对应的词语识别置信度分数。

[0652] 根据一些实现，替代假设列表2507最初可能是空的。如果是这样的话，格子重评分单元2502可以不执行重新评分，直到稍后的迭代。

[0653] 在该示例中，主题分析模块525包括格修剪单元2503。格修剪单元2503可以例如能够执行一种或多种类型的格修剪操作(诸如波束剪枝，后验概率修剪和/或格深度限制)，以

便降低输入语音识别格点的复杂性2501。

[0654] 图26示出了输入语音识别格的示例。如图26所示，未修剪的语音识别格可能相当大。图26中的圆圈表示语音识别格的节点。连接节点的曲线或“弧”对应于假设的词，其可以经由弧连接以形成假设词序列。

[0655] 包括图27A和27B的图27示出了修剪后的小语音识别格的一部分的示例。在这个例子中，修剪的语音识别格对应于话语的第一部分“I accidentally did not finish my beef jerky coming from San Francison to Australia(我不小心没有吃完从旧金山到澳大利亚的牛肉干)”。在这个例子中，针对同一假设词的替代词假设在编号节点之间的弧上被指示。可以遍历语音识别格的不同弧以形成替代假设词序列。例如，假设词序列“didn't finish”由连接节点2,6和8的弧表示。假设词序列“did of finish”由连接节点5,11,12和15的弧表示。假设词序列“did of finished”由连接节点5,11,12和14的弧表示。假设词序列“did not finish”由连接节点5,11和17-20的弧表示。假设词序列“did not finished”由连接节点5,11,17和18的弧表示。所有上述假设词序列对应于实际子话语“did not finish”。

[0656] 在一些语音识别系统中，给定语音识别器的声学输入特征，语音识别器可以用对数声学成本C_A来报告词语识别置信度分数，对数声学成本C_A是通过格子的这个路径上的这个假设词有多不可能正确的度量。给定语言模型，语音识别器还可以用对数语言成本C_L来报告词语识别置信度分数，该对数语言成本C_L是通过格子的该路径上的这个假设词的可能性有多不可能不正确的量度。对于格子中的每个弧可以报告声学和语言成本。

[0657] 对于图27所示的格子部分中的每个弧，例如，该弧的组合的声学和语言成本(C_A+C_L)在每个假设词旁边被示出。在该示例中，通过语音识别格的最佳假设词序列对应于从起始节点到结束节点的具有最低弧成本总和的路径。

[0658] 在图25所示的示例中，主题分析模块525包括词法单元2504。词法单元2504可以能够将假设词缩减为规范的基本形式。例如，在涉及将语音识别格的名词缩减为规范的基本形式的一些实现中，词法单元2504可能能够将名词的多个形式缩减为单数形式(例如，将“cars(多个汽车)”缩减为“car(汽车)”)。在一些涉及将语音识别格的动词缩减为规范的基本形式的实现中，词法单元2504可能能够将动词缩减为不定式(例如，缩减“running”为“run”，缩减“runs”为“run”)。

[0659] 词法单元2504的替代实现可以包括所谓的“词干分析器”，例如波特词干分析器(Porter Stemmer)。然而，这种类型的基本词干分析器可能不能精确地转换不规则名词或动词形式(例如将“mice”缩减为“mourse”)。对于这种转换可能需要更精确的词法实现，例如在Miller,George A,WordNet:A Lexical Database for English,in Communications of the ACM Vol.38,第11期,第39-41页(1995)中。

[0660] 图25的主题分析模块525包括术语频率度量计算器2505。在一些实现中，术语频率度量计算器2505可以能够确定语音识别格2501的假设词的术语频率度量。在一些这样的实现中，术语频率度量计算器2505可能能够确定在输入格中观察到的每个名词的术语频率度量(例如，词法单元2504可能能够确定哪些假设词是名词)。

[0661] 在一些实现中，术语频率度量计算器2505可以能够根据术语频率/逆文档频率(TF-IDF)函数来确定术语频率度量。在一个这样的例子中，每当在输入语音识别格中检测

到具有词典索引x的假设词时，术语频率度量TFx可以如下确定：

[0662] $$TFx = TFx' + \frac{C}{\mathrm{N.max}(\ln DF_x, MDF)} \quad （式45）$$

[0663] 在式45中，TF_x'表示词语x的先前术语频率度量。如果这是在当前迭代中第一次遇到词语x，TF_x'的值可能被设置为零。在式45中，DF_x表示文档频率度量，ln表示自然对数。如上所述，文档频率度量可以对应于词语将在会议中出现的预期频率。在一些示例中，预期频率可以对应于该词已经在两个或更多个在先会议中出现的频率。在一般性的商务电话会议系统的实例中，文档频率度量可以通过对该词语在大量商务电话会议上出现的频率进行计数而得出。

[0664] 作为替代地或者附加地，预期频率可以对应于主词候选在语言模型中出现的频率。本文公开的方法的各种实施方式可以与语音识别器一起使用，语音识别器可以将某种类型的词频度量应用作为其语言模型的一部分。因此，在一些实现中，用于语音识别的语言模型可以提供由术语频率度量计算器2505使用的文档频率度量。在一些实现中，这样的信息可以与语音识别格一起被提供或者被包含在会议元数据210中。

[0665] 在式45中，MDF表示指示最小对数文档频率的选定常数。在一些实施方式中，MDF值可以是-10至-4范围内的整数(例如-6)。

[0666] 在式45中，C表示在输入格子中由语音识别器报告的在范围[0-1]中的词语识别置信度分数。根据一些实现，可以根据以下来确定C：

[0667] $C = \exp(-C_A - C_L)$ (式46)

[0668] 在式46中，C_A表示对数声学成本，C_L表示对数语言成本，两者均使用自然对数表示。

[0669] 在式45中，N表示词义的数量。在一些实现中，N的值可以基于标准词典(例如特定字典)中的词语的定义的数量。

[0670] 根据一些替代实现，术语频率度量TF_X可以如下确定

[0671] $$TFx = TFx' + \frac{\alpha C + (1-\alpha)}{\mathrm{N.max}(\ln DF_x, MDF)} \quad （式47）：$$

[0672] 在式47中，α表示可以例如具有在0到1的范围内的值的权重因子。在式45中，以未加权的方式使用识别置信度C.在一些实例中，未加权的识别置信度C可能是非最优的，例如，如果假设词具有非常高的识别置信度，但出现的频率较低。因此，添加权重因子α可能有助于控制识别信心的重要性。可以看出，当α＝1时，式47等同于式45。然而，当α＝0时，不使用识别置信度，并且可以根据分母中的项的倒数确定术语频率度量。

[0673] 在图25所示的示例中，主题分析模块525包括替代词语假设修剪单元2506。当词语列表2508被创建时，系统通过对于相同的时间间隔分析通过格子的替代路径来为每个词语标记一组替代词假设。

[0674] 例如，如果会议参与者所说的实际词语是词pet，语音识别器可能已经报告了put和pat作为替代词假设。对于实际的词语pet的第二个实例，语音识别器可能已经报告了pat，pebble和parent作为替代词假设。在这个例子中，在分析了与会议中所有话语对应的所有语音识别格之后，用于词语pet的替代词假设的完整列表可以包括put，pat，pebble和parent。词语列表2508可以按照TF_x的降序被排序。

[0675] 在替代词假设修剪单元2506的一些实现中，可以从列表中去除在列表中进一步在后(例如，具有较低TF_x值)的替代词假设。删除的替代词可以被添加到替代词假设列表2507中。例如，如果假设词pet比其替代词假设具有更高的TF_x，则替代词假设修剪单元2506可以从词语列表2508中去除替代词假设pat，put，pebble和parent，并将替代词假设pat，put，pebble和parent添加到替代词假设列表2507。

[0676] 在该示例中，主题分析模块525至少临时地将替代词假设列表2507存储在存储器中。替代词假设列表2507可以如其他地方所述地通过多次迭代被输入到格子重评分单元2502。迭代次数可以根据具体实现而变化，并且可以是例如在1到20的范围内。在一个具体实现中，4次迭代产生令人满意的结果。

[0677] 在一些实现中，词语列表2508可以在每次迭代开始时被删除，并且可以在下一次迭代期间重新编译。根据一些实现，替代词假设列表2507可以在每次迭代开始时不被删除，因此替代词假设列表2507的尺寸可以随着迭代继续进行而增大。

[0678] 在图25所示的示例中，主题分析模块525包括主题评分单元2509。主题评分单元2509可以能够确定词语列表2508中的词语的主题分数。

[0679] 在一些示例中，主题分数可以至少部分地基于用于主题概括的本体2510，例如本文别处讨论的词网(WordNet)本体。因此，在一些实现中，产生主题列表可以涉及确定词语列表2508中的一个或多个词语的至少一个上位词。这样的实现可以涉及至少部分地基于上位词分数来确定主题分数。在一些实现中，上位词不需要已经被会议参与者说出以便成为主题分数确定过程的一部分。

[0680] 例如，“pet(宠物)”是“animal(动物)”的一个例子，它是一类organism(有机体)，它是一种living thing(生物)。因此，“动物”这个词可能被认为是“pet”这个词的第一级上位词。“organism”这个词可能被认为是“pet”这个词的第二级上位词以及是“animal”这个词的第一级上位词。“living thing”一词可能被认为是“pet”这个词的第三级上位词，“animal”这个词的第二级上位词和“organism”这个词的第一级上位词。

[0681] 因此，如果词语“pet”位于词语列表2508上，则在某些实现中，主题评分单元2509可能能够根据多个上位词“animal”，“organism”和/或“living thing”中的一个或多个确定主题分数。根据一个这样的示例，对于词语列表2508中的每个词语，主题评分单元2509可以遍历上位词树的N级(这里例如N＝2)，将每个上位词添加到主题列表2511(如果没有已经存在)，并且将该词的术语频率度量加到与上位词相关联的主题分数上。例如，如果“pet”在词语列表2508中存在，术语频率度量为5，则pet，animal和organism将被添加到主题列表中，术语频率度量为5。如果animal也位于词语列表2508中，术语频率度量为3，那么animal和organism的主题分数将被加3，总主题分数为8，living thing将被添加到词语列表2508中，术语频率度量为3。

[0682] 根据一些实现，方法2400的至少一些过程的多次迭代可以包括生成主题列表并确定主题分数的迭代。在一些这样的实现中，方法2400的块2525可以包括至少部分地基于主题分数将替代词假设包含在替代假设列表中。例如，在一些替代实现中，主题分析模块525可以能够基于术语频率度量计算器2505的输出来进行主题评分。根据一些这样的实现，除了替代词假设之外，替代词假设修剪单元2506还可执行主题的替代假设修剪。

[0683] 例如，假定主题分析模块525由于“pet(宠物)”的一个或多个实例的术语频率度量

为15，“dog(狗)”的实例的术语频率度量为5，“goldfish(金鱼)”的实例的术语频率度量为4，确定了“pet”的会议主题。进一步假定在会议中某处可能会有“cat”的单个话语，但是实际说出的词语是否是“cat”，“mat”，“hat”，“catamaran”，“catenary”，“caterpillar”等则非常不明确。如果主题分析模块525仅考虑了反馈循环中的词语频率，则词语列表2508将不会有助于消除这些假设的歧义的过程，因为只有“cat”的一个可能话语。但是，因为“cat”是“pet”的下位词，其利用其它说出的词语被标识为主题，所以主题分析模块525可能会更好地消除“cat”的可能话语的歧义。

[0684] 在该示例中，主题分析模块525包括元数据处理单元2515。根据一些实现，元数据处理单元2515可能能够产生至少部分地基于由主题分析模块525接收的会议元数据210的偏向词列表2512。偏向词列表2512可以例如能够包括可以用固定的术语频率度量直接插入到词语列表2508中的词语列表。元数据处理单元2515例如可以从与会议的主题或议题相关的先验信息(例如，从日历邀请，电子邮件等)中导出偏向词列表2512。偏向词列表2512可能使得主题列表构建过程偏向于更可能包含与会议的已知议题有关的主题。

[0685] 在一些实现中，可以根据多种语言模型生成替代词假设。例如，如果会议元数据将指示会议可能涉及法律和医疗问题，例如与基于由于医疗程序而导致的患者伤害或死亡的诉讼对应的医疗事故问题，则可以根据医疗和法律语言模型两者来生成替代词假设。

[0686] 根据一些这样的实现，多个语言模型可以由ASR过程在内部内插，使得在方法2400的块2405接收的语音识别结果数据和/或图25中接收的语音识别格2501基于多种语言模型。在替代实现中，ASR过程可以输出多组语音识别格，每组对应于不同的语言模型。可以为每种类型的输入语音识别格生成主题列表2511。多个主题列表2511可以根据所得到的主题分数被合并到单个主题列表2511中。

[0687] 根据本文公开的一些实现，主题列表2511可以用于有助于回放会议记录，搜索会议记录中的主题等的过程。根据一些这样的实现，主题列表2511可以用于提供与一些或所有会议记录相对应的主题的“词语云”。

[0688] 包括图28A和28B的图28示出了包括用于整个会议记录的词语云的用户界面的示例。用户界面606a可以在显示器上被提供，并且可以用于浏览会议记录。例如，如上文参考图6所述，用户界面606a可以在显示设备610的显示器上被提供。

[0689] 在该示例中，用户界面606a包括会议记录的会议参与者的列表2801。这里，用户界面606a以对应于会议参与者语音的时间间隔示出波形625。

[0690] 在该实现中，用户界面606a提供了用于整个会议记录的词语云2802。主题列表2511中的主题可以在词语云2802中按主题频率的降序(例如，从右到左)排列，直到没有其它空间可用(例如在给定最小字体大小的情况下)。

[0691] 根据一些这样的实现，每当用户调整缩放比率时，可以重新运行用于词语云2802的主题布置算法。例如，用户可以能够与用户界面606a交互(例如，经由触摸、手势、语音命令等)以便至少“放大”或者扩大图形用户界面606的一部分，以示出比整个会议记录的时间间隔更小的时间间隔。根据一些这样的示例，图6的回放控制模块605可以访问可以由会话动态分析模块510先前输出的会话动态数据文件515a-515n的不同实例，其与用户选定的时间间隔更接近地对应。

[0692] 包括图29A和29B的图29示出了包括用于多个会议段中的每一个的词语云的用户

界面的示例。如前面的例子那样，用户界面606b包括会议参与者的列表2801，并且以对应于会议参与者语音的时间间隔示出波形625。

[0693] 然而，在该实现中，用户界面606b为多个会议段1808A-1808J中的每一个提供了词语云。根据一些这样的实现，会议段1808A-1808J可以先前由分段单元确定，诸如上面参照图18B所描述的分段单元1804。在一些实现中，主题分析模块525可以针对会议的每个段1808被单独调用(例如，通过每次仅使用对应于来自一个段1808的话语的语音识别格2501)，对于每个段1808生成单独的主题列表2511。

[0694] 在一些实现中，用于在词语云中渲染每个主题的文本的大小可以与主题频率成比例。在图29A所示的实现中，例如，主题“kitten(小猫)”和“newborn(新生儿)”的字体大小可以稍大于主题“large integer(大整数)”，这指示在段1808C中主题“kitten”和“newborn”比主题“large integer”讨论得更多。然而，在一些实现中，主题的文本大小可能受到显示词语云可用的区域，最小字体大小(可能是用户可选择的)等约束。

[0695] 图30是简述本文公开的一些回放控制方法的块的流程图。与本文所述的其它方法一样，方法3000的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0696] 在一些实现中，方法3000可以至少部分地通过存储在非暂态介质上的指令(例如，软件)来实现。这种非暂态介质为诸如本文所描述的那些介质，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。在一些实现中，方法3000可以至少部分地通过装置(诸如图3A所示的装置)来实现。根据一些这样的实现，方法3000可以至少部分地由图6所示的回放系统609的一个或多个元件实现，例如由回放控制模块605实现。

[0697] 在该示例中，块3005涉及接收涉及多个会议参与者的会议的至少一部分的会议记录和会议主题的主题列表。在一些实现中，如图6所示，块3005可以包括由回放系统609接收诸如回放流401B-403B之类的各个回放流。根据一些这样的实现，块3005可以涉及接收其他数据，例如由图6的回放系统609接收的回放流索引401A-403A，分析结果301C-303C，段和词语云数据309，搜索索引310和/或会议概述信息311。因此，在一些示例中，块3005可以涉及接收包括会议段时间间隔数据和会议段分类的会议段数据。

[0698] 根据一些实现，块3005可以涉及经由接口系统接收会议记录和/或其他信息。接口系统可以包括网络接口，控制系统和存储系统之间的接口，控制系统与另一设备之间的接口，和/或外部设备接口。

[0699] 这里，块3010涉及提供用于控制显示器以展示被显示的用于会议的至少一部分的会议主题的指令。在该示例中，展示包括与至少一些会议主题对应的词语的图像，例如图28所示的词语云2802。在一些实现中，回放控制模块605可以在块3010中提供这样的用于控制显示器的指令。例如，块3010可以包括经由接口系统向诸如显示设备610的显示设备提供这样的指令。

[0700] 显示设备610可以例如是膝上型计算机，平板计算机，智能电话或能够在显示器上提供包括所显示的会议主题的词语云的图形用户界面的其它类型的设备，该图形用户界面诸如是图28的图形用户界面606a或图29的图形用户界面606b。例如，显示设备610可以执行用于根据来自回放控制模块605的指令提供图形用户界面的软件应用程序或“app”，接收用户输入，将与所接收的用户输入对应的信息发送到回放控制模块605等。

[0701] 在一些实例中，由回放控制模块605接收的用户输入可以包括用户(例如根据与“放大”或“缩小”命令相对应的用户输入)选择的所选会议记录时间间隔的指示。响应于这样的用户输入，回放控制模块605可以经由接口系统提供用于控制显示器以展示所显示的与所选择的会议记录时间间隔对应的会议主题的指令。例如，回放控制模块605可以选择会话动态数据文件(诸如被示出为由图5中的会话动态分析模块510输出的会话动态数据文件515a-515e之一)的不同实例，其最接近地对应于由用户选择的所选会议记录时间间隔，并向显示设备610提供相应的指令。

[0702] 如果块3005涉及接收会议段数据，则显示设备610可能能够控制显示器以展示一个或多个会议段的指示，并且展示指示在该一个或多个会议段中讨论的会议主题的所显示的会议主题，例如，如图29所示。显示设备610可以能够控制显示器来展示与会议参与者语音的实例对应的波形和/或与会议参与者对应的图像，诸如图28和29所示的那些。

[0703] 在图30所示的示例中，块3015涉及接收由用户从所显示的会议主题中选择的所选主题的指示。在一些示例中，块3015可以包括由回放控制模块605和经由接口系统接收来自显示设备610的用户输入。用户输入可已经经由用户与显示器的与所选主题相对应的部分的交互被接收，例如来自触摸传感器系统的用户在所显示的词语云的与所选主题对应的区域中的触摸的指示。另一个示例在图31中示出并在下面描述。在一些实现中，如果用户使光标悬停在所显示的词语云中的特定词语上，则可以回放与该词语关联的会议参与者语音的实例。在一些实现中，会议参与者语音可以以重叠的方式在空间上渲染和/或回放。

[0704] 在图30所示的示例中，块3020涉及选择包括会议记录的包括所选主题的一个或多个语音实例的回放音频数据。例如，块3020可以包括选择与所选主题相对应的语音实例，以及在所选主题之前和/或之后说出的至少一些词语，以便提供上下文。在一些这样的示例中，块3020可以涉及选择包括所选主题的话语。

[0705] 在一些实现中，块3020可以包括选择至少两个语音实例，包括由至少两个会议参与者中的每一个发出的至少一个语音实例。该方法可以包括将语音实例渲染到虚拟声学空间的至少两个不同的虚拟会议参与者位置，以产生渲染的回放音频数据，或者访问包括所选主题的先前渲染的语音的部分。根据一些实现，该方法可以包括调度语音实例的至少一部分同时回放。

[0706] 根据一些实现，块3015可以涉及接收由用户从多个会议参与者中选择的所选会议参与者的指示。一个这样的例子在图32中示出并在下面描述。在一些这样的实现中，块3020可以涉及选择包括会议记录的一个或多个语音实例的回放音频数据，该一个或多个语音实例包括所选会议参与者的关于所选主题的语音。

[0707] 这里，块3025涉及提供回放音频数据以供在扬声器系统上回放。例如，在块3025中，回放系统609可以经由接口系统向显示设备610提供混合和渲染的回放音频数据。作为替代地，在块3025中，回放系统609可以将回放音频数据直接提供给扬声器系统，诸如耳机607和/或扬声器阵列608。

[0708] 图31示出了从词语云选择主题的示例。在一些实施方式中，显示设备610可以在显示器上提供图形用户界面606c。在该示例中，用户从词语云2802中选择了词语“pet(宠物)”，并将该词语的表示拖到搜索窗口3105。作为响应，显示设备可以向回放控制模块605发送所选主题“pet”的指示。因此，这是可以在图30的块3015中被接收的“所选主题的指示”

的示例。作为响应，显示设备610可以接收对应于涉及宠物主题的一个或多个语音实例的回放音频数据。

[0709] 图32示出了从词语云选择主题以及从会话参与者的列表中选择会议参与者这两者的示例。如上所述，显示设备610可以在显示器上提供图形用户界面606c。在该示例中，在用户从词语云2802中选择了词语“pet”之后，用户将会议参与者Geogre Washington的表示拖到搜索窗口3105。显示设备610可以将所选主题“pet”和会议参与者Geogre Washington的指示发送到回放控制模块605。作为响应，回放系统609可以向显示设备610发送与会议参与者Geogre Washington关于宠物主题的一个或多个语音实例对应的回放音频数据。

[0710] 在查看大量的电话会议记录，甚至长时间的电话会议的单个记录时，手动定位所记得的电话会议的一部分可能是耗时的。先前已经描述了一些系统，通过该系统，用户可以通过输入他或她希望定位的关键词的文本来搜索语音记录中的关键词。这些关键词可以用于语音识别系统产生的文本的搜索。结果列表可以在显示屏幕上被呈现给用户。

[0711] 本文中公开的一些实现提供了用于呈现会议搜索结果的方法，可以涉及以被设计为允许听众注意到他或她感兴趣的那些结果的方式、非常快速地向用户播放会议记录的摘录(excerpt)。一些这样的实现可以被定制用于存储器扩容。例如，一些这样的实现可以允许用户搜索用户记得的会议(或多个会议)的一个或多个特征。一些实现可以允许用户非常快速地查看搜索结果以找到用户正在寻找的一个或多个特定实例。

[0712] 一些这样的示例涉及空间渲染技术，例如将每个会议参与者的会议参与者语音数据渲染到单独的虚拟会议参与者位置。如本文其他地方详细描述的，一些这样的技术可以允许收听者快速地听到大量内容，然后选择感兴趣的部分以供更详细和/或更慢的回放。一些实现可以涉及例如根据一组感知激发规则引入或改变会议参与者语音的实例之间的重叠。作为替代地或者附加地，一些实现可以涉及加速被回放的会议参与者语音。因此，这样的实现可以利用选择注意力的人才，以确保找到期望的搜索项，同时最小化搜索过程所需的时间。

[0713] 因此，不是返回与用户的搜索项非常相关的一些结果并且要求用户分别试听每个结果(例如，通过依次点击列表中的每个结果来进行播放)，一些这样的实现可以返回用户可以使用空间渲染和本文公开的其他快速回放技术快速(例如，在几秒钟内)试听的许多搜索结果。一些实现可以提供允许用户进一步研究(例如，以1:1回放速度的试听)用户搜索结果的所选实例的用户界面。

[0714] 然而，根据具体实现，这里公开的一些示例可以涉及或可以不涉及空间渲染，引入或改变会议参与者语音的实例之间的重叠，或者加速被回放的会议参与者语音。此外，一些公开的实现可以涉及除了内容之外或作为内容的替代，搜索一个或多个会议的其他特征。例如，除了在一个或多个电话会议中搜索特定词之外，一些实现还可以涉及对会议记录的多个特征执行并发搜索。在一些示例中，特征可以包括发言者的情感状态、发言者的身份、在说话时发生的会话动态的类型(例如陈述，讨论，问答环节等)、端点地点，端点类型和/或其他特征。

[0715] 涉及多个特征的并发搜索(有时在本文中称为多维搜索)可以提高搜索精度和效率。例如，如果用户只能执行关键词搜索，例如对于会议中的“销售”一词，则用户可能必须在从会议找到用户可能会记住的感兴趣的特定摘录之前收听许多结果。相比之下，如果用

户要对会议参与者Fred Jones所说的“销售”一词的示例执行多维搜索，那么用户可能会减少在找到感兴趣的摘录之前用户需要查看的结果的数量。

[0716] 因此，一些公开的实现提供了如下的方法和设备，其用于高效地指定用于一个或多个电话会议记录的多维搜索项，并且高效地查看搜索结果以定位感兴趣的特定摘录。

[0717] 图33是简述本文公开的某些主题分析方法的块的流程图。与本文所述的其它方法一样，方法3300的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0718] 在一些实现中，方法3300可以至少部分地通过存储在非暂态介质上的指令(例如，软件)来实现。这种非暂态介质为诸如本文所描述的那些介质，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。在一些实现中，方法3300可以至少部分地由控制系统实现，例如通过诸如图3A所示的装置的控制系统来实现。控制系统可以包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑或离散硬件组件中的至少一个。根据一些这样的实现，方法3300可以至少部分地由图6所示的回放系统609的一个或多个元件实现，例如通过回放控制模块605来实现。

[0719] 在该示例中，块3305涉及接收对应于涉及多个会议参与者的至少一个会议的记录的音频数据。在该示例中，音频数据包括被分别记录的来自多个端点的会议参与者语音数据，和/或来自对应于多个会议参与者的单个端点的会议参与者语音数据，其包括多个会议参与者的每个会议参与者的空间信息。

[0720] 在图33所示的示例中，块3310涉及基于一个或多个搜索参数来确定对音频数据的搜索的搜索结果。根据一些示例，确定搜索结果可能涉及接收搜索结果。例如，在一些实现中，诸如图6所示的回放系统609的回放系统的一个或多个元件可以执行方法3300的一些处理，而诸如服务器的另一个设备可以执行方法3300的其他处理。根据一些这样的实现，回放控制服务器650可以执行搜索，并且可以将搜索结果提供给回放系统609，例如提供给回放控制模块605。

[0721] 在其他示例中，在块3310中确定搜索结果可以涉及实际执行搜索。例如，在一些这样的实现中，回放系统609可能能够执行搜索。如下面更详细地描述的，回放系统609和/或另一设备可以能够根据用户输入执行搜索，在一些示例中可以经由在显示设备上提供的图形用户界面来接收该用户输入。

[0722] 在一些实现中，块3310可以涉及对在块3305中接收的音频数据的多个特征执行并发搜索。能够对音频数据的多个特征执行并发搜索可以提供许多潜在的优点，部分地是因为会议参与者将常常记住特定会议体验的许多不同方面。上面描述的一个例子涉及对于会议参与者Fred Jones所说的“销售”一词的实例的多维搜索。在更详细的例子中，会议参与者可能会记得Fred Jones在三个星期的时间间隔期间有时进行陈述的同时提到“销售”。会议参与者可能已经能够从Fred Jones的声音的声调确定他对这个话题感到兴奋。会议参与者可能会记得Fred Jones正在他在旧金山的办公室中戴上耳机进行交谈。这些单独搜索功能中的每一个可能在使用本身时可能不是非常具体，但是当它们组合在一起时，它们可能非常具体，并且可以提供非常集中的搜索。

[0723] 在一些示例中，特征可以包括词语，该词语可以根据关键词检索索引从语音识别

程序的内部语音识别格结构确定，其中的一些示例在下面详细描述。这样的实现可以允许关于在会议中说出了哪些词语而非常快速地搜索语音识别器提供的许多并发假设。作为替代地或者附加地，搜索中使用的词语可以对应于从语音识别格确定的会议主题，例如，通过使用上述的“词语云”方法。

[0724] 本文公开了确定会议段的各种方法，其可以基于会话动态。在一些实现中，多维搜索可以至少部分地基于搜索一个或多个类型的会议段。

[0725] 在一些实现中，多维搜索可以至少部分地基于会议参与者身份。对于诸如移动电话或基于PC的软客户端的单方端点，一些实现可以涉及从设备ID记录每个会议参与者的姓名。对于互联网协议电话(VoIP)软客户端系统，用户经常被提示输入他或她的姓名进入会议。这些姓名可能会被记录下来供将来参考。对于扬声器电话设备，可以使用声纹分析来从被邀请参会的那些人中识别设备周围的每个发言者(如果记录/分析系统例如基于会议邀请已经知晓受邀者列表)。一些实现可以允许基于关于会议参与者身份的一般分类的搜索，例如，基于会议参与者是美国英语的男性发言者这一事实的搜索。

[0726] 在一些示例中，时间可能是可搜索的特征。例如，如果会议记录与其开始和结束时间以及日期一起被存储，则一些实现可以允许用户搜索在指定的日期和/或时间范围内的多个会议记录。

[0727] 一些实现可以允许用户基于会议参与者的情感来搜索一个或多个会议记录。例如，分析引擎307可以对音频数据执行一种或多种类型的分析，以从音频记录确定会议参与者情绪特征(参见例如Bachorowski,J.A。,&Owren,M.J.(2007).Voice expression of emotion。Lewis,M.,Haviland-Jones,J.M.,&Barrett,L.F.(Eds.),The Handbook of Emotion,3rd Edition.New York:Guilford,(印刷中),其通过引用并入本文)，例如兴奋、攻击性、或压力/认知负荷。(参见例如Yap,Tet Fei.,Speech production under cognitive load:Effects and classification,Dissertation,The University of New South Wales(2012)，其通过引用并入本文)。在一些实现中，结果可以被索引，被提供给回放系统609并且用作多维搜索的一部分。

[0728] 在一些示例中，端点位置可以是可搜索的特征。例如，对于安装在特定房间中的端点，可能先验地知道该位置。一些实现可以涉及基于由车载GPS接收机提供的位置信息记录移动端点位置。在一些示例中，可以基于端点的IP地址来定位VoIP客户端的位置。

[0729] 一些实现可以允许用户基于端点类型搜索一个或多个会议记录。如果会议记录标记了关于每个参与者使用的电话设备的类型的信息(例如，电话的制作和/或模型，基于web的软客户端的用户代理字符串，设备的类别(耳机，手机或扬声器电话)等)，在一些实现中，该信息可以被存储作为会议元数据，提供给回放系统609并用作多维搜索的一部分。

[0730] 在一些示例中，块3310可以涉及执行与多个会议的记录对应的音频数据的搜索。下面介绍一些例子。

[0731] 在该示例中，在块3310中确定的搜索结果对应于音频数据中的会议参与者语音的至少两个实例。这里，会议参与者语音的至少两个实例包括由第一会议参与者发出的至少第一语音实例和由第二会议参与者发出的至少第二语音实例。

[0732] 在该实现中，块3315涉及将会议参与者语音的实例渲染到虚拟声学空间的至少两个不同的虚拟会议参与者位置，使得第一语音实例被渲染给第一虚拟会议参与者位置，并

且第二实例语音被渲染给第二虚拟会议参与者位置。

[0733] 根据一些这样的实现，回放系统的一个或多个元件，诸如回放系统609的混合和渲染模块604，可以执行块3315的渲染操作。然而，在一些实现中，块3315的渲染操作可以至少部分地由诸如图6所示的渲染服务器660的其它设备来执行。

[0734] 在一些示例中，是回放系统609还是其他设备(诸如渲染服务器660)执行块3315的渲染操作可以至少部分地取决于渲染过程的复杂性。例如，如果块3315的渲染操作涉及从一组预定虚拟会议参与者位置中选择虚拟会议参与者位置，则块3315可能不涉及大量的计算开销。根据一些这样的实现，块3315可以由回放系统609执行。

[0735] 然而，在一些实现中，渲染操作可能更复杂。例如，一些实现可以涉及分析音频数据以确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据，指示会议参与者双讲话(在其期间至少两个会议参与者同时发言)的实例的数据、和/或指示会议参与者会话的实例的数据。

[0736] 一些这样的示例可以涉及将会话动态数据应用作为如下向量的空间优化成本函数的一个或多个变量，该向量描述了虚拟声学空间中的每个会议参与者的虚拟会议参与者位置。一些实现可以涉及将优化技术应用于空间优化成本函数以确定局部最优解，并至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置。

[0737] 在一些这样的实现中，确定会话动态数据、将优化技术应用到空间优化成本函数等可以由除回放系统609以外的模块(例如，由回放控制服务器650)执行。在一些实现中，这些操作中的至少一些可以先前已经例如由回放控制服务器650或由联合分析模块306执行。根据一些这样的实现，块3315可以涉及接收这样的过程的输出，例如，通过混合和渲染模块604接收被分配的虚拟会议参与者位置，并将会议参与者语音的实例渲染到至少两个不同的虚拟会议参与者位置。

[0738] 在图33所示的示例中，块3320涉及调度会议参与者语音的实例的至少一部分以进行同时回放，以产生回放音频数据。在一些实现中，调度可以包括至少部分地基于搜索相关性度量来调度会议参与者语音的实例以进行回放。例如，不是根据例如会议参与者语音的每个实例的开始时间来调度会议参与者语音以进行回放，一些这样的实现可以涉及调度具有相对更高的搜索相关性度量的会议参与者语音，以便比具有相对较低的搜索相关性度量的会议参与者语音更早地回放。下面介绍一些例子。

[0739] 根据一些实现，块3320可以涉及调度先前在时间上不重叠的会议参与者语音的实例以在时间上重叠地回放，和/或调度先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地回放。在一些实例中，可以根据例如如本文别处所公开的感知激发规则的集合来执行这样的调度。

[0740] 例如，感知激发规则的集合可以包括指示单个会议参与者的两个讲话突发不应该在时间上重叠的规则，和/或指示如果两个讲话突发对应于单个会话参与者端点则该两个讲话突发不应该在时间上重叠的规则。在一些实现中，感知激发规则的集合可以包括如下规则，其中给定两个连续的输入讲话突发A和B，A已经在B之前发生，对应于B的输出讲话突发的回放可在对应于A的输出讲话突发的回放完成之前开始，但是不会在对应于A的输出讲话突发的回放已开始之前开始。在一些示例中，感知激发规则集合可以包括如下规则，该规则允许对应于B的输出讲话突发的回放不早于在对应于A的输出讲话突发的回放完成之前

的时间T开始，其中T大于零。

[0741] 根据一些实现，方法3300可以包括将回放音频数据提供给扬声器系统。作为替代地或者附加地，方法3300可以包括将回放音频数据提供给诸如图6的显示设备610的其他设备，其能够向扬声器系统(例如，耳机607，耳塞，扬声器阵列608等)提供回放音频数据。

[0742] 图34是示出搜索系统元件的示例的框图。在该实现中，搜索系统3420包括搜索模块3421，扩展单元3425，合并单元3426和回放调度单元3406。在一些实现中，搜索模块3421，扩展单元3425，合并单元3426和/或回放调度单元3406可以至少部分地通过存储在非暂态介质上的指令(例如，软件)来实现，这种非暂态介质为诸如本文所描述的那些介质，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。在一些实现中，搜索模块3421，扩展单元3425，合并单元3426和/或回放调度单元3406可以至少部分地被实现为控制系统的元件，例如通过如图3A所示的装置的控制系统实现。控制系统可以包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑或离散硬件组件中的至少一个。根据一些实现，搜索模块3421，扩展单元3425，合并单元3426和/或回放调度单元3406可以至少部分地由图6所示的回放系统609的一个或多个元件来实现，例如通过回放控制模块605来实现。

[0743] 在该示例中，搜索模块3421能够接收一个或多个搜索参数3422并根据搜索索引3423执行搜索过程，以产生搜索结果列表3424。根据一些实现，搜索索引3423可以相当于由图5的关键词检索和索引模块505输出的搜索索引310。下面提供了搜索索引的额外示例。在一些实现中，搜索过程可以是多阶段搜索过程，例如，如下所述。

[0744] 在一些示例中，搜索模块3421能够执行常规的“关键词检索”功能，诸如D.Can和M.Maraclar，“Lattice Indexing for Spoken Term Detection”，IEEE TRANSACTIONS ON AUDIO，SPEECH，AND LANGUAGE PROCESSING，Vol.19，No.8，November 2011(“格索引出版物”)，其通过引入而并入本文。作为替代地或者附加地，搜索模块3421可以执行涉及多个特征的多维搜索。这样的特征可以包括词语，会议段，时间，会议参与者情绪，端点位置和/或端点类型。本文提供了各种示例。

[0745] 在图34中，搜索模块3421被示出为接收可以从用户输入导出的搜索参数3422的列表。在一个示例中，如果用户输入pet animal(宠物动物)，搜索参数将包括pet(宠物)和animal(动物)，这意味着用户想要找到词语pet或词语animal的实例。搜索系统领域的普通技术人员已知的这些和/或其他搜索定义和过程可以由搜索模块3421实现。例如，“san francisco”在双引号中输入的情况下可作为双连词被搜索，并且可以对应于参数列表3422的单个条目。因此，搜索模块3421可以采用搜索参数的交集，而不是并集。在一些实现中，搜索参数可以包括其他类型的特征，例如，指示搜索应当被限制到特定类型的会议段、特定会议的语音、特定日期或日期范围等的搜索参数。

[0746] 搜索索引3423可以允许搜索参数3422与在一个或多个会议记录中找到的相应参数进行高速匹配。在一些示例中，搜索索引3423可以允许搜索模块3421实现有限状态转换机方法，诸如在格索引出版物中描述的方法。在一些实现中，搜索索引3423可以具有更简单的搜索索引数据结构，诸如散列表或二叉树的结构。对于搜索模块3421实现“关键词搜索”搜索的实现，搜索索引3423可以允许用户从输入语音识别格中找到描述对于会议中检测到的每个话语的语音识别引擎的假设的词语。对于其中搜索模块3421实现如本文所公开的多

维搜索的实现，搜索索引还可以提供加速的找到诸如会议段的其他特征的方式。

[0747] 在该示例中，搜索结果3424可以包括被假设为与搜索参数相关的会议摘录的列表。会议摘录可以包括与被包括在搜索参数中的一个或多个词语对应的会议参与者语音的实例。例如，搜索结果3424可以包括假设词的列表、以及所估计的每个假设词的词语识别置信度分数。在一些实现中，列表上的每个条目可以包括端点标识符，摘录的开始时间(例如，相对于会议开始时间)和摘录的结束时间。如果搜索索引包含多个会议，则列表中的每个条目可以包括会议标识符。

[0748] 在一些实现中，词语识别置信度分数可以与搜索相关性度量相对应。然而，一些实现可以涉及其他类型的相关性评估，例如，如上文参考会议主题确定和词语云生成实现所描述的。在一些实施例中，相关性度量可以被限制在从0到1的范围内。在其他实施例中，相关性度量可以被限制在不同的数值范围内。例如，相关性度量可以采取对数成本的形式，其可以类似于上面的成本C_A和C_L。在其他示例中，相关性度量可以是无约束的量，其可仅对于比较两个结果是有用的。在一些示例中，搜索结果3424可以按相关性降序排列。回放调度单元3406可以调度最相关的结果被首先回放。

[0749] 在一些实现中，搜索系统3420能够修改包括在搜索结果3424中的会议参与者语音的实例中的一个或多个的开始时间或结束时间。在该示例中，扩展单元3425能够扩展对应于会议参与者语音的实例的时间间隔，从而提供更多的上下文。例如，如果用户正在搜索词语“pet”，则扩展单元3425可能能够确保词语“pet”之前和之后的某些词语被包括在会议参与者语音的相应实例中。不是仅指示词语“pet”，而是所得到的会议参与者语音的实例例如可能包括诸如“I don’t have many pets(我没有很多宠物)”，“I have a pet dog named Leo(我有一只名叫Leo的宠物狗)”等上下文词语。因此，收听会议参与者语音的这种实例的用户可以更好地能够确定哪些实例相对更有可能或相对更不可能是感兴趣的，并且可以更准确地决定哪些实例值得更详细地收听。

[0750] 在一些实现中，扩展单元3425可以在摘录的开始时间不能早于包含它的讲话突发的开始时间的约束下，从会议参与者语音的实例的开始时间减去固定偏移量(例如2秒)。在一些实现中，扩展单元3425可以在摘录的结束时间不能晚于包含它的讲话突发的结束时间的约束下，向会议参与者语音的实例的结束时间加上固定偏移量(例如2秒)。

[0751] 在该实现中，搜索系统3420包括合并单元3426，该合并单元3426能够合并在扩展后在时间上重叠的与单个会议端点对应的会议参与者语音的两个或更多个实例。因此，合并单元3426可以确保在查看搜索结果时会议参与者语音的同一实例不会被多次听到。在一些示例中，当会议参与者语音的实例被合并时，合并后的结果被分配合并的实例的所有输入相关性分数中的最高分(最相关)。

[0752] 在该示例中，合并单元3426产生的修改后的搜索结果列表形成输入到回放调度器3406的输入讲话突发3401的列表。在一些实现中，输入讲话突发3401的列表可相当于上文参考图13所述的会议段1301。

[0753] 在该实现中，回放调度单元3406能够调度会议参与者语音的实例进行回放。在某些实现中，回放调度单元3406能够调度具有相对更高的搜索相关性度量的会议参与者语音的实例，以比具有相对较低的搜索相关性度量的会议参与者语音的实例更早地回放。

[0754] 根据一些示例，回放调度单元3406可能能够提供与上文参考图13描述的回放调度

器1306类似的功能。类似地，在一些实现中，回放安排3411可以与上文参考图13描述的输出回放安排1311相当。因此，回放调度部3406可以能够调度先前在时间上未重叠的会议参与者语音的实例以在时间上重叠地回放，和/或调度先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地回放。在一些实例中，可以根据例如如本文别处所公开的一组感知激发的规则来执行这样的调度。

[0755] 图35示出了回放调度单元，合并单元和回放调度单元功能的示例。在该示例中，搜索结果3424的搜索结果部分3501被示出为具有按输入时间排列的会议参与者语音的实例3507A-3510A。这些实例实际上在此示例中按照相关性降序被排序，如搜索结果3424中所示，每个实例被示出具有相应的搜索相关性度量。在此示例中，搜索相关性度量的值为从0到10。这里，底层搜索涉及单个会议记录，并且端点3501A和350BB是同一会议内的搜索模块3421对于其已经返回结果的两个不同示例性端点。

[0756] 在该实现中，搜索结果部分3501包括会议的讲话突发3504-3506。在该示例中，在端点3501A处发出讲话突发3504和3506，并且在端点350B发出讲话突发3505。

[0757] 在该示例中，会议参与者语音的实例3507A是在端点3501A处发出的讲话突发3504(例如，一个句子)的一部分(例如一个词)。会议参与者语音的实例3507A的搜索相关性度量为2。这里，会议参与者语音的实例3508A是在端点3501B处发出的讲话突发3505的一部分。会议参与者语音的实例3508A的搜索相关性度量为10。会议参与者语音的实例3509A和3510A是在端点3501A处发出的讲话突发3506的不同部分(例如，句子中的词语的两个不同实例)。会议参与者语音的实例3509A和3510A的搜索相关性度量分别为7和8。

