



(12) 发明专利申请

(10) 申请公布号 CN 103597543 A

(43) 申请公布日 2014. 02. 19

(21) 申请号 201280016075. 8

代理人 余刚 吴孟秋

(22) 申请日 2012. 01. 11

(51) Int. Cl.

(30) 优先权数据

11153211. 5 2011. 02. 03 EP

G10L 15/26 (2006. 01)

H04H 60/04 (2006. 01)

G10L 15/22 (2006. 01)

G06F 9/44 (2006. 01)

(85) PCT国际申请进入国家阶段日

2013. 09. 27

(86) PCT国际申请的申请数据

PCT/EP2012/050365 2012. 01. 11

(87) PCT国际申请的公布数据

W02012/104119 EN 2012. 08. 09

(71) 申请人 弗兰霍菲尔运输应用研究公司

地址 德国慕尼黑

(72) 发明人 克里斯蒂安·乌勒 于尔根·赫莱

哈拉尔德·波普 法尔科·里德鲁施

(74) 专利代理机构 北京康信知识产权代理有限

责任公司 11240

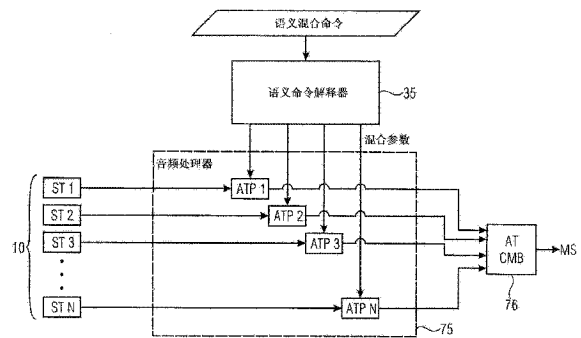
权利要求书2页 说明书23页 附图14页

(54) 发明名称

语义音轨混合器

(57) 摘要

一种用于将多个音轨混合成混合信号的混音器,包括:语义命令解释器(30、35),用以接收语义混合命令并且从该语义混合命令导出针对该多个音轨的多个混合参数;音轨处理器(70、75),用以根据该多个混合参数处理该多个音轨;以及音轨组合器(76),用以将由该音轨处理器处理的该多个音轨组合成该混合信号(MS)。一种相应的方法包括以下步骤:接收语义混合命令;从该语义混合命令导出针对该多个音轨的多个混合参数;根据该多个混合参数处理该多个音轨;以及组合由对多个音轨的处理产生的该多个音轨,以形成该混合信号。



1. 一种用于将多个音轨混合成混合信号(MS)的混音器,该混音器包括:
语义命令解释器(30、35),用于接收语义混合命令,并且从所述语义混合命令导出针对所述多个音轨的多个混合参数;
音轨处理器(70、75),用于根据所述多个混合参数处理所述多个音轨;以及
音轨组合器(76),用于将由所述音轨处理器处理的所述多个音轨组合成所述混合信号(MS)。
2. 根据权利要求1所述的混音器,其中,所述语义命令解释器(30、35)包括用于识别所述语义混合命令内的语义表达的词汇数据库(31)。
3. 根据权利要求1或2所述的混音器,还包括:音轨识别器(40、430),用于识别所述多个音轨中的目标音轨,所述目标音轨在所述语义混合命令中由音轨识别表达来指示。
4. 根据权利要求3所述的混音器,其中,所述音轨识别器(40、430)被配置为:
从音轨模板数据库(432)获取与所述音轨识别表达对应的数据记录,
执行对所述多个音轨之中至少一个音轨的音轨名称、音轨识别符、音色、节奏结构、频率范围、声音样本以及和声密度中的至少一项的分析,
将所述分析的结果与所述数据记录比较,从而产生至少一个匹配分数,以及
基于所述至少一个音轨与所述数据记录之间的所述至少一个匹配分数,来确定所述目标音轨。
5. 根据任一前述权利要求所述的混音器,还包括:时间段识别器(40、460),用于识别所述多个音轨内的目标时间段,所述目标时间段在所述语义混合命令中由时间段识别表达来指示。
6. 根据权利要求5所述的混音器,其中,所述时间段识别器(40、460)被配置为将所述多个音轨结构化多个时间段。
7. 根据权利要求5或6所述的混音器,其中,所述时间段识别器(40、460)被配置为执行对所述多个音轨的分析,以确定至少一个时刻,在该至少一个时刻,由所述多个音轨表示的音频信号的特征性质发生改变,并且将所述至少一个所确定的时刻用作两个邻近时间段之间的至少一个边界。
8. 根据任一前述权利要求所述的混音器,还包括:元数据(42、480)接口,用于接收关于所述多个音轨的元数据(12),所述元数据(12)指示音轨或该音轨的一部分的音轨名称、音轨识别符、时间结构信息、强度信息、空间属性、音色特性以及节奏特性中的至少一个。
9. 根据任一前述权利要求所述的混音器,还包括用于以语言格式接收所述语义混合命令的命令接口。
10. 根据任一前述权利要求所述的混音器,还包括:
示例接口(23、490),用于根据用户关于如何混合示例性混合信号的偏好,接收另一混合信号作为该示例性混合信号;以及
混合信号分析器(492),用于分析所述示例性混合信号,并且基于对所述示例性混合信号的分析产生所述语义混合命令。
11. 根据任一前述权利要求所述的混音器,其中,所述语义命令解释器(30、35)包括知觉处理器(63),该知觉处理器用于根据所述混合信号的听觉相关性质的知觉模型,将所述语义混合命令变换成所述多个混合参数。

12. 根据任一前述权利要求所述的混音器,其中,所述语义命令解释器(30、35)包括模糊逻辑处理器,该模糊逻辑处理器用于接收由所述语义命令解释器从所述语义混合命令导出的至少一个模糊规则,并且基于所述至少一个模糊规则产生所述多个混合参数。

13. 根据权利要求12所述的混音器,其中,所述模糊逻辑处理器被配置为接收由所述语义命令解释器准备的至少两个并行模糊规则,并且其中,所述混音器还包括用于在所述至少两个并行模糊规则中选择一个并行模糊规则的随机选择器。

14. 一种用于将多个音轨混合成混合信号的方法,该方法包括以下步骤:

接收语义混合命令;

从所述语义混合命令导出针对所述多个音轨的多个混合参数;

根据所述多个混合参数处理所述多个音轨;以及

组合由对所述多个音轨的处理产生的多个音轨,以形成所述混合信号。

15. 一种用于指示计算机执行根据权利要求14所述的方法的计算机程序。

16. 一种用于将多个音轨混合成混合信号(MS)的混音器,所述混音器包括:

语义命令解释器(30、35),用于接收语义混合命令,并且从所述语义混合命令导出针对所述多个音轨的多个混合参数;

音轨处理器(70、75),用于根据所述多个混合参数处理所述多个音轨;

音轨组合器(76),用于将由所述音轨处理器处理的所述多个音轨组合成所述混合信号(MS);以及

音轨识别器(40、430),用于识别所述多个音轨中的目标音轨,所述目标音轨在所述语义混合命令中由音轨识别表达来指示,所述音轨识别器被配置为分析所述音轨以确定大致匹配所述目标音轨的一个音轨或多个音轨。

17. 一种用于将多个音轨混合成混合信号的方法,所述方法包括:

接收语义混合命令;

从所述语义混合命令导出针对所述多个音轨的多个混合参数,所述多个混合参数包括针对目标音轨的混合参数;

通过分析所述音轨来识别所述多个音轨中的所述目标音轨以确定大致匹配所述目标音轨的一个音轨或多个音轨;

根据所述多个混合参数处理所述多个音轨;以及

组合由对所述多个音轨的处理产生的多个音轨,以形成所述混合信号。

18. 一种用于指示计算机执行根据权利要求17所述的方法的计算机程序。

语义音轨混合器

技术领域

[0001] 本发明的领域涉及用于根据用户规定混合多轨信号的混音器。本发明的领域涉及音频信号处理,特别是涉及根据用户定义标准的集合,混合多轨录音的任务。本发明的领域还涉及用于将多个音轨混合成混合信号的方法。本发明的领域还涉及用于指示计算机执行混合多个音轨的方法的计算机程序。

背景技术

[0002] 多媒体内容不断增长的可用性产生用户欣赏音乐并且与音乐交互的新的方式。这些可能性面临着开发用于在这样的活动中辅助使用者的工具的挑战。

[0003] 从信息获取的视角看,十多年前已应对该挑战,从而产生音乐信息获取的活跃的研究领域及众多商业应用。

[0004] 未解决到此程度的不同方面是与可以多轨格式获得的内容的交互。多轨格式可由每个声音对象(SO)或对象的群组(轨组;stem)的分离的并时间对准的信号(亦称为单音轨(ST))组成。根据一种定义,轨组为混合中出于再混合使用的目的而单独保存(通常保存至盘或磁带)的个别成分。

[0005] 在音乐产生的传统程序中,多个单音轨以复杂的方式组合成混合信号(mixture signal;MS),随后将该MS传递至最终使用者。数字音频技术的持续进化,例如,基于参数对象的音频的新的音频格式的开发在更大程度上实现与音乐的交互。使用者存取多轨录音并且可主动地控制混合程序。一些艺术家开始释放该艺术家的歌曲中的一些歌曲的轨组,为了使听众可以任何所要方式自由地再混合及再使用音乐。

[0006] 可以诸多方式使用以多轨格式发布的音乐或音频作品。使用者可控制不同轨的混合参数,因此强调选定的轨而衰减其他轨。例如出于伴唱机或同步演奏的目的,一个或多个轨可为无声的。诸如回音、混响、失真、合声等的音效可应用于选定轨而不影响其他轨。可从多轨格式摘录一个或多个轨,并且该所摘录的一个或多个轨可用于另一音乐作品或另一形式的音频作品中,诸如,音频书籍、演讲、播客等。在以下描述中,本文所公开的示教的应用以示例性方式论述经录音的音乐作品的主控。然而,应理解,本文所公开的示教旨在同等地解决并且覆盖涉及混合多个单音轨的任何经录音声音的处理。

[0007] 自动混合已为并且仍为若干研究计划的焦点。在2009年,Perez-Gonzalez等人描述多轨信号的自动均衡的方法(E.Perez-Gonzalez及J.Reiss,“Automatic Equalization of Multi-Channel Audio Using Cross-Adaptive Methods”,Proc.of the AES127th Conv.,2009年出版)。作者提供用于自动设定多轨信号的每个信号的衰减的方法。确定增益,以使得每个信号的响度等于所有信号的平均响度。相同作者的另一文章阐明“Automatic Gain and Fader Control for Live Mixing”并且出版于2009年的Proc.of WASPAA中。

[0008] 语义高保真度(HiFi)为欧洲计划IST-507913的名称(H.Vinet等人,“Semantic HiFi Final Report”,IST-507913的最终报告,2006年出版)。该报告主要涉及多媒体内容

的获取、浏览及共享。此报告包括数据库中的浏览及导览、播放表产生、轨内导览(使用如独唱与合唱识别的结构分析)及元数据共享。该报告亦解决交互/创作/编辑:产生包括同步(该同步为“序连(concatenating)”音频信号,而非混合多轨信号)、语音变换、节奏变换、语音控制乐器及音效是混合。