[0758] 在该示例中，搜索结果部分3501还示出了在扩展之后，例如在由图34的扩展单元3425处理之后，的会议参与者语音的实例。在该示例中，示出了会议参与者语音3507B-3510B的扩展实例。开始时间和结束时间已经被扩展，同时确保所得到的经扩展的会议参与者语音的实例3507B-3510B没有延伸超出其对应的讲话突发(例如，经扩展的会议参与者语音的实例3507B不会讲话突发3504的开始时间之前开始)。

[0759] 块3502示出了在扩展和合并之后的修改的示例搜索结果，为了清楚起见在输入时间中示出。会议参与者语音的实例实际上按相关性降序排序，如修改的搜索结果列表3512所示。在该示例中，从扩展和合并过程输出会议参与者语音的实例3507C，3508C和3510C。这里，实例3507C与实例3507B相同，因为扩展后没有发生合并。同样，在此示例中，实例3508C与实例3507C相同，因为扩展后没有发生合并。然而，实例3509B和3510B已经合并在一起，以形成实例3510C。这里，实例3509B和3510B已被合并，因为这两个会议参与者语音实例来自相同的端点并且在时间上重叠。在该示例中，两个搜索相关性度量中的较高者(8)被分配给所得到的实例3510C。

[0760] 在该示例中，块3503示出了在回放调度处理之后得到的输出回放安排3411的一部分。由于搜索结果3511和经修改的搜索结果3512按照相关性降序排序，所以会议参与者语音的实例3507D，3508D和3510D在输出时间上被调度以使得听众按相关性的降序听到输出。在该示例中，会议参与者语音的实例3507D，3508D和3510D中的每一个被调度为以比会议参与者语音的输入实例3507C，3508C和3510C更高的速率回放，因此相应的时间间隔被缩短。

[0761] 此外，在该示例中，已经在会议参与者语音的实例3508D和3510D之间引入了重叠。在该示例中，实例3510D被调度为在实例3508D被调度完成之前开始。这可以根据允许来自

不同端点的会议参与者语音的实例的这种重叠的感知激发规则被许可。在该示例中，实例3507D被调度为在实例3508D被调度完成时开始，以便消除居间时间间隔。但是，实例3507D没有被调度为在实例3508D调度完成之前开始，这是因为两个实例都来自同一端点。

[0762] 本文公开的各种实现涉及提供用于控制显示器以提供图形用户界面的指令。一些这样的方法可以包括接收对应于用户与图形用户界面的交互的输入，并且至少部分地基于该输入来处理音频数据。在一些示例中，输入可以对应于用于执行音频数据的搜索的一个或多个参数和/或特征。

[0763] 根据一些这样的实现，用于控制显示器的指令可以包括用于进行会议参与者的展示的指令。用于执行搜索的一个或多个参数和/或特征可以包括会议参与者的指示。在一些示例中，用于控制显示器的指令可以包括用于进行会议段的展示的指令。用于执行搜索的一个或多个参数和/或特征可以包括会议段的指示。根据一些实现，用于控制显示器的指令可以包括用于展示搜索特征的显示区域的指令。用于执行搜索的一个或多个参数和/或特征可以包括词语，时间，会议参与者情绪，端点位置和/或端点类型。本文公开了各种示例。

[0764] 图36示出了可以用于实现本公开的一些方面的图形用户界面的示例。在一些实现中，用户界面606d可以至少部分地基于由诸如图6所示的回放系统609的回放系统提供的信息而被呈现在显示器上。根据一些这样的实现，用户界面606d可以被呈现在诸如图6所示的显示设备610的显示设备的显示器上。

[0765] 在该实现中，用户界面606d包括会议参与者的列表2801。在此示例中，会议参与者的列表2801对应于多个单方端点，并且指示每个相应的会议参与者的姓名和图片。在该示例中，用户界面606d包括波形显示区域3601，该波形显示区域3601示出用于每个会议参与者的时间上的语音波形625。在该实现中，波形显示区域3601的时间尺度由波形显示区域3601内的垂直线指示，并且与会议记录的时间尺度相对应。该时间尺度在本文中可以称为“输入时间”。

[0766] 这里，用户界面606d还指示会议段1808K和1808L，它们分别对应于问答段和讨论段。在该示例中，用户界面606d还包括播放模式控件3608，用户可以在线性(输入时间)回放和非线性(调度的输出时间)回放之间切换。当回放被调度的输出时，在该实现中，点击回放模式控件3608允许用户更详细地查看结果(例如，以较慢的速度，具有附加的上下文)。

[0767] 这里，用户界面606d包括允许用户播放，暂停，倒带或快进内容的传输控件3609。在该示例中，用户界面606d还包括各种量过滤器3610，其控制返回的搜索结果的数量。在该示例中，量过滤器3610上指示的点越多，可能返回的搜索结果的数量越大。

[0768] 在该实现中，用户界面606d包括搜索窗口3105和用于输入搜索参数的文本字段3602。在一些示例中，用户可以将一个或多个显示的特征(例如会议段或会议参与者)“拖”到搜索窗口3105中和/或在文本字段3602中键入文本，以便指示该特征应被用于会议记录的搜索。在该示例中，搜索窗口3105的框3605指示用户已经发起了针对关键词“Portland”的实例的基于文本的搜索。

[0769] 在该示例中，用户界面606d还包括调度输出区域3604，其在此示例中具有输出时间(在本文中也称为“回放时间”)中的时间尺度。这里，线3606表示当前回放时间。因此，在该示例中，已经回放了会议参与者语音的实例3604A和3604B(分别具有最高和第二高搜索相关性度量)。在该实现中，调度输出区域3604中的会议参与者语音的实例3604A和3604B对

应于波形显示区域3601中所示的会议参与者语音的实例3601A和3601B。

[0770] 在该示例中，当前正在回放会议参与者语音的实例3604C和3604D。这里，会议参与者语音的实例3604C和3604D对应于波形显示区域3601中所示的会议参与者语音的实例3601C和3601D。在该实现中，会议参与者语音的实例3604E和3604F尚未被回放。在该示例中，会议参与者语音的实例3604E和3604F对应于波形显示区域3601中所示的会议参与者语音的实例3601E和3601F。

[0771] 在该示例中，会议参与者语音的实例3604A和3604B以及会议参与者语音的实例3604C和3604D被调度为在回放期间在时间上重叠。据一些实现，这根据如下的感知激发规则是可以接受的，该规则指示单个会议参与者或单个端点的两个讲话突发不应该在时间上重叠，但是该规则允许其它方式的重叠回放。然而，由于会议参与者语音的实例3604E和3604F来自同一端点和同一会话参与者，所以会议参与者语音的实例3604E和3604F没有被调度为重叠回放。

[0772] 图37示出了用于多维会议搜索的图形用户界面的示例。如图36所示的例子，框3605指示至少部分地基于对关键词“Portland”的搜索的用户对会议搜索的选择。然而，在该示例中，用户还将块3705a和3705b拖到搜索窗口3105中。块3705a对应于会议参与者Abigail Adams，并且块3705b对应于Q&A会议段。因此，已经对于在Q&A会议段期间由会议参与者Abigail Adams所说的词语“Portland”的实例执行了多维会议搜索。

[0773] 在该示例中，多维会议搜索已经返回了会议参与者语音的单个实例。该实例在波形显示区域3601中被示出为会议参与者语音的实例3601G，并且在调度输出区域3604中被示出为会议参与者语音的实例3604G。

[0774] 图38A示出了上下文增强语音识别格的示例部分。图38B和38C示出了可以通过使用如图38A所示的上下文增强语音识别格作为输入而生成的关键词检索索引数据结构的示例。例如，对于关键词检索索引3860a和3860b被示出的数据结构的示例可以用于实现涉及多个会议和/或多种类型的上下文信息的搜索。在一些实现中，关键词检索索引3860可以由图5所示的关键词检索和索引模块505输出，例如通过使用语音识别处理的结果(例如，语音识别结果401F-405F)作为输入。因此，关键词检索索引3860a和3860b可以是搜索索引310的实例。在一些示例中，上下文增强的语音识别格3850可以是由图4所示的自动语音识别模块405输出的语音识别结果的实例。在一些实现中，可以由大词汇量连续语音识别(LVCSR)过程基于加权有限状态转换器(WFST)来生成上下文增强的语音识别格3850。

[0775] 在图38A中，上下文增强的语音识别格3850的时间参考时间线3801被指示。图38所示的弧链接了上下文增强的语音识别格3850的节点或“状态”。例如，弧3807c链接两个状态3806和3808。如时间线3801中所示，开始时间3820和结束时间3822对应于弧3807c的时间跨度3809。

[0776] 在一些示例中，上下文增强的语音识别格3850可以包括用于每个弧的格式为“输入:输出/权重”的信息。在一些示例中，输入项可以对应于状态标识信息，如由用于弧3807b的状态标识数据3802所示。在一些实现中，状态标识数据3802可以是上下文相关的隐马尔可夫模型状态ID。输出项可以对应于词语标识信息，如由用于弧3807b的词语标识数据3803所示。在该示例中，“权重”项包括如本文其他地方所描述的词语识别置信度分数，其示例是弧3807b的分数3804。

[0777] 在该示例中，上下文增强的语音识别格3850的权重项还包括上下文信息，其示例是对于弧3807b示出的上下文信息3805。在会议期间，无论是面对面会议还是电话会议，及除了所说的词语和短语之外，会议参与者还可以观察和回忆上下文信息。在一些示例中，上下文信息3805可以例如包括从前端声学分析获得的音频场景信息。可以按不同的时间粒度以及通过各种模块来检索上下文信息3805。一些例子如下表所示：

[0778]

上下文信息	时间粒度	模块
端点类型	会议	系统硬件
发言者	会议	发言者标识
性别	会议	性别标识
位置	会议	车载GPS接收机，IP
会议段	段	分段单元1804
情绪	段	分析引擎307
可视线索	段	视频及屏幕分析器
距离	帧	音频场景分析
角度	帧	音频场景分析
扩散	帧	音频场景分析
信噪比	帧	前端处理

[0779] 表1

[0780] 在一些实现中，对于每个弧，不仅可以存储分数3804，而且可以存储上下文信息3805，例如以包含多个条目的“元组”的形式。可以基于相应时间跨度内的分数和上下文信息来分配值。在一些这样的实现中，可以针对整个会议或多个会议收集这样的数据。这些数据可以输入到统计分析中，以便获得诸如上下文分布之类的因素的先验知识。在一些示例中，这些上下文特征可以被归一化和聚集，并且可以通过矢量量化(VQ)过程对结果进行编码。

[0781] 图38B和38C中示出了关键词检索索引3860的数据结构的两个示例。在这两个示例中，用于上下文增强的语音识别格的每个弧的状态标识数据3802/词语标识数据3803对已经被变换为用于对应的关键词检索索引的每个弧的词语标识数据3803/词语标识数据3803A对。图38B和38C各自示出了关键词检索索引的非常小部分：在这些示例中，这些部分可以用于标记3元模型。

[0782] 在第一个例子中，如图38B所示，词语标识数据3803/词语标识数据3803A对被包括在对应的弧3830a-3832a中示出的对应的索引单元3810a-3810c的词语身份字段3812a-3812c中。在该示例中，分数3804，开始时间3820，结束时间3822和量化的上下文信息(本示例中为VQ索引3825a)存储在多维权重字段3813中。VQ索引有时在本文中称为“VQ ID”。该结构(在本文中可被称为“类型I”数据结构)具有至少三个潜在的优点。首先，将多维上下文信息变换为一维VQ索引3825a，这可以减少存储关键词检索索引3860所需的存储空间量。其次，索引结构可以与词语身份字段3812a-3812c中的输入和输出项一起存储，而不是例如词语和位置项。词语身份字段3812a-3812c的这个特征具有降低搜索复杂度的潜在优点。第三个优点是这种类型的数据结构(以及图38C中所示的“类型2”数据结构)有助于包括多个会议的记录的搜索和/或可涉及对于多种类型的上下文信息的并发搜索的搜索。

[0783] 类型1数据结构的一个潜在缺点是在一些示例中，用于搜索词语的附加后过滤处理之后可以是通过VQ索引过滤合格场景的过程。换句话说，基于具有类型1数据结构的关键词检索索引3860a的搜索可以是两阶段处理。第一阶段可以包括例如根据搜索查询的时间参数(例如开始时间和结束时间信息)来确定所希望的用于搜索的会议。第二阶段可以涉及根据其他搜索参数(可能包括基于上下文的查询)来检索搜索结果。

[0784] 图38C所示的类型2数据结构可以有助于更快的搜索。在该示例中，索引单元3811a-3811c包括相应的词语和VQ字段3814a-3814c，其包含词语/VQ元组。在该示例中，词语和VQ字段3814a-3814c包括包含词语标识数据3803和对应的VQ索引3825b的第一词语/VQ元组，以及包含词语标识数据3803A和相应的VQ指数3825c的第一词语/VQ元组。

[0785] 在该实现中，索引单元3811a-3811c中的每一个包括权重和时间字段3815，其包括分数3804，开始时间3820和结束时间3822。具有类型2数据结构的关键词检索索引3860b可以提供比具有类型1数据结构的关键词检索索引3860a更快的搜索。然而，具有类型2数据结构的关键词检索索引3860b可能需要比具有类型1数据结构的关键词检索索引3860a更多的存储空间。

[0786] 图39显示了被聚集的上下文特征的示例。此示例示出了两个显著上下文特征，设备类型和位置之间的关系。在本例中，竖轴表示位置，外部位置对应于“设备”轴下方的区域，而内部位置对应于设备轴下方区域。设备轴指示与移动设备、耳机，笔记本电脑和空间捕获设备(例如，空间会议电话)相对应的区域。在图39中，集群3901对应于在室内位置使用耳机的会议参与者，而集群3902和3905分别对应于使用笔记本电脑的室内和室外会议参与者。这里，集群3903对应于使用空间会议电话的室内会议参与者，而集群3904对应于使用移动设备的室外会议参与者。

[0787] 在一些实现中，可以在上下文索引的过程期间去除时间信息，这部分地是因为时间是连续的特殊上下文维度。此外，构建包括全局时间戳的大索引(例如，包括用于许多会议的音频数据)可能是有挑战性的。随着附加的会议被记录而对应的音频数据被处理，使用全局时间重建先前的索引可能是不可行的，因为该过程将为每个额外的会议记录引入额外的计算。

[0788] 图40是示出基于时间的分层索引的示例的框图。图40示出了分层索引4000，其中每个会议记录具有会议索引4001。在一天中可能存在多个会议记录，因此针对单个日索引4002指示多个会议索引4001。同样地，对于单个周指数4003指示多个日索引4002，并且针对单个月索引4004指示多个周索引4003。一些实现可以包括额外的层次级别，例如年索引，更少的层次级别和/或不同的层次级别。

[0789] 如图40所示，每当分层索引4000的任何层级的时间间隔结束时，相应的索引被建立，其将被全局时间戳散列表4005弄乱。例如，在每次会议结束时，在分层索引4000的最低层建立会议索引4001。例如，如果在特定一天中存在三个会议，则可以通过组合对于该三个会议中的每一个的关键词检索索引来创建对应的日索引4002。在该周末，可以产生周索引4003。可以在月底创建月索引4004。根据一些实现，开始和结束时间可以全局时间戳散列表4005以层级结构来保持。例如，高级时间戳散列表条目(例如，对于周索引4003)可以包括指向一个或多个低级索引(例如，日索引4002)中的每一个的指针。通过包含在每层中的相互关联的时间上下文信息，分层索引4000可以有助于跨多个会议记录的快速搜索。

[0790] 图41是示出上下文关键词搜索的示例的框图。在一些实现中，参考图41描述的处理可以至少部分地由诸如图34所示的搜索模块3421的并且如上所述的搜索模块来执行。在该示例中，接收到的查询4101被分成词语分量4103，时间分量4102和上下文分量4104。在一些实例中，词语分量4103可以包括一个或多个词语或短语。上下文分量4104可以包括一种或多种类型的上下文信息，包括但不限于上述表1所示的示例。

[0791] 在一些示例中，时间分量4102可以指示对应于单个会议的时间信息，而在其他示例中，时间分量4102可以指示对应于多个会议的时间信息。在该示例中，时间分量4102的时间信息被用于通过全局时间戳散列表4005(例如上文参照图40所述)对相应的索引进行过滤的过程(如图41中的过程4105所示)。下面参考图42描述过程4105的示例。

[0792] 在该示例中，将根据上下文分量4104中的信息来确定上下文索引。基于上下文索引，可以经由VQ码本4106搜索上下文输入以检索一组合格候选上下文VQ ID 4107。在一些实现中，一个或多个约束(例如距离限制(例如，欧几里德距离))可以应用于上下文输入搜索。

[0793] 在该示例中，依赖于关键词检索索引数据结构(其可以是图38所示的类型1或类型2数据结构)，可能存在不同类型的上下文索引单元。用于类型1数据结构的上下文索引单元可以具有基于词语的因子转换器索引，其与类型1数据结构的词语身份字段3812的数据结构相对应。因此，基于词语的因子转换器索引可以用于类型1上下文索引4109。用于类型2数据结构的上下文索引单元可以具有基于(词语，VQ ID)元组的因子转换器索引，其对应于词语的数据结构和类型2数据结构的VQ字段3814。因此，对于类型2上下文索引4108可使用基于(词语，VQ ID)元组的因子转换器索引。在一些实现中，检索过程可以涉及有限状态传感器复合操作。

[0794] 图42示出了自顶向下基于时间戳的散列搜索的示例。图42所示的示例可以是上文在图41的讨论中提及的过程4105的实例。在图42中，层级结构的每个级别对应于和(St，Ed)时间戳元组对应的不同时间间隔，(St，Ed)时间戳元组对应于开始时间和结束时间。每个块还包括指向处于不同级别的一个或多个块的指针“Pt”。在此示例中，级别4210是层级结构的最高级别。

[0795] 在该实现中，级别4210的每个块对应于1个月的时间间隔，而级别4220的每个块对应于1天的时间间隔。因此，可以看出，图42中的块的宽度不是准确地表示对应的时间间隔。级别4230的块对应于该示例中的各个会议。在一些这样的示例中，级别4230中的块的时间间隔可以根据每个会议的时间间隔而变化。在该示例中，如果查询的时间间隔(例如，如由接收到的查询4101的时间分量4102指示)不能跨越更高级别块的整个时间间隔，则搜索将转到较低的级别，以检索具有更详细的时间分辨率的相应的索引。

[0796] 例如，假设接收到的查询4101将包括时间分量4102，其对应于在从2014年10月1日至2014年11月2日的时间间隔内在太平洋标准时间下午2点举行的会议。在该示例中，块4201对应于2014年10月，并且块4202对应于2014年11月。因此，块4201的时间间隔将被接收到的查询4101的时间间隔完全涵盖。然而，块4202的时间间隔将不会被接收到的查询4101的时间间隔完全涵盖。

[0797] 因此，在该示例中，搜索引擎(例如，搜索模块3421)将该值提取到用于块4202的散列密钥，以获得指向较低级索引的指针Pt，该较低级索引在该实现中为级别4220。在该示例

中，块4203对应于2104年11月1日，并且块4204对应于2014年11月2日。因此，块4203的时间间隔将被接收到的查询4101的时间间隔完全涵盖，但是块4204的时间间隔将不会被接收到的查询4101的时间间隔完全涵盖。

[0798] 因此，在该示例中，搜索引擎将该值提取到块4204的散列密钥，以获得指向较低级索引的指针Pt，该较低级索引在该实现中是级别4230。在该示例中，2014年11月2日的前两次会议(对应于块4205和4206)的时间间隔被接收到的查询4101的时间间隔完全涵盖。在这种实例中，2014年11月2日的第三次会议的时间间隔(对应于块4207)是从下午1点至3点，并且因此不会被接收到的查询4101的时间间隔完全涵盖。然而，由于在此示例中，层级结构的最低级别对应于各个会议，所以对应于框4207的索引仍将被利用。然后，将使用全部选择的索引作为可在对其执行关键词检索的索引(类型1上下文索引4109或类型2上下文索引4108)数据库。

[0799] 如上所述，在一些实现中，检索过程可以涉及有限状态转换器复合操作。根据一些这样的示例，在获得结果之后，可以检索来自每个因子换能器弧的重量分量(例如，从索引单元3810的多维权重字段3813或从索引单元3811的权重和时间字段3815)。如图41所示，一些示例可以包括附加的后过滤过程4110，其用于基于类型1上下文索引的检索，以通过选择具有合格上下文ID的结果来过滤合格上下文。当使用基于类型2上下文索引的检索时，后过滤过程不是必需的，因此检索速度可能更快。

[0800] 与会议搜索有关的上述实现中的许多实现对于会议参与者的后续查看特别有用。现在将描述对于没有参加会议的人，例如对于无法参加的人员，尤其有用的各种实现。例如，查看会议记录的人可能希望获得会议的高级概述，以尽可能快地确定是否有可能讨论了听众感兴趣的任何材料。如果是这样，可能需要对会议录音(或至少其部分)进行更全面的查看。如果没有，不需要进一步查看。例如，听众可能希望确定谁参加了会议，讨论了哪些主题，谁做了大部分的发言等。

[0801] 因此，一些实现可能涉及仅选择总会议参与者语音的一部分以进行回放。“部分”可以包括会议参与者语音的一个或多个实例，例如一个或多个讲话突发和/或讲话突发的摘录。在一些示例中，选择过程可以涉及主题选择过程，讲话突发过滤处理和/或声学特征选择过程。一些示例可以涉及接收目标回放持续时间的指示。选择音频数据的部分可以包括使回放音频数据的持续时间在目标回放持续时间的阈值时间差之内。在一些示例中，选择过程可以包括仅保留一些讲话突发的一小部分和/或去除短的讲话突发，例如具有低于阈值持续时间的持续时间的讲话突发。

[0802] 图43是简述仅选择会议参与者语音的一部分以进行回放的方法的框图。与本文所述的其它方法一样，方法4300的块不一定按照所示的顺序执行。此外，这样的方法可以包括比所示和/或描述的块更多或更少的块。

[0803] 在一些实现中，方法4300可以至少部分地通过存储在非暂态介质上的指令(例如，软件)来实现，这种非暂态介质为诸如本文所描述的那些介质，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。在一些实现中，方法4300可以至少部分地由控制系统实现，例如由诸如图3A所示的装置的控制系统来实现。控制系统可以包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑或离散硬件组件中的至少一个。根据一些这样

的实现，方法4300可以至少部分地由图6所示的回放系统609的一个或多个元件实现，例如由回放控制模块605实现。作为替代地或者附加地，方法4300可以至少部分地由一个或多个服务器来实现。

[0804] 在该示例中，块4305涉及接收对应于会议记录的音频数据。在该示例中，音频数据包括对应于多个会议参与者中的每一个的会议参与者语音的数据。

[0805] 在图43所示的示例中，块4310涉及仅选择会议参与者语音的一部分作为回放音频数据。在一些实现中，图6所示的回放系统609的一个或多个元件(诸如回放控制模块605)可以执行块4310的选择过程。然而，在一些实现中，诸如服务器的另一设备可以执行块4310的选择过程。根据一些这样的实现，回放控制服务器650可以至少部分地执行块4310的选择处理。在一些这样的示例中，回放控制服务器650可以将选择处理的结果提供给回放系统609，例如提供给回放控制模块605。

[0806] 在该示例中，块4310涉及以下一个或多个：(a)主题选择过程，其根据所估计的会议参与语音与一个或多个会议主题的相关性来选择会议参与语音以供回放；(b)主题选择过程，其根据所估计的会议参与语音与会议段的一个或多个主题的相关性，选择会议参与者语音以供回放；(c)去除具有低于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发；(d)讲话突发过滤过程，去除具有等于或高于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发的一部分；和(e)根据至少一个声学特征选择会议参与语音以供回放的声学特征选择过程。如下面讨论的各种示例中所指出的，在一些实现中，选择可以涉及迭代过程。

[0807] 听众可能希望扫描涉及被估计为是最重要的会议话题的会议参与者讲话。例如，包括主题部分过程的一些实现可以涉及接收会议主题的主题列表并且确定所选择的会议主题的列表。主题列表可以例如先前由主题分析模块525生成，如上所述。所选择的会议主题的列表可能是主题列表的一个子集。确定所选择的会议主题的列表可能涉及主题排名过程。例如，一些这样的方法可以涉及接收主题排名数据，其指示主题列表上的每个会议主题的估计的相关性。在一些示例中，主题排名数据可以基于术语频率度量，例如本文别处公开的术语频率度量。确定所选择的会议主题的列表可以至少部分地基于主题排名数据。一些实现可以涉及用于多个会议段中的每一个的主题排名过程。

[0808] 作为替代地或者附加地，一些实现可以包括一种或多种类型的讲话突发过滤过程。在一些实现中，讲话突发过滤过程可以涉及去除至少一些输入讲话突发的初始部分。初始部分可以是从输入讲话突发开始时间到输出讲话突发开始时间的时间间隔。在一些实现中，初始部分可以是一秒，两秒等。一些这样的实现可以涉及去除在长讲话突发的开始附近的初始部分，例如至少具有阈值持续时间的讲话突发。

[0809] 这样的实现可能是有益的，因为人们经常以“插声停顿”(例如“嗯”，“哦”等)开始讲话突发。发明人有经验地确定，如果选择会议参与者语音的过程被影响以丢弃每个讲话突发的初始部分，则与选择过程保持每个语音突发开始时开始的语音相比，所得到的摘要(digest)倾向于包含更多的相关内容和更少的插声停顿。

[0810] 在一些实现中，讲话突发过滤过程可以包括至少部分地基于输入讲话突发持续时间来计算输出讲话突发持续时间。根据一些这样的实现，如果确定输出讲话突发持续时间超过输出讲话突发时间阈值，讲话突发过滤过程可以涉及为单个输入讲话突发生成会议参

与者语音的多个实例。在一些实现中，会议参与者语音的多个实例中的至少一个具有与输入讲话突发结束时间相对应的结束时间。下面更详细地描述讲话突发过滤过程的各种示例。

[0811] 涉及声学特征选择过程的一些实现可以包括根据音调变化，语音速率和/或响度选择会议参与语音以供回放。这种声学特征可以指示会议参与者的情绪，其可以对应于在相应的会议参与语音时被讨论的主旨的感知重要性。因此，根据这样的声学特征来选择会议参与语音以供回放可能是选择会议参与者语音的值得注意的部分的有用方法。

[0812] 如本文其他地方所述，在一些实现中，分析引擎307可以对音频数据执行更多类型的分析之一以确定会议参与者心情特征(参见例如Bachorowski，J.A。，&Owren，M.J.(2007).Voice expression of emotion。Lewis，M.，Haviland-Jones，J.M.，&Barrett，L.F.(Eds.)，The Handbook of Emotion，3rd Edition.New York:Guilford，(印刷中)，其通过引用并入本文)，例如兴奋、攻击性、或压力/认知负荷。(参见例如Yap，Tet Fei.，Speech production under cognitive load:Effects and classification，Dissertation，The University of New South Wales(2012)，其通过引用并入本文)。在一些实现中，分析引擎307可以在回放阶段之前执行这样的分析。一个或多个这样的分析的结果可以被索引，提供给回放系统609，并且被用作选择会议参与者语音以供回放的过程的一部分。

[0813] 根据一些实现，可以至少部分地根据用户输入来执行方法4300。例如，可以响应于用户与图形用户界面的交互来接收输入。在一些示例中，图形用户界面可以根据来自回放控制模块605的指令在诸如图6所示的显示设备610的显示器的显示器上被提供。回放控制模块605可以能够接收对应于用户与图形用户界面的交互的输入，以及至少部分地基于该输入来处理用于回放的音频数据。

[0814] 在一些示例中，用户输入可以涉及块4310的选择过程。在某些实例中，听众可能希望对所选择的会议参与者语音的回放时间设置时间限制。例如，听众只能在有限的时间内查看会议记录。听众可能希望尽可能快地扫描会议记录的突出部分，或许允许有一些额外的时间来查看感兴趣的部分。根据一些这样的实现，方法4300可以涉及接收包括目标回放持续时间的指示的用户输入。目标回放持续时间可以例如是在块4310中扫描所选择的会议参与者语音并且输出为回放音频数据所需的持续时间，。在一些示例中，目标回放持续时间可能不包括听众详细查看感兴趣的项目所需的额外时间。响应于用户与图形用户界面的交互，用户输入可被接收。

[0815] 在一些这样的示例中，框4310的选择过程可以包括根据目标回放持续时间来选择用于回放的会议参与者语音。选择过程例如可以包括使回放音频数据的持续时间在目标回放持续时间的阈值时间差之内。例如，阈值时间差可以是10秒，20秒，30秒，40秒，50秒，1分钟，2分钟，3分钟等。在一些实现中，选择过程可以包括使回放音频数据的持续时间在目标回放持续时间的阈值百分比之内。例如，阈值百分比可以是1％，5％，10％等。

[0816] 在一些实例中，用户输入可以涉及一个或多个搜索参数。这样的实现可以涉及至少部分地基于搜索相关性度量，选择会议参与者语音以用于回放和/或调度用于回放的会议参与者语音的实例。

[0817] 在该示例中，块4315包括将回放音频数据提供给用于回放的扬声器系统(例如，到耳机，耳塞，扬声器阵列等)。在一些示例中，块4315可以包括将回放音频数据直接提供给扬

声器系统，而在其他实现中，块4315可以涉及将回放音频数据提供给诸如图6所示的显示设备610的设备，该设备可以能够与扬声器系统通信。

[0818] 方法4300的一些实现可以涉及引入(或改变)会议参与者语音的实例之间的重叠。例如，一些实现可能涉及调度会议参与者语音的实例(其与会议参与者语音的另一实例先前在时间上没有重叠)以在时间上重叠地被回放，和/或调度会议参与者语音的实例(其与会议参与者语音的另一实例先前在时间上重叠)以在时间上进一步重叠地被回放。

[0819] 在一些这样的实现中，可以根据感知激发规则的集合来执行调度。例如，感知激发规则的集合可以包括指示单个会议参与者的两个讲话突发不应该在时间上重叠的规则，和/或指示如果两个讲话突发对应于单个端点，则该两个讲话突发不应该在时间上重叠的规则。在一些实现中，感知激发规则的集合可以包括如下规则，其中给定两个连续的输入讲话突发A和B，A已经在B之前发生，对应于B的会议参与者语音的实例的回放可在对应于A的会议参与者语音的实例的回放完成之前开始，但是不会在对应于A的会议参与者语音的实例的回放已开始之前开始。在一些示例中，感知激发规则集合可以包括如下规则，该规则允许对应于B的会议参与者语音的实例的回放不早于在对应于A的会议参与者语音的实例的回放完成之前的时间T开始，其中T大于零。

[0820] 方法4300的一些实现可以涉及通过利用空间渲染技术来减少回放时间。例如，音频数据可以包括被分别记录的来自多个端点的会议参与者语音数据，和/或来自对应于多个会议参与者的单个端点的并且包括多个会议参与者中的每个会议参与者的空间信息的会议参与者语音数据。一些这样的实现可以涉及将回放音频数据渲染在虚拟声学空间中，使得其语音被包括在回放音频数据中的每个会议参与者具有各自不同的虚拟会议参与者位置。

[0821] 然而，在一些实现中，渲染操作可能更复杂。例如，一些实现可以涉及分析音频数据以确定会话动态数据。会话动态数据可以包括指示会议参与者语音的频率和持续时间的数据，指示会议参与者双讲话(在此期间至少两个会议参与者同时发言)的实例的数据，和/或指示会议参与者会话的实例的数据。

[0822] 一些这样的示例可以涉及将会话动态数据应用作为描述各会议参与者在虚拟声学空间中的虚拟会议参与者位置的向量的空间优化成本函数的一个或多个变量。这样的实现可以包括将优化技术应用于空间优化成本函数，以确定局部最优解并且至少部分地基于局部最优解来在虚拟声空间中分配虚拟会议参与者位置。

[0823] 作为替代地或者附加地，一些实现可以涉及加速被回放的会议参与者语音。在一些实现中，回放音频数据的持续时间至少部分地通过将会议参与者语音的至少一些选定部分的持续时间乘以加速系数来确定。一些实现可以涉及将会议参与者语音的所有选定部分乘以加速系数。选定部分可以对应于单独的讲话突发，讲话突发的部分等。在一些实现中，选定部分可以对应于会议段的所有选择的会议参与者语音。下面介绍一些例子。

[0824] 图44示出了选择性摘要模块的示例。选择性摘要模块4400可能能够至少部分地执行上面参考图43描述的操作。在一些实现中，选择性摘要模块4400可以至少部分地通过存储在非暂态介质上的指令(例如，软件)来实现，这种非暂态介质为诸如本文所描述的那些介质，包括但不限于随机存取存储器(RAM)设备，只读存储器(ROM)设备等。在一些实现中，选择性摘要模块4400可以至少部分地通过控制系统来实现，例如通过如图3A所示的装置的

控制系统实现。控制系统可以包括通用单芯片或多芯片处理器，数字信号处理器(DSP)，专用集成电路(ASIC)，现场可编程门阵列(FPGA)或其他可编程逻辑器件，离散门或晶体管逻辑或离散硬件组件中的至少一个。根据一些这样的实现，选择摘要模块4400可以至少部分地由图6所示的回放系统609的一个或多个元件实现，例如通过回放控制模块605来实现。作为替代地或者附加地，选择性摘要模块4400可以至少部分地由一个或多个服务器来实现。

[0825] 选择性摘要模块4400可以例如仅选择包含在与一个或多个会议的记录相对应的所接收的音频数据中的会议参与者语音的一部分。在该示例中，选择性摘要模块4400能够从接收到的输入讲话突发的列表4430A中自适应地选择会议参与者语音的实例，使得当被调度时，对应于所选择的会议参与者语音的实例的回放音频数据的持续时间将接近接收到的目标回放时间持续时间4434的指示。会议参与者语音的实例可以例如包括讲话突发和/或讲话突发的部分，后者也可以在这里被称为“讲话突发摘录”。在一些实现中，选择摘要模块4400可能能够使回放音频数据的持续时间在目标回放时间段4434的阈值时间差或阈值时间百分比之内。

[0826] 在一些示例中，输入讲话突发的列表4430A可以包括会议中的所有讲话突发的列表。在替代示例中，输入讲话突发的列表4430A可以包括在会议的特定时间区域中的所有讲话突发的列表。在一些实现中，会议的时间区域可以与会议段对应。在一些示例中，输入讲话突发的列表4430A可以包括针对每个讲话突发的端点标识数据、开始时间和结束时间。

[0827] 在图44的示例中，选择性摘要4400被示出为输出所选择的讲话突发摘录的列表4424A。在一些实现中，所选择的讲话突发摘录的列表4424A可以包括针对每个所选摘录的端点标识数据、开始时间和结束时间。本文所描述的各种示例涉及输出用于回放的所选择的讲话突发摘录的列表，部分是因为这种讲话突发摘录可以被更快速地查看，并且在一些示例中可以包括相应的讲话突发的最显著部分。然而，一些实现涉及输出可以包括讲话突发和/或讲话突发摘录的会议参与者语音的所选实例的列表。

[0828] 在该示例中，选择性摘要4400还能够调度所选择的讲话突发摘录的列表4424A以进行回放。因此，选择性摘要4400还被示出为输出回放安排4411A。在该示例中，回放安排4411A描述如何回放会议的选择性摘要(会议参与者语音的所选实例的列表)或者电话会议的时间区域(例如，会议段)。在一些示例中，回放安排4411A可以类似于图34所示的输出回放安排3411，并且参考图34和35在上文被描述。

[0829] 图45示出了选择性摘要模块的元件的示例。在该示例中，选择性摘要模块4400包括选择器模块4531和回放调度单元4506。在该特定实现中，选择性摘要模块4400包括扩展单元4525和合并单元4526。然而，选择性摘要模块4400的替代实现可以包括或可以不包括扩展单元4525和/或合并单元4526。

[0830] 这里，选择器模块4531被示出为接收输入讲话突发的列表4430和目标回放持续时间的指示4434。在该示例中，选择器模块4531能够至少部分地基于目标回放持续时间4434和由实际持续时间复用器4532提供的经调度的回放持续时间4533，从输入讲话突发的列表4430中产生所选择的讲话突发摘录4424的候选列表。

[0831] 在该实现中，实际持续时间复用器4532确定当前迭代是否是第一次迭代，并且提供对应的经调度的回放持续时间。在一些实现中，在选择性摘要模块4400的操作的第一次迭代期间，将经调度的回放持续时间4533设置为零。这允许至少一次迭代，在该迭代期间，

扩展单元4525，合并单元4526和回放调度单元4506(或者在可能不包括扩展单元4525和/或合并单元4526的替代实现中，至少回放调度单元4506)可以对由选择器模块4531选择的讲话突发的摘录进行操作。在该例中，在后续的迭代中，由实际持续时间复用器4532提供给选择器模块4531的经调度的回放时间持续时间4533是由再现调度单元4506进行调度之后的实际的经调度的回放时间持续时间4535的值。这里，实际的经调度的回放时间持续时间4535对应于上述“回放音频数据的持续时间”。

[0832] 根据该示例，当经调度的回放持续时间4533处于目标回放持续时间4434的阈值范围内时，所选择的讲话突发摘录4424的候选列表被返回作为所选择的讲话突发摘录4424A的最终列表。在一个这样的示例中，阈值范围可以是+/-10％，这意味着经调度的回放持续时间4533必须小于或等于目标回放持续时间4434的110％，而大于或等于目标回放时间长度的90％。然而，在替代实例中，阈值范围可以是不同的百分比，例如1％，2％，4％，5％，8％，12％，15％等。在其他实现中，阈值范围可以是阈值时间差，例如10秒，20秒，30秒，40秒，50秒，1分钟，2分钟，3分钟等。

[0833] 在该示例中，扩展单元4525能够修改所选择的讲话突发4424的候选列表中的讲话突发摘录的开始和/或结束时间以提供附加的上下文。因此，在该示例中，扩展单元4525能够提供与如上参照图34所描述的扩展单元3425的功能相似的功能。因此，用户收听会议参与者语音的这种实例可以更好地确定哪些实例相对更可能或相对不太可能是感兴趣的，并且可以更准确地确定哪些实例值得更详细地收听。根据一些实现，扩展单元4525可以在讲话突发摘录的开始时间不早于包含它的讲话突发的开始时间的约束下，从讲话突发摘录的开始时间减去固定的偏移量t_{ex}(例如，1秒，2秒等)。根据一些示例，扩展单元4525可以在讲话突发摘录的结束时间可不晚于包含它的讲话突发的结束时间的约束下，将固定的偏移量t_{ex}(例如，1秒，2秒等)加到讲话突发摘录的结束时间上。

[0834] 在这种实现中，合并单元4526能够合并对应于单个会议端点和/或会议参与者的、在扩展之后在时间上重叠的会议参与者语音的两个或更多个实例。因此，合并单元4526可以确保在查看搜索结果时不会多次听到会议参与者语音的同一实例。在该示例中，合并单元4526能够提供与上文参考图34所述的合并单元3426的功能类似的功能。在该示例中，由合并单元4526产生的修改的讲话突发摘录的列表4501被断言给回放调度器4506。

[0835] 根据一些实施方式，回放调度单元4506可能能够提供诸如上文参照图13描述的回放调度器1306和/或上文参考图34和35所描述的回放调度单元3406的功能。因此，回放调度单元4506可以能够调度与会议参与者语音的另一实例先前在时间上没有重叠的会议参与者语音的实例(在此示例中，经修改的讲话突发摘录)以在时间上重叠地被回放，和/或调度与会议参与者语音的另一实例先前在时间上重叠的会议参与者语音的实例以在时间上进一步重叠地被回放。例如，回放调度单元4506可以根据一组感知激发规则来调度经修改的讲话突发摘录以进行回放。

[0836] 在该示例中，回放调度单元4506能够生成候选输出回放安排4411。例如，候选输出回放安排4411可以与上文参考图13描述的输出回放安排1311和/或以上参考图34和35描述的输出回放安排3411相当。在该实现中，当经调度的回放持续时间4533在目标回放持续时间4434的阈值范围内时，候选输出回放安排4411作为最终输出回放安排4411A被返回。

[0837] 在图45所示的例子中，回放调度单元4506返回实际的经调度的回放持续时间

4535,其对应于在通过回放调度部4506进行调度后的经修改的讲话突发摘录的回放时间。在替代实施方式中,可以在回放调度单元4506外确定实际的经调度的回放持续时间4535,例如通过将候选输出回放安排4411上的第一条目的输出开始时间与最后一个条目的输出结束时间进行比较。

[0838] 图46示出了用于将选择性摘要方法应用于分段会议的系统的示例。在一些实现中,选择性摘要系统4600可以至少部分地通过存储在非暂态介质上的指令(例如,软件)来实现,这种非暂态介质为诸如本文所描述的那些介质,包括但不限于随机存取存储器(RAM)设备,只读存储器(ROM)设备等。在一些实现中,选择性摘要系统4600可以至少部分地由控制系统实现,例如由诸如图3A所示的装置的控制系统来实现。控制系统可以包括通用单芯片或多芯片处理器,数字信号处理器(DSP),专用集成电路(ASIC),现场可编程门阵列(FPGA)或其他可编程逻辑器件,离散门或晶体管逻辑或离散硬件组件中的至少一个。根据一些这样的实施方式,选择性摘要系统4600可以至少部分地由图6所示的回放系统609的一个或多个元件实现,例如由回放控制模块605实现。作为替代地或者附加地,选择性摘要系统4600可以至少部分地由一个或多个服务器来实现。

[0839] 在一些实现中,选择性摘要系统4600可以包括比图46所示的元件更多或更少的元件。例如,在该实现中,选择性摘要系统4600包括多个选择性摘要模块4400A-4400C,每个会议段一个。然而,在一些替代实现中,对应于一些段(例如,混串音段和/或静默段)的音频数据将不被处理,并且将不存在对应的选择性摘要模块4400。在该示例中,示出正在处理来自仅三个会议段的音频数据,但是会议段1808B和1808C的表示之间的间隙预期表示一个或多个附加会议段。因此,在该示例中,输入音频数据4601表示用于整个会议记录的音频数据。其他示例可以涉及处理更多或更少的会议段,或者处理整个会议而不进行分段。

[0840] 在该示例中,选择性摘要模块4400A-4400C中的每一个接收分别与会议段1808A-1808C之一对应的输入讲话突发列表4430A-4430C中的对应的一个。这里,选择性摘要模块4400A-4400C中的每一个输出所选择的讲话突发摘录的逐段列表4624A-C(每个会议段一个)中的对应的一个。此外,每个选择性摘要模块4400A-4400C输出逐段输出回放安排4611A-4611C中的相应的一个。取决于具体实现,分段信息可以被包括也可以不被包括在选择性摘要模块4400A-4400C的输出中。