[0009] 另一计划被称为“结构音频”或 MPEG4 的名称。结构音频实现以低比特率并且在知觉上基于使用信号的符号及语义描述的声音数据的操作及存取,来传输音频信号(cf. B. L. Vercoe 及 W. G. Gardner 及 E. D. Scheirer, “Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations”, Proc. of IEEE, 第 86 卷, 第 922-940 页, 1998 年出版)。该计划的特征为描述了用于混合多个流及添加音效的参数声音后产生。参数描述确定如何合成声音。结构音频涉及合成音频信号。

[0010] 在国际公开号为 W02010/111373A1 的国际专利申请中,公开了一种背景获知、语音控制的接口和系统。语音控制的使用者接口系统包括:至少一个扬声器,用于将音频信号传送给用户;以及至少一个传声器,用于捕捉使用者的语音发言。一种接口装置接口连接扬声器和传声器,并提供多个音频信号至扬声器以被使用者听到。控制电路可操作地耦接至接口装置并被配置为选择多个音频信号中的至少一个作为通过传声器传送给使用者的前景音频信号。控制电路可操作于识别使用者的语音发言并使用所识别的语音发言控制前景音频信号的选择。

[0011] 美国专利申请公开号 US2002/0087310A1 公开了一种计算机执行的方法和系统,用于处理与使用者的语音对话。来自使用者的语音输入包括指向多种概念的词语。使用者语音输入包括进行服务的请求。使用者语音输入的语音识别被用于生成识别词语。对话模板被应用到所识别的词语。对话模板具有与预定概念相关的节点。节点包括不同的请求处理信息。在对话模板内基于哪个节点相关于最接近匹配所识别词语的概念的概念来识别概念区域。通过使用包括在所识别概念区域内的节点的请求处理信息来处理使用者的请求。

[0012] 文章“Transient Detection of Audio Signals Based on an Adaptive Comb Filter in the Frequency Domain”, M. Kwong 和 R. Lefebvre 提出适于音乐信号中的节奏检测的瞬态检测算法。在多个音频信号中,低能量瞬态被高能量静态音遮掩。这些被遮掩的瞬态以及更高能量和更明显的瞬态传达了关于音乐信号中的节奏和时间分段的重要信息。所提出的分段算法使用正弦模型结合频域上的适应性梳齿滤波来移除声音信号的静态分量。在滤波之后,残余信号的时间包络被分析以定位瞬态分量。结果显示提出的算法可精确地检测最低能量的瞬态。

[0013] 多轨录音的混合通常为常由专家、混合工程师执行的创作任务。如交互的音频格式这样的多媒体的当前发展,产生了多轨录音需要由非专家指导以自动方式或以半自动方式混合的应用。期望自动导出的混合信号具有可与由人类专家产生的混合信号相当的主观声音质量。

发明内容

[0014] 本文所公开的示教针对该总体目标。示教涉及音频信号处理,特别是涉及出于收听的(最终)目的,根据使用者定义的录音标准的集合,混合多轨的任务。根据本文所公开的

示教的混音器及用于将多个音轨混合成混合信号的方法建立非专家的充分美学设想与所得混合信号之间的联系。

[0015] 这些目标中的至少一个目标和 / 或可能的其他目标借助于根据权利要求 1 或权利要求 16 所述的混音器、根据权利要求 14 或权利要求 17 所述的用于混合多个音轨的方法及根据权利要求 15 或权利要求 18 所述的计算机程序来获得。

[0016] 根据本文所公开的示教,一种用于将多个音轨混合成混合信号的混音器包括语义命令解释器、音轨处理器及音轨组合器。语义命令解释器经配置用于接收语义混合命令并从该语义混合命令导出针对该多个音轨的多个混合参数。音轨处理器经配置用于根据该多个混合参数处理该多个音轨。音轨组合器经配置用于将由音轨处理器处理的多个音轨组合成混合信号。

[0017] 根据所公开示教,用于将多个音轨混合成混合信号的方法包括以下步骤:接收语义混合命令;自该语义混合命令导出针对该多个音轨的多个混合参数;根据该多个混合参数处理该多个音轨;以及组合由该多个音轨的处理产生的该多个音轨,以形成混合信号。

[0018] 计算机程序包括或表示用于使计算机或处理器能执行混合多个音轨的方法的指令。计算机程序可在计算机可读取媒体上实施,该计算机可读取媒体上储存有当在计算机上执行时用于执行权利要求第 14 项中叙述的方法的计算机程序。

[0019] 语义混合命令可基于使用者定义标准,该使用者定义标准提供所要获得的混合信号的语义描述。根据本文所公开的示教,语义音频分析、心理声学及音频信号处理可彼此结合,以基于语义描述自动导出混合信号。此过程可被称为“语义混合”。

[0020] 语义混合可视为一种使计算机能根据用户给定的规定混合多轨录音方法,。通常以语义描述的形式给定规定。在给定此语义描述的情况下,可通过考虑一个或多个单音轨的特性及人类听觉,来确定混合参数。

[0021] 根据本文所公开的示教的混音器因此通常包括计算机或处理器,或该混音器与计算机 / 处理器交互。音轨处理器及音轨组合器可组合为单一单元。

[0022] 从语义混合命令导出该多个混合参数可涉及以下步骤:分析语义混合命令或部分该语义混合命令的意义。语义混合命令的一部分可为语义表达,诸如,用语或用语的群组。一个或多个语义表达可随后翻译成用于该多个音轨的特定混合参数的集合。因此,语义混合命令借助于与语义混合命令的意义对应的特定混合参数来实施。对语义混合命令和 / 或该语义混合命令的构成语义表达的翻译动作可包括以下步骤:例如,评估翻译函数或查询查找表。翻译函数的参数或查找表中的数据记录通常被预先定义并且表征(例如)有经验的混合工程师的专业知识的集合。例如,可通过记录由艺术家或音乐生产者给予该艺术家或音乐生产者的混合工程师的口头指令以及由该混合工程师执行的设定,来随时间收集专业知识。因此,翻译函数和 / 或查找表可由专家混合工程师训练。

[0023] 根据本文所公开的示教的一个方面,语义命令解释器可包括用于识别语义混合命令内语义表达的词汇数据库。借助于词汇数据库,语义命令解释器可识别例如同义词。进一步可能的是将语义混合命令中含有的用语或用语的群组映像至特定值。举例而言,可将用于识别乐器的用语(“吉他”)映像至特定信道号码或识别符,该乐器被记录在该特定信道号码或识别符上。词汇数据库可进一步包括识别音乐部分的某一部分的条目,诸如,开始(例如,“导奏”)、合唱(“合唱”)或结束(例如,“尾声”或“尾曲”)。词汇数据库的又一可能

的使用为用于辨识并且分配语义表达的混合参数或风格,诸如,“大声的”、“轻柔的”、“清楚的”“消音的”“远距的”“接近的”等。

[0024] 在本文所公开的示教的实施方式中,混音器可进一步包括用于识别该多个音轨中的目标音轨的音轨识别器。目标音轨可在语义混合命令中由音轨识别表达来指示。若该多个音轨未经清楚地标记或识别关于该音轨含有哪个部分或轨组,则可用到音轨识别器。举例而言,可将音轨简单地编号为“轨 1”、“轨 2”、…“轨 N”。音轨识别器可随后分析该多个音轨中的每个音轨,以确定大致匹配由音轨识别表达所识别的音轨的一个或者若干音轨或无音轨匹配。

[0025] 音轨识别器可配置为从音轨模板数据库获取对应于音轨识别表达的数据记录,音轨识别器可配置为执行该多个音轨之中至少一个音轨的音轨名称、音轨识别符、音色、节奏结构、频率范围、声音样本及和声密度中的至少一个的分析,音轨识别器可配置为将分析的结果与数据记录比较,从而产生至少一个匹配分数,并且音轨识别器可配置为基于至少一个音轨与数据记录之间的至少一个匹配分数,来确定目标音轨。由音轨识别器执行的任务为识别该多个音轨之中的目标音轨。目标音轨对应于音轨识别表达,亦即,若音轨识别表达为“吉他”,则在由音轨识别器成功识别的后,目标音轨应通常含有音乐作品的吉他部分。音轨模板数据库可包括对应于乐器“吉他”的数据记录,该数据记录本身包括吉他的特性的值和 / 或信息。举例而言,数据记录可包括典型的吉他声音的频率模型和 / 或典型的吉他声音的上升 - 衰减模型。数据记录亦可含有吉他的声音样本,吉他的该声音样本可用于音轨识别器的类似分析。

[0026] 根据本文所公开的示教的一个方面,混音器可进一步包括用于识别该多个音轨内目标时间段的时间段识别器,该目标时间段在语义混合命令内由时间段识别表达来指示。若用户希望以不同于音频信号(例如,音乐作品)的第二段的方式混合该相同音频信号的第一段,则混音器通常需要知道音频信号的各个段开始及结束于何处,以将特定混合参数应用于音频信号的这些段。

[0027] 时间段识别器可配置为将该多个音轨构造成多个时间段。尤其音乐作品通常具有受音乐惯例影响的某一结构,例如,具有交替的独唱及合唱段的歌曲形式。通过首先确定由该多个音轨表示的音频信号是否遵循某一音乐结构,之后将音频信号的时间段分配至音乐结构的时间段,这种知识可被时间段识别器来开发。为此,时间段识别器可包括辨识音频信号内重复和 / 或相似模式的模式辨识器。例如,模式辨识可基于旋律分析、和声分析及节奏分析等。

[0028] 时间段识别器可配置为执行该多个音轨的分析,以确定由该多个音轨表示的音频信号的特征性质发生改变的至少一个时刻,并且将该至少一个确定的时刻用作两个邻近时间段之间的至少一个边界。

[0029] 混音器可进一步包括用于接收关于该多个音轨的元数据的元数据接口,该元数据指示音轨或部分该音轨的音轨名称、音轨识别符、时间结构信息、强度信息、空间属性、音色特性及节奏特性中的至少一个。元数据可由该多个音轨的生产者产生,并且元数据提供对混音器或用于混合该多个音轨的方法有用的信息。元数据的可用性使混音器或方法不必执行音频信号广泛分析,以识别各个音轨和 / 或时间段。元数据接口亦可用于储存分析结果(乐器、时间结构、…)供将来再次使用。因此,该多个音轨的可能的冗长分析仅需要执行一

次。此外,亦可储存对于自动确定的分析结果的任何手动修正,以使得使用者不必再三修正相同问题。使用者在具有所储存分析结果后,可使用相同元数据产生与相同多个音轨不同的混合版本。

[0030] 根据本文所公开的示教的一个方面,混音器可进一步包括用于以语言格式接收语义混合命令的命令接口。该语言格式允许用户基本上借助于普通语言来表达该用户关于混音器执行的混合结果的期望。可将具有语言格式的语义混合命令作为口语使用麦克风的或使用(例如)键盘的书面语输入至混音器。

[0031] 根据本文所公开的示教的另一方面,混音器可进一步包括用于接收混合信号的示例接口、以及用于分析该示例性混合信号并且基于分析该示例性混合信号而产生语义混合命令的混合信号分析器。使用经由示例接口提供的示例性混合信号,混合信号分析器可确定哪些特征表征示例性混合信号。举例而言,混合信号分析器可辨识对于较少强调旋律的(快速重复的)鼓声部分及低音部分的强调。这些侦测到的特征暗示所谓的舞蹈混合(Dance-Mix),亦即,混合的某一风格。可将此信息从混合信号分析器提供至语义命令解释器。基于此信息,例如,语义命令解释器可增加鼓声部分及低音部分相对于其他部分的音量。语义命令解释器甚至能够以(例如)通常用于所要舞蹈混合风格的合成鼓声部分来替代鼓声部分。

[0032] 示例接口可进一步配置为接收从中获取示例性混合信号的多个示例音轨。混合信号分析器可配置为将示例音轨与示例性混合信号比较,以确定之前用于获得所得示例性混合信号的混合参数。则由混合信号分析器产生的语义混合命令可包括在示例音轨混合在一起形成示例性混合信号之前该示例音轨被如何修改的描述。举例而言,语义混合命令可包括,诸如“鼓声显着较大;声乐适度较轻柔、较远,用高通滤波器过滤”的表达。语义命令解释器可随后从该语义混合命令导出该多个混合参数。

[0033] 根据本文所公开的示教的另一方面,语义命令解释器可包括知觉处理器,该知觉处理器用于根据混合信号的听觉相关的性质的知觉模型,将语义混合命令变换成多个混合参数。知觉模型通常实施心理声学规则,该心理声学规则描述应如何选择某些混合参数,以实现听众的所要效果。举例而言,为传达距离的印象,可涉及若干声音处理动作,诸如,混响、频率滤波及衰减。通常基于心理声学发现的知觉模型辅助确定用于实现所要效果的适合的混合参数。

[0034] 根据本文所公开的示教的另一方面,语义命令解释器包括模糊逻辑处理器,用于接收由语义命令解释器从语义混合命令导出的至少一个模糊规则并且基于该至少一个模糊规则而产生多个混合参数。模糊逻辑处理器较好地适合于处理处于至少一个模糊规则形式的语义混合命令。该至少一个模糊规则在基本语义域中将模糊逻辑处理器的输入量映像至模糊逻辑处理器的输出量,亦即,从第一语义格式的量映像至第二语义格式的量。

[0035] 模糊逻辑处理器可配置为接收由语义命令解释器准备的至少两个并行模糊规则,并且其中,混音器进一步包括随机选择器,用于在该至少两个并行模糊规则之中选择一个并行模糊规则。通过使从两个或两个以上并行模糊规则中的模糊规则的选择随机化,可产生人工自由的错觉,以使得由混音器产生的混合信号不会倾向于发出对于混合风格而言基本上类似的声音,否则就会出现混音器遵循关于模糊规则的更严格方案的情况。

[0036] 关于用于混合多个音轨的方法,可查询用于识别语义混合命令中语义表达的词汇

数据库。

[0037] 该方法可进一步或可选地包括以下步骤：识别该多个音轨之中的目标音轨，该目标音轨在语义混合命令中由音轨识别表达来指示。为此，可从音轨模板数据库获取对应于该音轨识别表达的数据记录。随后，可执行该多个音轨之中至少一个音轨的音轨名称、音轨识别符、音色、节奏结构、频率范围、声音样本及和声密度中的至少一个的分析。可将分析的结果与数据记录比较，从而产生至少一个匹配分数。随后，可基于至少一个音轨与数据记录之间的至少一个匹配分数，来确定目标音轨。

[0038] 该方法亦可包括用于识别该多个音轨内目标时间段的动作，该目标时间段在语义混合命令内由时间段识别表达来指示。用于识别目标时间段的动作可配置为将该多个音轨构造成多个时间段。时间段识别可包括以下步骤：执行该多个音轨的分析，以确定至少一个时刻，在该至少一个时刻，由该多个音轨表示的音频信号的特征性质发生改变，并且将该至少一个所确定的时刻用作两个邻近时间段之间的至少一个边界。

[0039] 根据本文所公开的示教的另一方面，该方法可进一步包括以下步骤：在元数据接口处接收关于该多个音轨的元数据。元数据可指示音轨或部分该音轨的音轨名称、音轨识别符、时间结构信息、强度信息、空间属性、音色特性及节奏特性中的至少一个。

[0040] 该方法可进一步包括以下步骤：在相应混音器的命令接口处接收语言格式的语义混合命令。

[0041] 根据本文所公开的示教的另一方面，该方法可进一步包括以下步骤：在示例接口处接收示例性混合信号；借助于混合信号分析器分析该示例性混合信号；以及基于该示例性混合信号的分析，而产生语义混合命令。

[0042] 从语义混合命令导出该多个混合参数的动作可包括以下步骤：根据混合信号的听觉相关的性质的知觉模型，将语义混合命令变换成多个混合参数。

[0043] 根据本文所公开的示教的一个方面，导出该多个混合参数的动作可包括以下步骤：接收通过语义命令解释器从语义混合命令导出的至少一个模糊规则，并且基于该至少一个模糊规则而产生该多个混合参数。可通过模糊逻辑处理器来执行接收至少一个模糊规则及基于该至少一个模糊规则而产生该多个混合参数。

[0044] 该方法可进一步包括以下步骤：接收至少两个并行模糊规则，并且在至少两个并行模糊规则之中随机选择一个并行模糊规则。

[0045] 本文所公开的示教在以下诸点中不同于以上提及的先前技术：

[0046] -Perez-Gonzalez 等人建议的方法不考虑语义描述来控制处理。

[0047] -语义高保真度计划未解决多轨格式的处理。该语义高保真度计划未解决根据语义描述混合信号的问题。该语义高保真度计划未解决计算实现语义描述的混合信号所需要的知觉方面。

[0048] -“结构化音频”计划关于合成音频信号。相比之下，本文所公开的示教(语义混合)关于混合音频信号。

[0049] 为简要地概述本文所公开的示教的核心方面中的一些核心方面，混合多轨录音为创作任务。语义混合的目的在于开发用于混合通过语义描述引导的多轨录音的解决方案。语义混合将语义音频分析、心理声学及音频信号处理的技术组合。语义混合适用于各种应用，如音乐创作、空间对象音频编码(Spatial Object Audio Coding; SAOC)、家庭视频创

作、虚拟现实及游戏。

[0050] 可用以下(部分为可选的)特征简短描述语义混合：

[0051] - 语义混合为使用者交互提供手段。

[0052] - 语义混合在较大程度上解决知觉成分。这亦可包括对环境的适应、播放系统及用户偏好。

[0053] - 语义混合组合语义部分及心理声学部分。任何语义处理需要考虑知觉方面。语义混合聚焦于音频信号处理而非聚焦于语义分析的传统应用(音乐信息获取、播放表产生)。语义混合的目的在于提供与内容交互的新的方式。

[0054] - 语义混合关于多轨录音的处理。

[0055] 本文所公开的示教、尤其关于一种用于根据用户规定混合多轨信号的方法。本文所公开的示教关于音频信号处理,特定言之关于根据用户定义标准的集合,混合多轨录音的任务。用户定义标准提供所得混合信号的语义描述。本文所公开的示教可结合音频分析、心理声学及音频信号处理,以基于语义描述自动导出混合信号。

附图说明

[0056] 本文所公开的示教的以上特征及其他特征将通过以下说明显而易见,参照所附的示意性图式仅以举例的方式进行说明,其中：

[0057] 图 1 示出混音器的示意性方块图；

[0058] 图 2 示出通常在流行音乐中采用的歌曲结构中音乐作品的示例性时间结构；

[0059] 图 3 示出古典音乐中已知的奏鸣曲形式的音乐作品的另一示例性时间结构；

[0060] 图 4 示出流行音乐录音的示例音轨布局；

[0061] 图 5 示出根据本文所公开示教的混音器的示意性方块图；

[0062] 图 6 示出模糊逻辑处理器的示意性方块图；

[0063] 图 7 示出模糊集合的示例性隶属函数；

[0064] 图 8 示出包括模糊逻辑处理器的混音器的示意性方块图；

[0065] 图 9 示出根据本文所公开示教的混音器的另一配置的示意性方块图；

[0066] 图 10 示出根据本文所公开示教的一个方面的语义混合命令及该语义混合命令的分解；

[0067] 图 11 示出根据本文所公开示教的一个方面的另一语义混合命令及该语义混合命令的分解；

[0068] 图 12 示出根据本文所公开示教的混音器的另一配置的示意性方块图；

[0069] 图 13 示出根据本文所公开示教的包括音轨识别器的混音器的配置的示意性方块图；

[0070] 图 14 示出根据本文所公开示教的包括时间段识别器的混音器的配置的示意性方块图；

[0071] 图 15 示出根据本文所公开示教的包括元数据接口的混音器的配置的示意性方块图；

[0072] 图 16 示出根据本文所公开示教的包括用于接收示例性混合信号的示例接口的混音器的配置的示意性方块图；

[0073] 图 17 示出根据本文所公开示教的包括知觉处理器及知觉模型的混音器的配置的示意性方块图 ;以及

[0074] 图 18 示出根据本文所公开示教的用于将多个音轨混合成混合信号的方法的示意流程图。

具体实施方式

[0075] 图 1 图示混音器的示意性方块图。混音器允许组合多个单音轨 ST,以使得形成混合信号 MS。为了控制单音轨 ST 的组合,通常将每个单音轨馈送至个别信号处理器。一个单音轨的个别信号处理器可包括(例如)均衡器 EQ、平移控制 PAN、混响器 REVERB、音量控制 VOL 及可能进一步的音效。混音器的中心作用是调整该多个单音轨 ST 中的每个单音轨 ST 的音量,以使得混合信号是由该多个单音轨 ST 提供的音频信号的良好平衡的迭加。通常由混合工程师作出决策:音效及单音轨 ST 的音量的哪种特定设定构成良好平衡的迭加。多个个别信号处理器修改多个音轨信号。经修改的音轨信号随后在信号组合器 Σ 中组合,以产生混合信号 MS。

[0076] 图 2 示出属于流行音乐类型的典型歌曲的时间结构。图 2 内用作示例的歌曲以导奏(INTRO)开始,继之以独唱段(VERSE1)、合唱(CHORUS)、具有基本上相同音乐但不同歌词的第二独唱(VERSE2)段、合唱的重复、过渡(BRIDGE)、合唱的另一重复及尾声或尾曲(OUTRO)。尽管存在此基本方案的众多变化,但对于大多数人而言区分流行音乐歌曲的各个区段通常是有可能的。举例而言,通常在整个歌曲中的不同位置上具有基本上相同歌词及旋律地重复合唱,使得该合唱可轻易地由听众辨识。

[0077] 图 3 示出以奏鸣曲形式组合的音乐作品的时间结构。奏鸣曲形式已由众多古典音乐的作曲家使用。如名称所展示,奏鸣曲形式广泛用于奏鸣曲中,通常用于该奏鸣曲的第一乐章中。许多交响乐的第一乐章亦通常呈奏鸣曲形式。奏鸣曲形式的特性段是呈现部分、发展部分及再现部分,在该再现部分中基本上相同的音乐素材呈现有各种修改,尤其相对于和弦进行有各种修改。可选地,导奏及尾声可分别存在于音乐作品的开始及结束处。尽管区分奏鸣曲形式的各个时间段可能需要一些经验,但这对于人类听众而言一般是可能的。

[0078] 混合工程师可能想要以不同方式处理音乐作品的不同时间段。原因可能为实现某一艺术效果的期望,或通过补偿可能发生在该多个音轨的录音期间的潜在缺陷来使混合信号 MS 声音更均匀的期望。关于音乐作品或一般录音(例如音频书籍、演讲等)的时间结构的知识可辅助混合工程师找到录音中各个时间段的起始点及终点。

[0079] 图 4 示出流行音乐类型中歌曲的录音的示例音轨布局。单音轨 ST 存在于以下乐器中:主奏吉他、节奏吉他、声乐部分、钢琴及低音。利用用于鼓组的各个部分的若干单音轨对鼓组录音:强音钹、打点钹、高帽钹、高音鼓、小军鼓及低音鼓。使用用于鼓组的不同部分的若干音轨 ST 允许混合工程师比当仅将单音轨用于整个鼓组时更佳地平衡鼓声部分。