[0841] 在该实现中,选择性摘要系统4600包括时间乘法器4602A-4602C,每个正在处理音频数据的会议段一个。在一些示例中,通过将每个段的输入持续时间乘以系数α来计算每个段的目标回放时间,该系数反映了要加速回放的期望因子。在一些示例中,α可以在从0到1的范围内。在实验原型中成功使用的α的一些示例值分别包括0.5,0.333,0.25和0.1,分别对应于回放速率中的2x,3x,5x和10x加速。根据一些实现,α的值可以对应于关于回放速率中的期望加速的用户输入,或用户对回放速率中的最大容许加速的指示。

[0842] 在该示例中,选择性摘要系统4600包括拼接单元4603。这里,拼接单元4603能够将所选择的讲话突发摘录的逐段列表4624A-C(例如,按照每个会议段的开始时间的顺序)拼接成所选择的讲话突发摘录的最终列表4624D。在一些实现中,可以丢弃逐段输出回放安排4611A-4611C,而在其他实现中,可以保留逐段输出回放安排4611A-4611C。取决于具体实现,分段信息可以包括也可以不包括在级联单元4603的输出中。

[0843] 在该实现中,选择性摘要系统4600包括最终回放调度单元4606。在一些实现中,最

终回放调度单元4606可以具有与系统1700的功能类似的功能，系统1700包括段调度器单元1710并且在上文参考图17进行了描述。因此，最终回放调度单元4606可能能够调度来自连续段的所选择的讲话突发摘录在时间上重叠。

[0844] 在一些示例中，最终回放调度单元4606可以具有与在上文参考图45所述的回放调度单元4506的功能类似的功能。在一些这样的示例中，最终回放调度单元4606可以能够调度每个段的所选择的讲话突发摘录以在输出时间内彼此跟随。虽然可以安排一些讲话突发摘录来重叠回放，但是这样的实现可能不涉及调度整个会话段的所选择的讲话突发摘录以重叠回放。

[0845] 在该示例中，最终回放调度单元4606输出最终回放安排4611D，该最终回放安排4611D是在此示例中用于会议的所有选择的讲话突发摘录的安排。在一些实现中，最终回放安排4611D对应于与乘以系数α的电话会议的输入持续时间近似成比例的经调度的回放时间持续时间。然而，在替代实现(诸如涉及会议段的同时回放的那些)中，经调度的回放持续时间可能不与乘以系数α的电话会议的输入持续时间成比例。

[0846] 图47示出了根据一些实现的选择器模块的块的示例。在该示例中，选择器模块4531能够提供主题选择功能。例如，选择器模块4531可以基于所估计的与会议或段的整体主题的相关性来确定要选择会议参与者语音的哪些实例。

[0847] 在该示例中，选择器模块4531被示出为接收输入讲话突发列表4430和主题列表4701。在一些实现中，输入讲话突发列表4430和主题列表4701可以对应于整个会议，而在其他实现中，输入讲话突发列表4430和主题列表4701可以对应于会议段。主题列表4701可以例如对应于上文参考图25描述的主题列表2511。在一些实现中，主题列表4701中的主题可以按估计的重要性(例如根据术语频率度量)的降序存储。对于主题列表4701上的每个主题，可能会有会议参与者语音的一个或多个实例。会议参与者语音的每个实例可以具有端点指示，开始时间和结束时间。

[0848] 在该实现中，选择器模块4531被示出为接收目标回放持续时间4434和经调度的回放持续时间4533。可以根据来自用户界面的用户输入来接收目标回放持续时间4434，例如如上文参照图43和44所述。可以从回放调度单元4506接收经调度的回放持续时间4533，例如如上文参考图45所述。在该示例中，选择器模块4531能够在迭代过程中操作以调整要从主题列表4701保持的词语数N，直到经调度的回放持续时间4533在目标回放持续时间4434的预定范围内(例如，百分比或绝对时间范围)。如上所述，本文所用的术语“word(词语)”还可以包括短语，例如“living thing(生物)”。(在上面描述的一个例子中，短语“living thing”被描述为词”pet“的第三级上位词，词“animal”的第二级上位词和词“organism”的第一级上位词)

[0849] 在该示例中，选择器模块4531包括前N个词语选择器4702，其能够选择主题列表4701的N个最重要的词语，例如，根据术语频率度量来估计。例如，前N个词选择器4702可以按所估计的重要性的降序经过主题列表4701。对于遇到的每个主题，前N个词语选择器4702可以按降序取词，直到前N个词语的列表4703已经被编译。

[0850] 在该实现中，N的最终值根据由包括搜索调整单元4705和N初始化器4706的调整模块4710执行的迭代过程来确定。对于第一迭代，N初始化器4706将N设置为适当的初始值N_0。在该示例中，状态变量4707被示出在调整模块4710内，其是被存储的并且从迭代到迭代被

更新的可变值N。

[0851] 在该示例中，搜索调整单元4705能够基于N的先前值以及目标回放持续时间4434与经调度的回放持续时间4533之间的差，产生N的更新估计。如果经调度的回放持续时间4533太低，则搜索调整单元4705可以添加更多内容(换句话说，可以增加N的值)，而如果经调度的回放持续时间4533太高，则搜索调整单元4705可以删除内容(换句话说，可以降低N的值)。

[0852] 依赖于具体实现，搜索调整单元4705可以根据不同的方法来调整N的值。在一些示例中，搜索调整单元4705可以执行线性搜索。例如，搜索调整单元4705可以从$N(0)=N_0=0$开始。在每次迭代中，搜索调整单元4705可以将N增加固定量(例如，5或10)，直到目标回放持续时间4434与经调度的回放持续时间4533之间的差在预定范围内。

[0853] 在一些实现中，搜索调整单元4705可以执行不同类型的线性搜索。例如，搜索调整单元4705可以从$N(0)=N_0=0$开始。对于每次迭代，搜索调整单元4705可以增加N，使得来自主题列表4701上的下一个主题的所有词语被包括。搜索调整单元4705可以重复该过程，直到目标回放持续时间4434和经调度的回放持续时间4533之间的差在预定范围内。

[0854] 在替代实现中，搜索调整单元4705可以执行二分搜索。例如，在每次迭代期间，搜索调整单元4705可以保持N_{min}(N的下限)，和N_{max}(N的上限)。例如，搜索调整单元4705可以以$N_{min}(0)=0$，$N_{max}(0)=N_{total}$，$N(0)=N_0=\alpha N_{total}$，其中$N_{total}$表示主题列表4701的所有主题所包含的词语总数。对于每次迭代k，如果经调度的回放持续时间4533低于目标回放持续时间4434，则搜索调整单元4705可以如下地设置N_{min}和N_{max}：

[0855] $N_{min}(k)=N(k-1)$，$N_{max}(k)=N_{max}(k-1)$，

[0856] $$N(k)=\left\lfloor\frac{N_{min}(k)+N_{max}(k)}{2}\right\rfloor$$

[0857] 然而，如果经调度的回放持续时间4533高于目标回放持续时间4434，则搜索调整单元4705可以如下地设置N_{min}和N_{max}：

[0858] $$N_{min}(k)=N_{min}(k-1),\ N_{max}(k)=N(k-1),\ N(k)=\left\lfloor\frac{N_{min}(k)+N_{max}(k)}{2}\right\rfloor$$

[0859] 搜索调整单元4705可以重复该过程，直到目标回放持续时间4434和经调度的回放持续时间4533之间的差在预定范围内。

[0860] 在通过调整模块4710确定N的最终值之后，可以将N的最终值提供给前N个词语选择器4702。在该示例中，前N个词语选择器4702能够选择主题列表4701的N个最重要的词语，并输出前N个词语的列表4703。

[0861] 在该实现中，将前N个词语的列表4703提供给讲话突发过滤器4704。在该示例中，讲话突发过滤器4704仅保留在输入讲话突发列表4430和前N个词语的列表4703两者中存在的讲话突发的摘录。例如，保留词语可以按照它们在输入讲话突发的列表4430中被指定的顺序(例如按照时间顺序)返回到所选择的讲话突发摘录的列表4424中。虽然在图47中未示出，但是在一些示例中，所选择的讲话突发摘录的列表4424可以由扩展单元4525来处理，以便为讲话突发摘录提供更多的上下文。在一些实现中，所选择的讲话突发摘录的列表4424也可以由合并单元4526处理。

[0862] 图48A和48B示出了根据一些替代实现的选择器模块的块的示例。在该示例中，选

择器模块4531能够提供启发式选择功能。例如，选择器模块4531可以能够去除具有低于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发。作为替代地或者附加地，选择器模块4531可以能够去除具有等于或高于阈值输入讲话突发持续时间的输入讲话突发持续时间的至少一些输入讲话突发的一部分。在一些实现中，选择器模块4531能够仅保持每隔一个讲话突发的部分、每第三个讲话突发的部分，每第四个讲话突发的部分等。在一些实施方式中，选择器模块4531可能能够在没有关于会议主题的信息的情况下提供启发式选择功能。

[0863] 能够提供启发式选择功能的选择器模块4531的一些实现也可以包括扩展单元4525。在一些这样的实现中，当选择器模块4531提供启发式选择功能时，可以限制或取消扩展单元4525的效果，例如通过将t_{ex}设置为零或小的值(例如，0.1秒，0.2秒，0.3秒等)。根据一些这样的实现，讲话突发摘录的最小尺寸可以由下面描述的t_{speck}参数来控制。

[0864] 在该示例中，选择器模块4531被示出为接收输入讲话突发的列表4430。在一些实施方式中，输入讲话突发的列表4430可以对应于整个会议，而在其他实现中，输入讲话突发的列表4430和主题列表4701可以对应于会议段。在该实现中，选择器模块4531还被示出为接收目标回放持续时间4434和经调度的回放持续时间4533。可以根据来自用户界面的用户输入来接收目标回放持续时间4434，例如，如上文参照图43和44所述。可以从回放调度单元4506接收经调度的回放持续时间4533，例如，如上文参考图45所述。

[0865] 在该实现中，选择器模块4531能够应用迭代启发式选择过程来调整所选择的讲话突发的回放时间，直到所选择的讲话突发摘录的输出列表4424的经调度的回放持续时间4533在目标回放持续时间4434的预定范围(例如，百分比或绝对时间范围)内。

[0866] 在该示例中，选择器模块4531包括过滤器4801和调整模块4802。在一些实现中，过滤器4801可以应用两个参数K和t_{speck}。在一些这样的实现中，K可以表示参数，例如在0到1的范围内，其表示每个讲话突发的应该被保持的比例。根据一些这样的实现，t_{speck}可以表示例如可以以秒为单位测量的持续时间阈值(例如，讲话突发或讲话突发摘录的最小持续时间)。

[0867] 根据一些示例，对于每次迭代k，调整模块4802可以基于先前的值K(k-1)和t_{speck}(k-1)以及以及经调度的回放持续时间4533和目标回放持续时间4434之间的差来确定参数K(k)和t_{speck}(k)的新值。在一些这样的示例中，比t_{speck}短(在以K缩放之后)的讲话突发摘录可以被过滤器4801去除。

[0868] 在一些实现中，调整模块4802可以应用以下的一组启发式规则。在第一次迭代中，可以将K设置为最大值(例如，1)，并且将t_{speck}设置为零秒，使得保持所有内容。在随后的迭代中，K的值可以减小和/或t_{speck}的值可以增加，从而逐渐去除更多的内容，直到经调度的回放持续时间4533和目标回放持续时间4434之间的差在预定范围内，例如根据以下启发式规则。首先，如果t_{speck}小于阈值(例如，3秒，4秒，5秒等)，则一些实现涉及增加t_{speck}的值(例如，每次迭代0.1秒，0.2秒或0.3秒等)。根据一些这样的实现方式，在去除长讲话突发的部分的过程之前，将去除短的讲话突发(在阈值持续时间之下的那些)。

[0869] 如果在去除低于阈值持续时间的讲话突发之后，经调度的回放持续时间4533和目标回放持续时间4434之间的差仍然不在预定范围内，则一些实现涉及降低K的值。在一些示例中，可以通过应用公式K(k)＝β*K(k-1)来减小K的值，其中β在(0,1)范围内(例如，0.8，

0.85,0.9,0.95等)。根据这样的例子,内容将被去除,直至经调度的回放持续时间4533和目标回放持续时间4434之间的差在预定范围内。

[0870] 根据一些实现,来自输入讲话突发的列表4430的讲话突发可以被顺序地(例如按照时间顺序)呈现给过滤器4801。如图48B所示,对于给定的具有初始持续时间t_0的输入讲话突发4803,在一些示例中,过滤器4801或者产生相应的输出讲话突发摘录4804,其被添加到所选择的讲话突发摘录的列表4424中,或者消耗输入讲话突发4803而不产生相应的输出讲话突发摘录4804。

[0871] 根据一些示例,掌控过滤器4801的这种操作的启发式规则如下。在一些这样的示例中,过滤器4801将根据$t_1=Kt_0$计算候选输出讲话突发的输出持续时间t_1。根据一些这样的示例,如果$t_1<t_{speck}$,则过滤器4801将不会产生输出讲话突发。在一些示例中,过滤器4801可以根据以下来计算相对于输入讲话突发(4803)的开始时间的候选输出讲话突发的开始时间t_s:

[0872] $$t_s=\begin{cases} t_{um}, (t_{\mathrm{um}}+t_1)\le t_0 \\ t_0-t_1, \text{其它} \end{cases} \qquad (式\ 48)$$

[0873] 在式48中,t_{um}表示在一些示例中可能在[0,2]秒范围内的系数。在某些实现中,t_{um}的值可被选择为使得通常保留在长讲话突发的开头附近的语音,而不是在长讲话突发的刚开始处的语音。这种选择的动机是人们常常开始用诸如“嗯”,“哦”等插声停顿来开始讲话突发。本发明人通过实验确定,如果选择器偏向于忽略在长讲话突发刚刚开始时的语音(例如,在每个讲话突发的前1秒期间,在每个讲话突发的前1.5秒期间,在每个讲话突发的前2秒期间,等等)则与如果选择器模块4531保持在每个讲话突发的刚开始处开始的语音相比,所得到的摘要包含更多相关的内容和更少的插声停顿。

[0874] 在一些实现中,过滤器4801可以为单个输入讲话突发4803产生多个讲话突发摘录。根据一些这样的实现方式,多个讲话突发摘录中的至少一个可能具有与输入讲话突发结束时间相对应的结束时间。

[0875] 在一些这样的示例中,当候选输出讲话突发t_1的持续时间超过第一阈值t_2(例如,8秒,10秒,12秒等)但是小于阈值t_3(例如,15秒,20秒,25秒,30秒等),过滤器4801可以产生两个输出讲话突发摘录。例如,第一输出讲话突发摘录可以相对于输入讲话突发的开始时间在时间t_s开始,并且可以具有持续时间$t_1/2$。在一些这样的示例中,第二输出讲话突发摘录也可以具有持续时间$t_1/2$,并且可以在输入讲话突发4803结束之前$t_1/2$的时间开始,使得第二输出讲话突发摘录的结束时间段对应于输入讲话突发的结束时间。

[0876] 根据一些这样的实现方式,当候选输出讲话摘录t_1的长度超过阈值t_3时,过滤器4801可以产生三个输出讲话突发摘录。例如,第一输出讲话突发摘录可以相对于输入讲话突发的开始时间在时间t_s处开始,并且可以具有持续时间$t_1/3$。第三输出讲话突发摘录也可以具有持续时间$t_1/3$,并且可以在输入讲话突发4803结束之前$t_1/3$的时间开始,使得第三输出讲话突发摘录的结束时间对应于输入讲话突发的结束时间。根据一些这样的示例,第二输出讲话突发摘录也可以具有持续时间$t_1/3$,并且可以在时间$((t_0+t_s)-t_1/3))/2$开始。因此,第二输出讲话突发摘录的开始时间可被选择为使得第二输出讲话突发摘录在第一和第三输出讲话突发摘录之间。

[0877] 在一些实施方式中,过滤器4801可以产生四个或更多个输出讲话突发摘录。根据

一些这样的实现方案，多个输出讲话突发摘录中的至少一个可能具有与输入讲话突发的结束时间对应的结束时间。在一些这样的示例中，输出讲话突发摘录可以对应于从输入讲话突发4803以规则的间隔取得的样本，从而长输入讲话突发4803的语音被规则地采样。

[0878] 图49示出了根据其他替代实现的选择器模块的框的示例。在该示例中，选择器模块4531能够提供声学特征选择功能。例如，选择器模块4531可以基于为每个讲话突发计算的声学特征(诸如音调方差，语音速率，响度等)来确定要选择会议参与者语音的哪个实例，该声学特征可以指示哪个讲话突发是相对的更令人兴奋的。这种功能是基于经验观察的，该经验观察表明当讲话者关于一个主题更激动时，存在可用于检测到这种兴奋的相应声学特征。可以假设当讲话者更加兴奋的时候，听众也可能对这个话题更有兴趣。

[0879] 在该示例中，选择器模块4531被示出为接收输入讲话突发的列表4430和声学特征列表4901。在一些实施方式中，输入讲话突发的列表4430和声学特征列表4901可以对应于整个会议，而在其他实现中，输入讲话突发的列表4430和声学特征列表4901可以对应于会议段。例如，分析引擎307可以在先执行对会议记录的音频数据的更多类型的分析之一以确定会议参与者情绪特征，例如兴奋、攻击性、或压力/认知负荷。上面描述了一些示例。声学特征列表4901可以是这种分析的结果。声学特征列表4901上的每个条目可以是会议参与者语音的实例，例如讲话突发或讲话突发摘录。会议参与者语音的每个实例可以具有端点指示，开始时间和结束时间。

[0880] 在一些实现中，声学特征列表4901可以按所估计的重要性(例如根据兴奋度量)的降序被存储。兴奋度量可以例如是音调方差，语速和/或响度的函数。然而，一些类型的“激动的言论”，如笑声，可能很容易发现，而不一定对应于重要的话题。相反，笑声可能对应于个人评论，非主题笑话等。因此，一些实现可以涉及将相对较低的重要性水平(例如，通过分配相对较低的兴奋度量)分配给检测到的会议参与者笑声的实例。

[0881] 根据一些实现，对于声学特征可能变化很大的长讲话突发，讲话突发可以被分成几个单独的条目，每个条目根据本地声学特征进行排名。例如，具有超过20秒的持续时间的讲话突发可以被分成不超过10秒长的一系列讲话突发，每个具有单独计算的声学特征。

[0882] 在一些示例中，声学特征列表4901可以基于音调方差。在一个示例中，兴奋度量可以如下计算。可以使用已知的音调跟踪技术(例如根倒谱技术)为每个音频帧提取基频估计(F0)。然后，F0的值可以转换为半音，以消除男性和女性讲话者之间的变化。可以针对每个讲话突发或讲话突发摘录计算半音值的标准偏差。标准偏差可以用作该讲话突发或讲话突发摘录的兴奋度量。声学特征列表4901可以通过根据兴奋度量以降序排序讲话突发和/或讲话突发摘录而被创建。

[0883] 在该实现中，选择器模块4531被示出为接收目标回放持续时间4434和经调度的回放持续时间4533。可以根据来自用户界面的用户输入来接收目标回放持续时间4434，例如，如上文参照图43和44所述。可以从回放调度单元4506接收经调度的回放持续时间4533，例如，如上文参考图45所述。在该示例中，选择器模块4531能够在迭代过程中进行操作，以调整要从声学特征列表4901保留的讲话突发(或讲话突发)的数量N，直到经调度的回放持续时间4533在目标回放持续时间4434的预定范围(例如百分比或绝对时间范围)内。

[0884] 在该示例中，选择器模块4531包括能够选择声学特征列表4901的N个最重要的讲话突发(或讲话突发摘录)的前N个讲话突发选择器4902，例如，根据术语频率度量来估计。

前N个讲话突发选择器4902可以例如按估计的重要性的降序通过声学特征列表4901，直到编译了前N个讲话突发(或讲话突发摘录)的列表4903。

[0885] 在该实现中，N的最终值根据由包括搜索调整单元4905和N初始化器4906的调整模块4910执行的迭代过程来确定。在一些实施方式中，调整模块4910可以具有诸如上文参考图47的调整模块4710所描述的功能。对于第一迭代，N初始化器4906将N设置为适当的初始值N_0。在该示例中，状态变量4907被示出在调整模块4910内，其是被存储的并且从迭代到迭代被更新的可变值N。

[0886] 在该示例中，搜索调整单元4905能够基于N的先前值以及目标回放持续时间4434与经调度的回放持续时间4533之间的差，产生N的更新估计。一般来说，如果经调度的回放持续时间4533太低，则搜索调整单元4905可以添加更多内容(换句话说，可以增加N的值)，而如果经调度的回放持续时间4533太高，则搜索调整单元4905可以删除内容(换句话说，可以降低N的值)。

[0887] 依赖于具体实现，搜索调整单元4905可以根据不同的方法来调整N的值。在一些示例中，搜索调整单元4905可以执行线性搜索或者二分搜索，如上文参照图47的搜索调整单元4705所描述的。

[0888] 在通过调整模块4910确定N的最终值之后，可以将N的最终值提供给前N个讲话突发选择器4902。在该示例中，前N个讲话突发选择器4902能够选择声学特征列表4901的N个最重要的讲话突发(或者讲话突发摘录)，并输出前N个讲话突发(或者讲话突发摘录)的列表4903。

[0889] 在该实现中，列表4903提供给讲话突发过滤器4904。在该示例中，讲话突发过滤器904仅保留在输入讲话突发列表4430和列表4903两者中存在的讲话突发(或讲话突发摘录)。例如，保留的讲话突发(或讲话突发摘录)可以按照它们在输入讲话突发的列表4430中被指定的顺序(例如按照时间顺序)返回到所选择的讲话突发(或讲话突发摘录)的列表4424中。虽然在图49中未示出，但是在一些示例中，讲话突发摘录可以由扩展单元4525来处理以便提供更多的上下文。在一些实现中，讲话突发摘录也可以由合并单元4526处理。

[0890] 在本公开中描述的实现的各种修改对于本领域普通技术人员来说是显而易见的。在不脱离本公开的范围的实例中，本文定义的一般原理可以应用于其他实现。例如，一些替代实现不涉及根据TF-IDF算法来确定术语频率度量。一些这样的实现可以涉及使用简约的语言模型来生成主题列表。

[0891] 一些实现可以涉及将讲话突发过滤过程与声学特征选择过程相组合。根据一些这样的实现，至少部分地基于讲话突发持续时间的语音突发过滤过程可以与至少部分地基于音调变化的声学特征选择过程组合。例如，如果K为0.5(对应于保留了输入讲话突发的一半的示例)，则可以保持具有较大间距变化的半讲话突发。

[0892] 在涉及将讲话突发过滤处理与声学特征选择过程组合的另一种实现中，可以识别基于音调变化和讲话突发长度的输入讲话突发的排名，并且可以通过使用加权因子来生成组合排名。在一个这样的例子中，相等的权重(0.5)可以被分配用于音调变化和讲话突发长度。排名阈值可以位于实现期望的压缩比之处(换句话说，目标回放持续时间4434和经调度的回放持续时间4533之间的差在预定范围内的阈值)。组合排名低于阈值的讲话突发可能被删除。

[0893] 作为替代地或者附加地，一些实现可以涉及将主题选择过程与声学特征选择过程组合。根据一些这样的实现，可以根据声学特征选择过程(例如，根据诸如音调变化的兴奋度量)对与同一主题相关的会议参与者语音的实例进行排名。在其他实现中，用于输入讲话突发的排名可以基于声学特征选择过程和主题选择过程。可以通过使用加权因子来生成根据这两个过程的组合排名。

[0894] 一些实现可能涉及将会话动态分析与声学特征选择过程相结合。根据一些这样的实现，可以根据兴奋度量(例如音调变化)的突然增加和/或说话之后的双讲话的突然增加来识别与对于话语的兴奋响应相对应的会议参与者语音的实例。在一些例子中，可以通过说话后的静默时间间隔和/或兴奋度量的突然增加和/或静默时间间隔之后双讲话的突然增加，来识别与说话之后的“死寂”相对应的会议参与者语音的实例。

[0895] 因此，权利要求不旨在限于本文所示的实施方式，而是应被给予与本公开，本文公开的原理和新颖特征相一致的最宽范围。

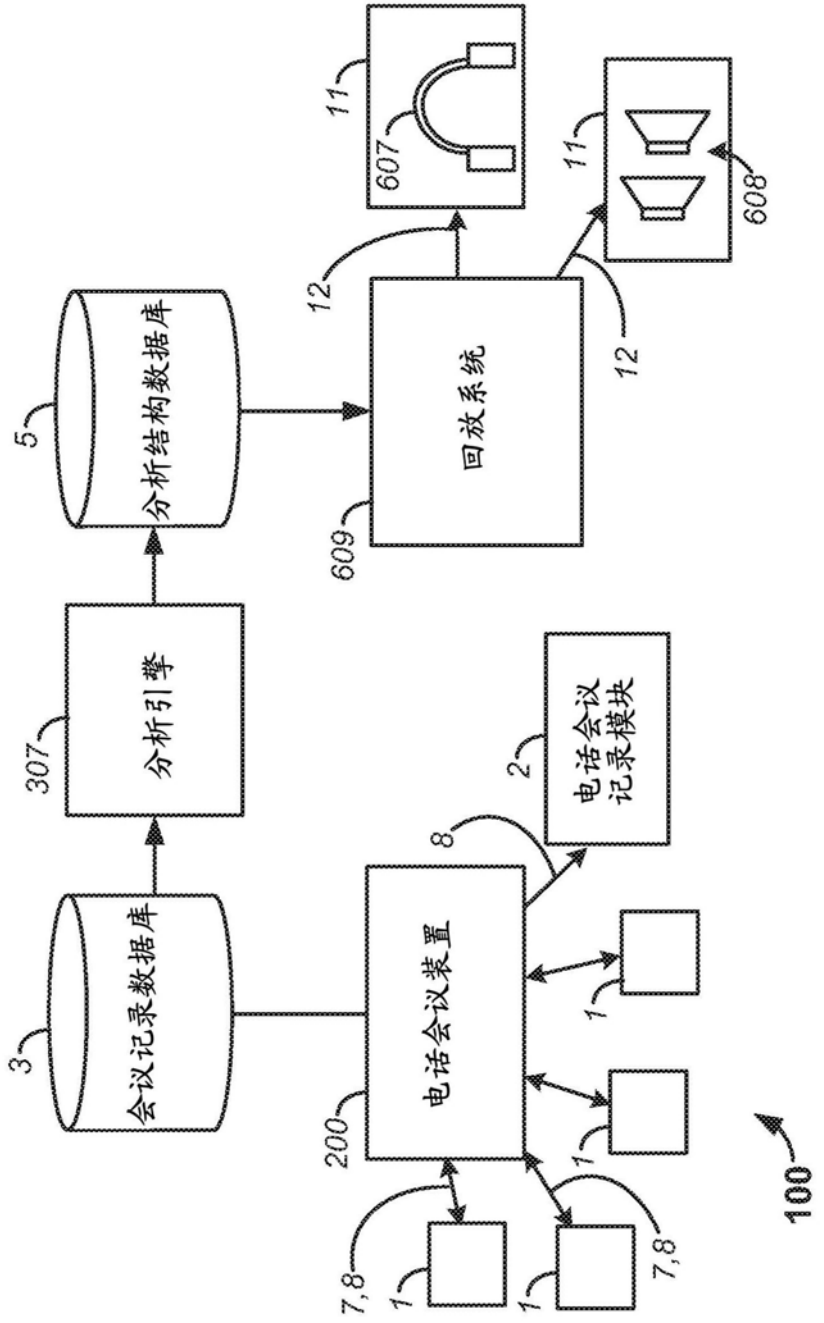

11
607
11
608
12
12
5
分析结构数据库
回放系统
609
307
分析引擎
2
电话会议
记录模块
8
3
会议记录数据库
电话会议装置
200
1
1
1
7,8
1
7,8
100

图1A

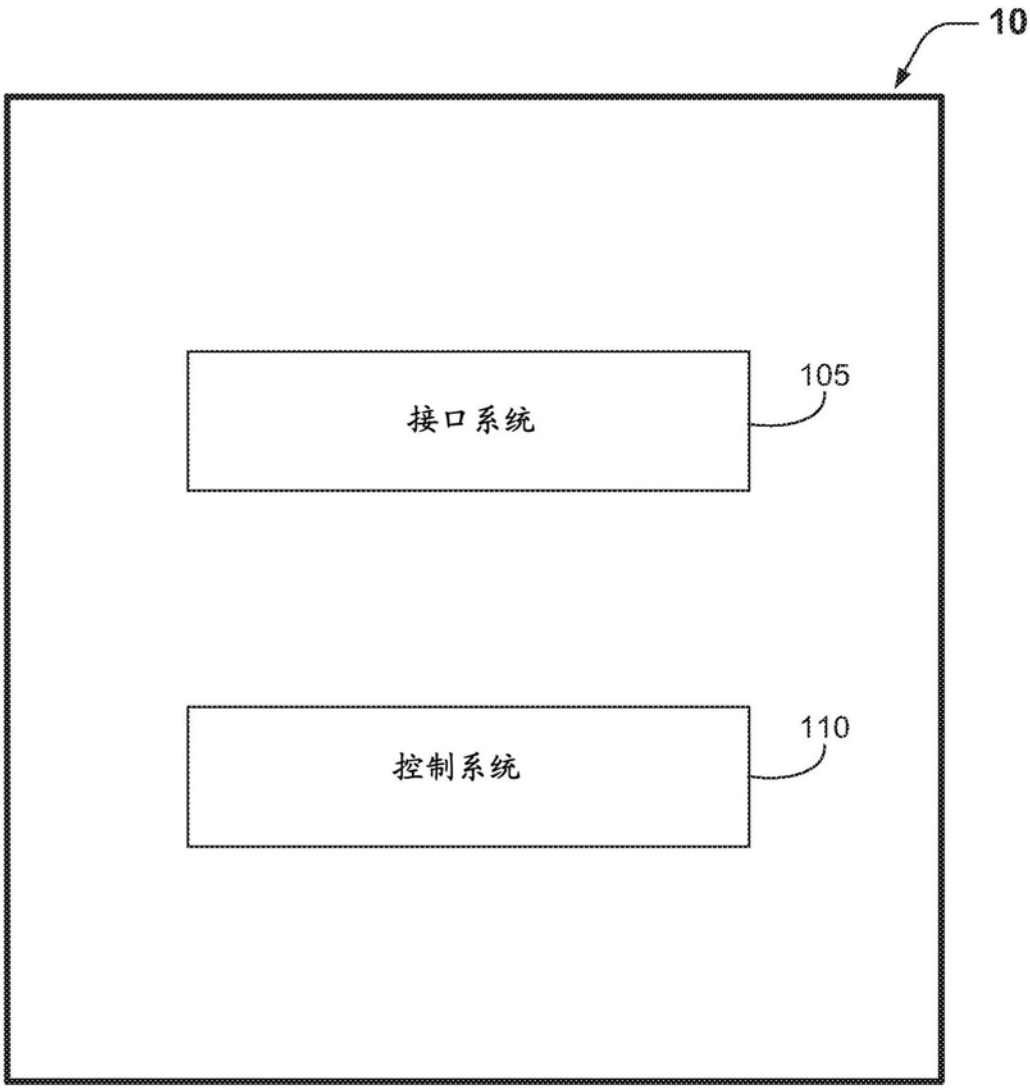


图1B

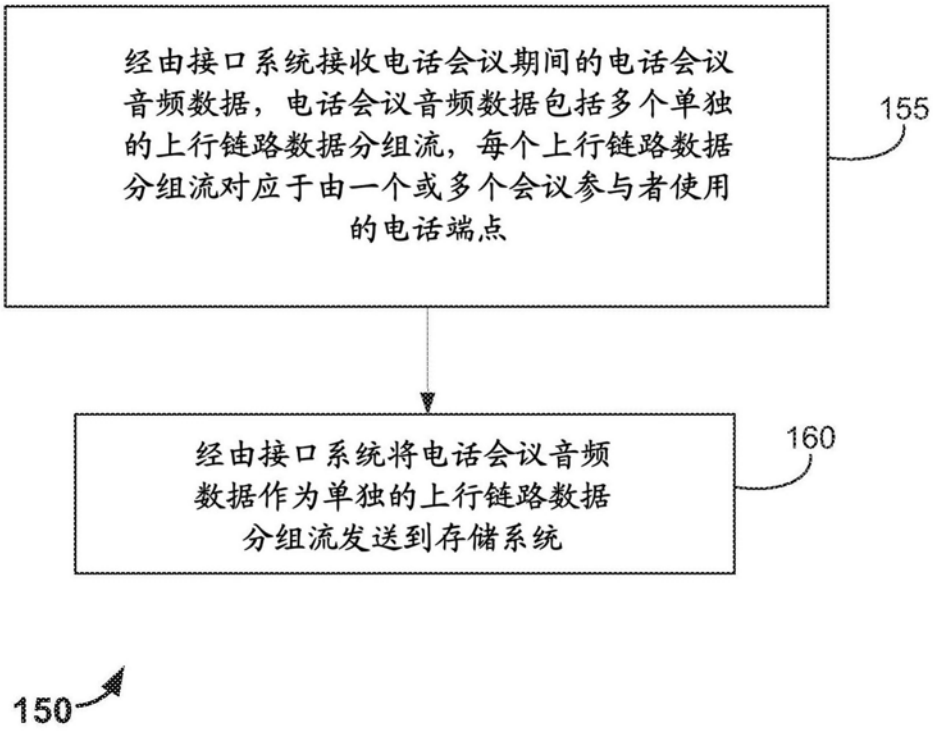

经由接口系统接收电话会议期间的电话会议音频数据，电话会议音频数据包括多个单独的上行链路数据分组流，每个上行链路数据分组流对应于由一个或多个会议参与者使用的电话端点
155
经由接口系统将电话会议音频数据作为单独的上行链路数据分组流发送到存储系统
160
150

图1C

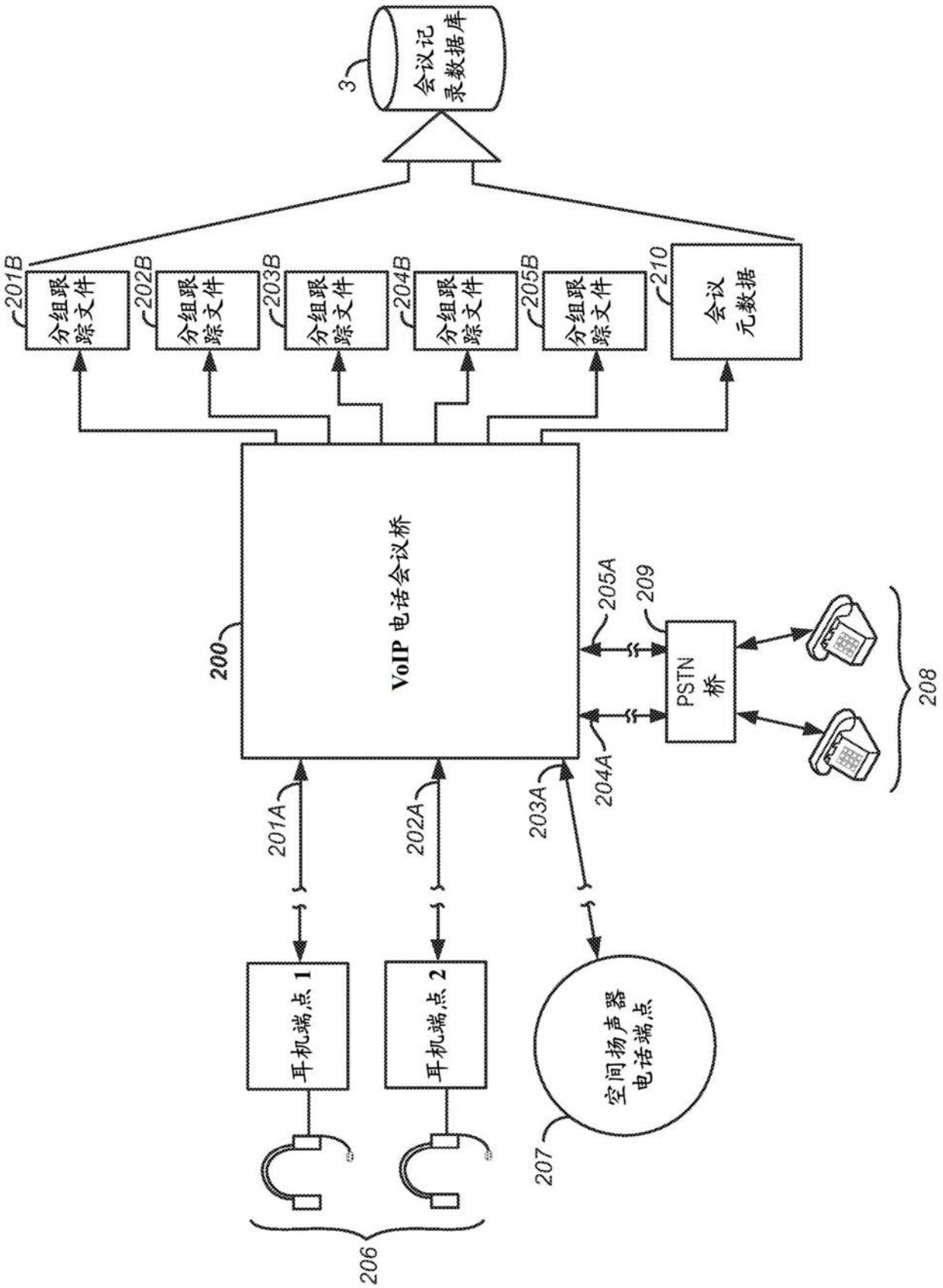

3
会议记录数据库
201B
分组跟踪文件
202B
分组跟踪文件
203B
分组跟踪文件
204B
分组跟踪文件
205B
分组跟踪文件
210
会议元数据
200
VoIP 电话会议桥
205A
209
PSTN 桥
208
204A
201A
202A
203A
耳机端点 1
耳机端点 2
空间扬声器电话端点
206
207

图2A

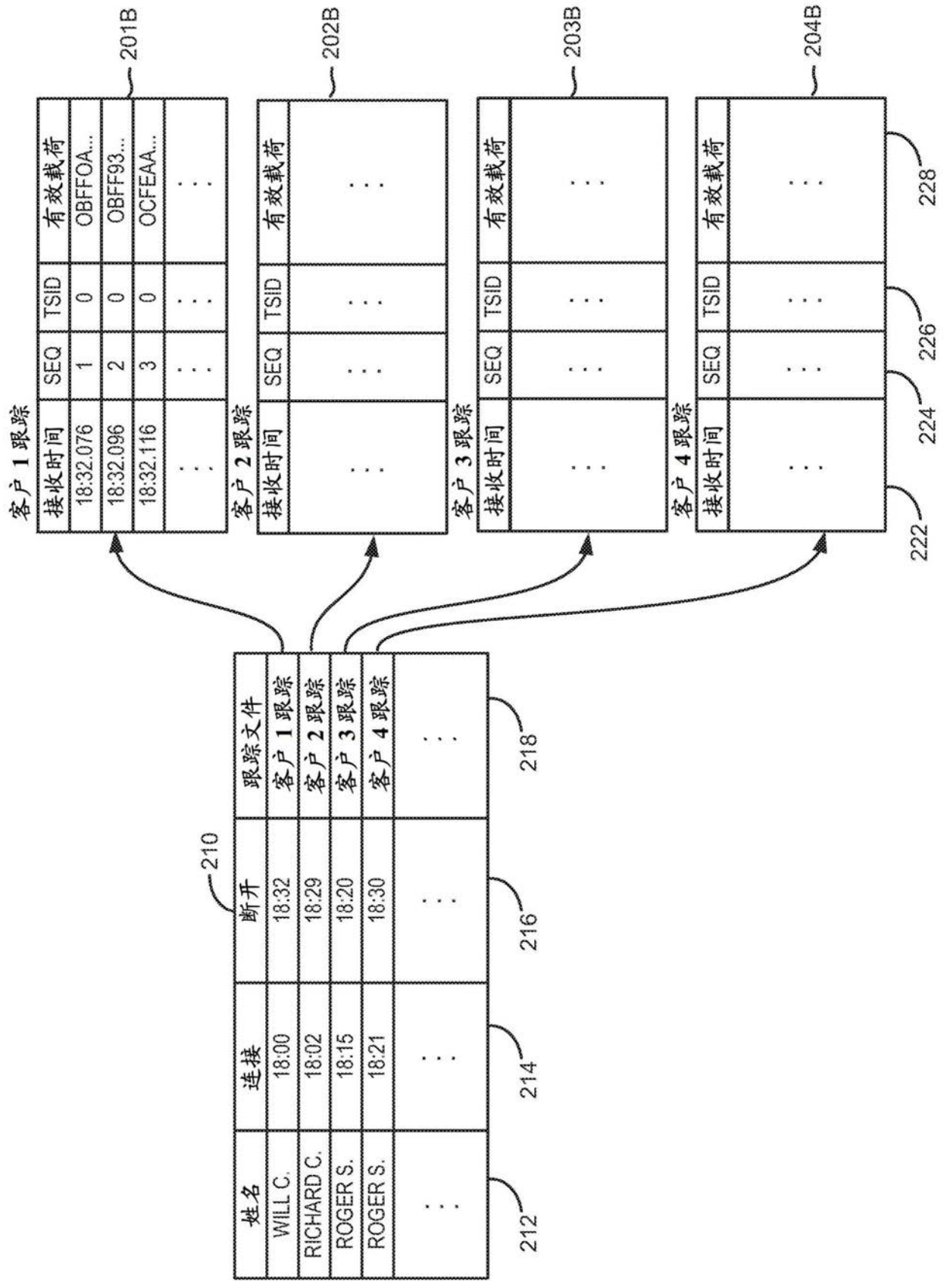

客户 1 跟踪
接收时间 SEQ TSID 有效载荷
18:32.076 1 0 OBFFOA...
18:32.096 2 0 OBFF93...
18:32.116 3 0 OCFEAA...
201B
客户 2 跟踪
接收时间 SEQ TSID 有效载荷
202B
客户 3 跟踪
接收时间 SEQ TSID 有效载荷
203B
客户 4 跟踪
接收时间 SEQ TSID 有效载荷
204B
222
224
226
228
210
姓名 连接 断开 跟踪文件
WILL C. 18:00 18:32 客户 1 跟踪
RICHARD C. 18:02 18:29 客户 2 跟踪
ROGER S. 18:15 18:20 客户 3 跟踪
ROGER S. 18:21 18:30 客户 4 跟踪
212
214
216
218

图2B

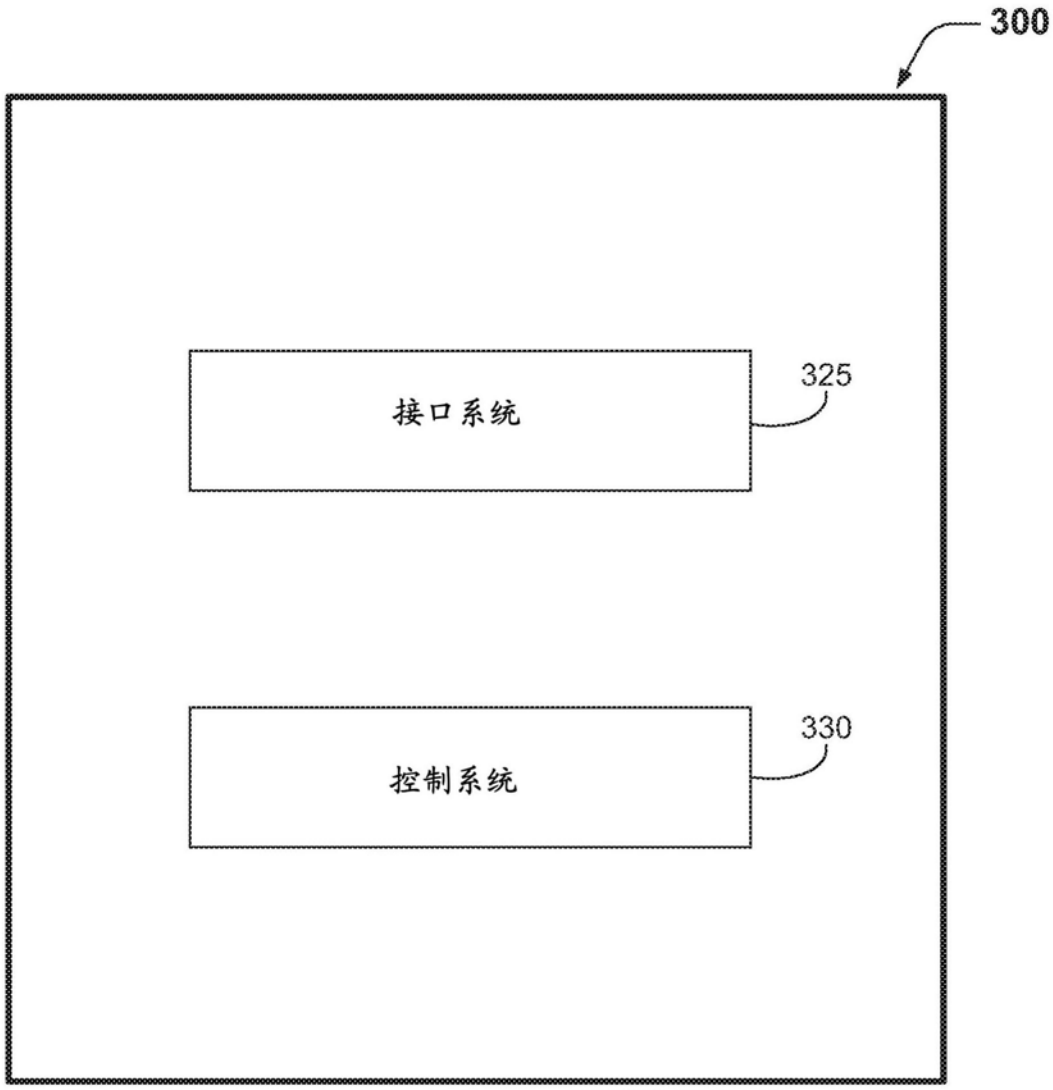


图3A

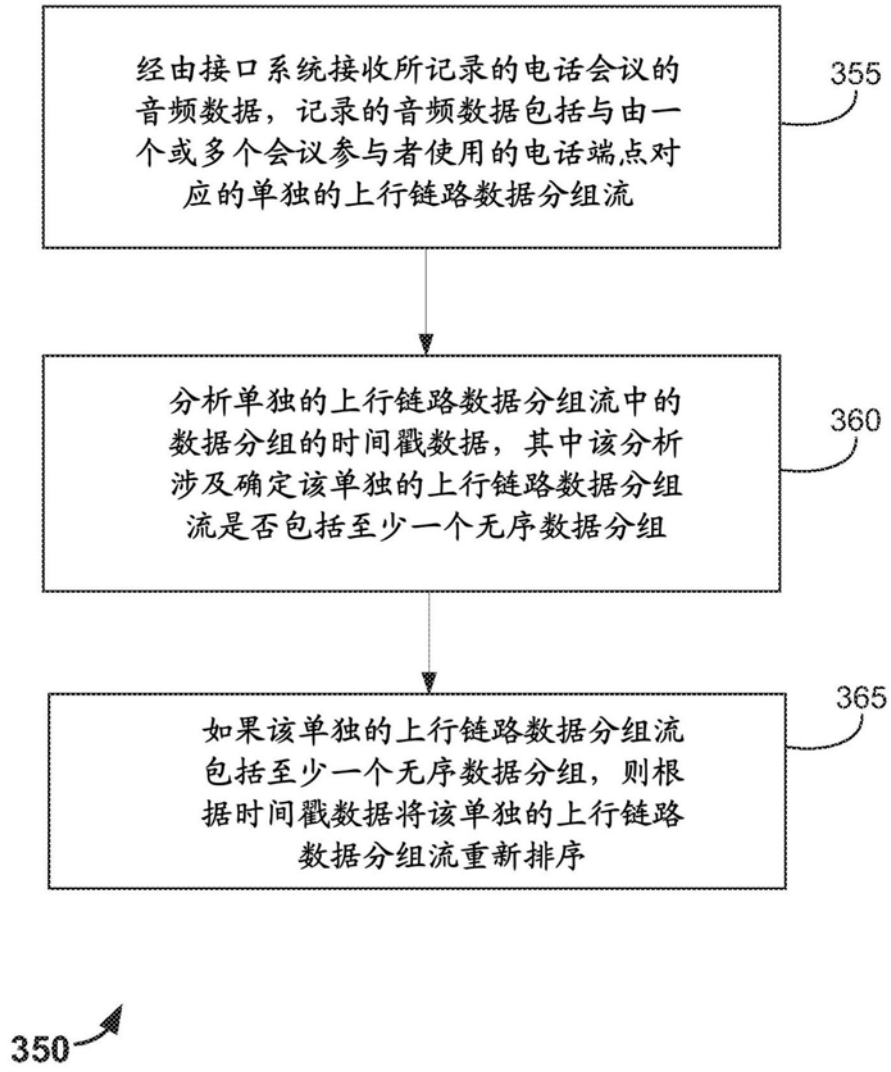

经由接口系统接收所记录的电话会议的音频数据，记录的音频数据包括与由一个或多个会议参与者使用的电话端点对应的单独的上行链路数据分组流
355
分析单独的上行链路数据分组流中的数据分组的时间戳数据，其中该分析涉及确定该单独的上行链路数据分组流是否包括至少一个无序数据分组
360
如果该单独的上行链路数据分组流包括至少一个无序数据分组，则根据时间戳数据将该单独的上行链路数据分组流重新排序
365
350

图3B

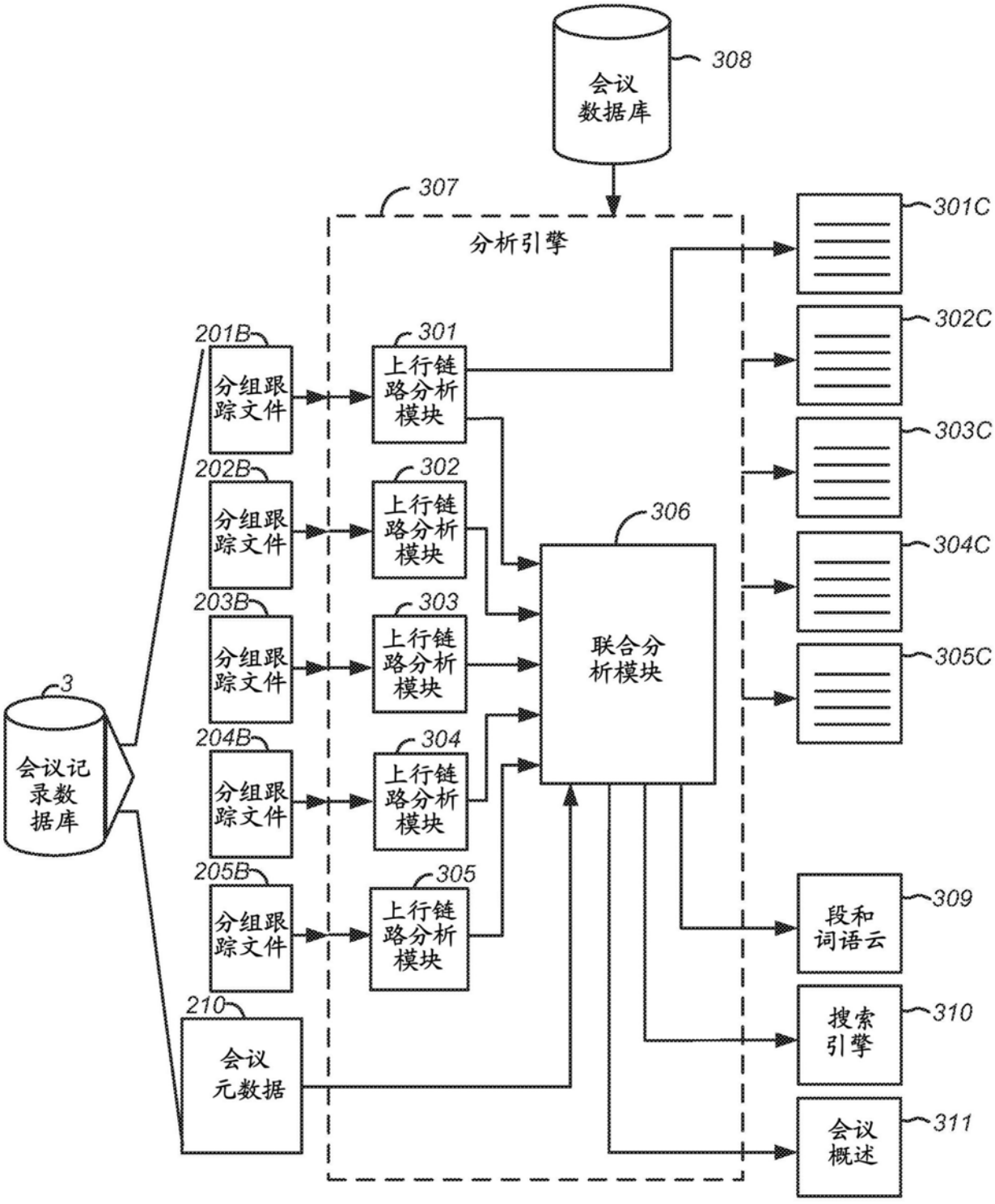

会议数据库
308
307
分析引擎
301C
302C
303C
304C
305C
201B
分组跟踪文件
301
上行链路分析模块
202B
分组跟踪文件
302
上行链路分析模块
203B
分组跟踪文件
303
上行链路分析模块
204B
分组跟踪文件
304
上行链路分析模块
205B
分组跟踪文件
305
上行链路分析模块
306
联合分析模块
3
会议记录数据库
210
会议元数据
309
段和词语云
310
搜索引擎
311
会议概述

图3C

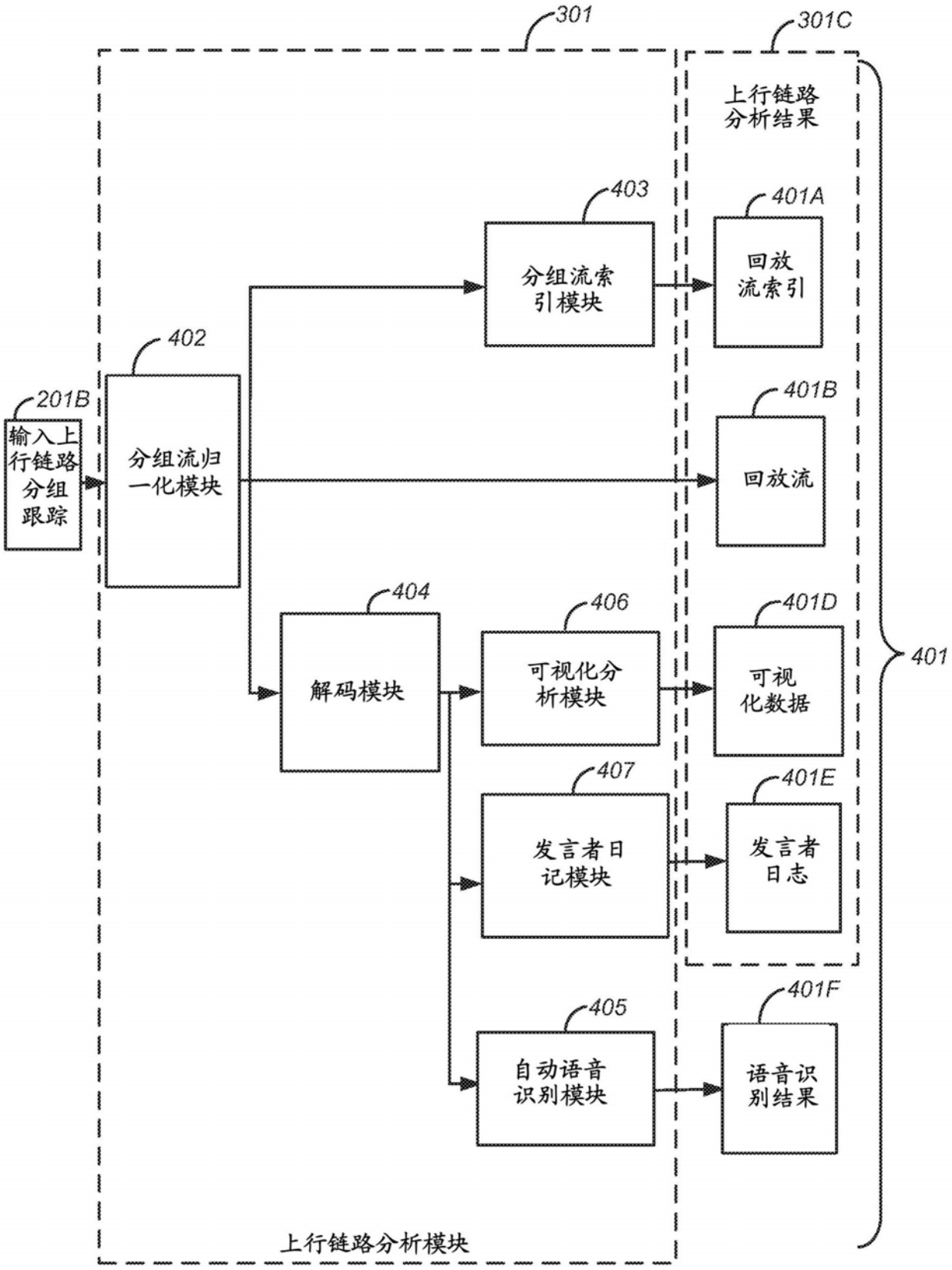

301
301C
上行链路
分析结果
403
401A
分组流索
引模块
回放
流索引
402
201B
输入上
行链路
分组
跟踪
分组流归
一化模块
401B
回放流
404
406
401D
解码模块
可视化分
析模块
可视
化数据
401
407
401E
发言者日
记模块
发言者
日志
405
401F
自动语音
识别模块
语音识
别结果
上行链路分析模块

图4

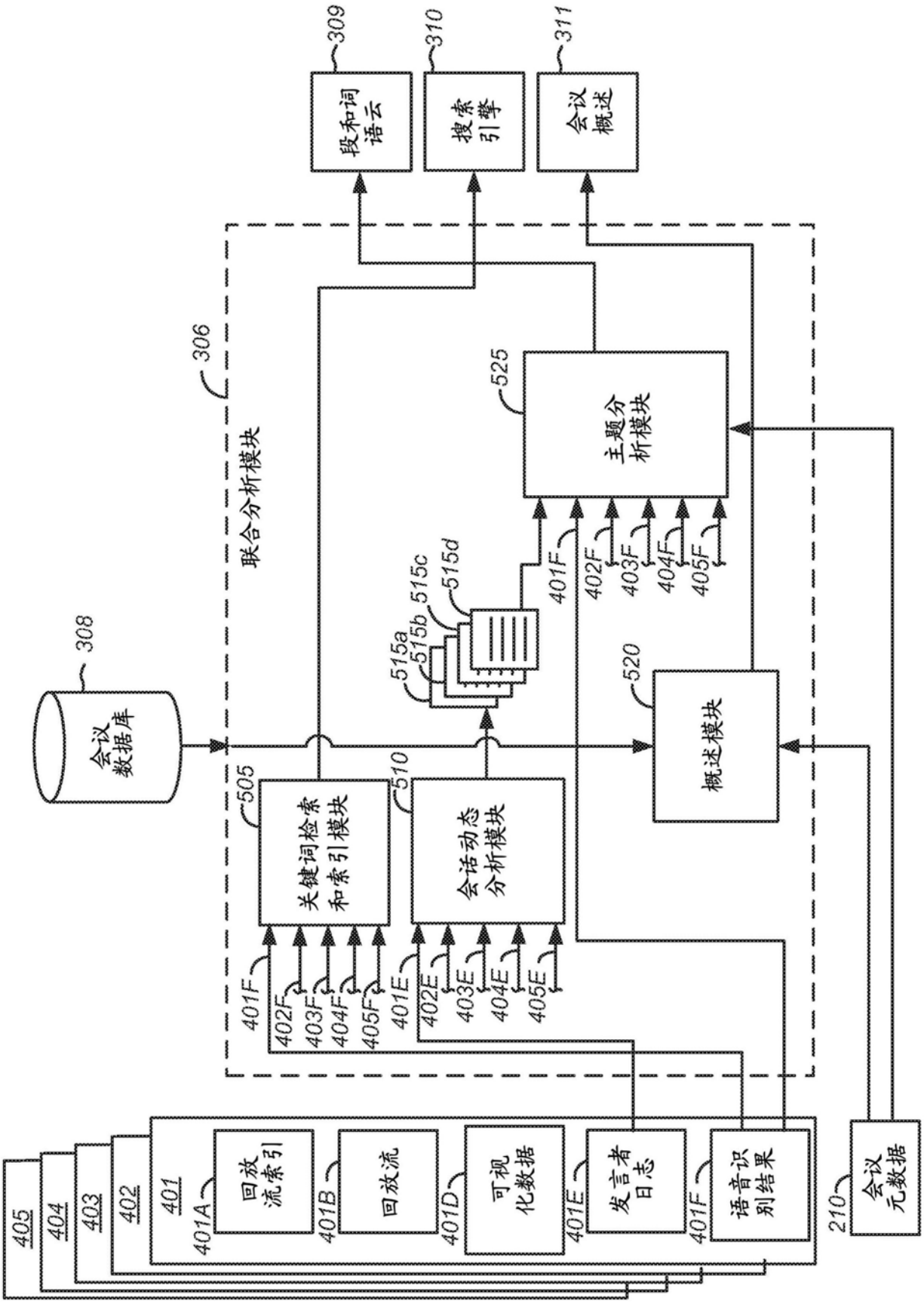

309
段和词语云
310
搜索引擎
311
会议概述
306
联合分析模块
525
主题分析模块
401F
402F
403F
404F
405F
515a
515b
515c
515d
308
会议数据库
520
概述模块
505
关键词检索和索引模块
510
会话动态分析模块
401E
402E
403E
404E
405E
405
404
403
402
401
401A
回放流索引
401B
回放流
401D
可视化数据
401E
发言者日志
401F
语音识别结果
210
会议元数据

图5

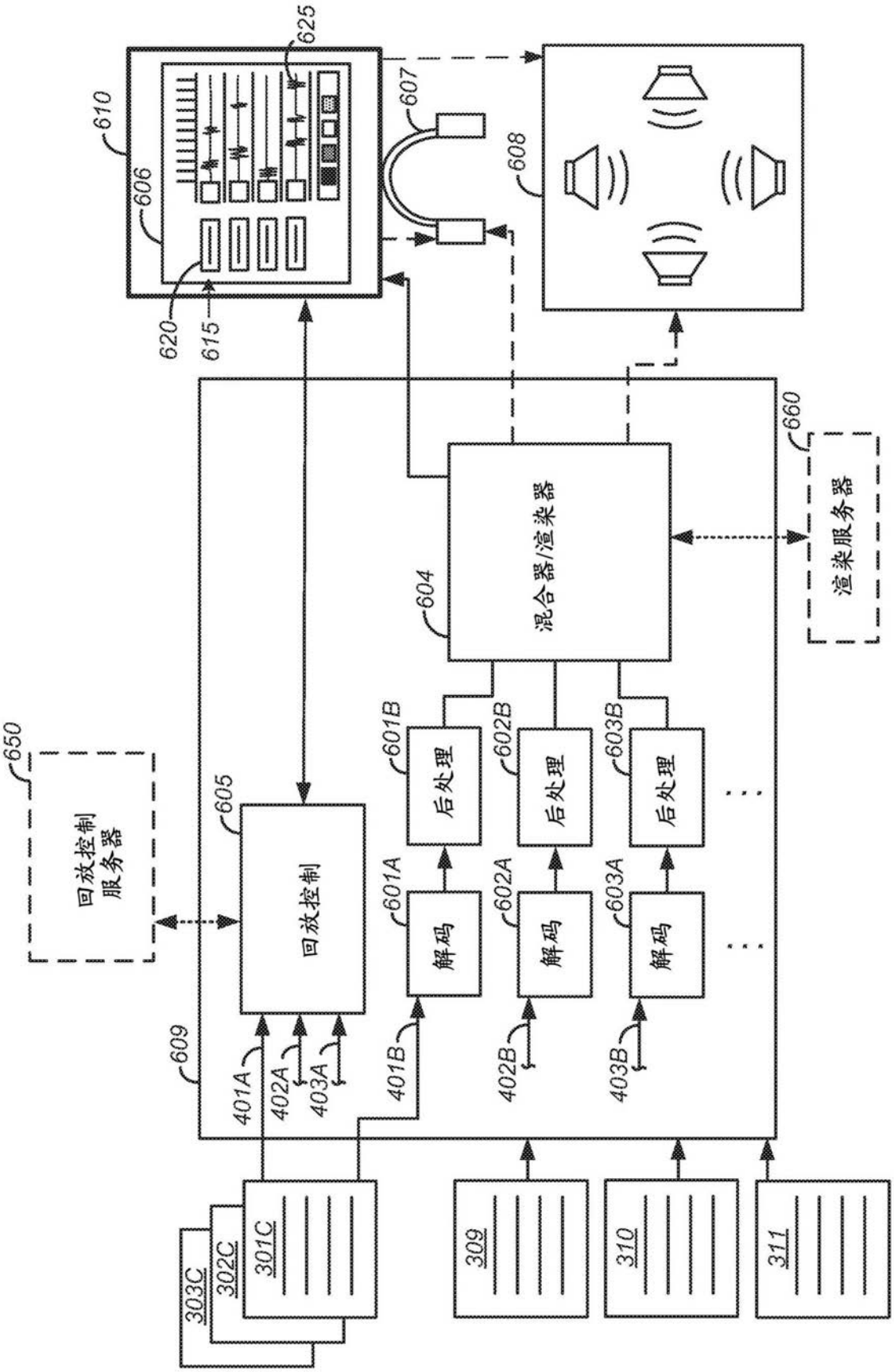

625
610
606
607
608
620
615
660
渲染服务器
混合器/渲染器
604
650
回放控制
服务器
605
回放控制
601B
602B
603B
后处理
601A
602A
603A
解码
609
401A
402A
403A
401B
402B
403B
303C
302C
301C
309
310
311

图6

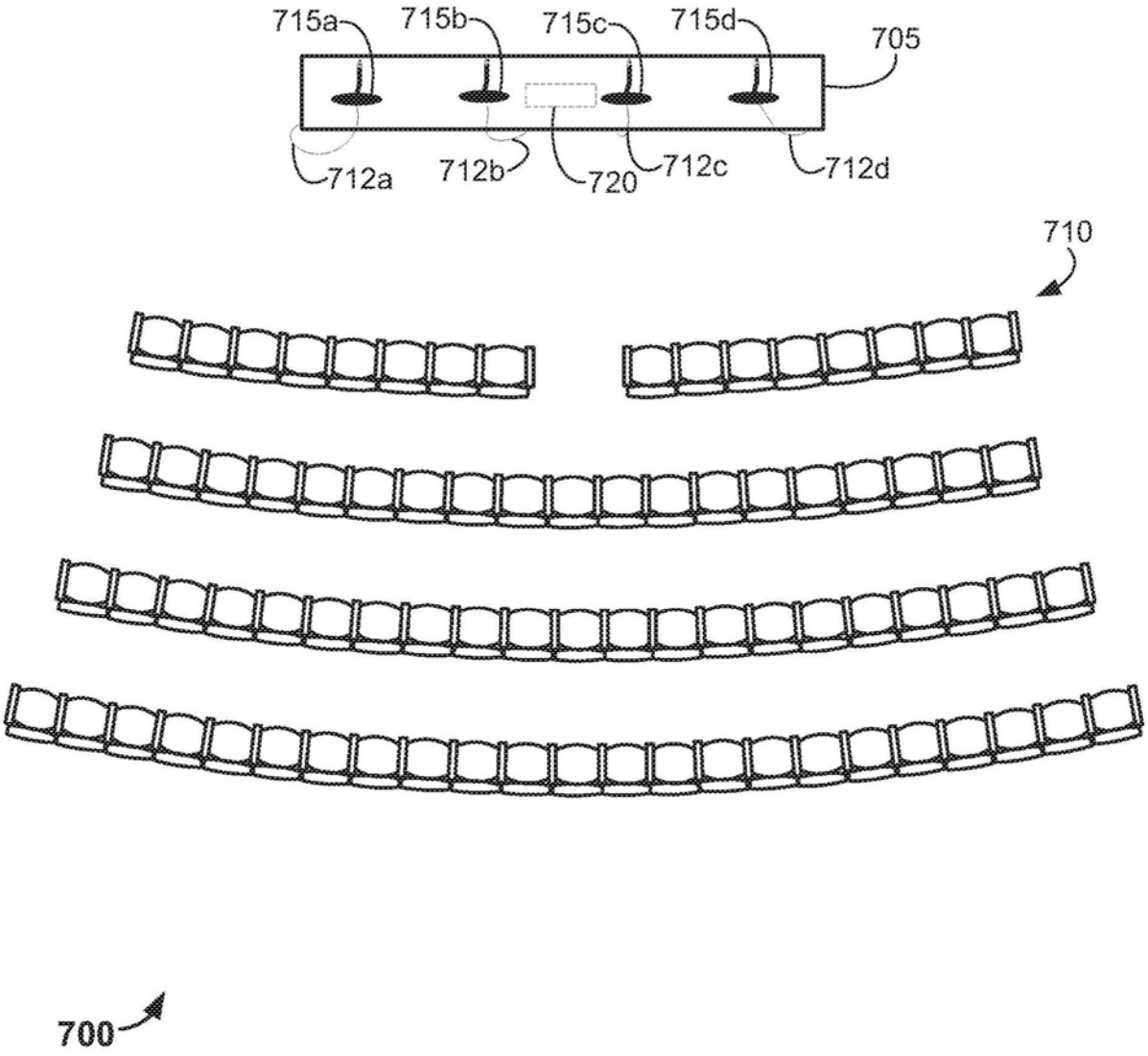


图7

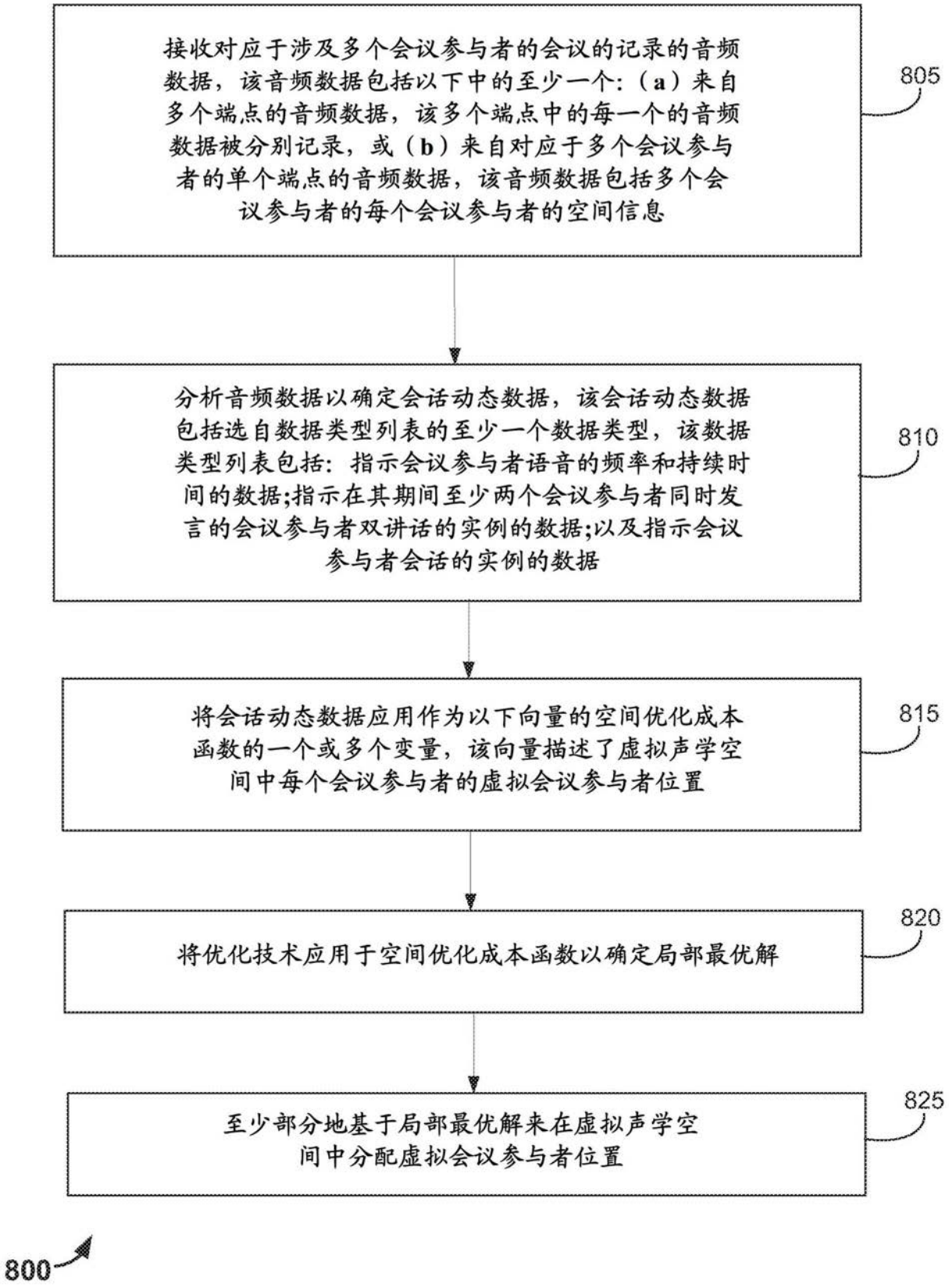

接收对应于涉及多个会议参与者的会议的记录的音频数据，该音频数据包括以下中的至少一个:（a）来自多个端点的音频数据，该多个端点中的每一个的音频数据被分别记录，或（b）来自对应于多个会议参与者的单个端点的音频数据，该音频数据包括多个会议参与者的每个会议参与者的空间信息
805
分析音频数据以确定会话动态数据，该会话动态数据包括选自数据类型列表的至少一个数据类型，该数据类型列表包括：指示会议参与者语音的频率和持续时间的数据;指示在其期间至少两个会议参与者同时发言的会议参与者双讲话的实例的数据;以及指示会议参与者会话的实例的数据
810
将会话动态数据应用作为以下向量的空间优化成本函数的一个或多个变量，该向量描述了虚拟声学空间中每个会议参与者的虚拟会议参与者位置
815
将优化技术应用于空间优化成本函数以确定局部最优解
820
至少部分地基于局部最优解来在虚拟声学空间中分配虚拟会议参与者位置
825
800

图8

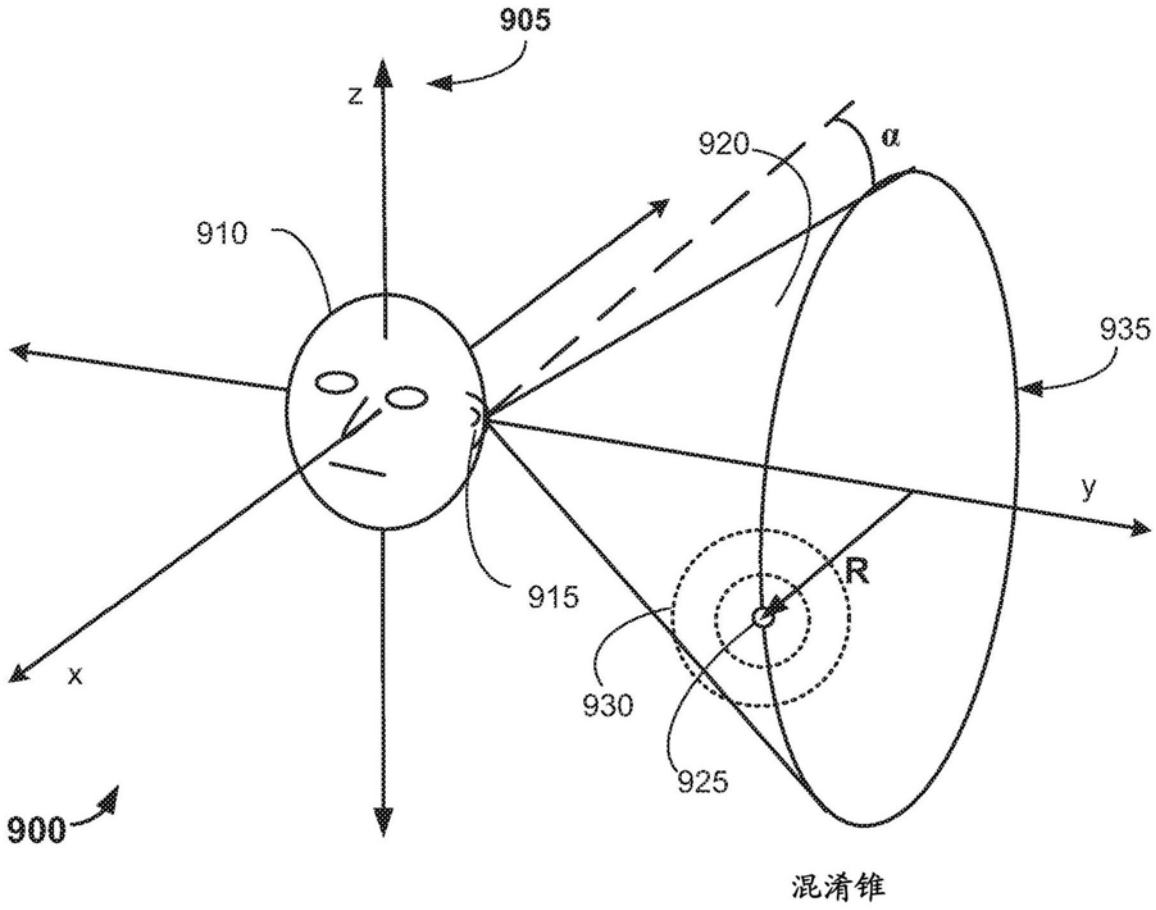


图9

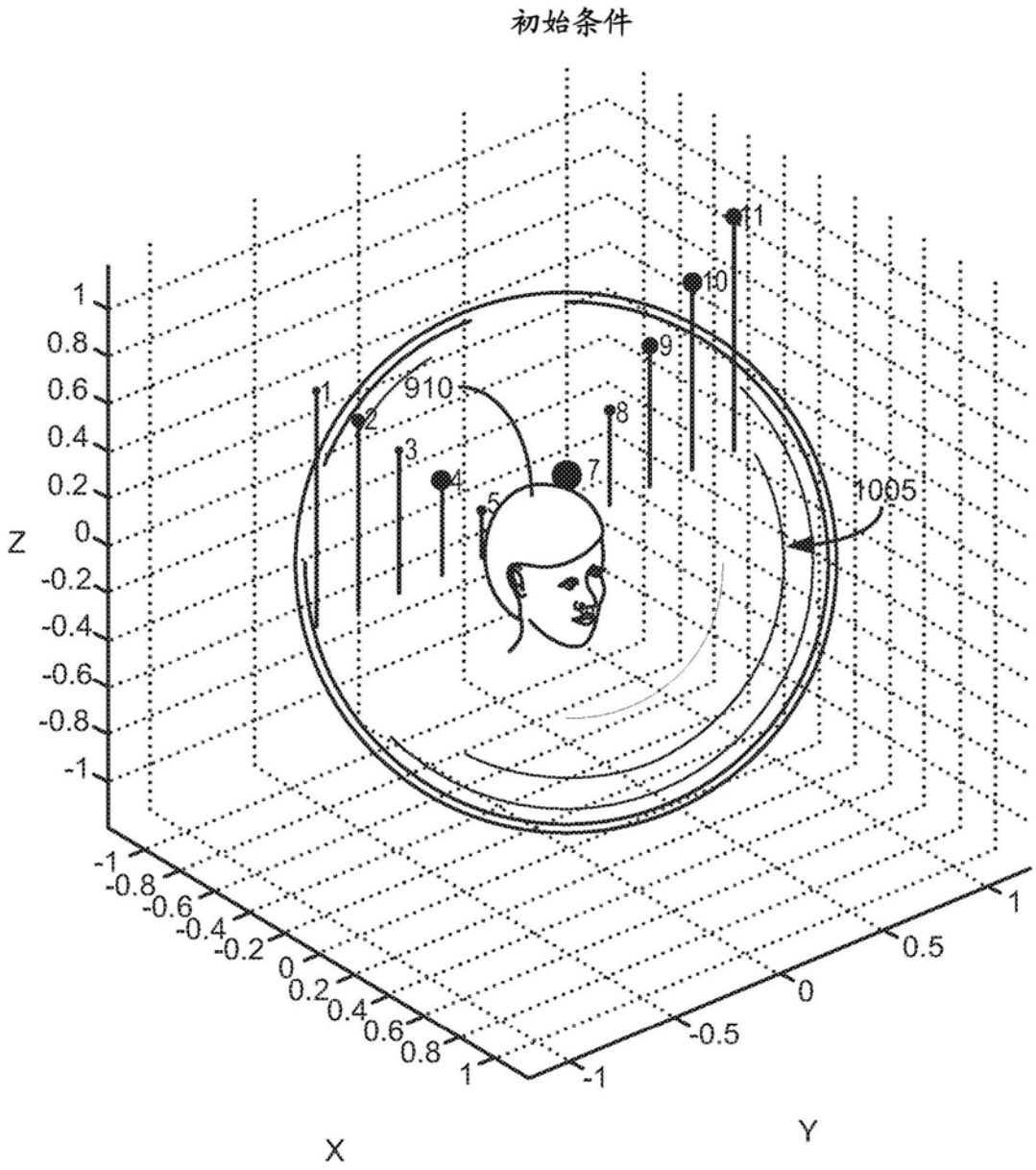

初始条件
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Z
1
2
3
4
5
7
8
9
10
11
910
1005
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
X
1
0.5
0
-0.5
-1
Y