[0080] 利用所具有的该多个单音轨,必要时可由表演艺术家和 / 或混合工程师主控音乐作品。特定言之,通过改变用于该多个音轨 ST 的混合参数,可明显改变音乐作品的特性或“情绪”。为混合的消费者提供多个音轨给该消费者提供较大的自由度。然而,特定言之由于混合参数之间复杂的互连及交互,所以许多使用者缺乏适当地选择混合参数的经验。为了实现似乎影响单音轨的某一效果,可能有必要调整若干乃至所有音轨的混合参数。

[0081] 图 5 示出根据本文所公开示教的具有第一可能配置的混音器的示意性方块图。

[0082] 通常,用户(或听众)具有混合信号应如何发出声音的某一想法,但使用者(或听众)不知道应如何调整混合参数来实现此想法。

[0083] 根据本文所公开的示教的混音器建立以简明形式描述使用者的想法或期望的语义表达与将多个单音轨 ST 混合成混合信号 MS 所需要的实际混合参数之间的联系。

[0084] 导引混合程序的语义描述的简单但说明性的示例为以下:“在吉他独奏期间,突出混合该吉他并且略微地移动键盘至背景中”。

[0085] 为完成该目的,通常需要解决下文列出的各个子任务中的至少一些子任务:

[0086] - 需要使用适当的用户接口获取由用户给定的语义描述。

[0087] - 需要将使用者输入翻译成机器可读取形式。

[0088] - 需要执行对音乐音频信号的语义分析(例如,识别吉他轨及键盘轨,找到吉他独奏的开始及结束)。

[0089] - 需要通过考虑人类听觉机制来确定物理混合参数(例如,确定增益,以使得混合中吉他的知觉的响度比任何其他乐器更大,但不能太大声;对于键盘,确定增益、延迟并发送音轨至所期望距离知觉的混响效果的增益)。

[0090] - 使用计算出的物理混合参数导出混合。参数包括单音轨 ST 及输出信道的每个组合的增益因子及时间延迟。此外,物理混合参数控制数字音频效果处理器(DAFx),例如,人工混响及动态范围处理。语义描述可(例如)指定

[0091] - 混合信号 MS 中每个声音对象 SO 的知觉位置及响度

[0092] - 每个轨的 DAFx 的参数

[0093] - 混合信号 MS 的特性(例如,混响的量、动态范围)。

[0094] 在示出根据本文所公开的示教的混音器的可能配置的图 5 的示意性方块图中,以上提及的子任务通过混音器的模块来完成。混音器包括用户接口(UI)20、命令解释器(CI)30、语义音频分析(SAA)40、目标描述符分配单元(DAU)50、知觉处理器 60 及信号处理器 70。

[0095] 用户接口 20 提供用于从混音器的使用者获取输入的设施。如作为用户接口 20 的部分的多个子模块所示,存在用户输入的实施的不同的选择。示例为:

[0096] - 默认的集合(子模块 22)中的一个预设的选择;

[0097] - 一组 n 维控制器(子模块 21),将该 n 维控制器分配至单音轨及所得混合信号 MS 的不同特性;

[0098] - 自然语言输入(子模块 24);

[0099] - 混合信号 MS 的示例或与相应混合信号 MS 一起的多轨的示例的输入(子模块 23)。随后将分析给定的示例,以导出混合信号 MS 的语义描述。通过这种使用者输入控制的混音器的操作模式将在后续描述中称为“通过示例混合”。

[0100] 命令解释器 30 连接至用户接口 20,并且命令解释器 30 将输入(该输入为人类可读的或通过示例给定)翻译成机器可读取命令。这些命令通常具有限制性的词汇及已知的文法,该已知的文法通过词汇文法子模块 31 提供和/或控制。

[0101] 在表 1 中针对不同用户接口设计给出命令解释器 30 的实施示例。

[0102]

输入	命令解释器 30 的实施
预设	使用查找表选择规则的集合
n 维控制器	将控制器输入的函数映像至命令
通过示例混合	音频信号的分析
自然语言	语音识别及理解

[0103] 除使用者输入外,混音器亦接收包括单音轨 ST 的数据 10 作为输入。若数据 10 对应于音乐作品的音轨,则数据 10 可包括音乐容器(container)11 及可选的元数据容器 12。可经由适合的接口(未示出于图 5 中)向混音器提供数据 10。

[0104] 将数据 10 馈送至语义音频分析(SAA)40。语义音频分析 40 通常为自动程序,该自动程序计算音轨 ST 中的每个音轨 ST 的元数据的集合。此外,可计算描述多轨(亦即,该多个音轨)的元数据(例如,音乐类型)。元数据为表征音频信号的语义描述符。

[0105] 语义音频分析 40 可包括:

[0106] - 乐器识别

[0107] - 结构分析(独唱、合唱及每个信号的其他部分的标示)

[0108] - 演奏风格(独奏、伴奏、旋律的、和声及节奏一致性)的识别

[0109] - 节奏分析(例如,用于拍子同步音效的拍子追踪)

[0110] - 旋律及和声分析

[0111] - 音色的特性(例如,亮度、粗度、锐度)

[0112] - 单音轨 ST 之间的相似性(相对于音色、演奏风格、形式)的特性

[0113] - 音乐类型

[0114] 这些元数据可用于经由混合参数向单音轨 ST 中的每个单音轨 ST 分配适当的信号处理。

[0115] 每次执行语义混合的程序时,可执行语义音频分析。可选地,可执行一次语义音频分析(在产生/创作期间),并且所得元数据可与多媒体项目一起储存并且传输。

[0116] 可选地,语义音频分析 40 可通过使用者输入来引导,亦即,若用户不满意语义音频分析的自动导出的结果中的一个或多个结果,则该用户可辅助语义音频分析 40 或该用户可输入元数据。可通过语义音频分析储存这些修正的使用者输入,以供在未来分析期间考虑,以使得语义音频分析 40 可适应于使用者的偏好,亦即,借助于使用者输入随时间训练语义音频分析 40。

[0117] 语义音频分析 40 可包括第一子模块 41,用于基于多个音轨 ST 中含有的音频信号来计算元数据。另外或可选地,语义音频分析 40 可包括第二子模块 42,用于读取与该多个音轨 ST 一起提供的元数据。

[0118] 连接命令解释器 30 与语义音频分析 40 的是目标描述符分配单元(DAU)50。在给定来自命令解释器 30 的命令以及从语义音频分析 40 获取的元数据的情况下,目标描述符分配单元 50 选择音频信号的一部分(该部分确定对应于对其存在命令的声音对象的音轨及起始时间和停止时间)并且将适当的知觉目标描述符(PTD)分配至该部分。

[0119] 知觉目标描述符可描述：

[0120] - 声音对象的知觉强度(响度)

[0121] - 声音对象的空间属性(侧角、高度、距离、扩散度、宽度)

[0122] - 声音对象的音色特性(例如,亮度、锐度、粗度)

[0123] - 关于数字音频效果(DAFx)的特性

[0124] 若使用语言变量给定命令,则目标描述符分配单元 50 可使用将语言学变量转换成明确值(crisp value)的模糊逻辑。

[0125] 提供知觉目标描述符的目标描述符分配单元 50 的输出连接至知觉处理器(PP)60 的输入。在给定分配的知觉目标描述符的情况下,知觉处理器 60 计算用于混合及额外的信号处理(例如,DAFx)的物理参数(混合参数)。此举通常为需要考虑心理声学 61 及专业知识 62 的高要求任务。

[0126] 使用以下示例说明此举:针对特定音频信号,例如,吉他音轨,将知觉水平的描述符设定成“高”。简单的解决方案为将吉他增益增加固定量,例如,6dB。由于响度的知觉以高度复杂的方式取决于经处理音频信号及混合信号 MS 的频谱及时间特性,故此简单的解决方案可能并非在所有情况下具有所要效果。

[0127] 可通过设定所有水平,以使得例如通过使用响度及泛音(partial)响度的知觉模型来调整混合中由听众感知的吉他的响度,来获取更好的结果。泛音响度是存在于声音混合物中的信号的响度,其中所关注信号被部分遮掩。

[0128] 除响度的知觉的外,通常需要解决人类听觉及声音知觉的不同方面。这些方面为混响的量、声音定位的知觉及空间属性的知觉。

[0129] 心理声学部分对将语义描述(例如,“使此声音略微更大声”)翻译成物理参数(例如,“升高 4.5dB”)较为重要。

[0130] 知觉处理器 60 经由该知觉处理器 60 的输出中的一个输出连接至信号处理器(SP)70 的输入。信号处理器 70 可包括处理物理混合参数的模块 71、一个或多个数字音频效果的模块 72 及格式化模块 73。信号处理器 70 使用用于混合及信号处理的物理参数,计算混合信号 MS。

[0131] 在会议论文“Automatic Music Production System Employing Probabilistic Expert Systems”(Audio Engineering Society,在第 129 次会议提供,2010 年 11 月 4 日—7 日)中,作者 R. Gang 等人建议采用机率性的图形模型嵌入专业的音频工程知识并且基于自音频档案获取的音乐信息来推断自动化生产决策。可从人类音频工程师的操作数据学习或根据领域知识人工构建表示为机率性图形模型的生产模式。知觉处理器 60 和 / 或语义命令解释器 30 可实施此会议论文中建议的技术特征。上文会议论文的全部内容以引用的方式并入本文。

[0132] 混合多轨录音包括以下步骤：

[0133] - 每个单音轨的水平及平移(panning)位置的调整(处理物理混合参数 71 的模块)

[0134] - 均衡(针对单音轨 ST 及混合信号 MS)

[0135] - 动态范围处理(DRP)(针对单音轨 ST 及混合信号 MS)

[0136] - 人工混响

[0137] - 应用音效(DAFx72)

[0138] 这些操作中的每个操作由知觉处理器 60 计算的物理参数来控制。

[0139] 可选地需要格式化 73 来应对实体限制(例如,应用自动增益控制)及格式转换(音频编码/译码)。

[0140] 以下部分详述处理方块中的每个处理块的示例性实施。

[0141] 用户接口 20 可实施为默认的集合。每个预设表示具有一组特性的“混合类型”。这些特性可以“混合规则”的形式给定为语义表达,并且在以下命令解释器 30 的描述的背景中描述这些特性。

[0142] 例如,混合类型可为“舞蹈混合”、“环境混合”、“摇滚吉他混合”等。

[0143] 这些名称以高度简洁的方式给出目标混合信号 MS 的描述,然而使用者可解译该名称(或该名称的子集)。使用者解译预设的名称的能力是基于惯例及广泛使用的风格分类。举例而言,使用者可使特定演奏风格和/或声音与某一艺术家的名称相关联。

[0144] 在命令解释器 30 的背景下,使用查找表将混合规则的集合分配至默认中的每个预设。以如此处所示的模糊逻辑(J. M. Mendel, “Fuzzy Logic Systems for Engineering: A Tutorial”, Proc. of IEEE, 第 83 卷, 第 345-377 页, 1995 年出版)中的 IF-THEN 声明的形式将混合规则描绘为逻辑暗示,其中声音对象描述符 <SOD> 为前提,并且混合操作描述符 <MOD> 为后续:

[0145] IF <SOD> THEN <MOD>

[0146] 该 IF-THEN 声明指定

[0147] - 在混合信号 MS 中如何呈现声音对象 SO,这被表达为混合操作描述符(MOD)。根据通过声音对象描述符(SOD)给定的声音对象的特性选择 MOD。

[0148] - 与特定混合操作描述符 MOD 无关的混合信号 MS 的特性,并且该 IF-THEN 声明指定混合信号 MS 的操作的参数。

[0149] 声音对象描述符 SOD 可表示为(数据)结构,例如:

[0150]

SO.ID	声音对象识别符,例如,表演者的名称
SO.INSTR	声音对象 SO 的乐器种类
SO.BRIGHTNESS	声音对象 SO 的感知亮度
SO.PERCUSSIVENESS	SO 的敲击计量符
SO.CHARACTERISTIC	声音对象 SO 的另一特性

[0151] 混合操作描述符 MOD 描述声音对象 SO 的水平(亦即,音量)、平移位置、距离及可在混合信号 MS 中感知到的其他特性。应用于声音对象 SO 的混合操作描述符 MOD 可通过在数据结构内的 SO.MOD 来指定。混合操作描述符 MOD 亦可应用于混合信号 MS。这些混合操作描述符 MOD 由 MT.MOD 来指定。通常,这些混合操作描述符 MOD 控制应用于所有音频信号或应用于混合信号的信号处理,例如,混响或动态范围处理 DRP。

[0152] 混合操作描述符 MOD 可由知觉属性及分配至该知觉属性的值组成。混合操作描述符可实施为语言变量。

[0153] 知觉属性的列表可含有以下(除其他的外):

[0154]

知觉属性	描述
LOUDLEVEL	知觉水平
LATERALDISPLACEMENT	声音位置从中心偏离的侧角
PDISTANCE	感知声音对象 SO 的距离
FX1LEVEL	DAFx1 的感知量
FX2LEVEL	DAFx2 的感知量
REVERBLEVEL	混响的感知量
BRIGHTNESS	音色描述符
DIFFUSENESS	描述声音混合的扩散程度与直接程度

[0155]

[0156] 知觉属性可为语言变量。经分配的值可为以下值中的一个值: {“非常低”、“低”、“中等”、“高”、“非常高”}。

[0157] 将未由混合操作描述符 MOD 设定的知觉属性设定成默认值。

[0158] 随后混合规则可能看起来像以下:

[0159] IF<SO. INSTR=value>AND<SO. C1=value>...AND<SO. Cn=value>

[0160] THEN <SO. MOD1=value> AND <SO. MOD2=value> ... AND<SO. MODn=value>

[0161] 应注意,使用连接词(亦即,“AND”)足够,并且可将分离(亦即,“OR”)表达为单独的规则。

[0162] 示例性规则集合:针对舞蹈混合的示例给出所具有的使用情况的混合规则的集合:

[0163] 针对乐器种类指定这些混合规则:

[0164]


```

1. IF <SO.INSTR="kick drum">
  THEN <SO.PLOUDLEVEL = "high"> AND
<SO.LATERALDISPLACEMENT = "zero">
  AND <SO.DISTANCE = "near">
2. IF <SO.INSTR="bass">
  THEN <SO.PLOUDLEVEL = "high"> AND
<SO.LATERALDISPLACEMENT = "zero">
  AND <SO.DISTANCE = "near">
3. IF <SO.INSTR = "percussion"> AND <SO.ENTROPY = "high">
  THEN <SO.PLOUDLEVEL = "high"> AND <SO.FXLEVEL = "high">
4. IF <SO.INSTR = "percussion"> AND <SO.ENTROPY = "low">
  THEN <SO.PLOUDLEVEL = "low">

```

[0165] 针对与乐器种类无关的特性指定以下混合规则：

[0166] 5. IF<SO. INSTR="*" >AND<SO. ENTROPY=" low" >

[0167] THEN<SO. LATERALDISPLACEMENT=" far left" >

[0168] 6. IF<SO. INSTR="*" >AND<SO. CREST=" low" >

[0169] THEN<SO. PLOUDLEVEL=" low" >

[0170] 可选地,可针对混合信号 MS 指定混合规则。该混合规则与声音对象 SO 的特性没有关联。若在混合规则的 IF 部分中未指定任何声音对象,则将所得操作应用于所有声音对象 SO。

[0171] 7. IF*

[0172] THEN<MS. REVERBLEVEL=" low" >

[0173] 8. IF*

[0174] THEN<MS. FX1LEVEL=" high" >

[0175] 此外,在规则的 IF 部分中,亦可将属性与相对值而非绝对值进行比较。此举意谓例如使用如“最大化”或“最小化”的操作,可将一个声音对象 SO 的属性与所有其他声音对象 SO 的相同属性比较。

[0176] 9. IF<SO. INSTR="*" >AND<SO. ENTROPY=" maximum" >

[0177] THEN<SO. FX2LEVEL=" high" >

[0178] 应注意,以上列出的属性及规则为示例并且并不意味着是特定混合默认完整集合。

[0179] 根据本文所公开的示教的方面,可执行规则集合的变化。特定言之,规则集合可被实现为含有并行规则(具有相同前提但不同后续的规则),任意地(随机化的)选择该并行规则中的一个并行规则。此举将变化引入至结果中并且由此增加使用者满意度。这在产生规则集合的产生过程中没有一致的规则集合可获得同意的情形下亦有益。

[0180] 现转向语义音频分析 40 的示例性实施,其中,应用语义音频分析 40 来收集关于多个音轨 ST 及可能的多轨的信息,该信息可用于确定将哪些混合参数应用于该多个音轨 ST

中的哪个音轨 ST。语义音频分析单独处理该多个音轨中的每个音轨 ST,并且语义音频分析还处理多轨 MT 的表征。多轨 MT 的表征可(例如)以来源于以单位增益下混所有音轨 ST 而获得的混合信号的形式获取。

[0181] 可将结果表征为结构数组(其中每个数组元素含有一个音轨 ST 的元数据)及含有多轨的元数据的额外结构。结构元素的变量类型可为字符串(例如,针对乐器名称)、标量值(例如,针对拍子、一致性)或数组(例如,针对演奏风格的描述的起始时间及停止时间)或结构元素本身的专用结构(例如,描述乐曲形式的结构)。

[0182] 分析结果可伴随有置信测度,该置信测度表示各别结果的可靠程度。

[0183] 通过语义音频分析 40 产生的结果的表示的示例:

[0184] ST (1).ID="TR909"

[0185] ST (1).INSTRUMENT="kick drum"

[0186] ST (1).INSTRUMENT_CONFIDENCE=0.93

[0187] ST (1).ENTROPY=0.12

[0188] ST (2).ID="lead guitar"

[0189] ST (2).INSTRUMENT="guitar"

[0190] ST (2).INSTRUMENT_CONFIDENCE=0.68

[0191] ST (2).SOLO=[[123.4 234.5][567.7789.0]]

[0192] ST (3).ID="background vocals"

[0193] ST (3).INSTRUMENT="human singing"

[0194] ST (3).INSTRUMENT_CONFIDENCE=0.8

[0195] ST (3).BRIGHTNESS=0.12

[0196] MT.TEMPO="126"

[0197] MT.GENRE="electro"

[0198] MT.FORM=<form structure>

[0199] 语义音频分析 40 可用于通过将唯一的识别符分配至音轨 ST 并且分配至音乐作品的各个时间段,来标准化所提供的多轨音频材料。特定言之,多轨音频材料通常不是遵循某一惯例的预定格式。换言之,混音器无法依赖特定音轨(例如,“轨 1”)始终含有某一乐器(例如,“吉他”)。然而,由语义音频分析产生的元数据可提供关于多轨信号的组织及内容的基本上标准化信息,该标准化信息辅助混音器的其他模块完成其他模块的各自任务。由语义音频分析执行的标准化是有用的,因为该标准化允许由命令解释器 30 提供的混合命令与多轨音频信号所遇到的情形有关。因此,命令解释器 30 与语义音频分析 40“讲相同的语言”。

[0200] 目标描述符分配单元 DAU50 处理由语义音频分析 40 提供的元资料及来自命令解释器 30 的混合规则,以将混合操作描述符分配至多个音轨 ST 或分配至音轨 ST 的各段。这些描述符声明在目标混合信号 MS 中音轨 ST 的各段中占主导的每个声音对象 SO 是如何被感知的。

[0201] 假定在每个音轨 ST 中,每次仅一个声音对象占主导。在此假定的情况下,将自语义音频分析 40 导出的属性(对每个音轨 ST 计算该属性)处理为声音对象 SO 的属性。或者,若音轨 ST 含有多个声音对象,尤其如果若干声音对象 SO 在音轨 ST 内在时间上彼此接续,

此状况意味着若干声音对象 S_0 可相对容易地分离,则语义音频分析可输出每个音轨 ST 的多于一个的属性结构。另一可能性为,第一声音对象 S_{01} 主要存在于立体声信号的左信道中,而第二声音对象 S_{02} 主要存在于右通道中。又一可能性将为,若干声音对象在频率域中可借助于低通、高通和 / 或带通滤波器分离。

[0202] 若输入变量为明确值,则可应用模糊逻辑,但使用模糊属性(例如,“低”或“高”)来制定规则集合。举例而言,可将演奏乐器的变化程度表达为介于 0 与 1 之间的范围内的标量值。此外,语义音频分析 40 可与置信值(例如,机率)一起输出元数据,该置信值描述计算经估计的元数据的置信度。

[0203] 模糊逻辑允许通常结合专业知识来模型化复杂任务。模糊逻辑利用提供用于将精确值前后转换成模糊描述的直接机制的模糊集合。

[0204] 实施为模糊逻辑系统时的处理的综述图示于图 6 的方块图中(Mendel, 1995 年)。模糊逻辑系统包括模糊化模块 622、推断模块 624、规则集合 626 及去模糊化模块 628。模糊化模块 622 (例如)从语义音频分析 40 接收明确输入(*crisp input*)的集合。模糊化 622 基于明确输入产生模糊输入集合,该模糊输入集合被馈送至推断模块 624。推断模块 624 借助于被同等地馈送至该推断模块 624 的规则集合 626 评估模糊输入集合。规则集合 626 可由命令解释器 30 提供。推断模块 624 产生模糊输出集合并且将该模糊输出集合馈送至去模糊化模块 628。在去模糊化模块 628 中,将模糊输出集合翻译成明确输出(*crisp output*),该明确输出可随后用作为混合参数或作为中间量。

[0205] 现更详细地转向模糊化,基于由命令解释器 30 确定的规则集合的 IF 部分中描述的标准,执行将混合操作描述符 MOD 分配至单音轨 ST 。若来自语义音频分析 40 的各个元数据被给定为实数或与置信值一起给定为字符串(例如,作为乐器分类的结果),则使用模糊化将该实数翻译成语言变量。模糊集合为元素具有隶属度的集合。此隶属度可为区间 $[0, 1]$ 中的任何实数(与经典的集合论相反,在经典的集合论中,隶属度为 0 或者 1)。

[0206] 使用图 7 中示例性地所示的模糊集合的隶属函数执行模糊化。在模糊化中,针对每个实值的输入变量,确定相应的模糊集合(I. A. Zadeh, “Fuzzy Sets”, *Information and Control*, 第 8 卷, 第 338-353 页, 1965 年出版)及隶属度。举例而言,给定 0.25 的亮度,相应的模糊集合是隶属为 0.5 的“非常低”;及隶属为 0.5 的“低”。