图10

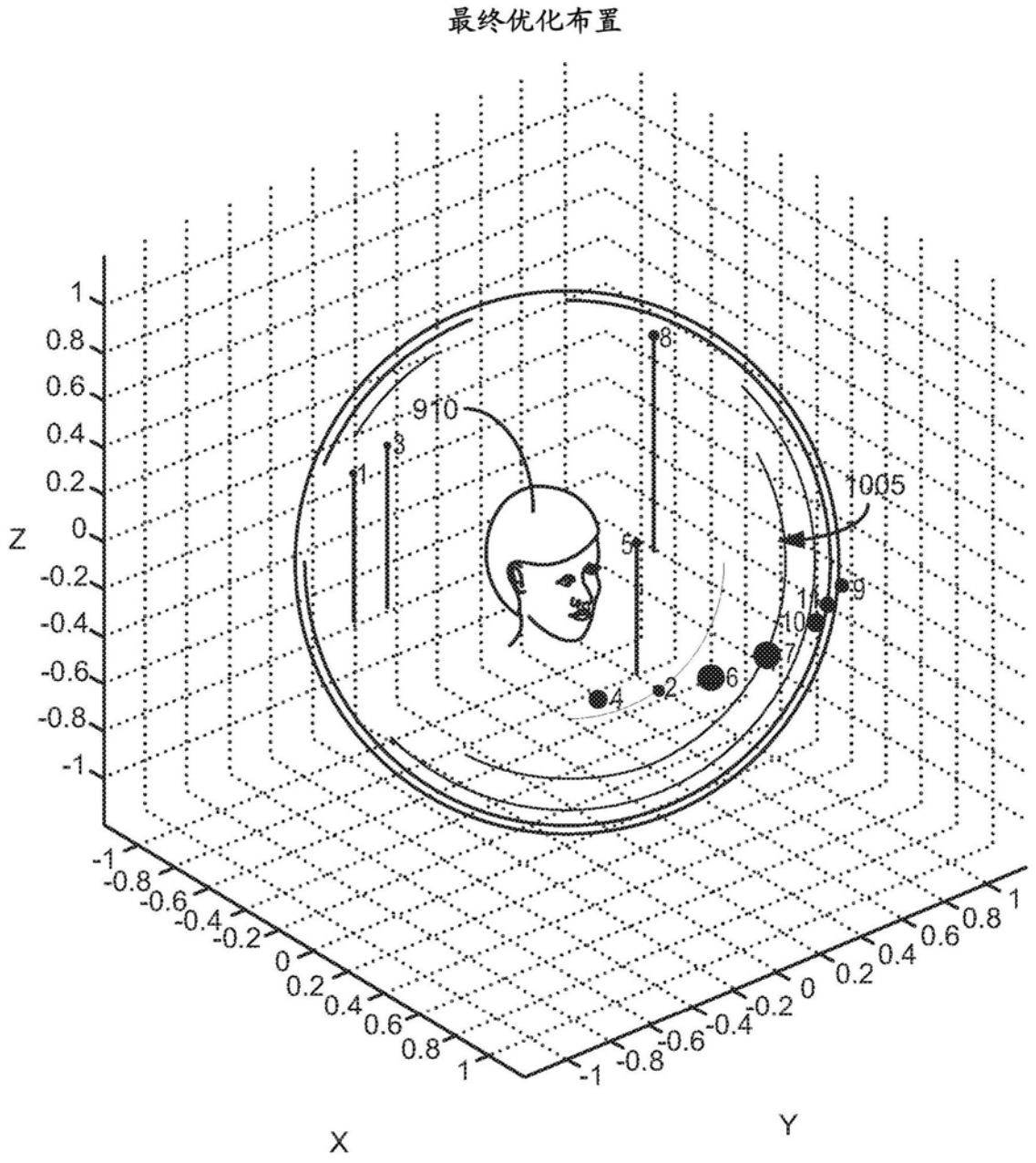


图11

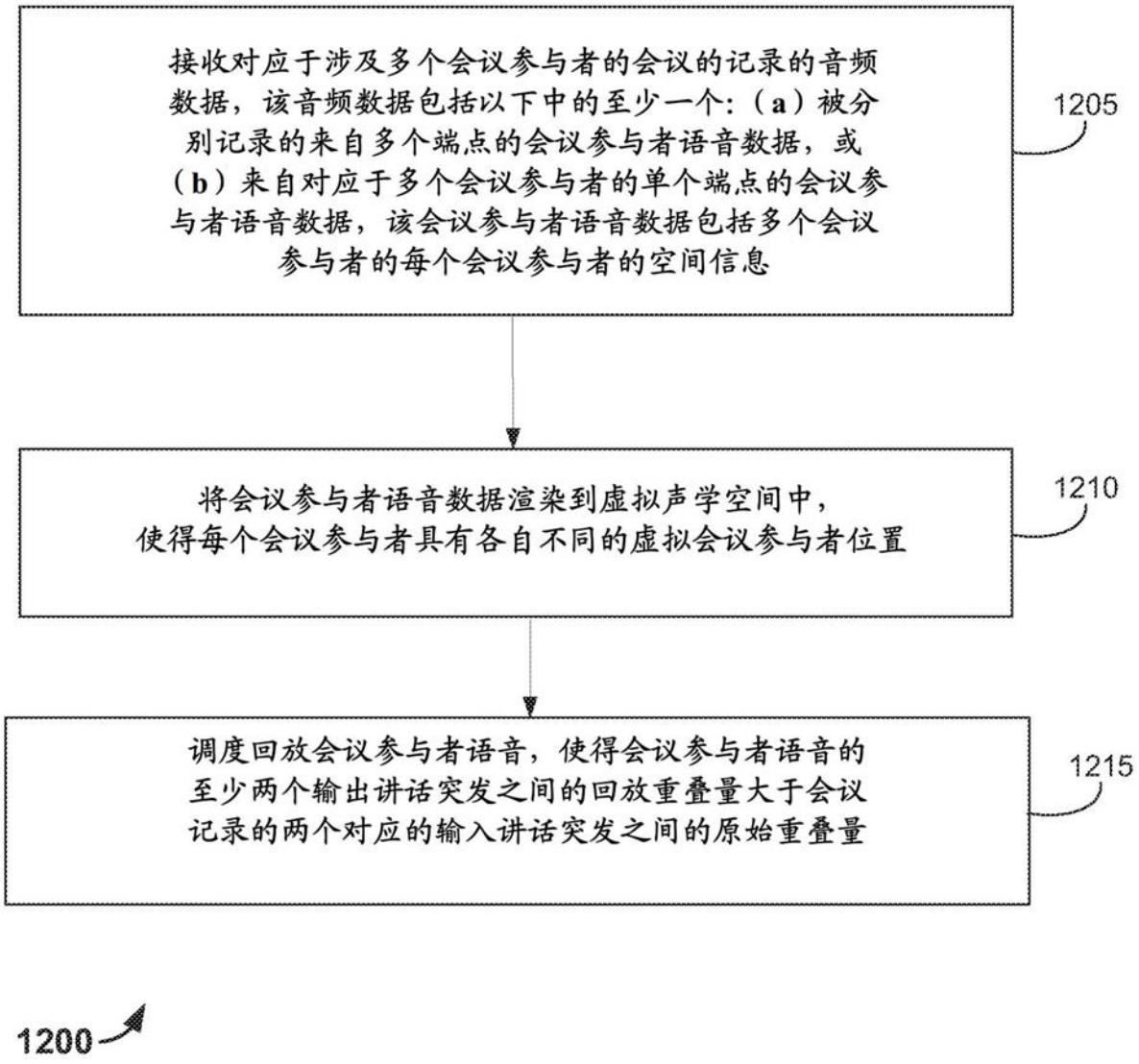

接收对应于涉及多个会议参与者的会议的记录的音频数据，该音频数据包括以下中的至少一个:（a）被分别记录的来自多个端点的会议参与者语音数据，或（b）来自对应于多个会议参与者的单个端点的会议参与者语音数据，该会议参与者语音数据包括多个会议参与者的每个会议参与者的空间信息
1205
将会议参与者语音数据渲染到虚拟声学空间中，使得每个会议参与者具有各自不同的虚拟会议参与者位置
1210
调度回放会议参与者语音，使得会议参与者语音的至少两个输出讲话突发之间的回放重叠量大于会议记录的两个对应的输入讲话突发之间的原始重叠量
1215
1200

图12

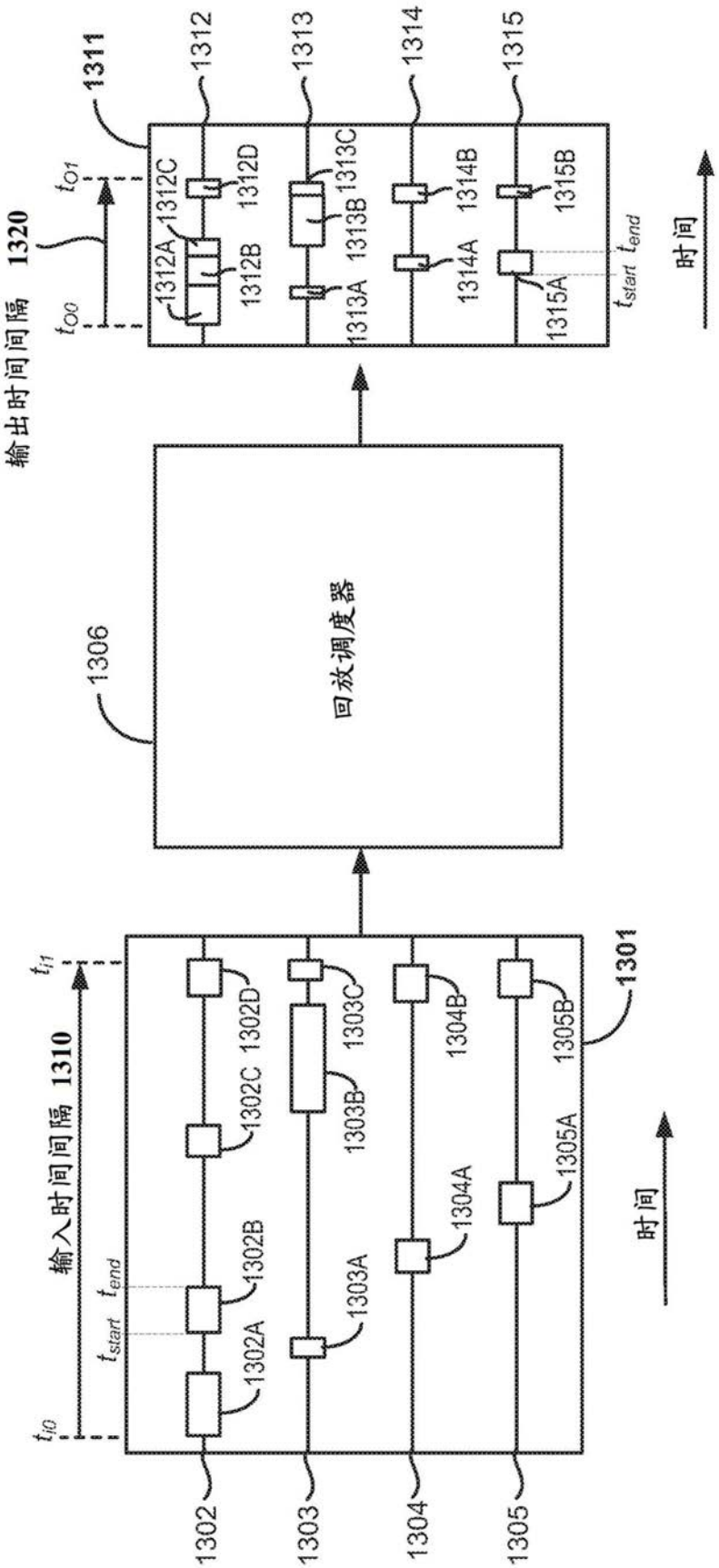

输出时间间隔 1320
t_{O0}
t_{O1}
1311
1312
1313
1314
1315
1312A
1312B
1312C
1312D
1313A
1313B
1313C
1314A
1314B
1315A
1315B
t_{start} t_{end}
时间
1306
回放调度器
输入时间间隔 1310
t_{I0}
t_{I1}
t_{start} t_{end}
1301
1302
1303
1304
1305
1302A
1302B
1302C
1302D
1303A
1303B
1303C
1304A
1304B
1305A
1305B
时间

图13

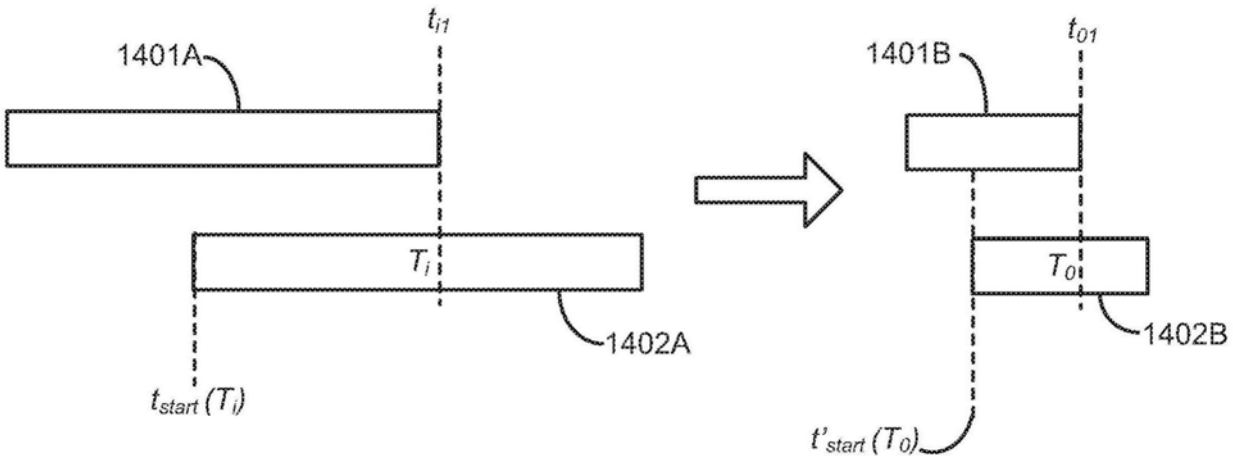

t_{i1}
1401A
T_i
1402A
$t_{start}(T_i)$
t_{01}
1401B
T_0
1402B
$t'_{start}(T_0)$

图14

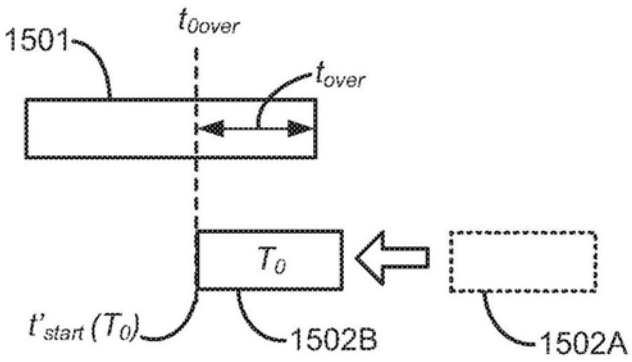

t_{0over}
1501
t_{over}
T_0
$t'_{start}(T_0)$
1502B
1502A

图15

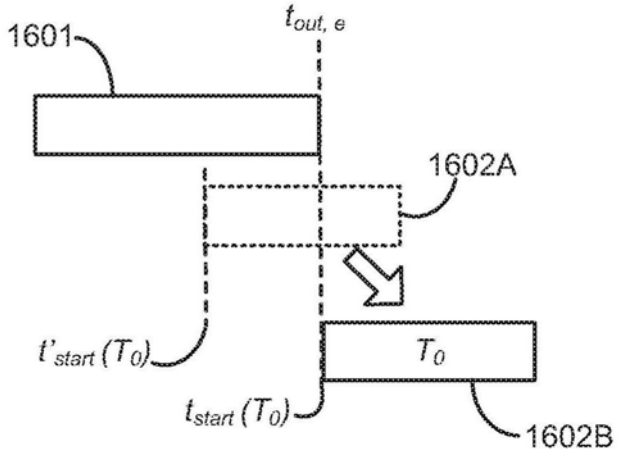

$t_{out,\ e}$
1601
1602A
$t'_{start}(T_0)$
T_0
$t_{start}(T_0)$
1602B

图16

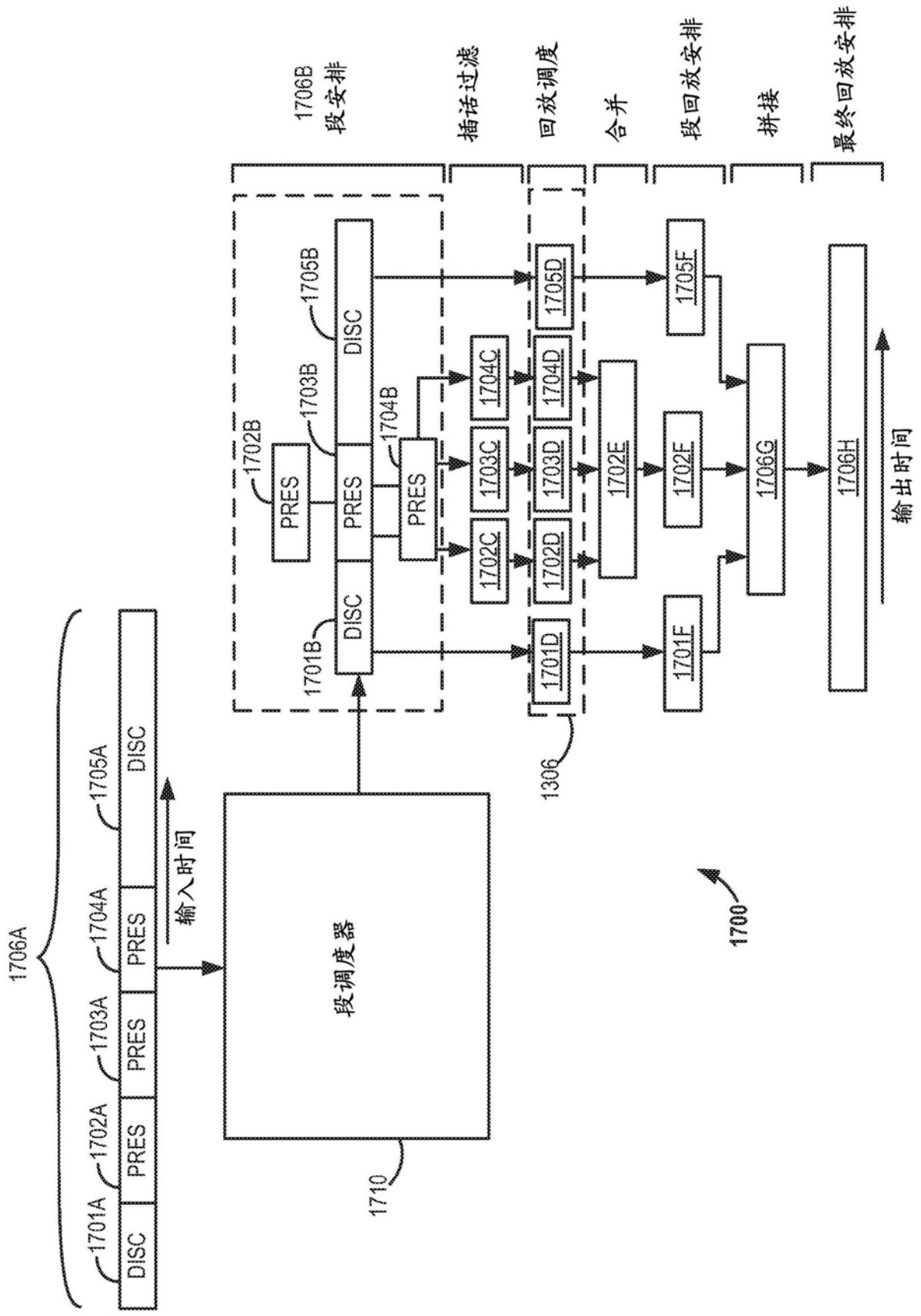

1706A
1701A
DISC
1702A
PRES
1703A
PRES
1704A
PRES
1705A
DISC
输入时间
段调度器
1710
1706B
段安排
插话过滤
回放调度
合并
段回放安排
拼接
最终回放安排
1701B
DISC
1702B
PRES
1703B
PRES
1704B
PRES
1705B
DISC
1701D
1702D
1703D
1704D
1705D
1702C
1703C
1704C
1306
1701F
1702E
1702F
1705F
1706G
1706H
1700
输出时间

图17

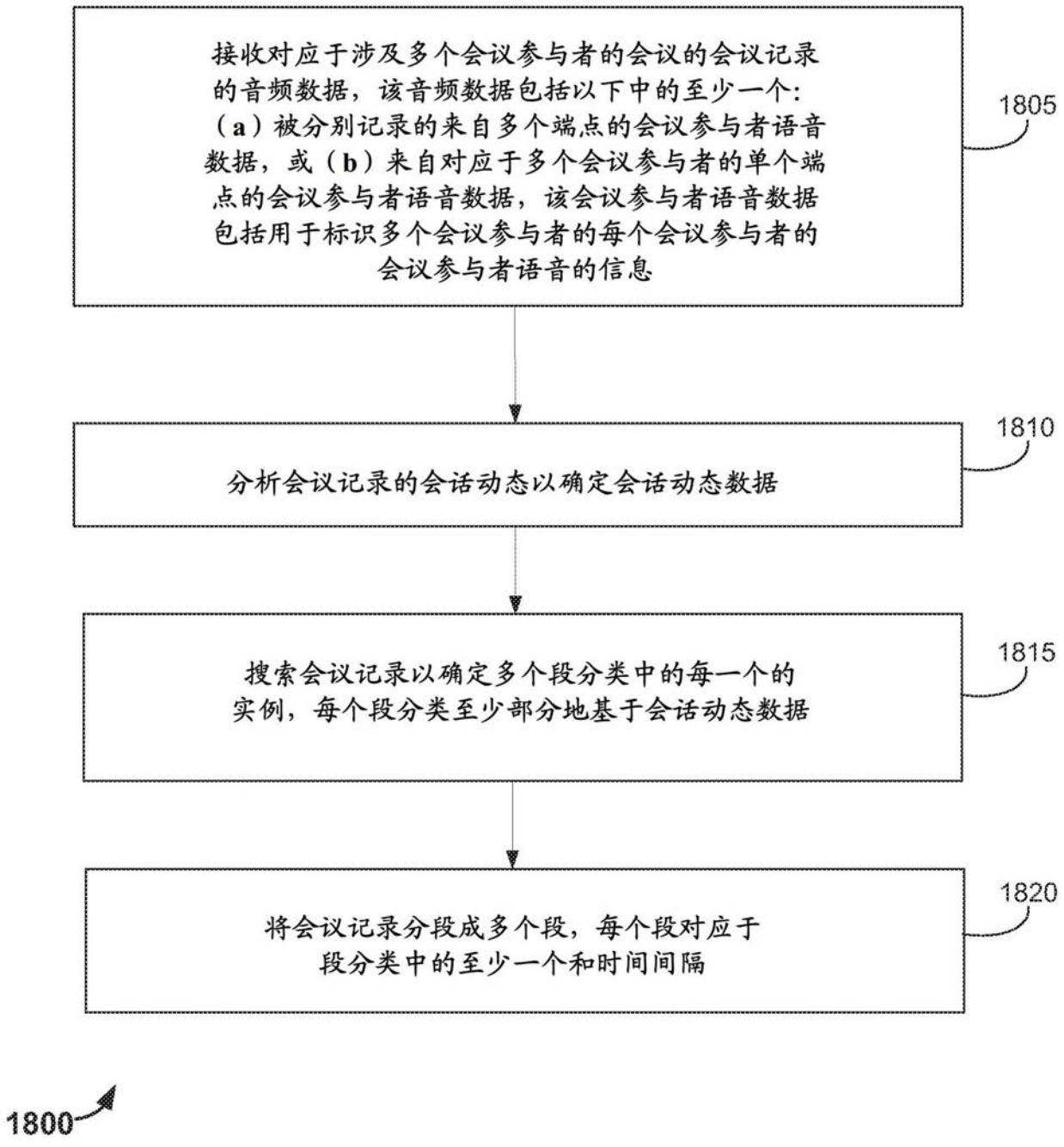

接收对应于涉及多个会议参与者的会议的会议记录的音频数据，该音频数据包括以下中的至少一个：（a）被分别记录的来自多个端点的会议参与者语音数据，或（b）来自对应于多个会议参与者的单个端点的会议参与者语音数据，该会议参与者语音数据包括用于标识多个会议参与者的每个会议参与者的会议参与者语音的信息
1805
分析会议记录的会话动态以确定会话动态数据
1810
搜索会议记录以确定多个段分类中的每一个的实例，每个段分类至少部分地基于会话动态数据
1815
将会议记录分段成多个段，每个段对应于段分类中的至少一个和时间间隔
1820
1800

图18A

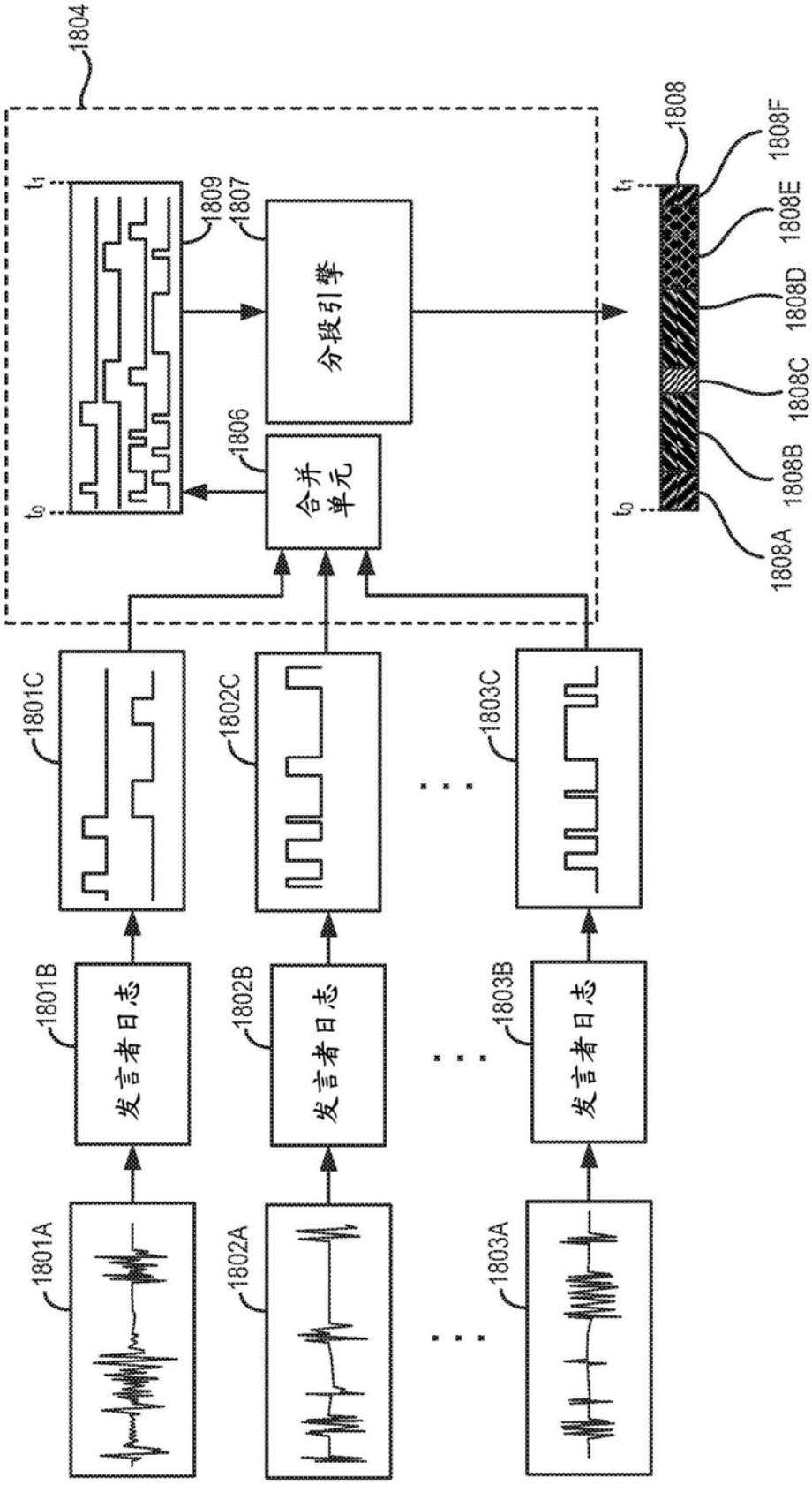

1804
1809
1807
分段引擎
1806
合并单元
1808
1808F
1808E
1808D
1808C
1808B
1808A
1801A
1801B
1801C
1802A
1802B
1802C
1803A
1803B
1803C
发言者日志

图18B

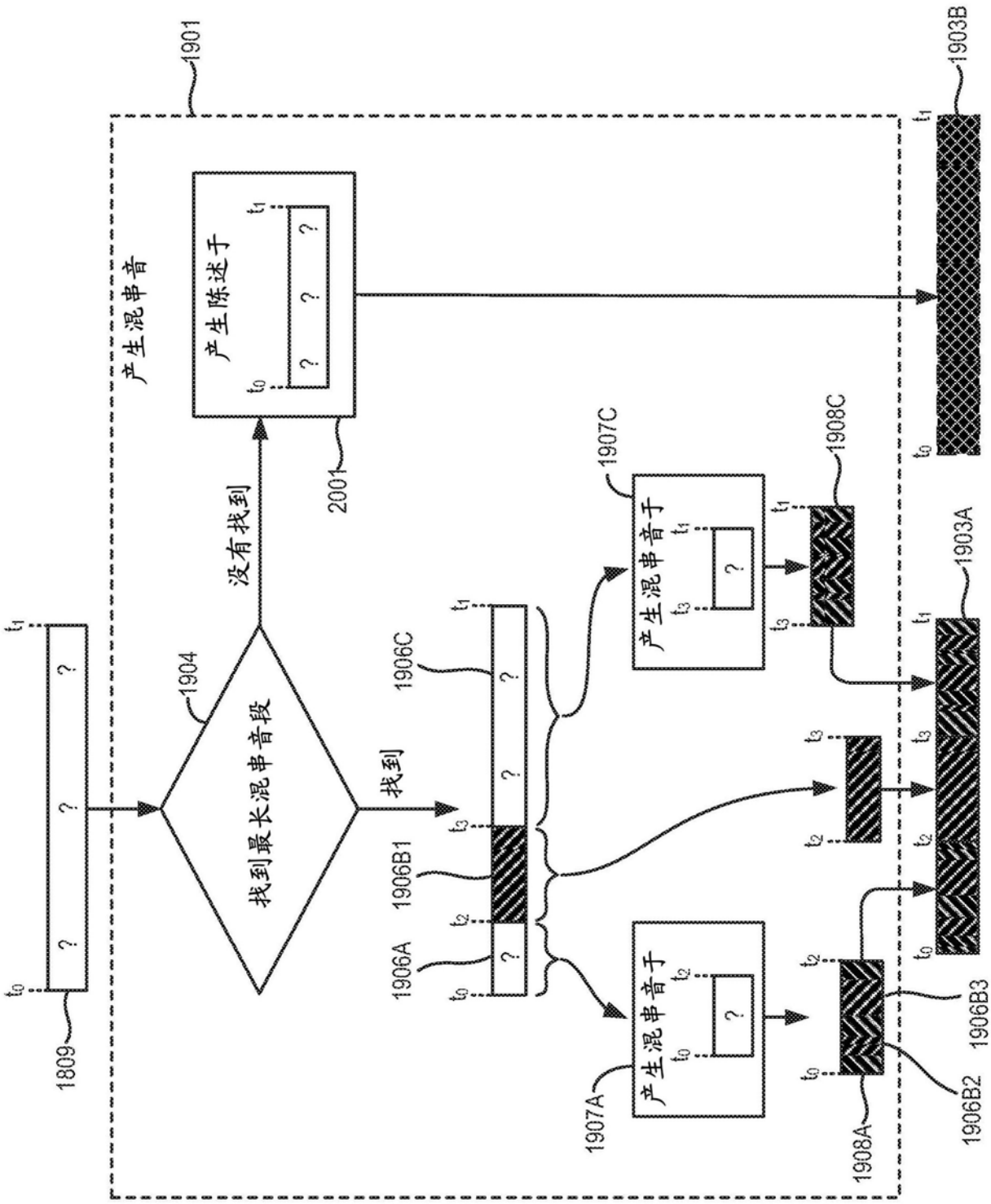

1901
产生混串音
1809
1904
找到最长混串音段
没有找到
找到
2001
产生陈述于
1906A
1906B1
1906C
1907A
1907C
1908A
1908C
1906B2
1906B3
1903A
1903B
产生混串音于

图19

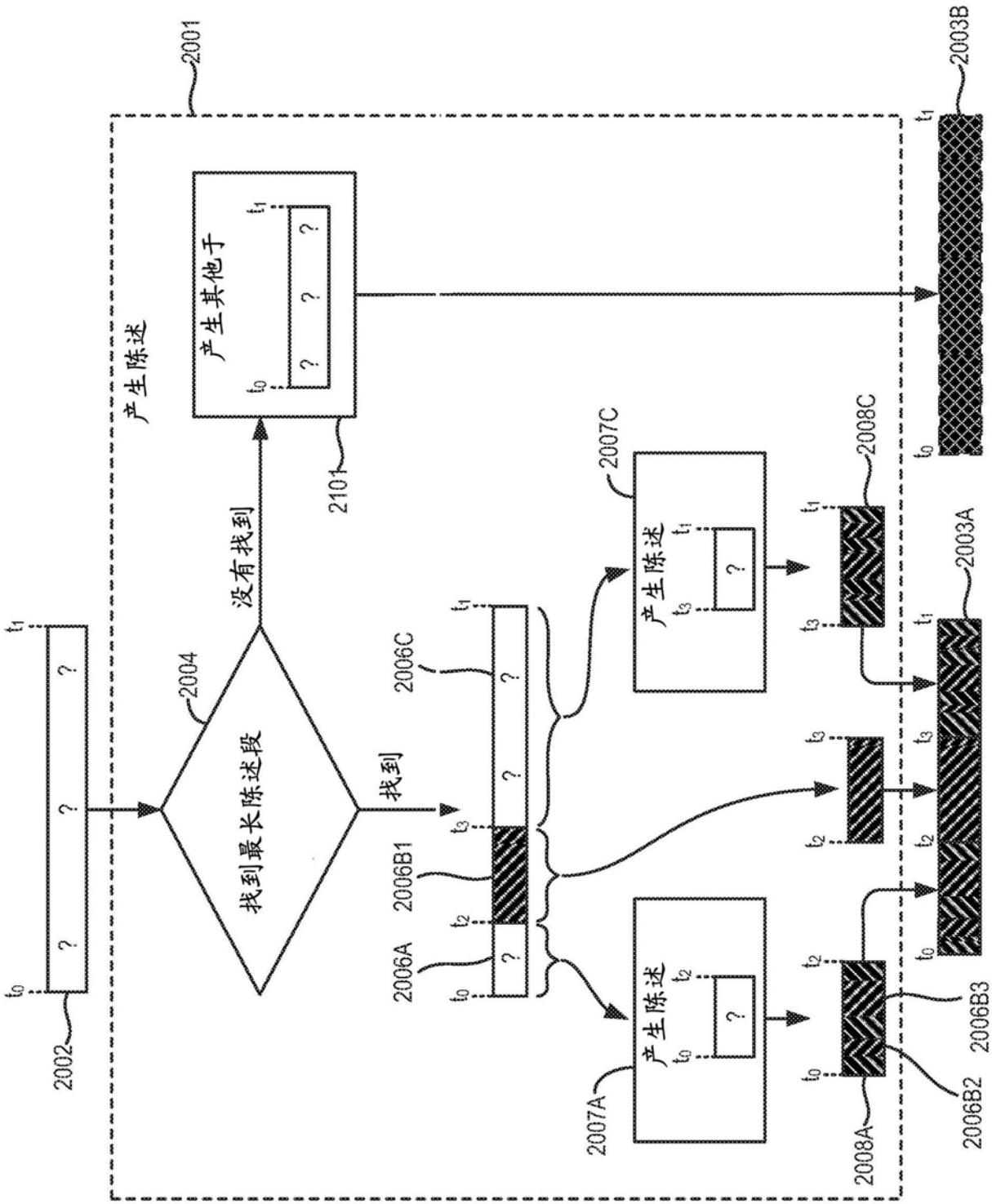

2001
产生陈述
2002
2004
找到最长陈述段
没有找到
找到
2101
产生其他子
2003B
2003A
2006A
2006B1
2006C
2007A
2007C
2008A
2008C
2006B2
2006B3
产生陈述

图20

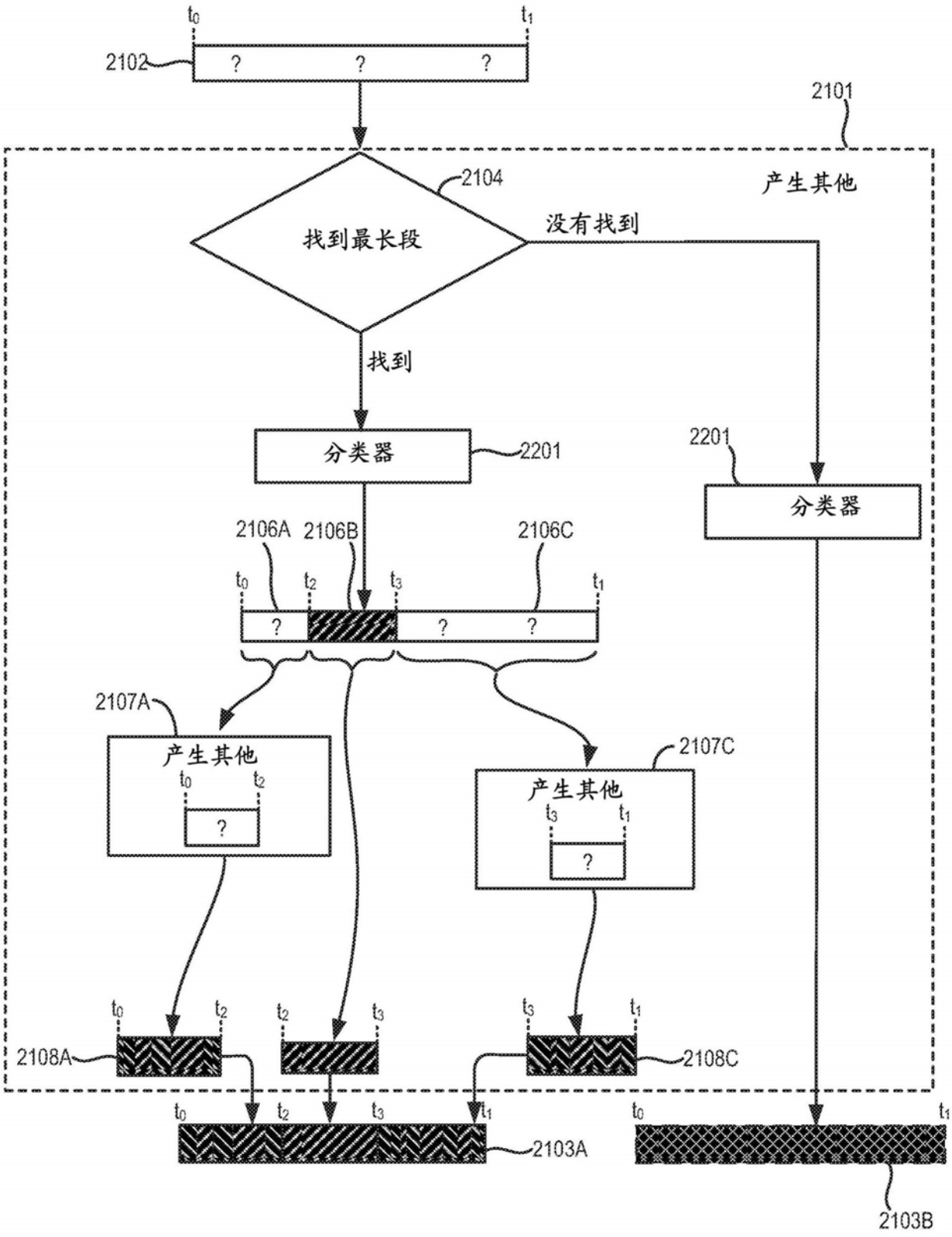

t0
t1
2102
? ? ?
2101
2104
找到最长段
产生其他
没有找到
找到
分类器
2201
2201
分类器
2106A 2106B
2106C
t0 t2 t3 t1
? ? ?
2107A
产生其他
t0 t2
?
2107C
产生其他
t3 t1
?
t0 t2
t2 t3
t3 t1
2108A
2108C
t0 t2 t3 t1
2103A
t0 t1
2103B

图21

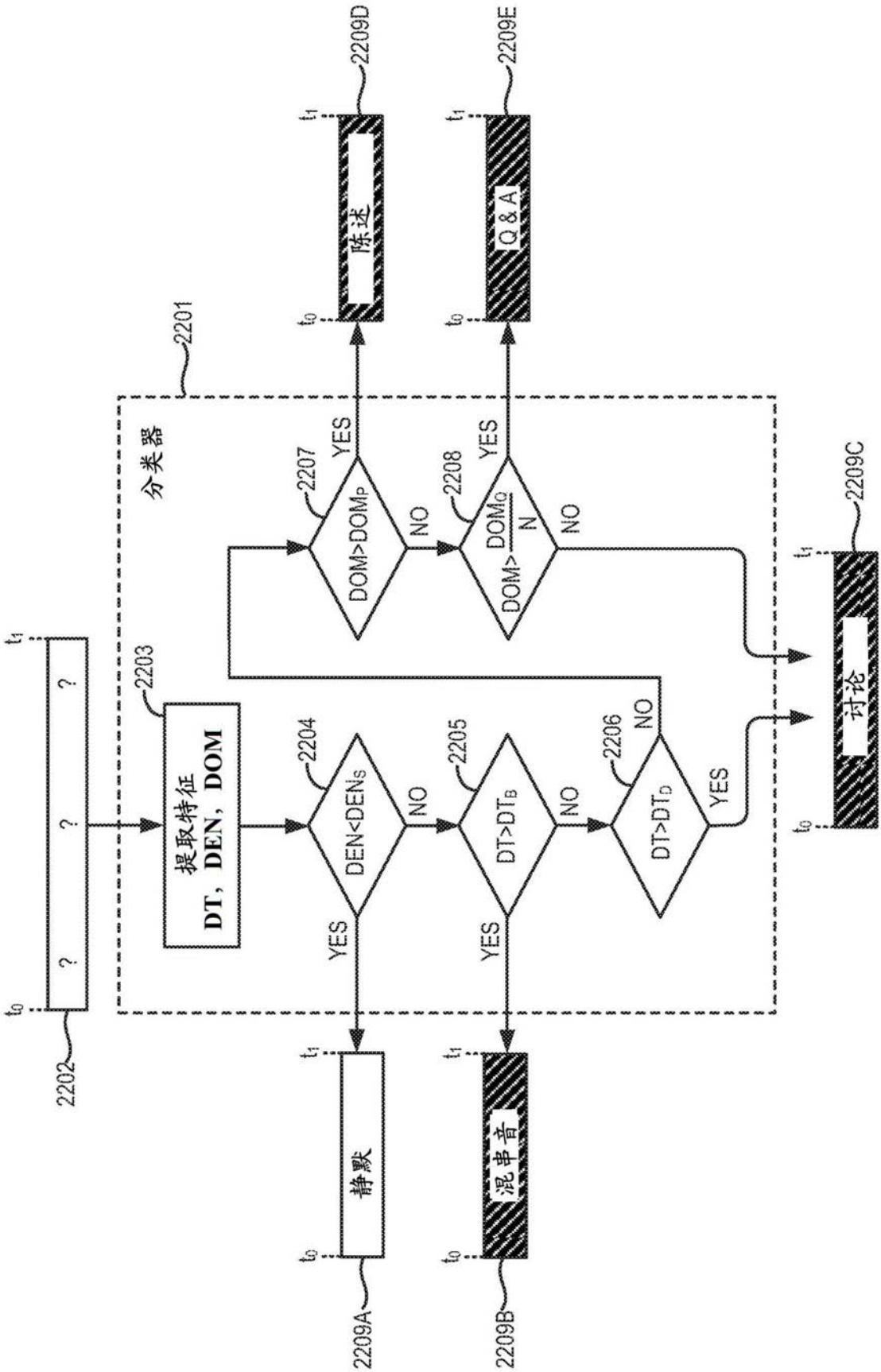

2209D
2209E
t1
t0
陈述
Q&A
2201
分类器
2207
YES
DOM>DOMP
NO
2208
YES
DOMQ
DOM>
N
NO
2209C
讨论
2203
提取特征
DT，DEN，DOM
2204
DEN<DENS
NO
2205
DT>DTB
NO
2206
NO
DT>DTD
YES
YES
YES
2202
2209A
2209B
静默
混串音