[0207] 在推断步骤或模块 624 中,使用规则集合 626 将输入变量的模糊集合映像至输出变量的模糊集合。结果再次为知觉属性的语言变量的集合(与相应的置信隶属度一起)。

[0208] 在随后的步骤或去模糊化模块 628 中,使用输出变量的相应模糊集合将推断的结果转换成该输出变量的明确值。亦即,知觉属性的以上表中列出的变量具有带明确值的相对应部分。

[0209] 关于知觉处理器 60,命令解释器 30 及目标描述符分配单元 50 的输出确定声音对象 S_0 中的每个声音对象 S_0 应如何出现于混合信号 MS 中。到目前为止,借助于知觉值给出此规定。

[0210] 通过考虑信号特性及人类听觉机制,知觉处理器 60 将知觉值翻译成物理混合参数。以下段落说明某些知觉值的处理,该知觉值即,声级、给定侧角的平移系数、混响级及时间延迟、DAFx 参数、均衡及动态范围处理。

[0211] 可使用知觉响度模型来计算声音对象 S_0 的声级,该知觉响度模型例如, Glasberg

在 2002 年描述的模型。

[0212] 可选地, Moore 在 1996 年描述的响度模型可用以计算声音信号的混合物内声音信号的响度(B. C. J. Moore 及 B. R. Glasberg, “A Revision of Zwicker’s Loudness Model”, *Acustica-Acta Acustica*, 第 82 卷, 第 335-345 页, 1996 年出版)。

[0213] 计算每个音轨 ST 的增益因子以使得音轨 ST (或混合信号 MS) 中声音对象 SO 的知觉响度匹配通过混合操作描述符 MOD 表达的语义描述。

[0214] 给定侧角的平移系数: 声音对象 SO 的侧向位置的知觉通过耳朵入口处的耳间级差(inter-aural level differences; ILD)及耳间时差(inter-aural time differences; ITD)来确定(Lord Rayleigh, “On our perception of Sound Direction”, *Philosophical Magazine*, 第 6 卷, 第 214-232 页, 1907 年出版)。在知觉处理器 60 的背景下, 针对每个播放通道确定时间延迟及级差, 以唤起侧面化的知觉。

[0215] 混响级及时间延迟: 确定人工混响处理器的级别, 以使得混响的知觉量匹配由使用者给定的语义描述。针对每个声音对象单独定义混响级和 / 或针对混合信号 MS 定义混响级。可针对每个声音对象单独调整混响级, 以唤起特定声音对象 SO 的距离知觉。通过级别、时间延迟、均衡曲线及侧向位置来额外控制距离知觉。

[0216] DAFx 参数: 设定数字音频效果的参数取决于特定 DAFx 处理器。使用响度模型(例如, Moore, 1996 年) 计算经 DAFx 处理的信号的级别。

[0217] 均衡: 设定均衡参数, 以使得经处理信号匹配相对于混合信号 MS 或声音对象的“亮度”的知觉属性。

[0218] 动态范围处理: 将动态范围处理 DRP 的参数设定成匹配动态范围的知觉属性。

[0219] 图 8 示出包括模糊处理器 37 的混音器的部分的示意性方块图。模糊处理器 37 的输入连接至语义音频分析 40, 并且模糊处理器 37 的输入被配置为经由此连接接收音轨分析值。音轨分析值可为明确值或者语言变量。模糊处理器 37 还具有用于从语义命令解释器 35 接收规则或规则集合的输入。如以上所阐释的, 模糊处理器 37 使用规则来处理音轨分析值, 该音轨分析值产生可向音轨处理器 75 提供的明确(crisp)混合参数。

[0220] 由语义命令解释器 35 基于由使用者提供的语义混合命令产生规则。

[0221] 知觉模型 64 向模糊逻辑处理器 37 提供模糊化及去模糊化参数。模糊化及去模糊化参数建立数值与相应语义描述之间的联系。举例而言, 模糊化及去模糊化参数可指定向听众呈现轻柔、中等或大声的音频信号的响度范围。

[0222] 此外, 知觉模型 64 可指定当期望某一效果时, 涉及哪些混合参数。这些混合参数的相应值亦可由知觉模型 64 指定。可向语义命令解释器 35 提供这些指定作为指导方针。语义命令解释器 35 在产生模糊规则时可遵循这些指导方针。

[0223] 混音器可包括可选的随机模糊规则选择器 38, 当两个并行模糊规则已由语义命令解释器 35 产生并且仅一个并行模糊规则可由模糊逻辑处理器 37 实施时, 使用该随机模糊规则选择器 38。适度的随机度可增加使用者满意度, 因为混合程序似乎更自然并且更“人性化”。毕竟, 人类混合工程师亦可能偶然动作略微随机, 这可能被混合工程师的用户视为“具有艺术性”。

[0224] 图 9 示出根据本文所公开示教的混音器的可能的基本配置的示意性方块图。以多个单音轨 ST 的形式提供数据 10。混音器包括语义命令解释器 35、音轨处理器 75 及音轨组

合器(AT CMB) 76。

[0225] 语义命令解释器 35 大体上对应于图 5 的命令解释器 30。此外,语义命令解释器 35 可包括目标描述符分配模块 50 及知觉处理器 60 的一些功能。语义命令解释器 35 接收语义混合命令作为输入并且从语义混合命令导出一个混合参数或多个混合参数。向音轨处理器 75 或更精确而言向音轨处理器 75 的个别音轨处理器 ATP1、ATP2、ATP3、ATP N 提供多个混合参数。混合参数通常呈明确值的形式,该明确值可较容易地由多个个别音轨处理器 ATP1 至 ATP N 来实施。

[0226] 多个个别音轨处理器 ATP1 至 ATP N 根据混合参数修改由音轨 ST1 至 ST N 中的相应音轨提供的音频信号。

[0227] 经修改的音频信号由音轨组合器 76 组合,以获取混合信号 MS。

[0228] 在图 9 中所示的配置中,语义命令解释器 35 能够将语义混合命令内的特定语义内容分配给相应的个别音轨处理器 ATP1 至 ATP N 的适当的混合参数。语义命令解释器 35 的该能力可基于以下事实:根据达成一致的标准组织该多个单音轨 ST1 至 STN,以使得该语义命令解释器 35 可已知道哪个音轨对应于哪个乐器。在图 11 至图 14 中,在此描述的对应部分中图示并且描述混音器的可选配置,此混音器能够从数据本身获得关于多轨录音的组织 and / 或被录音的音乐作品的时间结构的信息。

[0229] 图 10 示出语义混合命令。语义混合命令包括呈英语句子的形式的语言表达。当然,亦可使用其他语言。该句子写的是:“During the guitar solo,mix the guitar prominently”。此句子的语义分析公开该句子可分解成三个部分。第一部分含有表达“During the guitar solo”并且可识别为指定语义混合命令的目标时间段的表达。第二部分含有表达“the guitar”并且可识别为指定目标音轨的表达。第三部分含有表达“mix[···] prominently”并且可识别为指定所要混合操作的表达。

[0230] 图 11 示出语义混合命令的延伸的示例。延伸的混合命令基于来自图 10 的语义混合命令。此外,已添加第二目标音轨的第二混合操作,亦即“[···]move the keyboards slightly into the background”。一个关联用于指定第一混合操作 / 第一目标音轨与第二混合操作 / 第二目标音轨之间的关系。在所示情况下,连接为用语“and”,以使得第一混合操作及第二混合操作两者皆在该第一混合操作及该第二混合操作的各自目标音轨上同时执行。

[0231] 图 12 示出根据另一个可能配置的混音器的部分的示意性方块图。特定言之,图 12 示出由多个音频信号 ST1 至 ST N 及由默认混合信号 MT (“多音轨”)提供的数据可如何用于获取关于音乐作品的音轨布置和 / 或时间结构的有用信息。除非另有说明,否则该多个音轨的引用应包括默认混合信号 MT 的引用。

[0232] 向语义音频分析 40 提供多个音轨 ST1 至 MT。通过分析该多个音轨,可获取向语义至明确转换模块 65 提供的音轨信息及时间结构信息。

[0233] 语义混合命令包括多个表达,每个表达包括指定目标时间段 26、目标音轨 27 及混合操作 28。

[0234] 语义至明确转换模块 65 近似对应于图 5 的目标描述符分配单元 50。语义至明确转换模块 65 也从语义混合命令接收信息作为输入。语义至明确转换模块 65 基于所提供的输入产生一个或多个知觉目标描述符 PTD 及相应的混合参数。知觉目标描述符 PTD 可含有

受影响音轨 ST1 至 ST N 的音轨识别符,以及时间段信息,仅以防一个或多个受影响音轨的时间段受混合命令的影响。注意到,混合参数可为明确值或者将在随后阶段解析的语言变量。

[0235] 语义音频分析 40 可可选地接收目标时间段指定 26 和 / 或目标音轨指定 27 作为输入,以使得该语义音频分析 40 可分析多个音轨 ST1 至 MT,特别聚焦于所提供指定上。

[0236] 图 13 示出根据本文所公开示教的混音器的另一可能配置的示意性方块图。此配置以音轨识别器 430 为特征结构。

[0237] 图 13 中所示的配置的基本结构基本上与图 9 中相同;然而,为清晰起见,已省略一些部分。

[0238] 因为并非始终立即清楚音轨 ST1 至 ST N 中哪个音轨含有哪个乐器或声乐部分,所以音轨识别器 430 可用以确定该信息。音轨识别器 430 可为语义音频分析 40 的一部分。

[0239] 语义混合命令包括已相对于图 12 提及的目标音轨识别 27。向音轨模板数据库 432 提供目标音轨识别 27。音轨模板数据库 432 提供对应于目标音轨识别 27 的一个或多个数据记录,并且音轨模板数据库 432 向音轨识别器 430 提供该一个或多个数据记录。该数据记录可包括形式为量测值、声音样本等的关于(例如)乐器的信息。音轨识别器 430 可随后将数据记录中含有的信息与该多个音轨 ST1 至 ST N 中的每个音轨的音频信号比较。为此,音轨识别器可(例如)执行来自数据记录的声音样本与音轨信号的短段的交叉相关。另一选择是确定音轨信号的泛音的位置及幅度并比较该结果与数据记录中的相应数据。又一选择为分析及比较音轨信号的上升-衰减-维持-释放行为。

[0240] 音轨识别器产生向音轨处理器 75 提供的音轨识别信息,以使得该音轨处理器 75 可根据由(例如)语义混合命令内的乐器名称的指示,处理每个单音轨 ST1 至 STN。

[0241] 图 14 图示混音器的另一可能配置的示意性方块图,在该配置中,时间段识别器 460 从该多个音轨 ST1 至 MT 获取时间段信息。时间段识别器 460 连接至该多个音轨 ST1 至 MT 并且被配置为分析由该音轨 ST1 至 MT 呈现的音乐作品的时间结构。特定言之,时间段识别器 460 可在音乐作品内寻找相似或基本上相同的段。若音乐作品属于流行音乐类型,则这些相似或基本上相同的段可能为歌曲的合唱。时间段识别器 460 亦可计数音乐作品的拍子或小节,这可改良时间段识别的精度。

[0242] 向语义命令解释器 35 提供时间段信息,该语义命令解释器 35 使用该时间段信息来将语义混合命令内使用的语义时间段表达翻译成明确(crisp)的区段开始及结束时间值。