图22

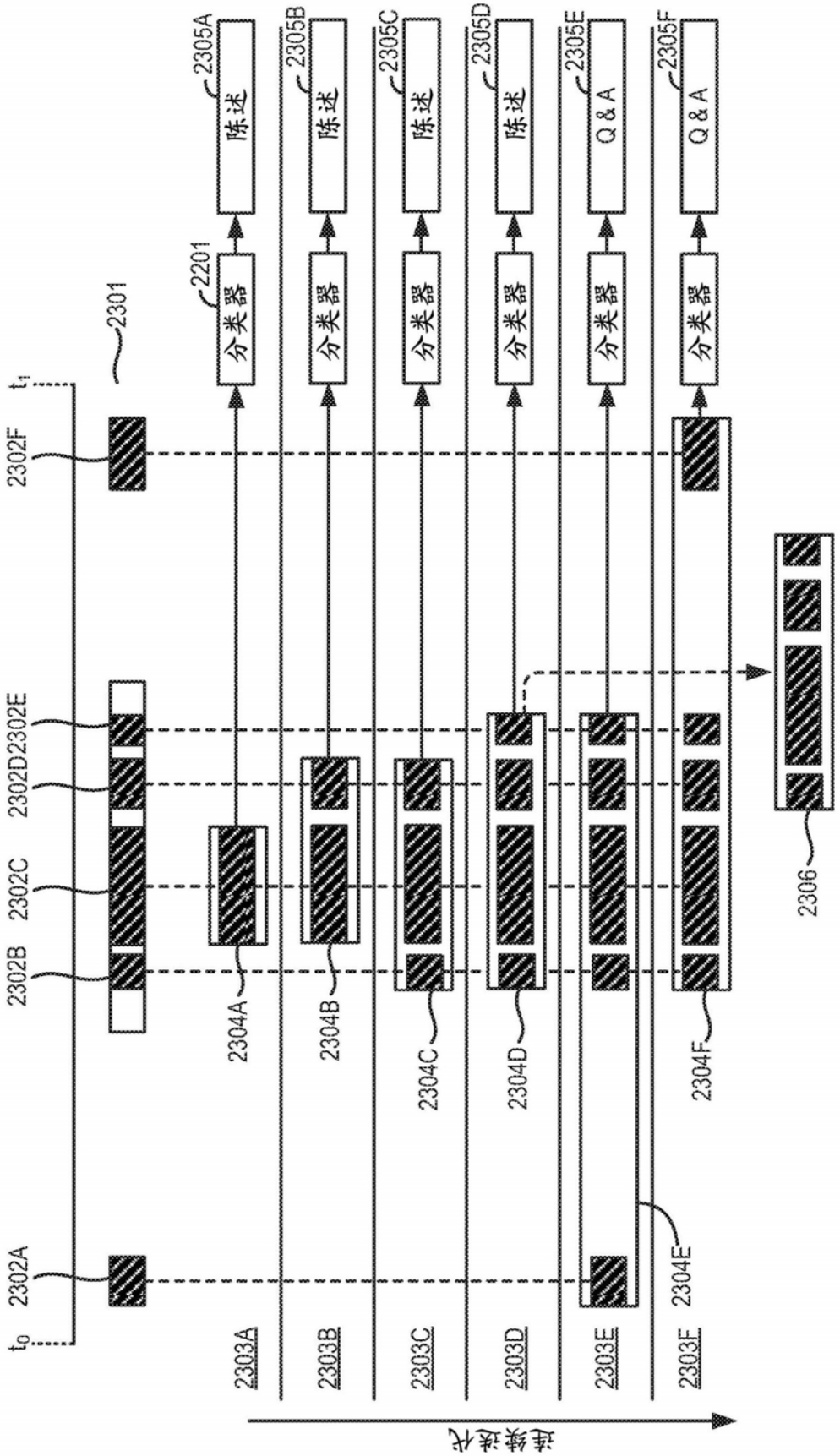

2305A
陈述
2305B
陈述
2305C
陈述
2305D
陈述
2305E
Q & A
2305F
Q & A
2201
分类器
分类器
分类器
分类器
分类器
分类器
2301
t1
2302F
2302D2302E
2302C
2302B
2302A
t0
2304A
2304B
2304C
2304D
2304E
2304F
2306
2303A
2303B
2303C
2303D
2303E
2303F
连续迭代

图23

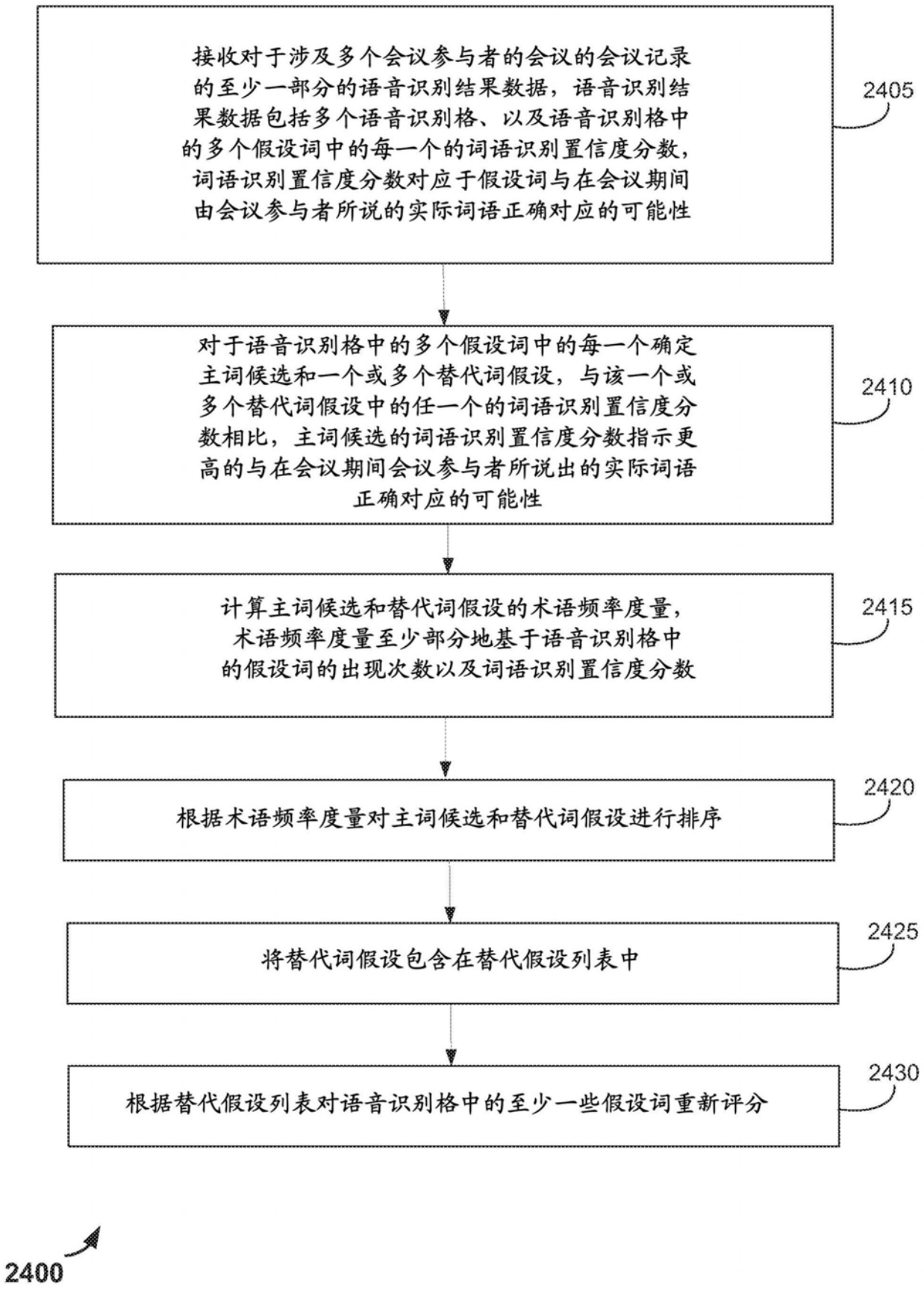

接收对于涉及多个会议参与者的会议的会议记录的至少一部分的语音识别结果数据，语音识别结果数据包括多个语音识别格、以及语音识别格中的多个假设词中的每一个的词语识别置信度分数，词语识别置信度分数对应于假设词与在会议期间由会议参与者所说的实际词语正确对应的可能性
2405
对于语音识别格中的多个假设词中的每一个确定主词候选和一个或多个替代词假设，与该一个或多个替代词假设中的任一个的词语识别置信度分数相比，主词候选的词语识别置信度分数指示更高的与在会议期间会议参与者所说出的实际词语正确对应的可能性
2410
计算主词候选和替代词假设的术语频率度量，术语频率度量至少部分地基于语音识别格中的假设词的出现次数以及词语识别置信度分数
2415
根据术语频率度量对主词候选和替代词假设进行排序
2420
将替代词假设包含在替代假设列表中
2425
根据替代假设列表对语音识别格中的至少一些假设词重新评分
2430
2400

图24

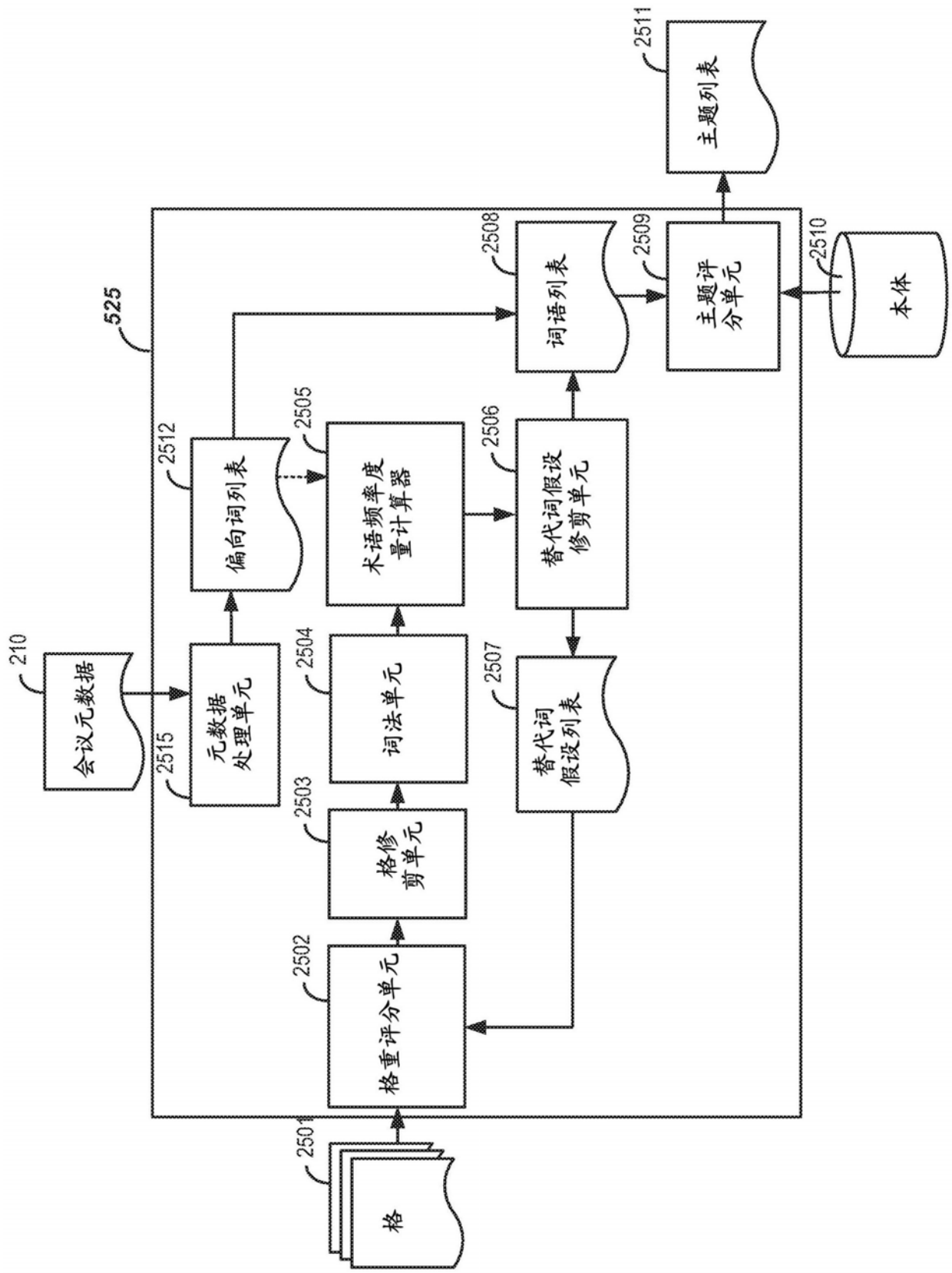

2511
主题列表
2508
词语列表
2509
主题评分单元
2510
本体
525
2512
偏向词列表
2505
术语频率度量计算器
2506
替代词假设修剪单元
210
会议元数据
2515
元数据处理单元
2504
词法单元
2507
替代词假设列表
2503
格修剪单元
2502
格重评分单元
2501
格

图25

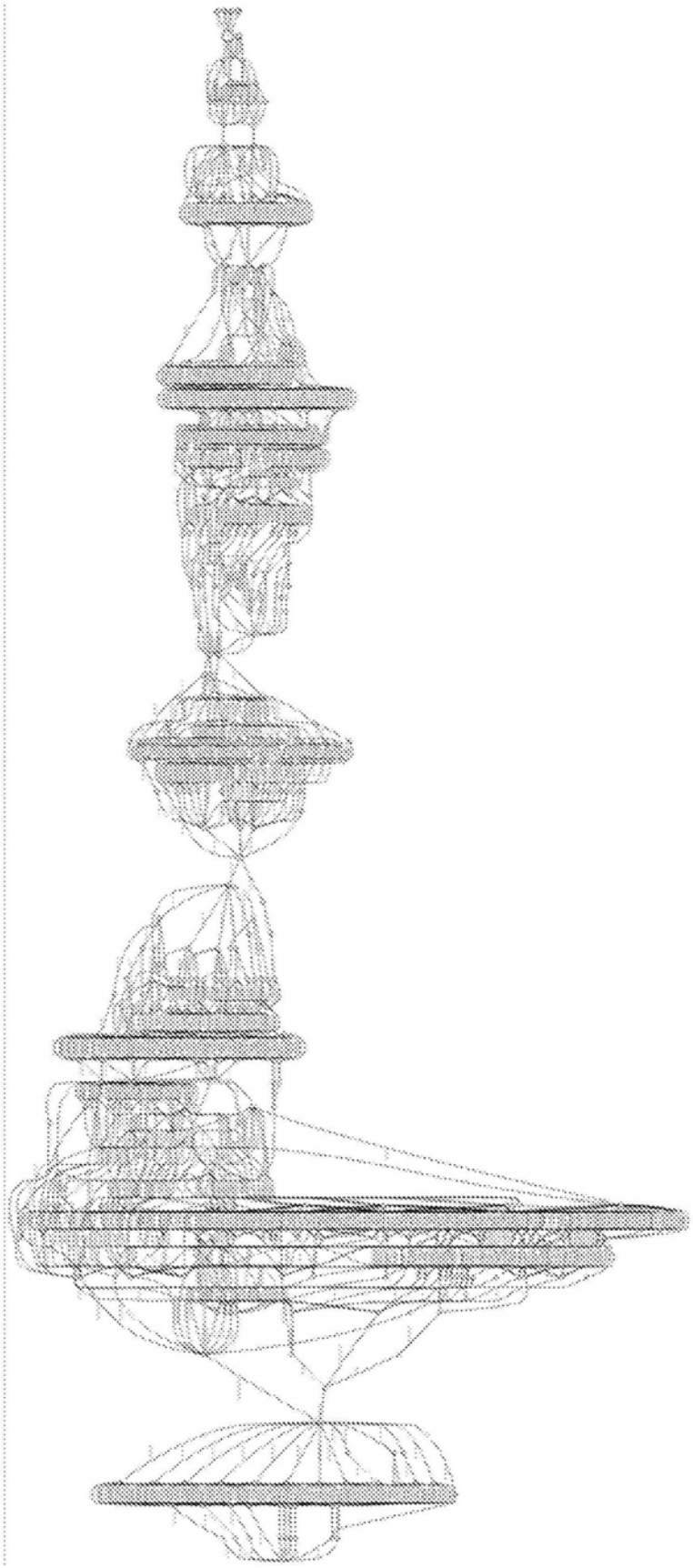


图26

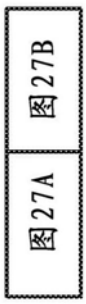


图27

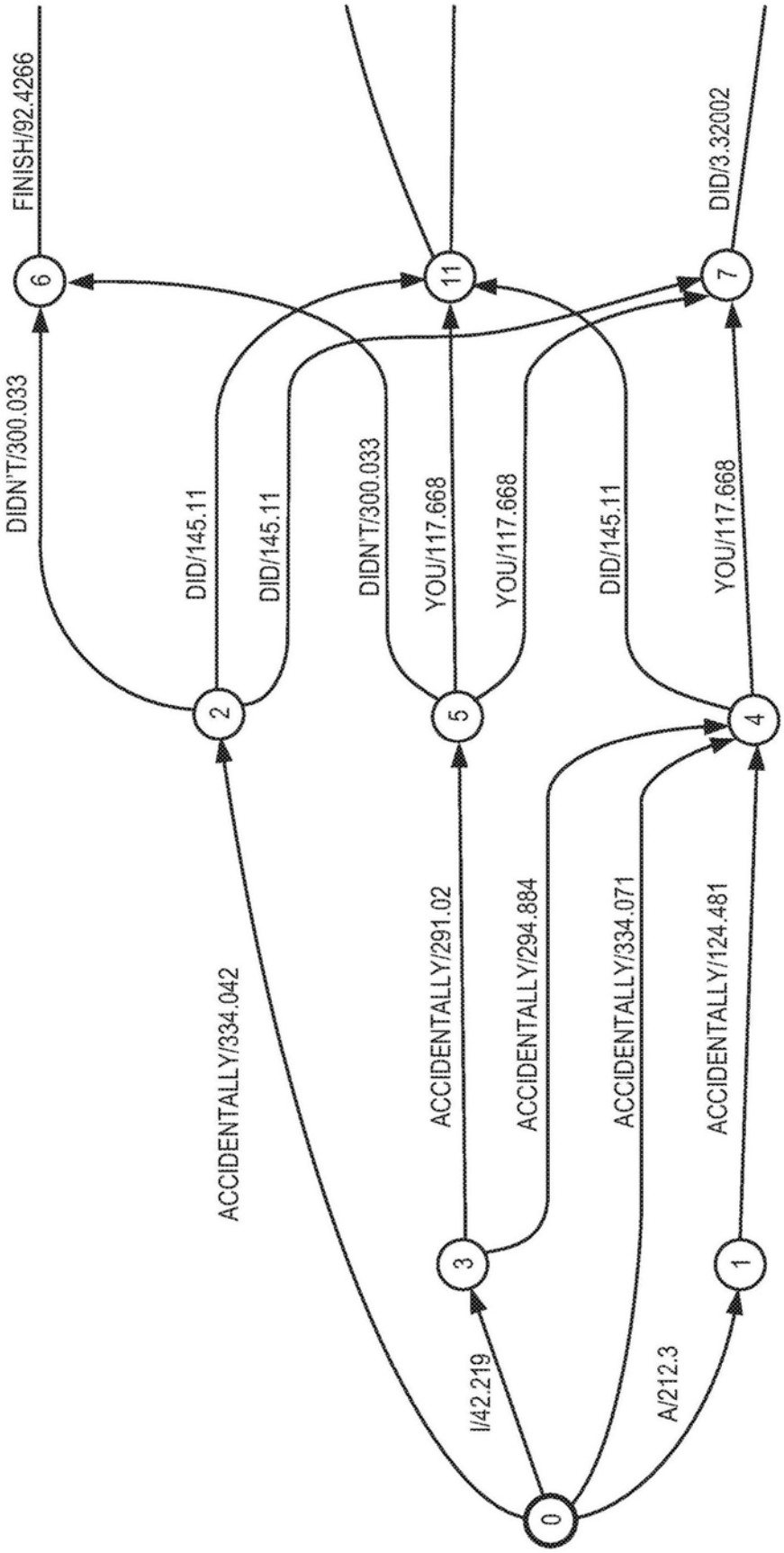

FINISH/92.4266
DID/3.32002
DIDN'T/300.033
DID/145.11
DID/145.11
DIDN'T/300.033
YOU/117.668
YOU/117.668
DID/145.11
YOU/117.668
ACCIDENTALLY/334.042
ACCIDENTALLY/291.02
ACCIDENTALLY/294.884
ACCIDENTALLY/334.071
ACCIDENTALLY/124.481
I/42.219
A/212.3
6
11
7
2
5
4
3
1
0

图27A

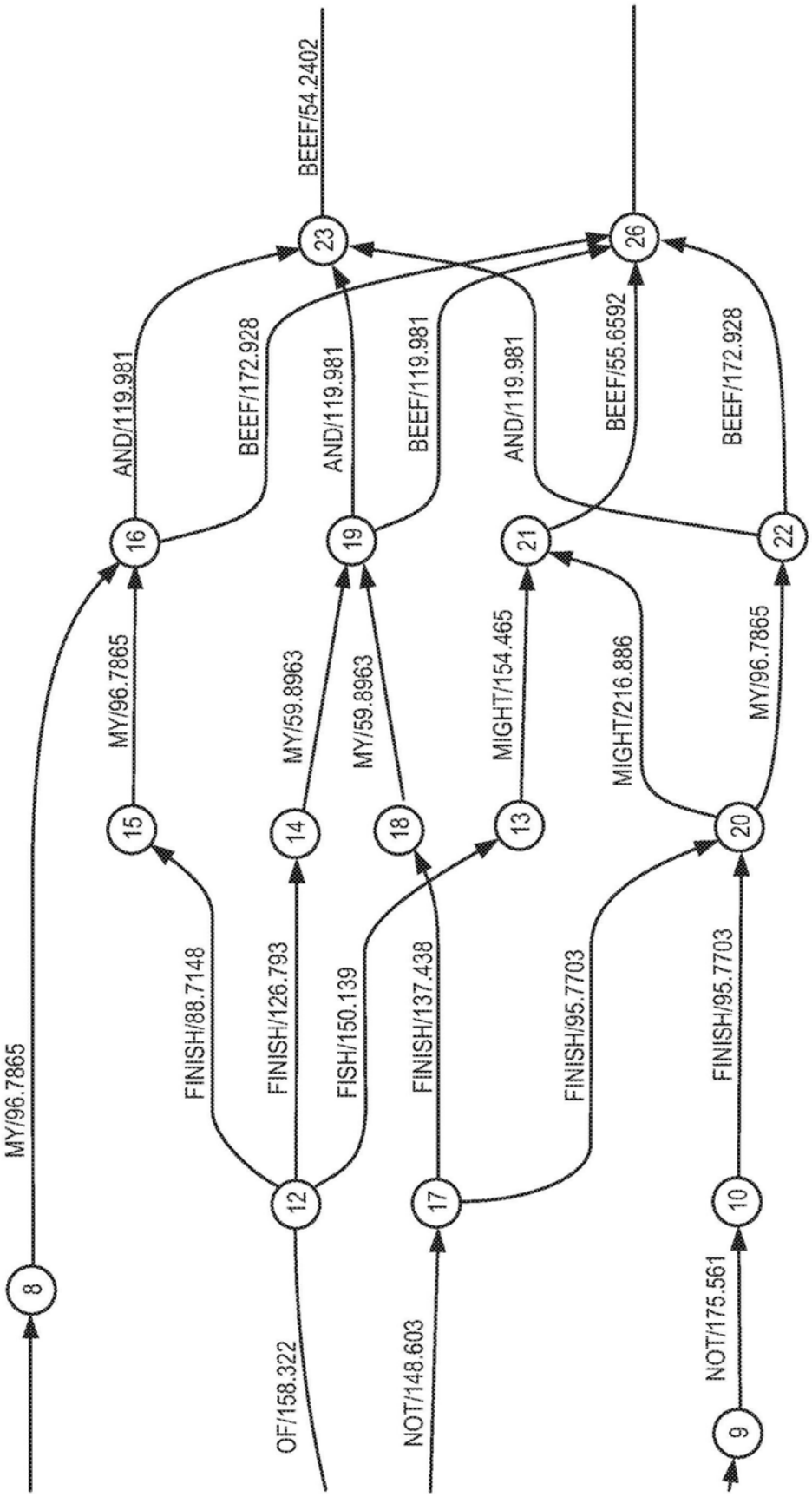

BEEF/54.2402
23
26
AND/119.981
BEEF/172.928
AND/119.981
BEEF/119.981
AND/119.981
BEEF/55.6592
BEEF/172.928
16
19
21
22
MY/96.7865
MY/59.8963
MY/59.8963
MIGHT/154.465
MIGHT/216.886
MY/96.7865
15
14
18
13
20
MY/96.7865
FINISH/88.7148
FINISH/126.793
FISH/150.139
FINISH/137.438
FINISH/95.7703
FINISH/95.7703
8
12
17
10
OF/158.322
NOT/148.603
NOT/175.561
9

图27B

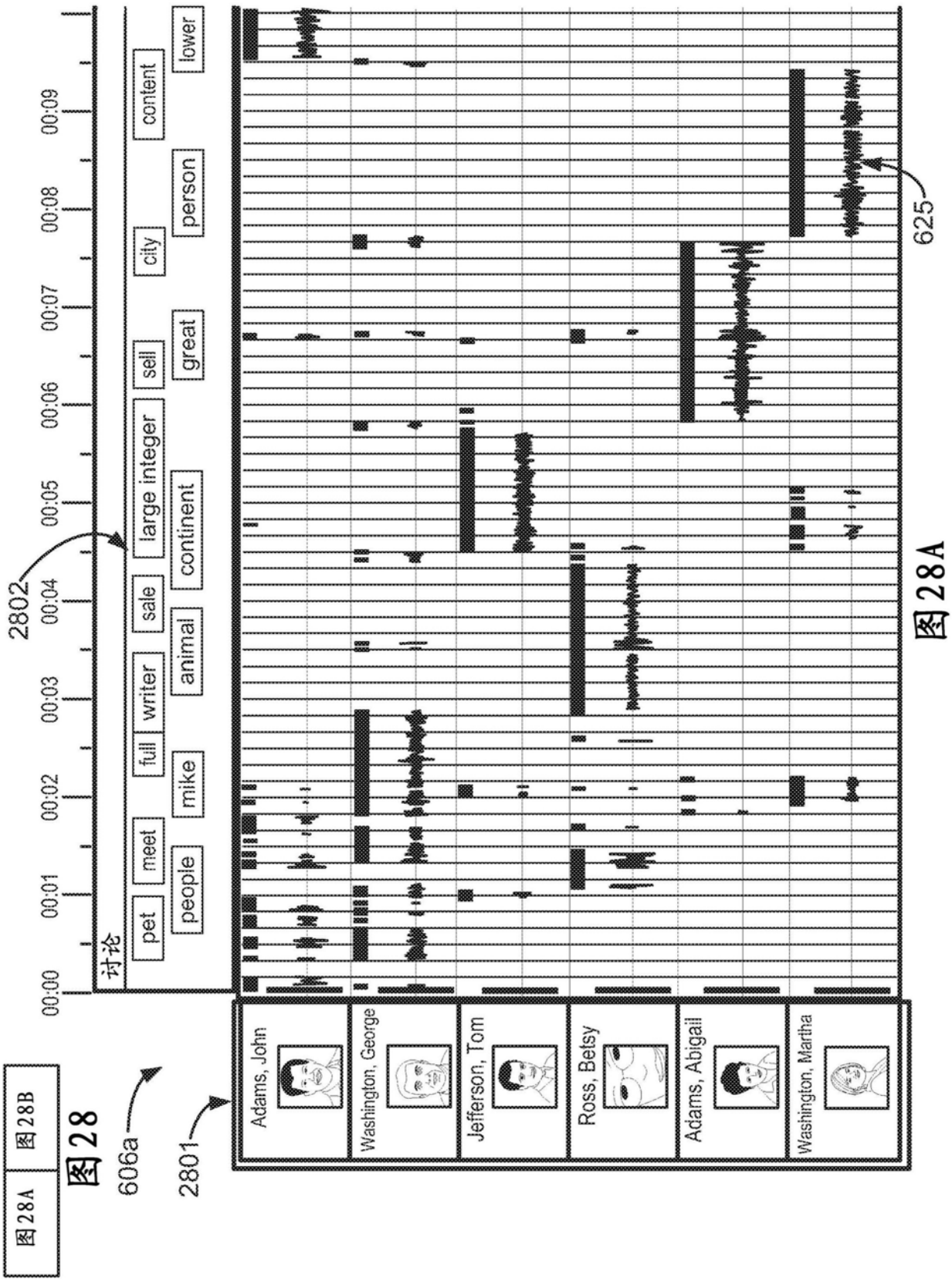


图28A
图28B
图28
606a
2801
2802
625
讨论
00:00
00:01
00:02
00:03
00:04
00:05
00:06
00:07
00:08
00:09
pet
meet
full
writer
sale
large integer
sell
city
content
people
mike
animal
continent
great
person
lower
Adams, John
Washington, George
Jefferson, Tom
Ross, Betsy
Adams, Abigail
Washington, Martha

图28A

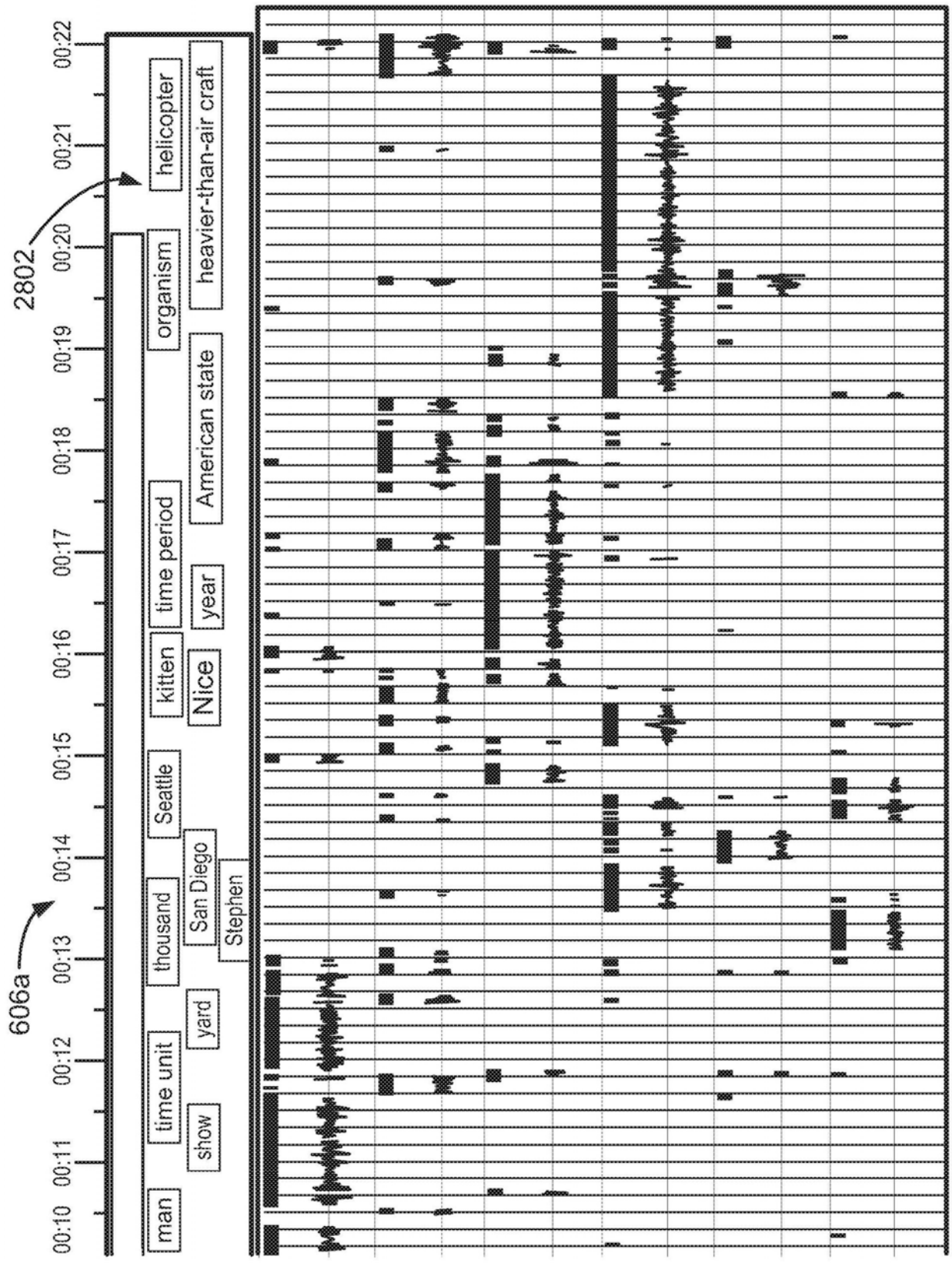

2802
606a
00:10
00:11
00:12
00:13
00:14
00:15
00:16
00:17
00:18
00:19
00:20
00:21
00:22
man
time unit
show
yard
thousand
San Diego
Stephen
Seattle
kitten
Nice
time period
year
American state
organism
heavier-than-air craft
helicopter

图28B

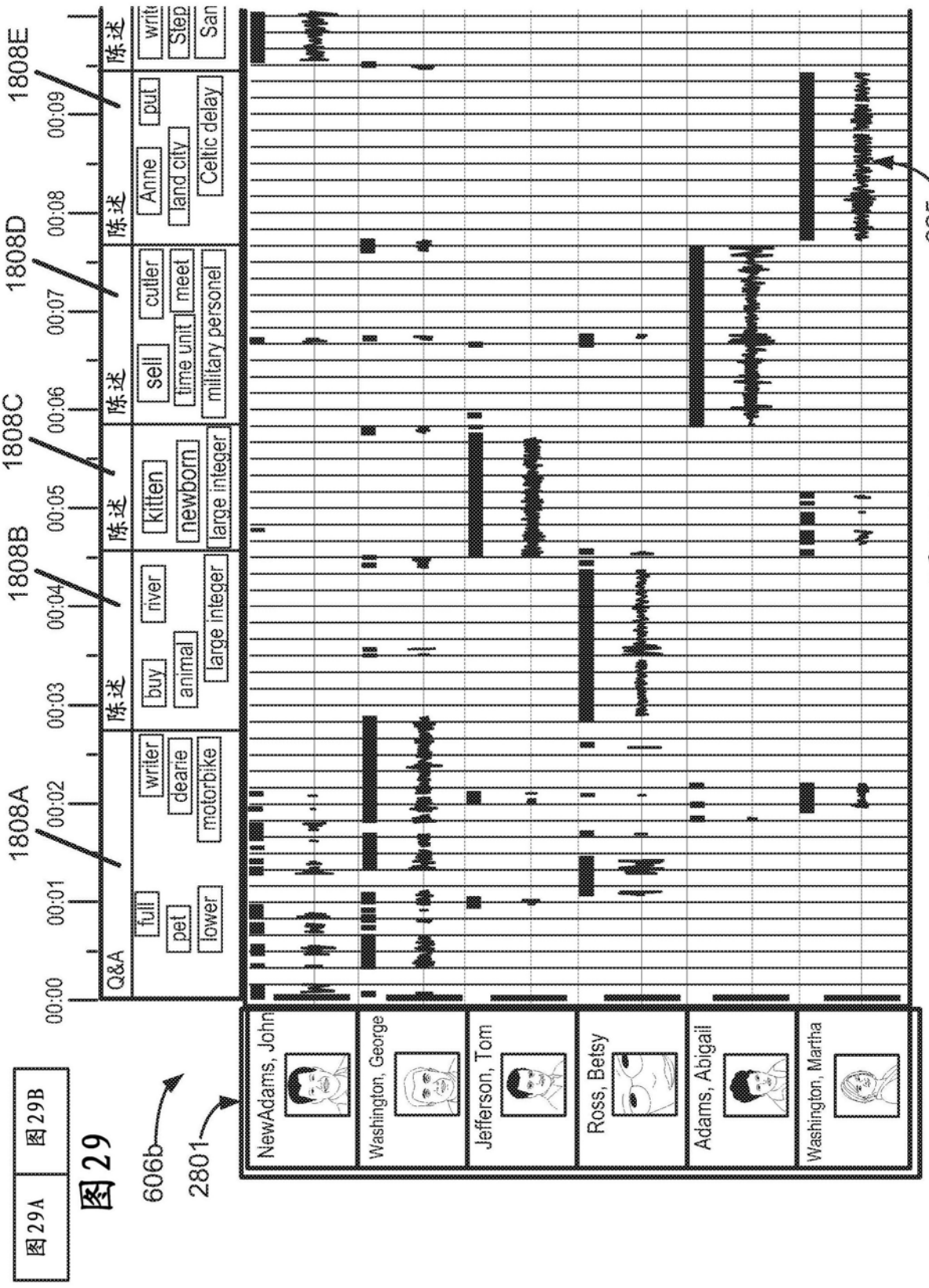


图29A
图29B
图29
606b
2801
1808A
1808B
1808C
1808D
1808E
00:00
00:01
00:02
00:03
00:04
00:05
00:06
00:07
00:08
00:09
Q&A
full
pet
lower
writer
dearie
motorbike
陈述
buy
animal
large integer
river
kitten
newborn
large integer
sell
time unit
military personel
cutler
meet
Anne
land city
Celtic delay
put
NewAdams, John
Washington, George
Jefferson, Tom
Ross, Betsy
Adams, Abigail
Washington, Martha
625

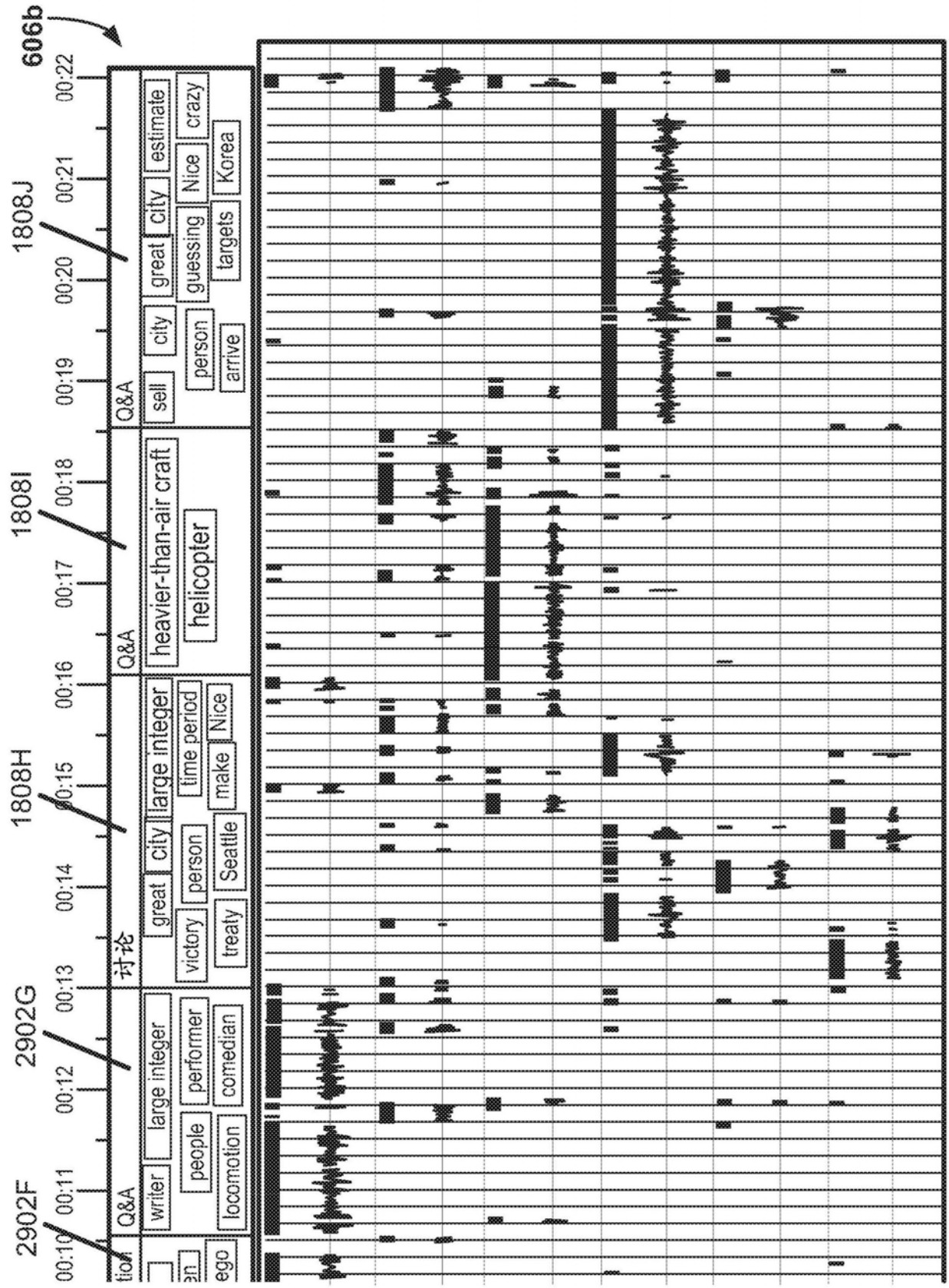

606b
2902F
2902G
1808H
1808I
1808J
00:10
00:11
00:12
00:13
00:14
00:15
00:16
00:17
00:18
00:19
00:20
00:21
00:22
Q&A
writer
large integer
people
performer
locomotion
comedian
讨论
great
city
large integer
victory
person
time period
treaty
Seattle
make
Nice
Q&A
heavier-than-air craft
helicopter
Q&A
sell
city
person
arrive
great
city
estimate
guessing
Nice
crazy
targets
Korea