[0243] 通过时间段识别器执行的音乐作品的时间结构分析可采用过去各个研究者建议的方法中的一个或多个方法。在 ICASSP2005, Xi Shao 等人的文章“Automatic Music Summarization Based on Music Structure Analysis”中,建议基于音乐结构分析的音乐概述的新颖方法,该文章的全部内容以引用的方式并入本文。特定言之,首先从音频信号获取起始音符,以获取歌曲的时间节奏。可基于此节奏信息执行音乐结构分析。在将音乐内容结构化诸如导奏(Intro)、独唱、合唱、尾曲(Outro)等的不同区域的后,可产生具有合唱及音乐乐句的最终音乐概述,在选定的合唱之前或之后包括该合唱及音乐乐句,以获得最终概述的所要长度。音乐结构分析区分基于旋律的相似性区域(独唱)与基于内容的相似性区域(合唱)。

[0244] 在“Chorus Detection with Combined Use of MFCC and Chroma Features and Image Processing Filters”, Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 2007 年 9 月 10 日至 15 日出版中, 作者 Antti Eronen 描述用于检测流行摇滚乐中合唱段的计算上有效率的方法。该方法利用距离矩阵表征, 该距离矩阵表征通过合计使用美尔频率倒谱系数 (mel-frequency cepstral coefficient) 及音高半音特征计算的两个单独的距离矩阵来获取。Eronen 的文章的全部内容以引用的方式并入本文。

[0245] Mark Levy 等人为文章“Extraction of High-Level Musical Structure from Audio Data and its Application to Thumbnail Generation”, ICASSP2006 的作者, 该文章的内容以引用的方式全部并入本文。在该文章中, 介绍了用于用阶层式音色模型将音乐音频分段的方法。提供新证据以显示音乐分段可重作为音色特征的丛集, 并且描述新的丛集算法。

[0246] 在“A Chorus Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station”, IEEE Transactions on Audio, Speech, and Language Processing, 第 14 卷, 第 5 期, 2006 年 9 月出版中, 作者 Masataka Goto 描述用于获取流行音乐的压缩盘片录音中重复合唱 (“引人乐段 (hook)”) 段的列表的方法。首先, 自输入信号的每个帧获取称为半音向量的十二维特征向量并且随后计算这些向量之间的相似性, 该半音向量相对于伴奏的变化是稳固的。列出并且整合识别为正在重复段的段。该方法甚至可以通过甚至在调变的后, 引入以知觉方式刺激的声学特征及实现重复合唱区段的侦测的相似性, 来侦测经调变的合唱段。该文章的全部内容以引用的方式并入本文。

[0247] Bee Suang Ong 在 Bee Suang Ong 的论文“Structural Analysis and Segmentation of Music Signals”, Universitat Pompeu Barcelona, 2007 年出版, ISBN978-84-691-1756-9 中编辑当时所知的自动音乐结构分析方法的综述, 该论文的全部内容以引用的方式并入本文。

[0248] 图 15 示出混音器的另一可能配置的示意性方块图, 在该配置中提供元数据接口 480, 以利用与多轨信号一起提供的元数据 12。元数据可包括图 12 及图 13 的背景下说明的关于音轨组织的信息或时间段信息。

[0249] 元数据 12 (若存在) 使混音器不必从多轨信号确定音轨信息、时间段信息或其他有用信息。此确定可涉及可能耗费相对较长时间的计算密集型数据处理任务。此外, 通过混音器本身执行的确定的结果可能比由多轨音频信号的发起者产生并且提供的元数据可靠性更差。

[0250] 元数据接口 480 被配置为从多轨录音的数据 12 获取元数据 12。在输出侧, 元数据接口 480 连接至语义命令解释器 35 的输入。在图 15 中所示的配置中, 语义命令解释器 35 被配置为将由元数据接口 480 提供的元数据 12 用于从语义混合命令导出多个混合参数的过程中。

[0251] 图 16 示出混音器的另一可能配置的示意性方块图, 在该配置中, 提供示例接口 490 及示例性混合信号分析器 492, 以基于该示例性混合信号产生语义混合命令。

[0252] 示例接口 490 被配置为接收示例性混合信号。示例性混合信号可 (例如) 储存于内存中或经由网络获取。使用者可根据该使用者的偏好从示例性混合信号的集合选择示例性

混合信号,该偏好例如,因为该使用者喜欢混合特定混合信号的方式。一般而言,任何音频信号可用作示例性混合信号,但若该示例性混合信号具有类似于多轨录音的结构及风格,则通常预期结果更好。举例而言,若示例性混合信号的配器基本上与将通过混音器混合的多轨信号的配器相同,则该示例性混合信号可能是有用的。

[0253] 示例接口 490 将示例性混合信号转发至混合信号分析器 492。混合信号分析器 492 可配置为识别示例性混合信号中的乐器及声乐部分。此外,混合信号分析器 492 可整体确定所识别乐器部分、所识别声乐部分和 / 或示例性混合信号的相对响度级别和 / 或频率曲线。还可确定诸如混响的音频效果的量。基于所确定的值,混合信号分析器 492 可建立示例性混合信号和 / 或语义混合命令的特征。举例而言,由混合信号分析器 492 执行的分析可揭示示例性混合信号的鼓声音轨及低音音轨相对占主导,而其他音轨较轻柔。因此,语义混合命令可包括声明鼓声音轨及低音音轨应在将由混音器产生的整个混合信号 MS 占主导的表达。

[0254] 示例接口 490 还可配置为与示例性混合信号一起接收示例音轨。示例音轨在图 16 中由虚线菱形标记的“示例 ST'”表示。由示例接口 490 向混合信号分析器 492 提供示例音轨。示例音轨对应于示例性混合信号,因为该示例音轨用以产生该示例性混合信号。使用可用的示例音轨,混合信号分析器 492 可将示例性混合信号与该示例音轨中的每个示例音轨比较,以找出如何在被混合成示例性混合信号之前,某示例性混合信号被如何修改。以此方式,可通过混合信号分析器 492 以语义形式或半语义形式确定音轨相关的混合参数。

[0255] 图 17 图示混音器的另一可能配置的示意性方块图,在该配置中,知觉处理器 63 及知觉模型 64 用于将语义混合命令转换成混合参数的过程中。在图 17 的配置中将知觉处理器 63 及知觉模型 64 示出为语义命令解释器 35 的一部分。如以上所述,通过考虑信号特性及人类听觉机制,知觉处理器 63 将知觉值翻译成物理混合参数。描述人类听觉机制的参数由知觉模型 64 来提供。可将知觉模型 64 组织为数据库或知识库。数据库的条目可包括听觉相关的现象的语义描述及呈音频效果、响度、相对响度、频率内容等的参数的形式的相应实施。听觉相关的现象可(例如)通过表达来描述,该表达诸如,“远”、“近”、“平”、“完全”、“亮”、“偏向低频”、“偏向高频”等。相应实施可包括指示应如何选择该多个音轨 ST 中的一个或多个音轨 ST 的混合参数来实现所要效果的数值。从语义描述至混合参数的相应值的这种映像通常基于专门知识及心理声学。可能已在复杂的科学测试及研究期间获取了该专门知识及心理声学。

[0256] 图 8 及图 11 至图 16 中所示的配置可以任何组合与彼此组合。举例而言,通过组合图 12 及图 13 中所示的配置,可提供包括音轨识别器 430 及时间段识别器 460 的混音器。

[0257] 图 18 示出用于将多个音频信号混合成混合信号的方法的示意流程图。在 102 处,方法开始之后,接收语义混合命令,如方块 104 所示。语义混合命令可由用户使用键盘以本文形式输入、口头上作为发音命令输入、作为自多个预设的选择输入,通过调整一个或多个参数输入、作为示例性混合信号输入或以其他方式输入。

[0258] 在由方块 106 表示的动作处,从语义混合命令导出多个混合参数。此动作可涉及专门知识及心理声学,以使得混合参数产生使用者所要的结果。

[0259] 由方块 108 表示的动作的背景下,根据混合参数处理多个音轨。该多个音轨的处理可包括设定响度级别、平移位置、音频效果、频率滤波(均衡)及其他修改。

[0260] 在由方块 110 表示的动作处,在方法结束于方块 112 处之前,将由处理产生的音轨组合以形成混合信号。

[0261] 尽管已在设备的背景下描述一些方面,但显然这些方面也表示相应方法的说明,其中,方块或装置对应于方法步骤或方法步骤的特征。类似地,在方法步骤的背景下描述的方面也表示相应方块或项目或相应设备的特征结构的说明。该方法步骤中的一些或全部步骤可由(或使用)硬件装置来执行,该硬件装置例如:微处理器、可编程计算机或电子电路。在一些实施方式中,最重要方法步骤中的某一个或多个方法步骤可由这样的装置执行。

[0262] 取决于某些实施要求,可在硬件中或在软件中实施本发明的实施方式。可使用数字储存介质(例如软盘、DVD、蓝光光盘、CD、ROM、PROM、EPROM、EEPROM 或闪存)执行实施,该数字储存媒体上储存有电子可读取的控制信号,该电子可读取的控制信号与可编程计算机系统合作(或能够合作),以执行各个方法。因此,数字储存介质可以是计算机可读取的。

[0263] 根据本发明的某些实施方式包括具有电子可读取控制信号的数据载体,该电子可读取控制信号能够与可编程计算机系统合作,以执行本文所述的方法中的一种方法。

[0264] 大体而言,本发明的实施方式可实施为具有程序代码的计算机程序产品,当计算机程序产品在计算机上执行时,该程序代码可操作用于执行该方法中的一种方法。例如,该程序代码可储存在机器可读取载体上。

[0265] 其他实施方式包括在机器可读取载体上储存的用于执行本文所述方法中的一种方法的计算机程序。

[0266] 因此,换而言之,本发明方法的实施方式为计算机程序,该计算机程序具有用于当其在计算机上执行时执行本文所述的方法中的一种方法的程序代码。

[0267] 因此,本发明方法的另一实施方式为数据载体(或数字储存介质或计算机可读取介质),该数据载体包括该数据载体上记录的用于执行本文所述方法中的一种方法的计算机程序。数据载体、数字储存介质或记录介质通常为有形的和/或非暂时性的。

[0268] 因此,本发明方法的另一实施方式是表示用于执行本文所述方法中的一种方法的计算机程序的数据流或信号的序列。数据流或信号的序列可例如配置为经由数据通信连接传输,例如经由因特网。

[0269] 另一实施方式包括处理构件,例如计算机或可编程逻辑装置,该处理构件被配置为或适用于执行本文所述方法中的一种方法。

[0270] 另一实施方式包括计算机,该计算机上安装有用于执行本文所述方法中的一种方法的计算机程序。

[0271] 根据本发明的另一实施方式包括被配置为将用于执行本文所述方法中的一种方法的计算机程序(例如,以电子学方式或以光学方式)转移至接收器的设备或系统。例如,该接收器可为计算机、行动装置、内存装置或类似物。例如,设备或系统可包括用于将计算机程序传输至接收器的文件服务器。

[0272] 在一些实施方式中,可编程逻辑设备(例如现场可编程门阵列)可用以执行本文所述方法的功能中的一些或全部。在某些实施方式中,现场可编程门阵列可与微处理器合作,以执行本文所述方法中的一种方法。大体而言,方法优选由任何硬件装置来执行。

[0273] 以上描述的实施方式仅说明本发明的原理。应理解,本文所述的配置及细节的修改及变化对本领域技术人员而言显而易见。因此,本发明旨在仅受到已授权的专利权利要