图29B

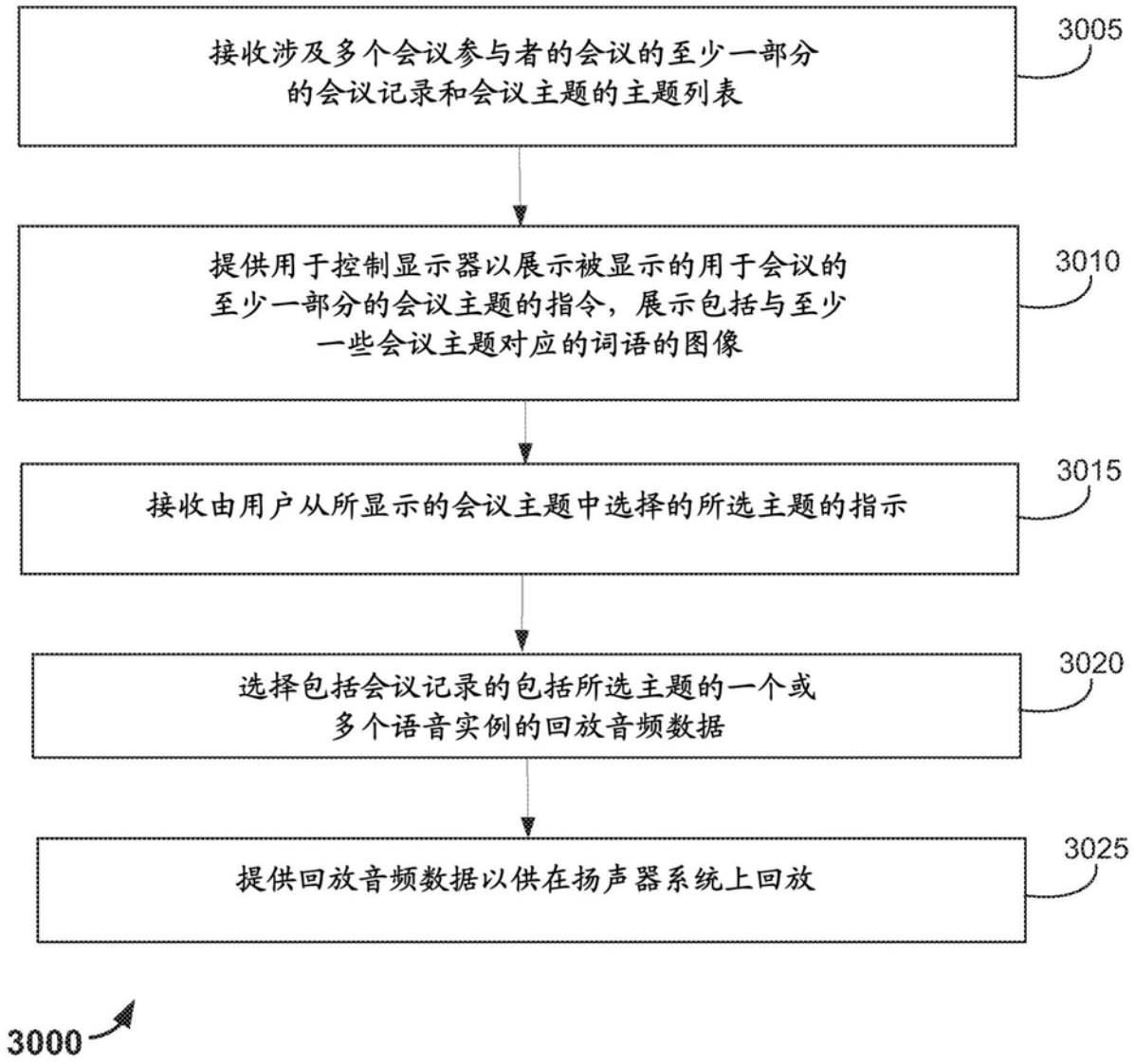

接收涉及多个会议参与者的会议的至少一部分的会议记录和会议主题的主题列表
3005
提供用于控制显示器以展示被显示的用于会议的至少一部分的会议主题的指令，展示包括与至少一些会议主题对应的词语的图像
3010
接收由用户从所显示的会议主题中选择的所选主题的指示
3015
选择包括会议记录的包括所选主题的一个或多个语音实例的回放音频数据
3020
提供回放音频数据以供在扬声器系统上回放
3025
3000

图30

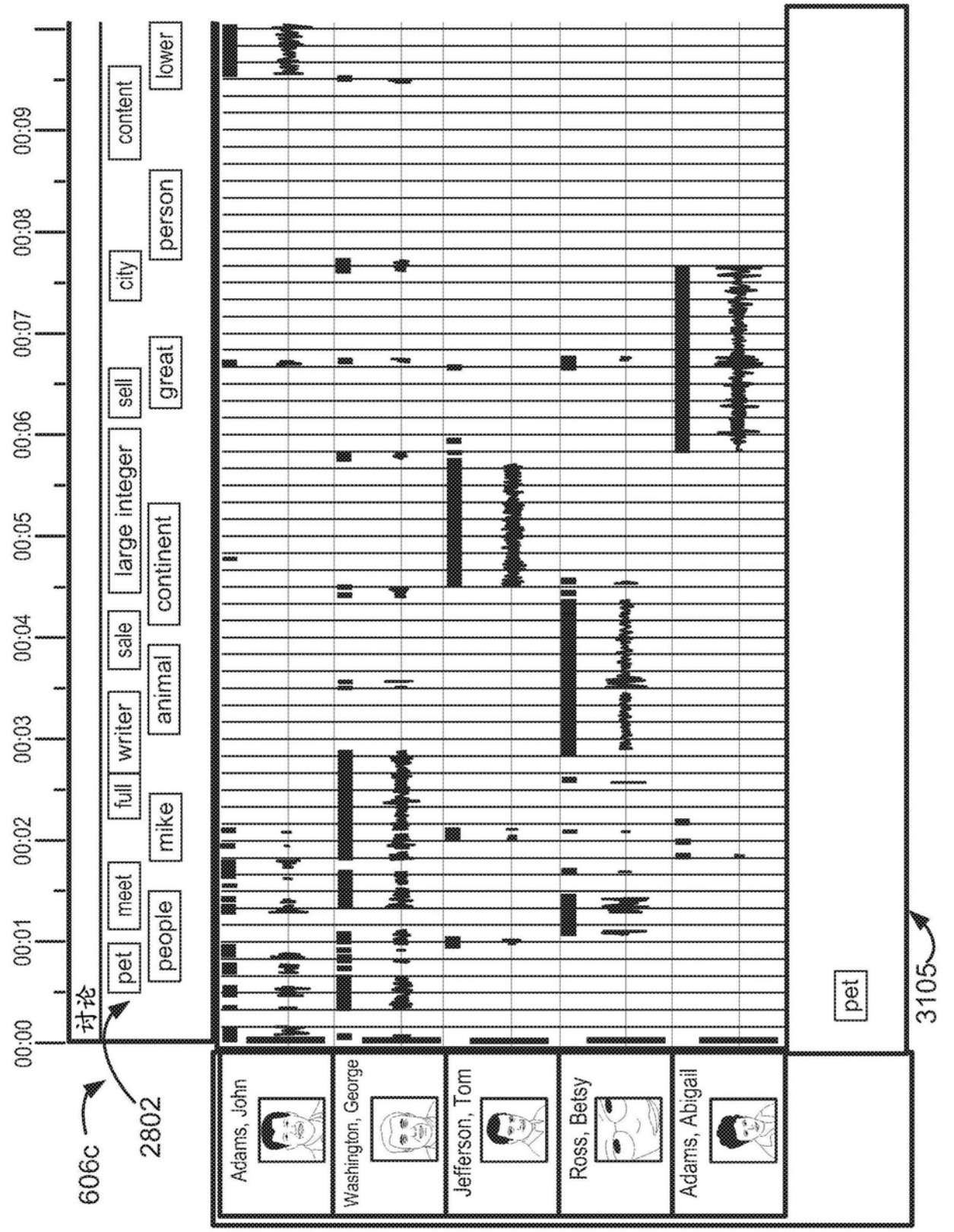

00:00
00:01
00:02
00:03
00:04
00:05
00:06
00:07
00:08
00:09
讨论
pet
meet
full
writer
sale
large integer
sell
city
content
people
mike
animal
continent
great
person
lower
606c
2802
Adams, John
Washington, George
Jefferson, Tom
Ross, Betsy
Adams, Abigail
pet
3105

图31

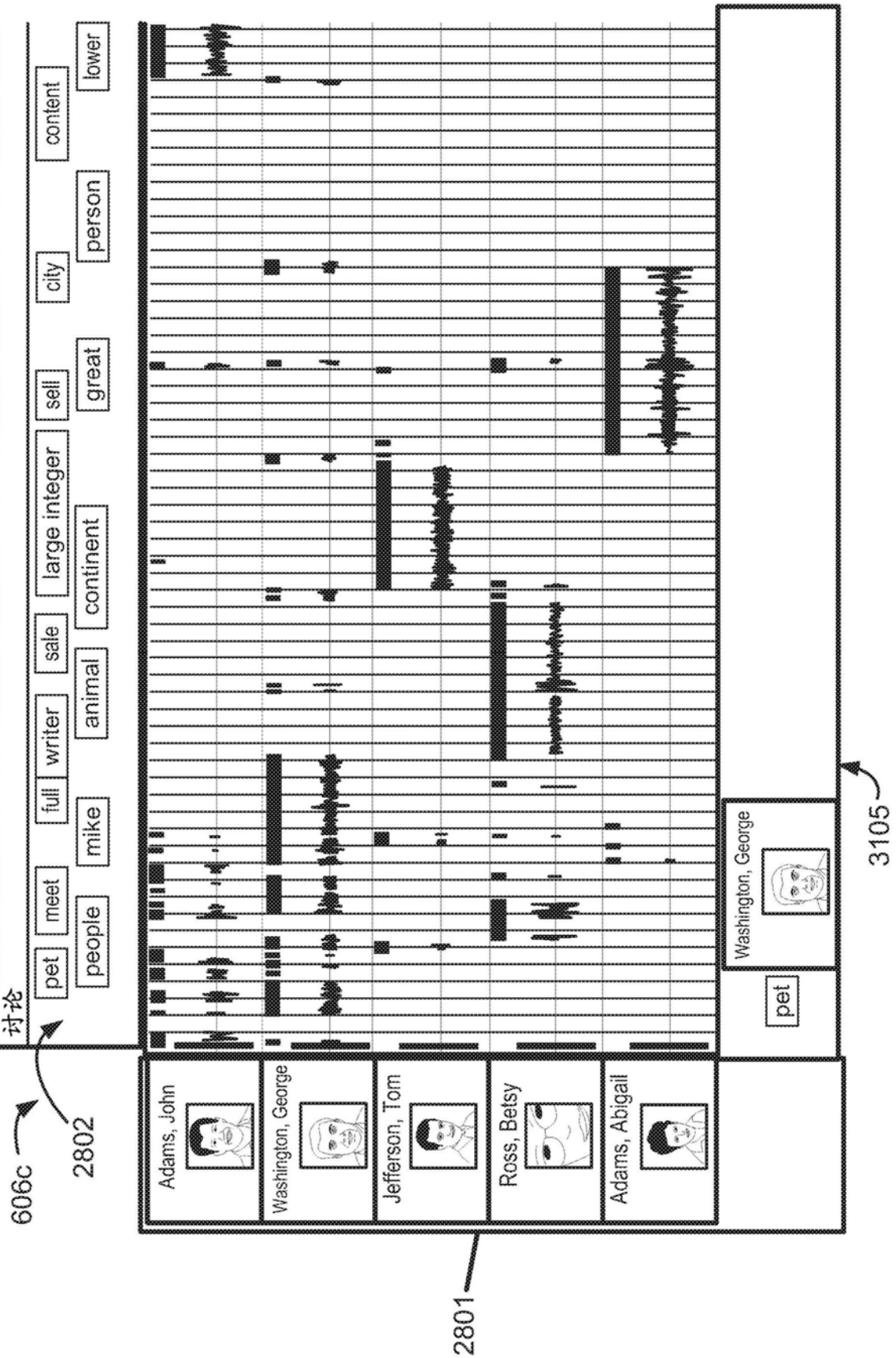

00:00
00:01
00:02
00:03
00:04
00:05
00:06
00:07
00:08
00:09
讨论
pet
meet
people
mike
full
writer
animal
sale
continent
large integer
sell
great
city
person
content
lower
606c
2802
2801
3105
Adams, John
Washington, George
Jefferson, Tom
Ross, Betsy
Adams, Abigail
Washington, George
pet

图32

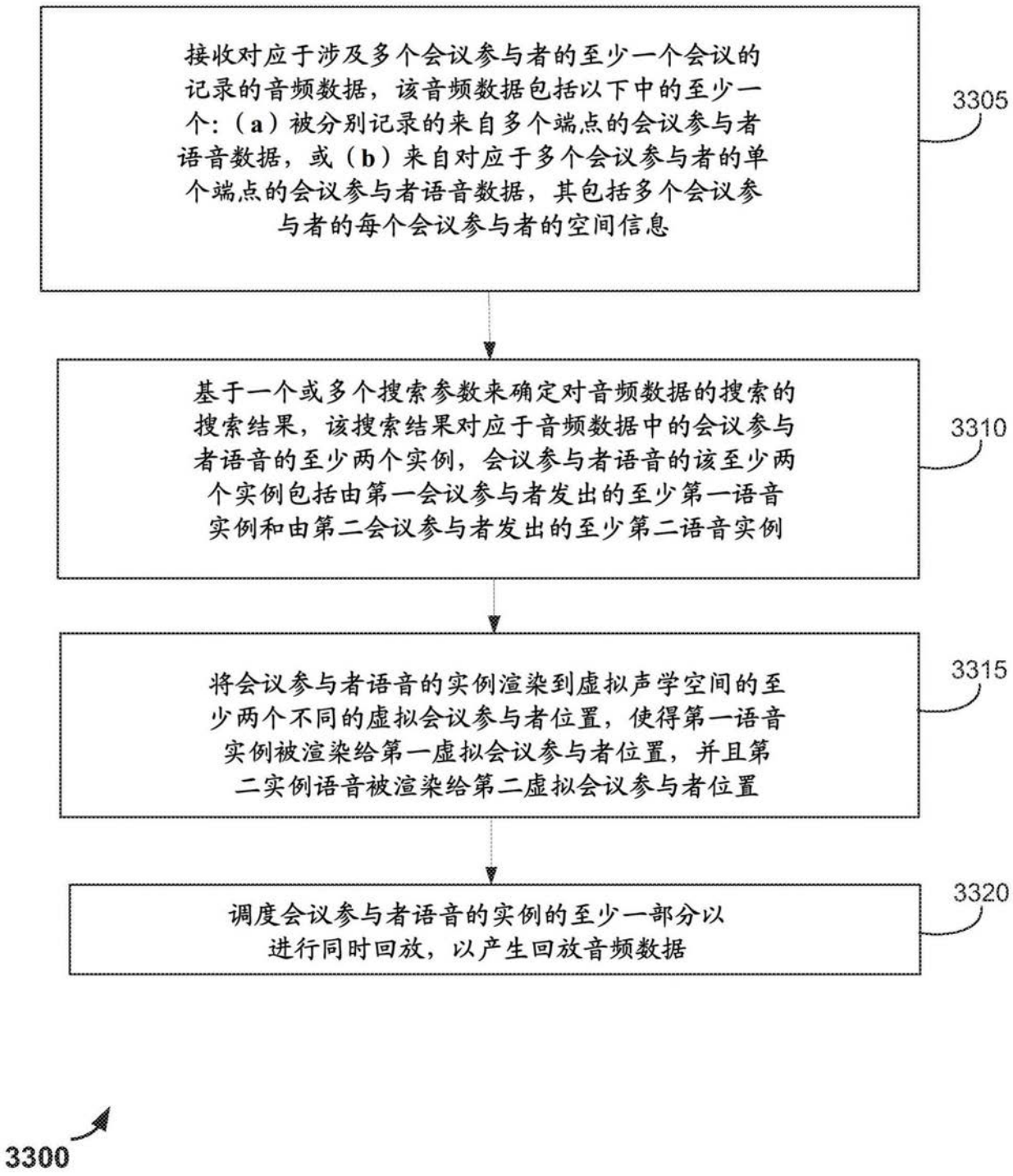

接收对应于涉及多个会议参与者的至少一个会议的记录的音频数据，该音频数据包括以下中的至少一个:（a）被分别记录的来自多个端点的会议参与者语音数据，或（b）来自对应于多个会议参与者的单个端点的会议参与者语音数据，其包括多个会议参与者的每个会议参与者的空间信息
3305
基于一个或多个搜索参数来确定对音频数据的搜索的搜索结果，该搜索结果对应于音频数据中的会议参与者语音的至少两个实例，会议参与者语音的该至少两个实例包括由第一会议参与者发出的至少第一语音实例和由第二会议参与者发出的至少第二语音实例
3310
将会议参与者语音的实例渲染到虚拟声学空间的至少两个不同的虚拟会议参与者位置，使得第一语音实例被渲染给第一虚拟会议参与者位置，并且第二实例语音被渲染给第二虚拟会议参与者位置
3315
调度会议参与者语音的实例的至少一部分以进行同时回放，以产生回放音频数据
3320
3300

图33

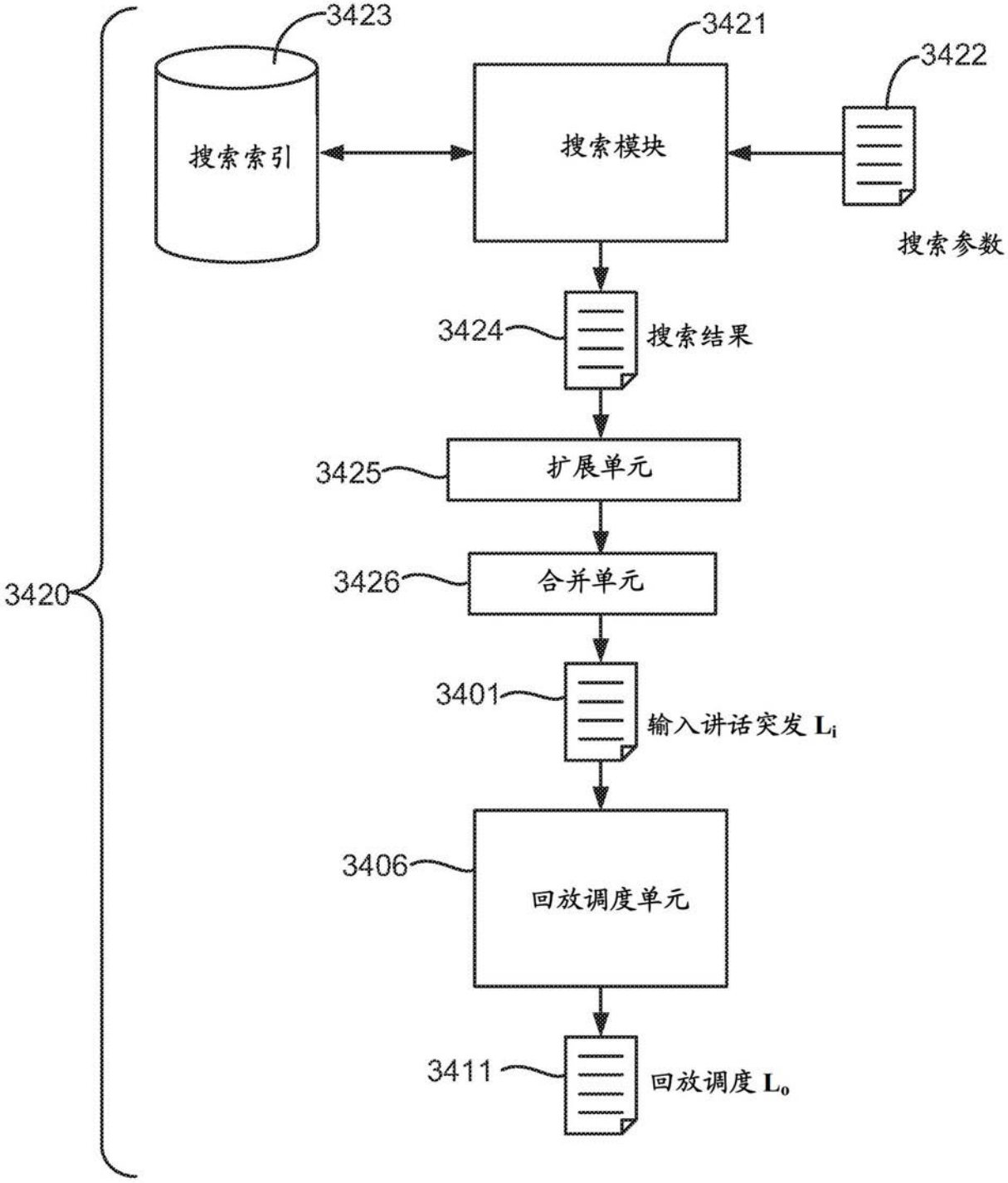

3423
3421
3422
搜索索引
搜索模块
搜索参数
3424
搜索结果
3425
扩展单元
3426
合并单元
3420
3401
输入讲话突发 L_i
3406
回放调度单元
3411
回放调度 L_o

图34

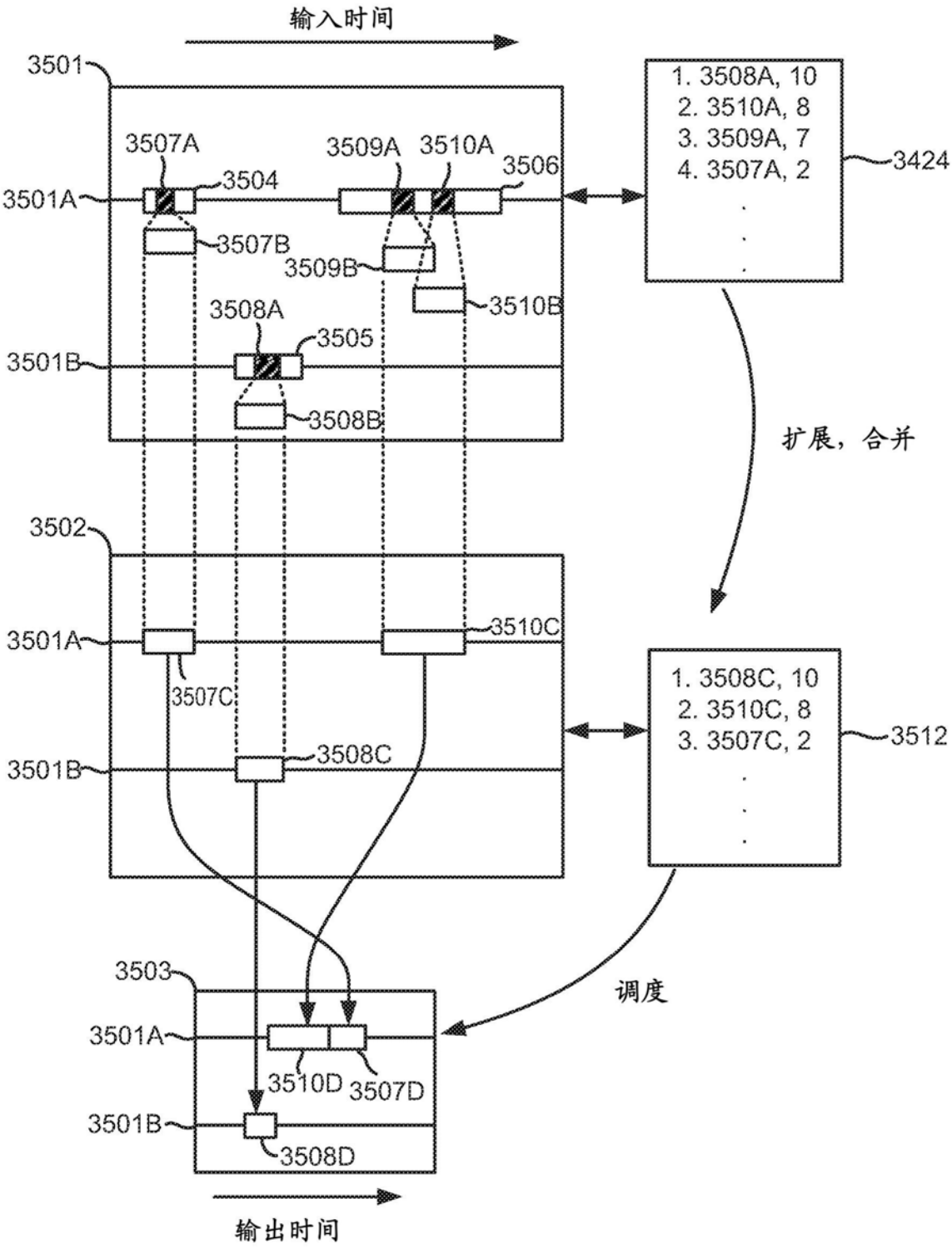

输入时间
3501
3507A
3504
3509A
3510A
3506
3501A
3507B
3509B
3510B
3508A
3505
3501B
3508B
1. 3508A, 10
2. 3510A, 8
3. 3509A, 7
4. 3507A, 2
3424
扩展，合并
3502
3501A
3510C
3507C
3501B
3508C
1. 3508C, 10
2. 3510C, 8
3. 3507C, 2
3512
调度
3503
3501A
3510D
3507D
3501B
3508D
输出时间

图35

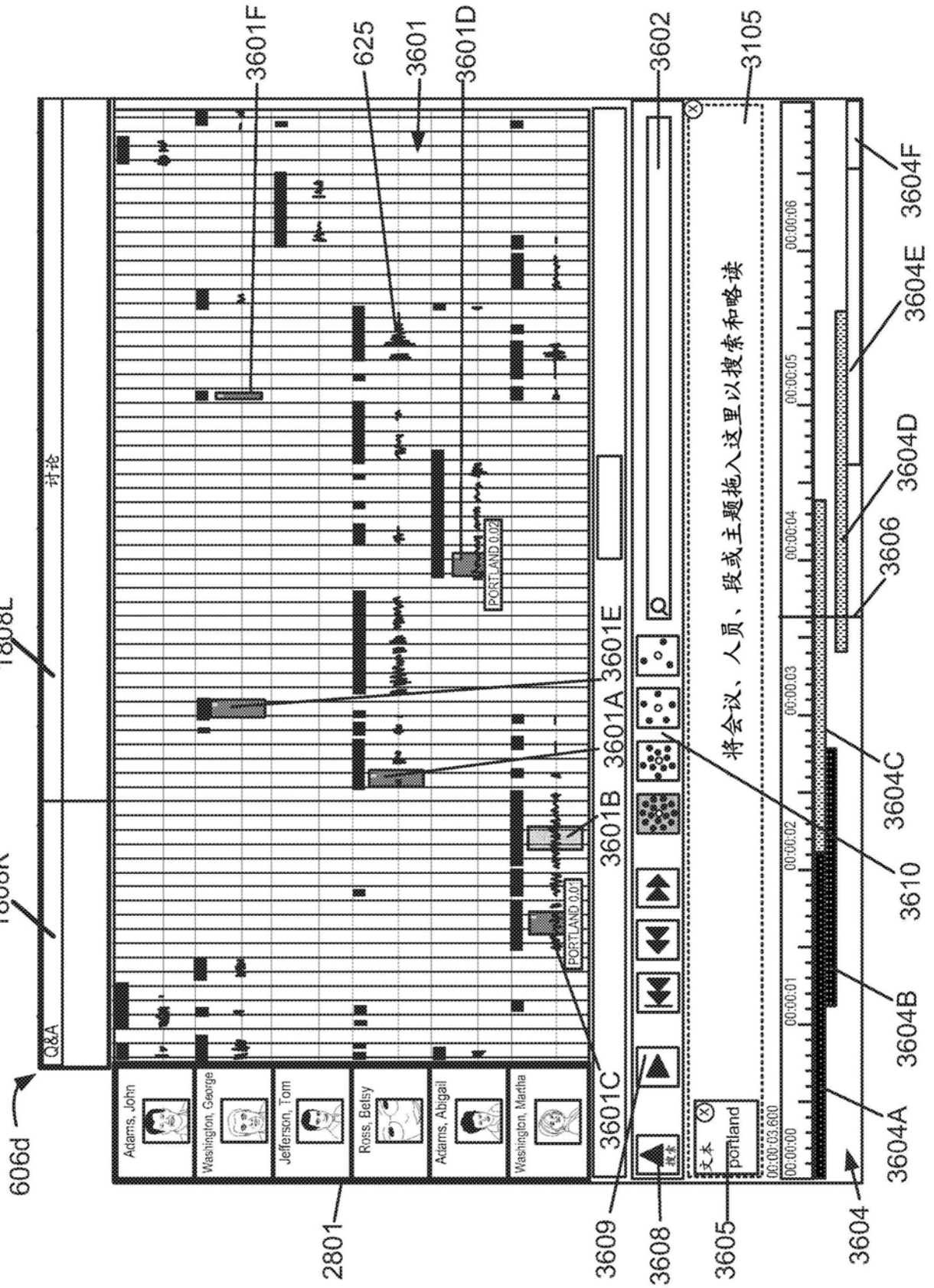

3601F
625
3601
3601D
3602
3105
3604F
3604E
3604D
3606
3604C
3610
3604B
3604A
3604
1808L
1808K
606d
2801
3609
3608
3605
3601C
3601B
3601A
3601E
讨论
Q&A
Adams, John
Washington, George
Jefferson, Tom
Ross, Betsy
Adams, Abigail
Washington, Martha
PORTLAND 0.02
PORTLAND 0.01
将会议、人员、段或主题拖入这里以搜索和略读
portland
00:00:00
00:00:01
00:00:02
00:00:03
00:00:04
00:00:05
00:00:06

图36

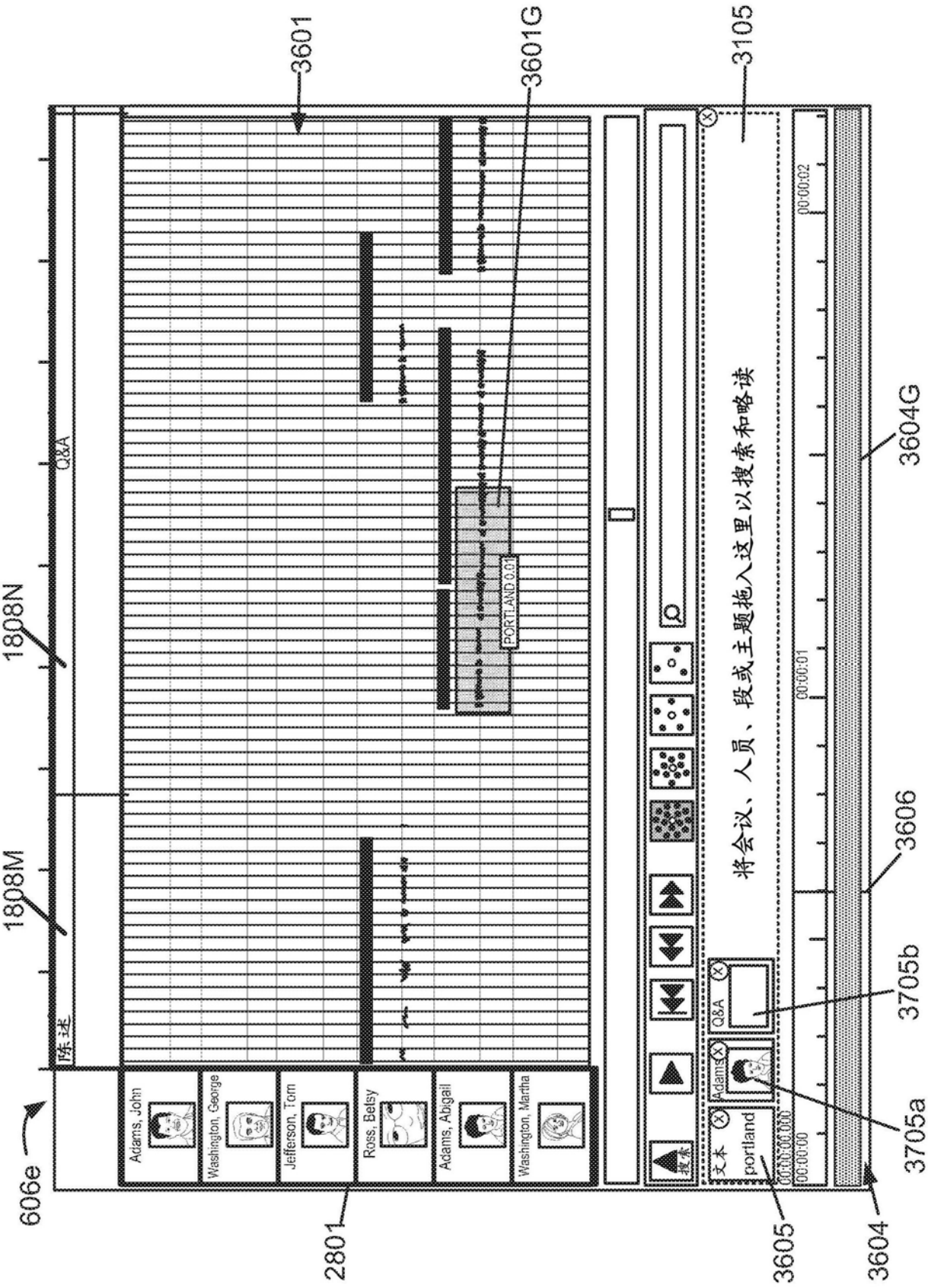

3601
3601G
3105
1808N
1808M
606e
2801
3605
3604
3705a
3705b
3606
3604G
Q&A
陈述
Adams, John
Washington, George
Jefferson, Tom
Ross, Betsy
Adams, Abigail
Washington, Martha
PORTLAND 0:01
将会议、人员、段或主题拖入这里以搜索和略读
00:00:01
00:00:02
00:00:00
portland

图37

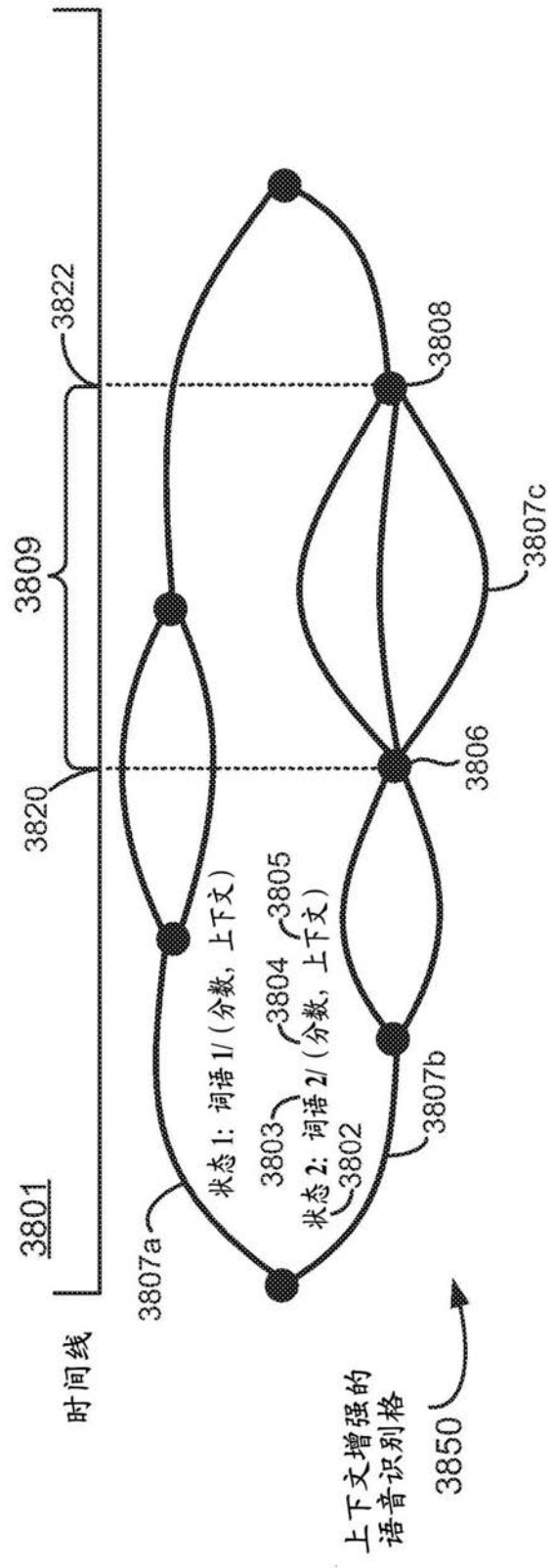

3822
3808
3809
3807c
3820
3806
3801
状态1: 词语1/(分数, 上下文)
3805
3804
状态2: 词语2/(分数, 上下文)
3803
3802
3807b
3807a
时间线
上下文增强的
语音识别格
3850

图38A

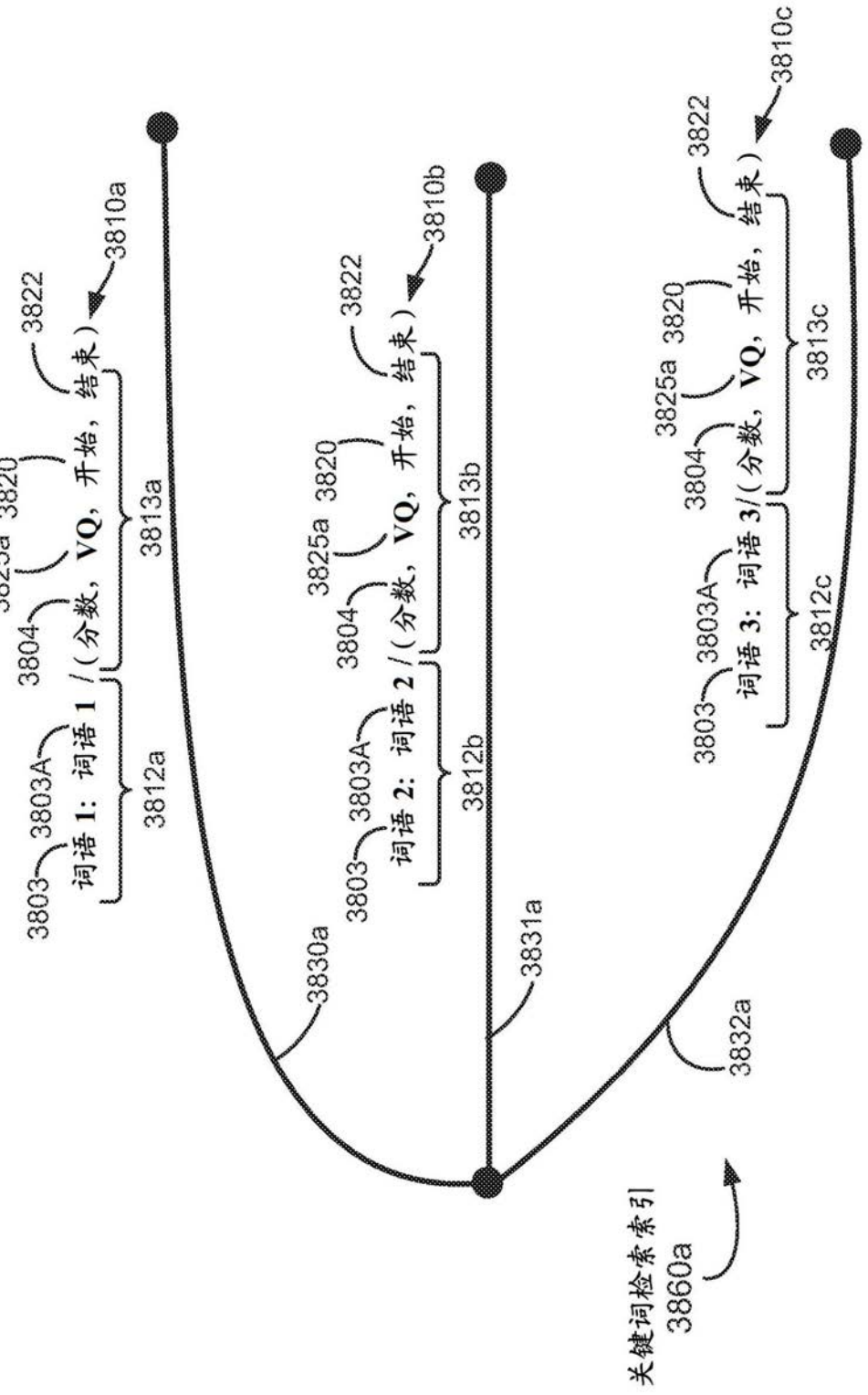

关键词检索索引
3860a
3830a
3831a
3832a
3803
3803A
词语 1：词语 1
3804
/（分数，VQ，开始，结束）
3825a
3820
3822
3810a
3812a
3813a
词语 2：词语 2
3810b
3812b
3813b
词语 3：词语 3
3810c
3812c
3813c

图38B

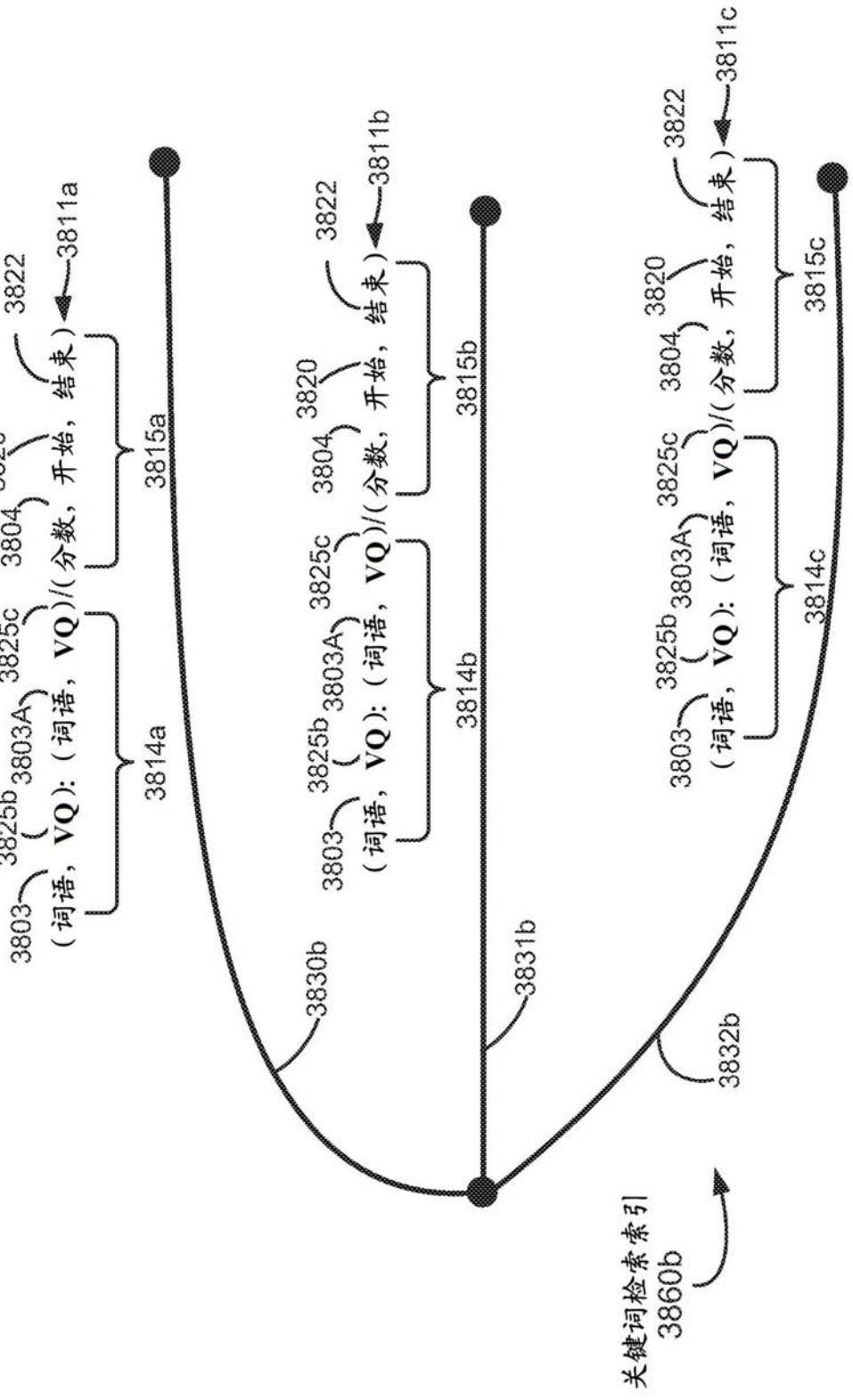

关键词检索索引
3860b
3830b
3831b
3832b
3803
(词语，VQ)
3825b
3803A
3825c
3804
(分数，开始，结束)
3820
3822
3811a
3811b
3811c
3814a
3814b
3814c
3815a
3815b
3815c

图38C

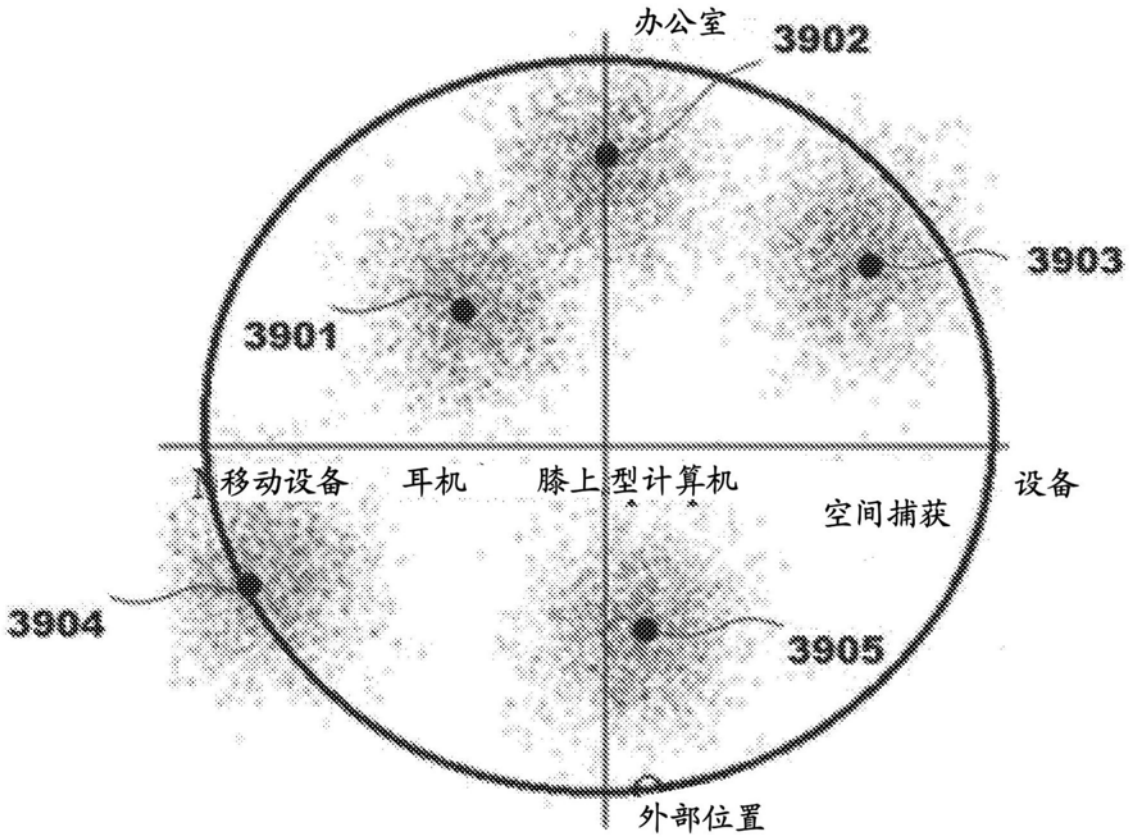


图39

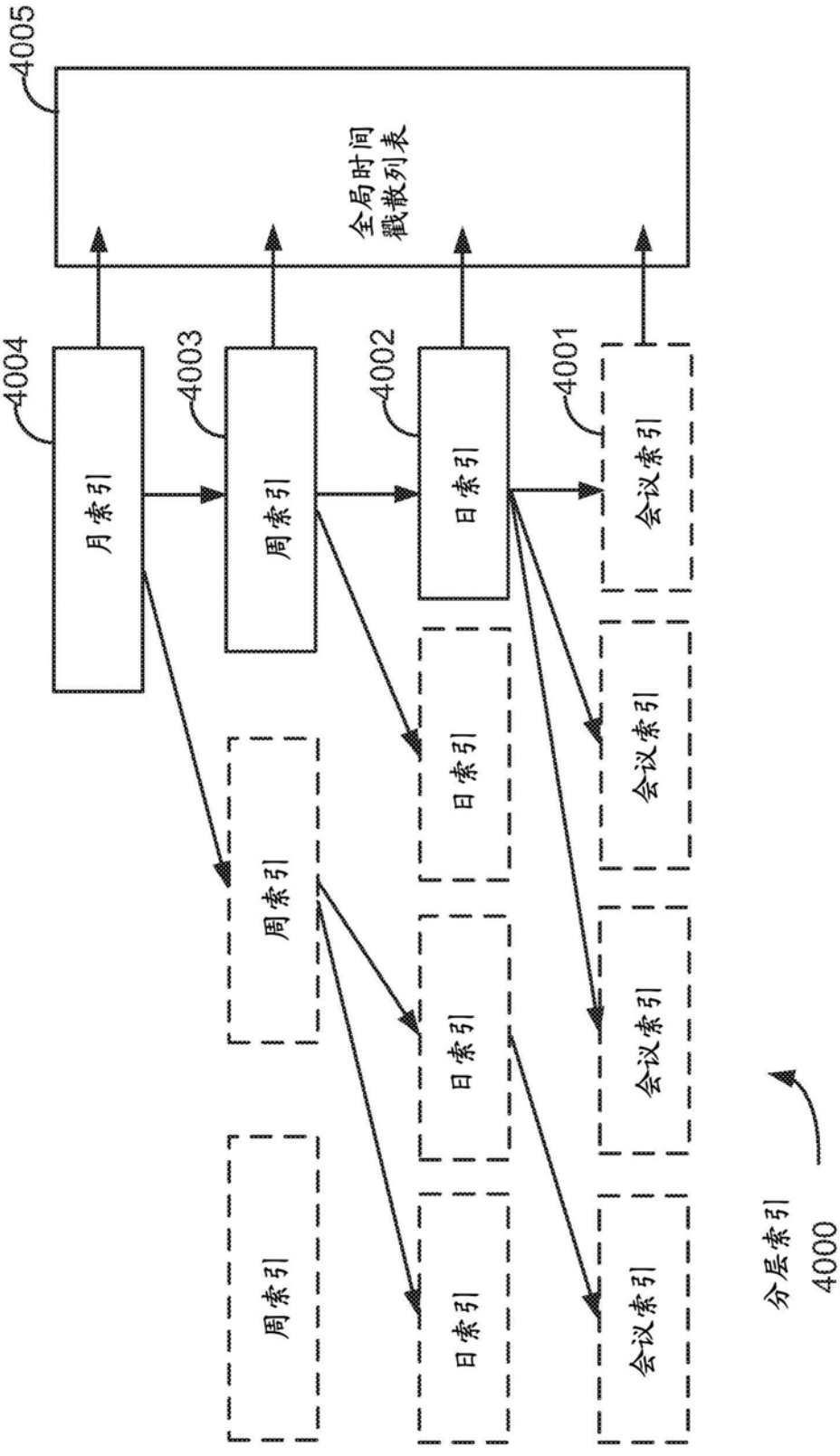

4005
全局时间戳散列表
4004
月索引
4003
周索引
4002
日索引
4001
会议索引
周索引
日索引
会议索引
周索引
日索引
会议索引
日索引
会议索引
分层索引
4000

图40

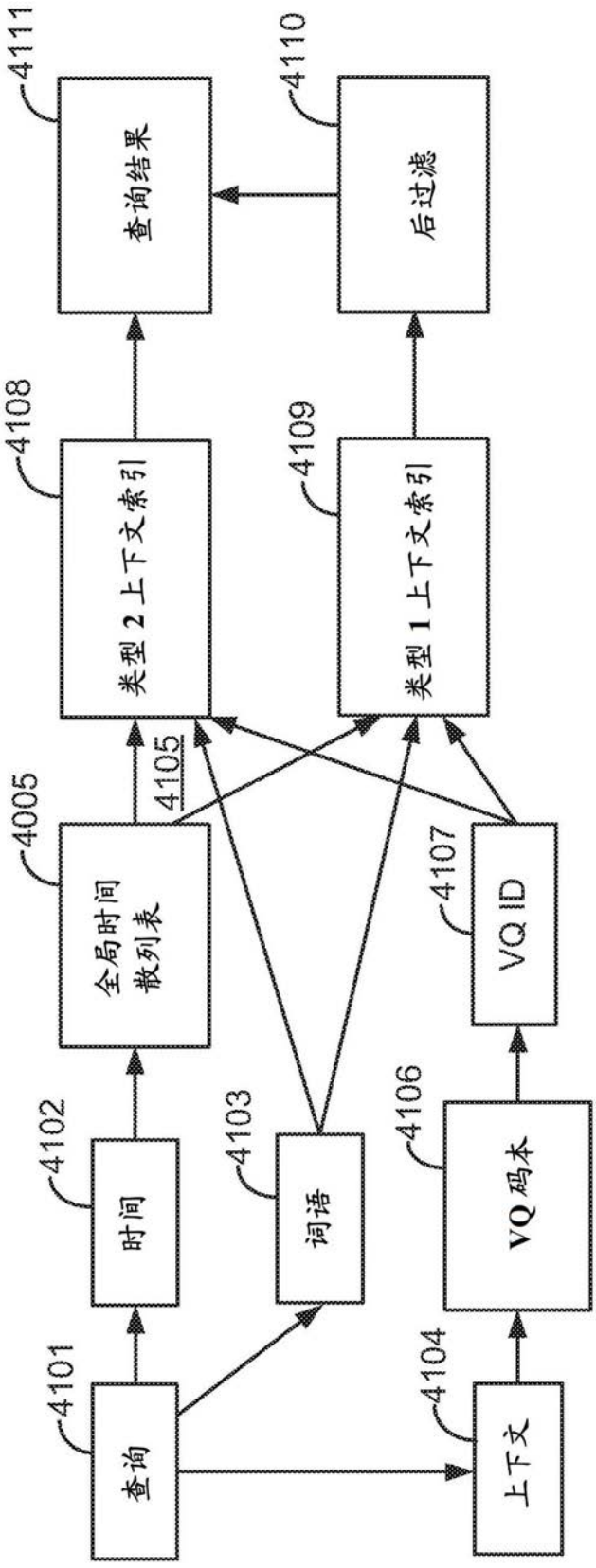

4111
查询结果
4110
后过滤
4108
类型 2 上下文索引
4109
类型 1 上下文索引
4005
全局时间
散列表
4105
4107
VQ ID
4102
时间
4103
词语
4106
VQ 码本
4101
查询
4104
上下文

图41

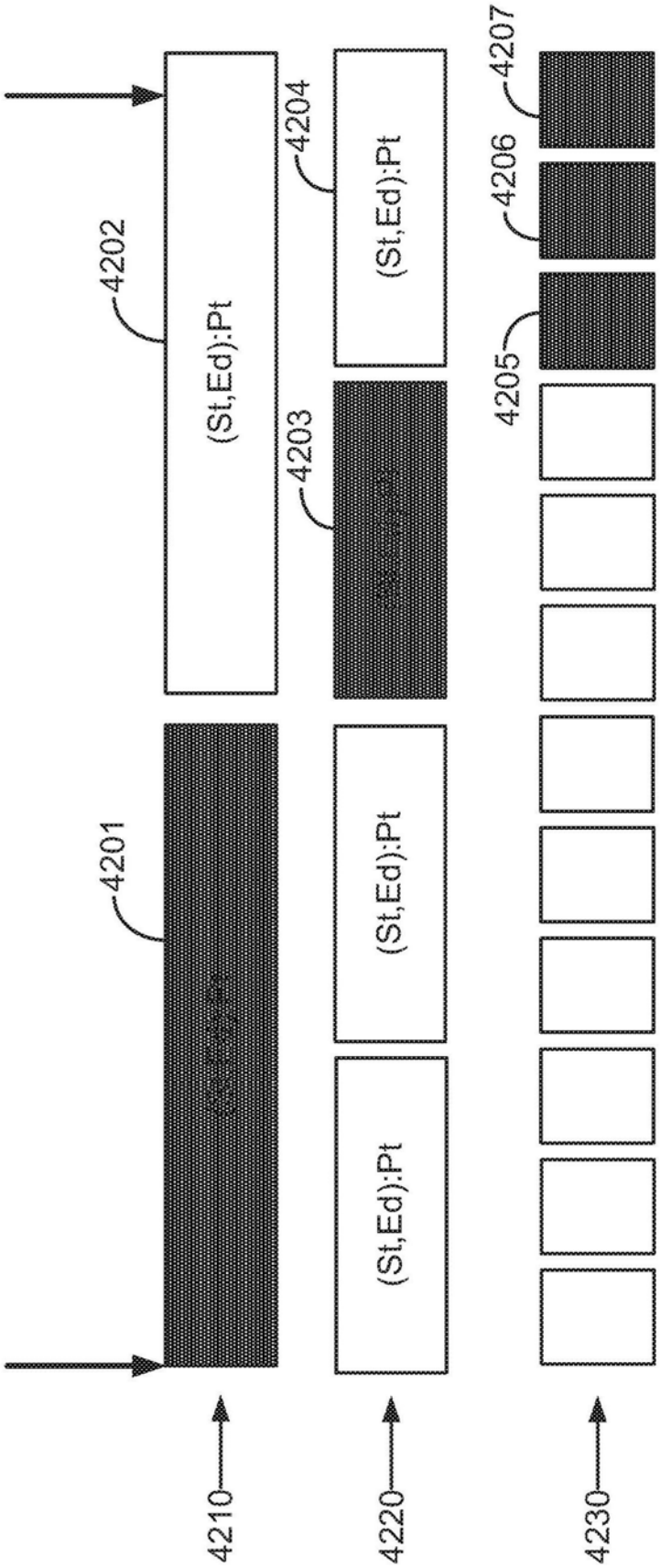

4201
4202
4203
4204
4205
4206
4207
(St,Ed):Pt
(St,Ed):Pt
(St,Ed):Pt
(St,Ed):Pt
4210
4220
4230

图42

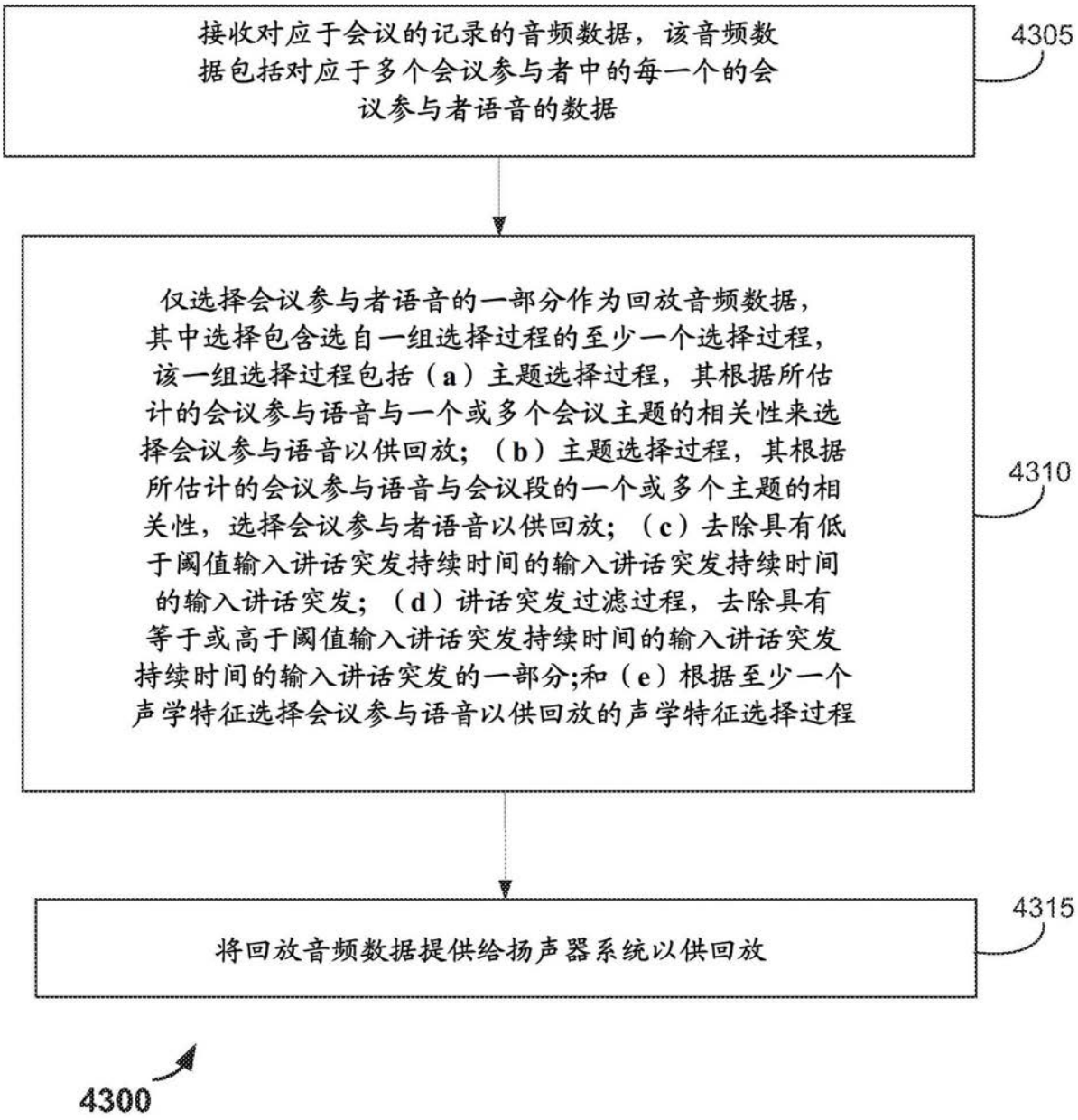

接收对应于会议的记录的音频数据，该音频数据包括对应于多个会议参与者中的每一个的会议参与者语音的数据
4305
仅选择会议参与者语音的一部分作为回放音频数据，其中选择包含选自一组选择过程的至少一个选择过程，该一组选择过程包括（a）主题选择过程，其根据所估计的会议参与语音与一个或多个会议主题的相关性来选择会议参与语音以供回放；（b）主题选择过程，其根据所估计的会议参与语音与会议段的一个或多个主题的相关性，选择会议参与者语音以供回放；（c）去除具有低于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发；（d）讲话突发过滤过程，去除具有等于或高于阈值输入讲话突发持续时间的输入讲话突发持续时间的输入讲话突发的一部分;和（e）根据至少一个声学特征选择会议参与语音以供回放的声学特征选择过程
4310
将回放音频数据提供给扬声器系统以供回放
4315
4300

图43

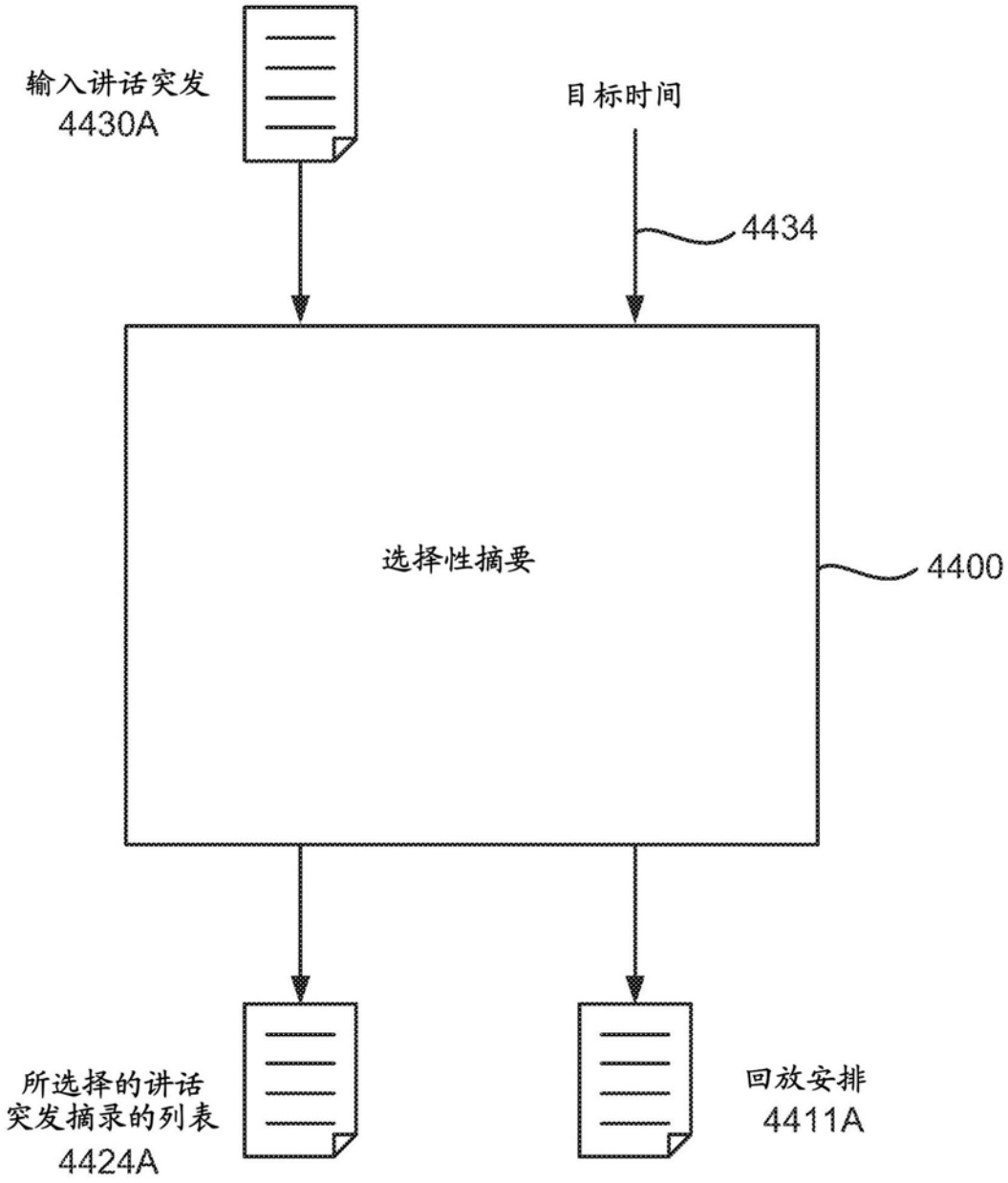

输入讲话突发
4430A
目标时间
4434
选择性摘要
4400
所选择的讲话
突发摘录的列表
4424A
回放安排
4411A

图44

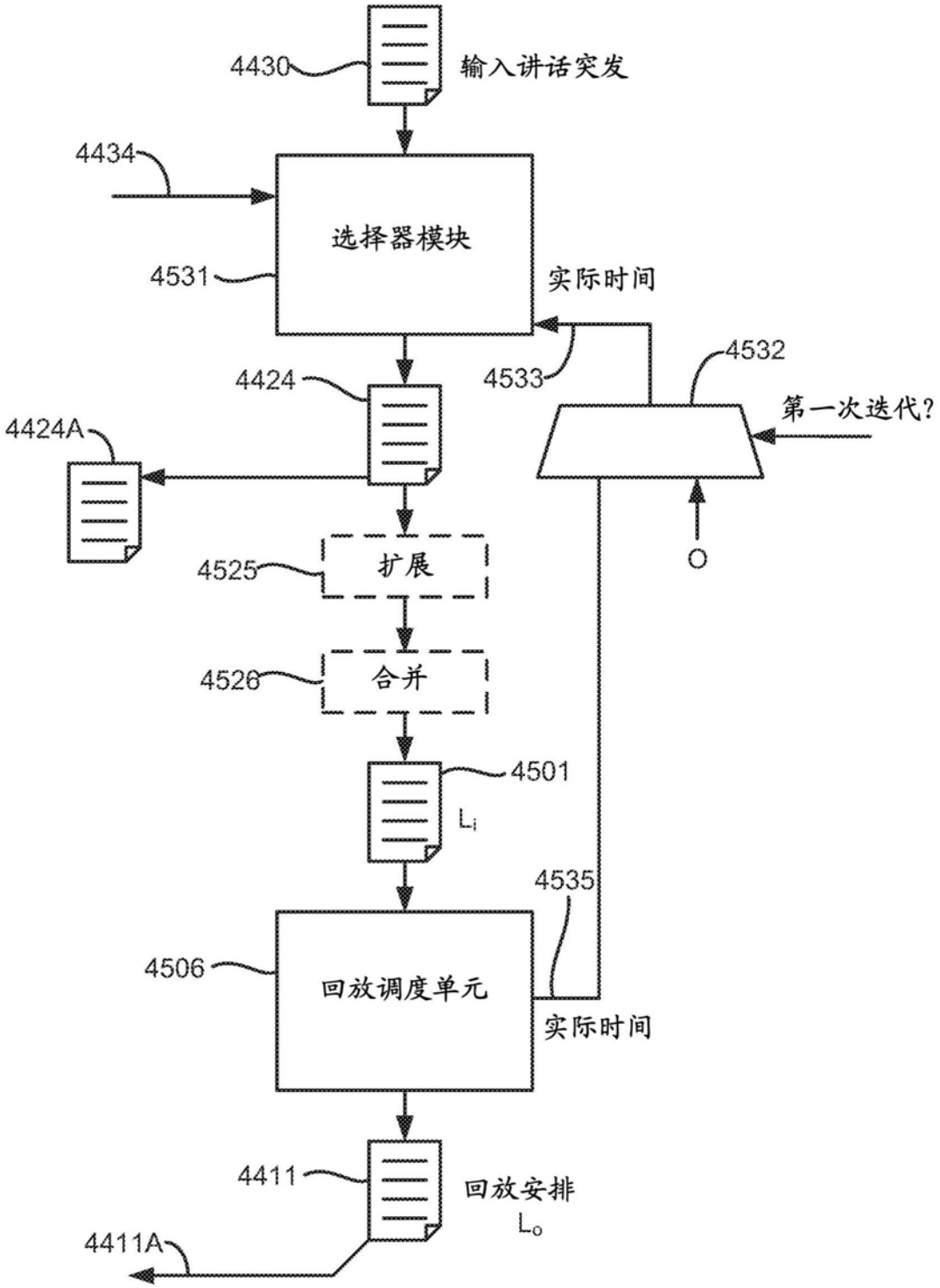

4430
输入讲话突发
4434
选择器模块
4531
实际时间
4533
4532
第一次迭代?
0
4424
4424A
4525
扩展
4526
合并
4501
L_i
4535
4506
回放调度单元
实际时间
4411
回放安排
L_o
4411A

图45

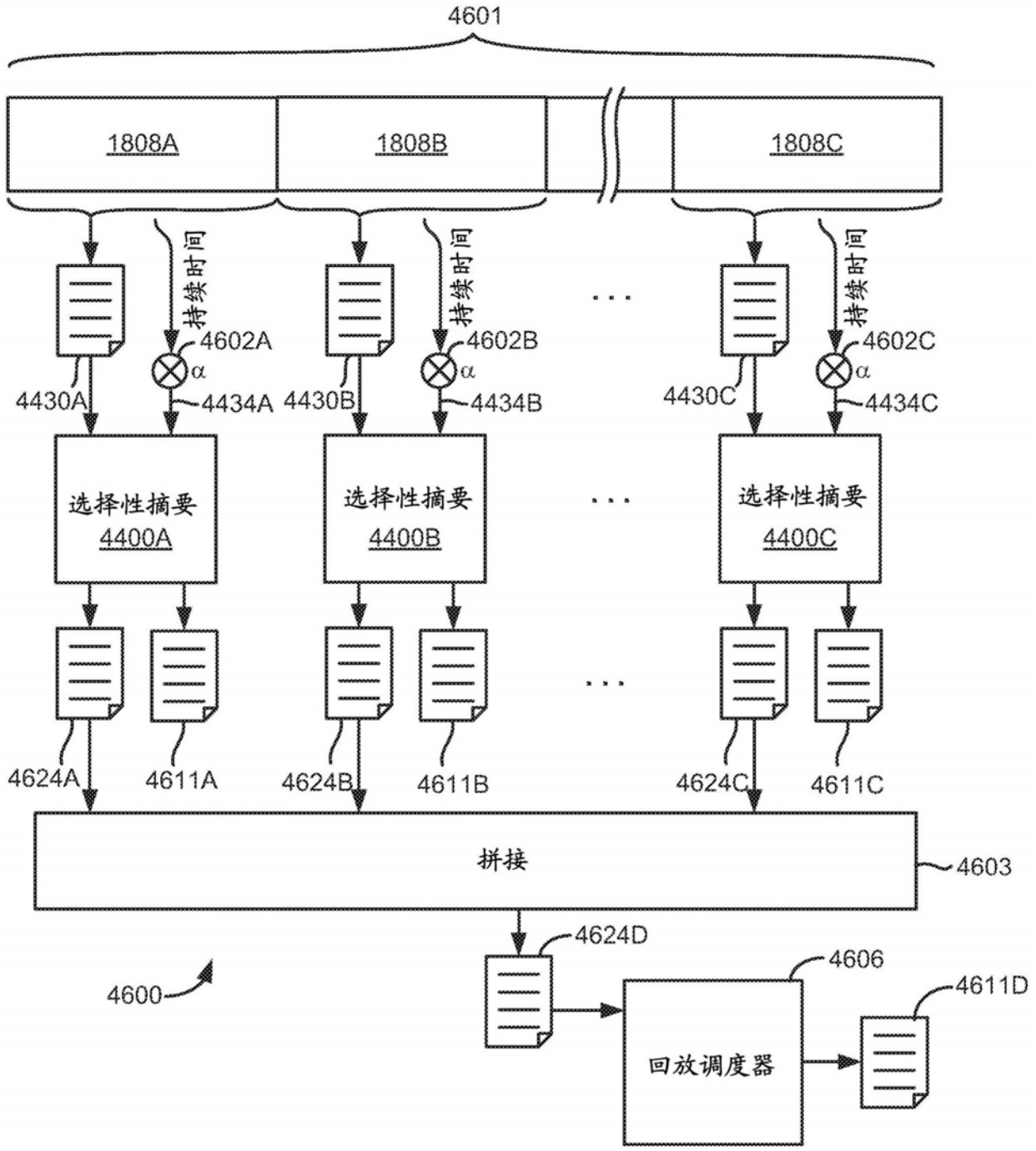

4601
1808A
1808B
1808C
持续时间
4602A
α
4434A
4430A
持续时间
4602B
α
4434B
4430B
持续时间
4602C
α
4434C
4430C
选择性摘要
4400A
选择性摘要
4400B
选择性摘要
4400C
4624A
4611A
4624B
4611B
4624C
4611C
拼接
4603
4624D
4606
4611D
回放调度器
4600

图46

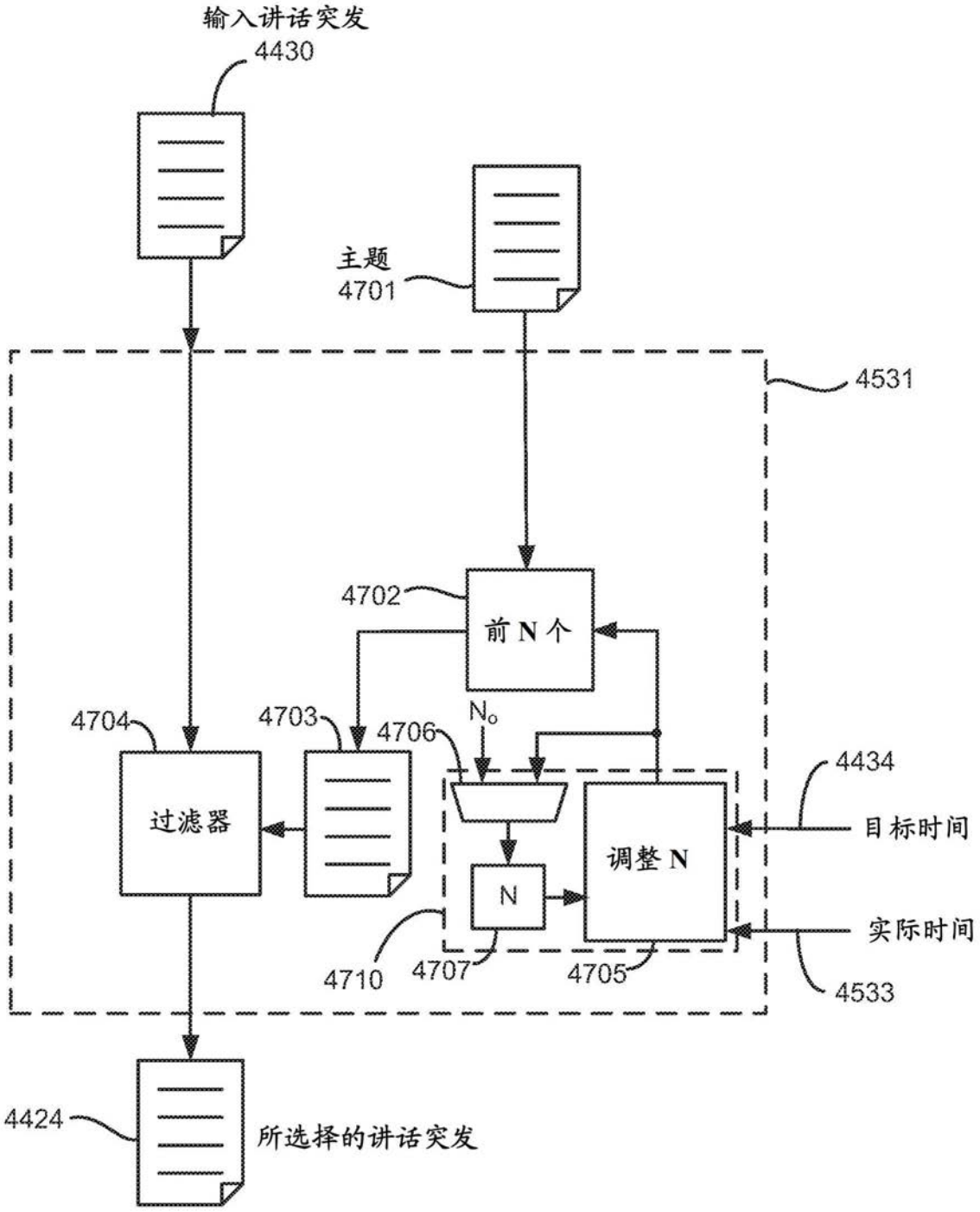

输入讲话突发
4430
主题
4701
4531
4702
前 N 个
4704
4703
4706
N_o
4434
过滤器
目标时间
调整 N
N
实际时间
4710
4707
4705
4533
4424
所选择的讲话突发

图47

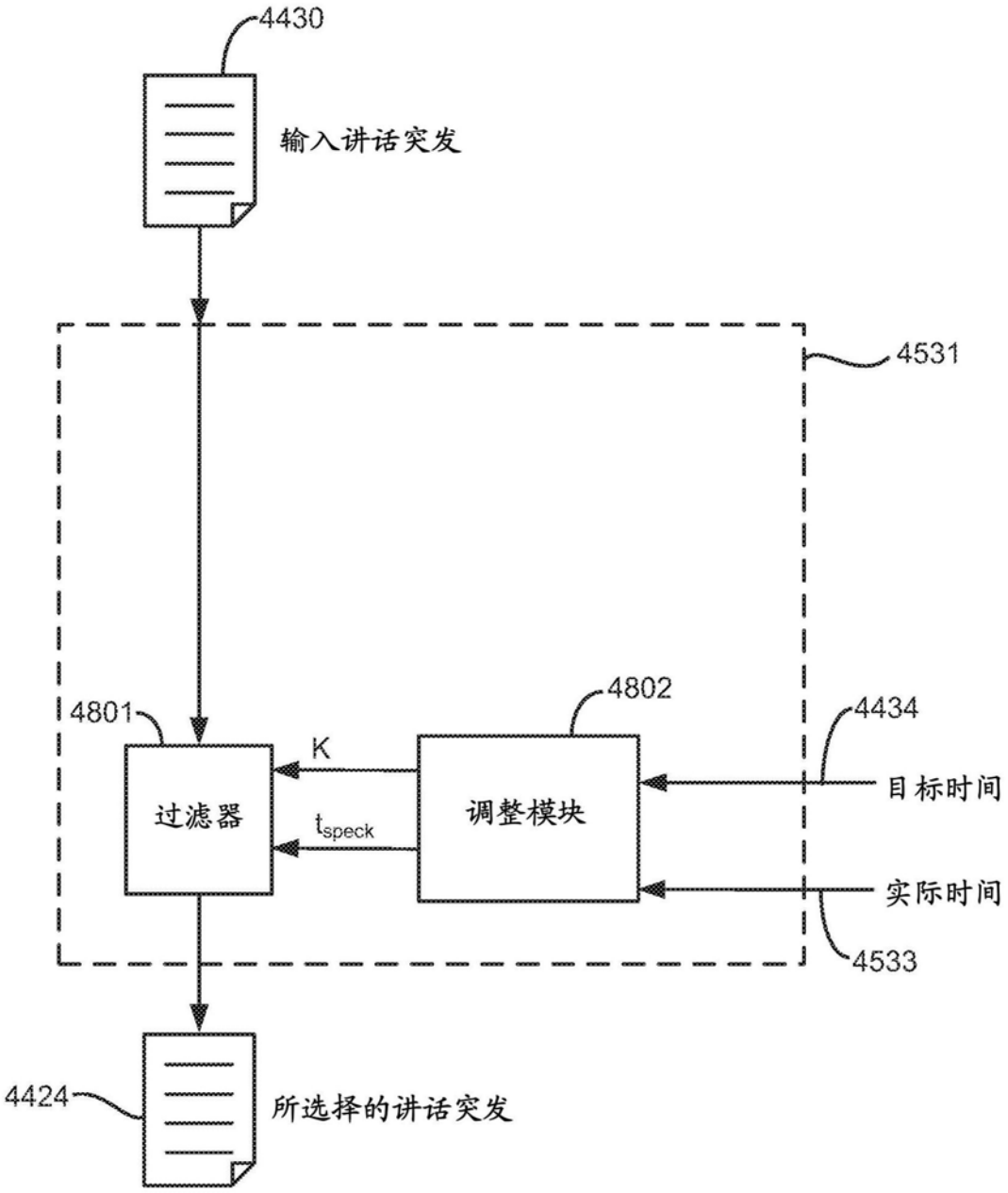

4430
输入讲话突发
4531
4801
4802
4434
K
过滤器
t_{speck}
调整模块
目标时间
实际时间
4533
4424
所选择的讲话突发

图48A

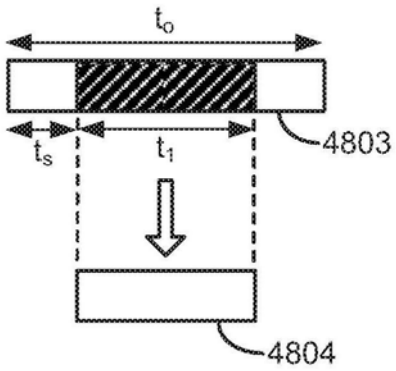


图48B

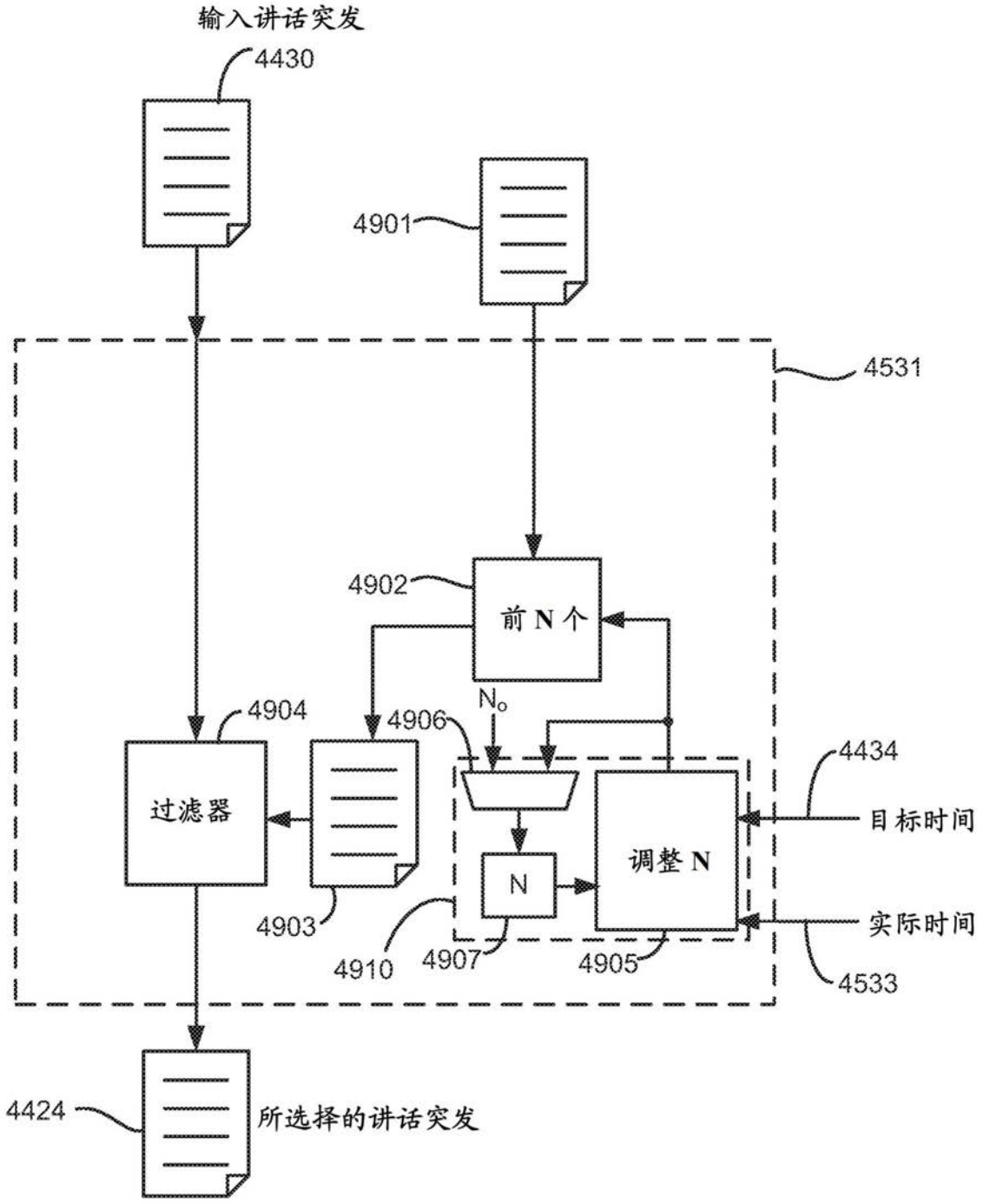

输入讲话突发
4430
4901
4531
4902
前 N 个
N_o
4904
4906
过滤器
4434
目标时间
N
调整 N
4903
4910
4907
4905
实际时间
4533
4424
所选择的讲话突发

图49