求的范围的限制,而不受以本文实施方式描述及说明的方式呈现的具体细节的限制。

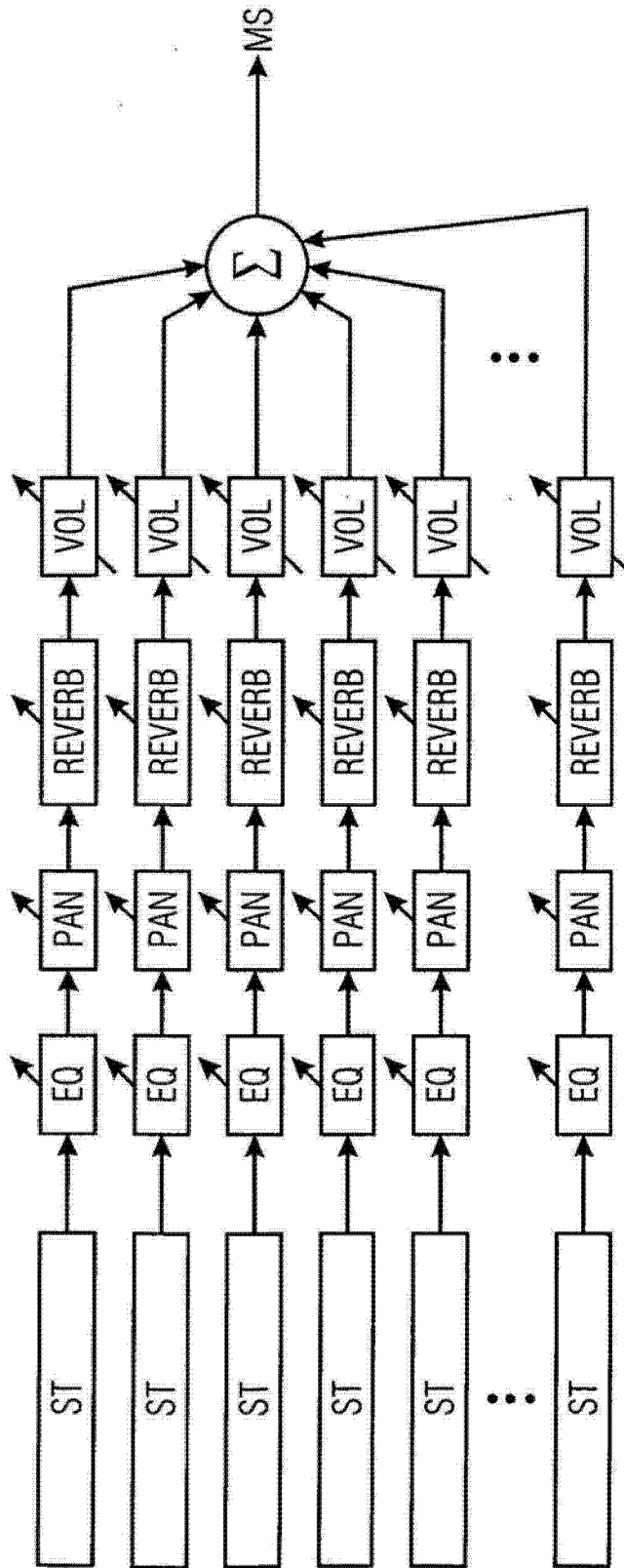


图 1

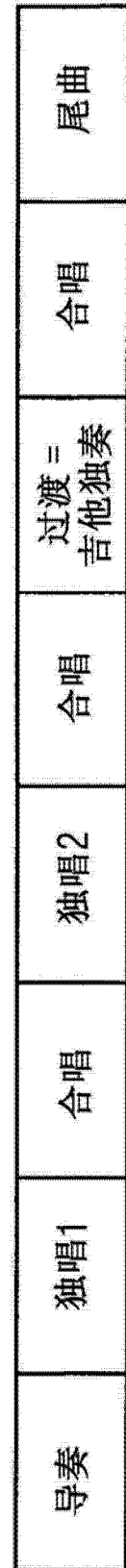


图 2



图 3

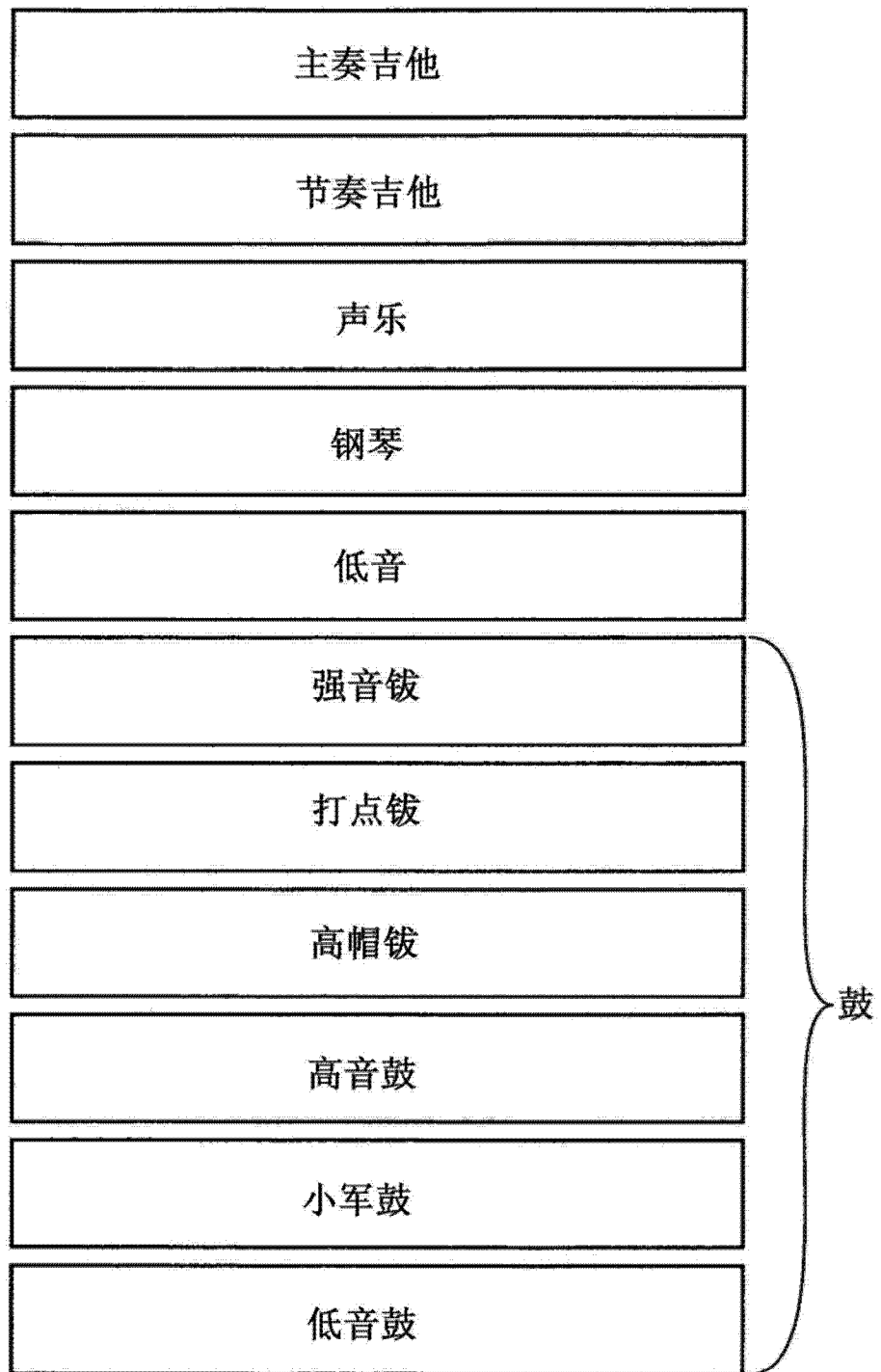


图 4

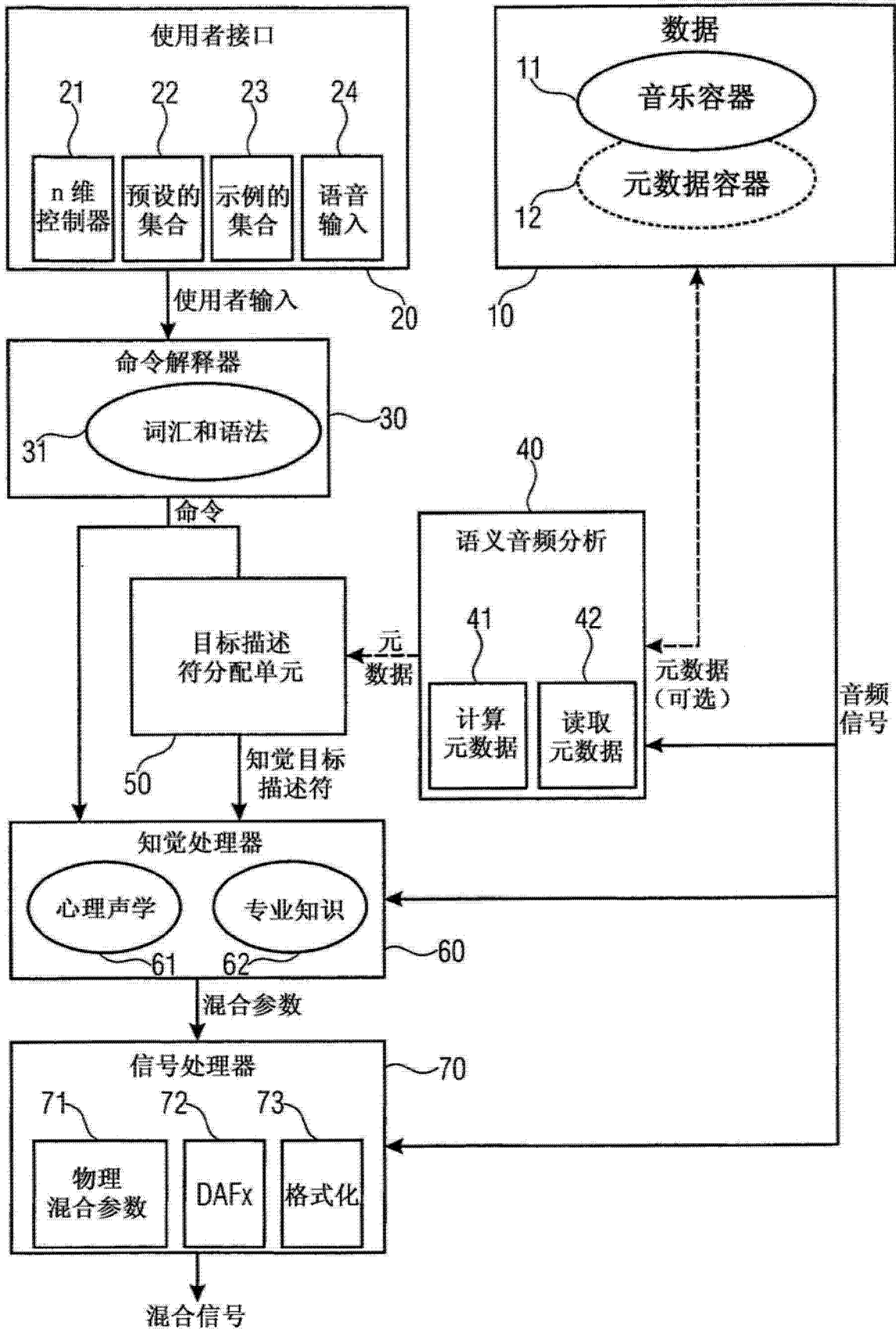


图 5

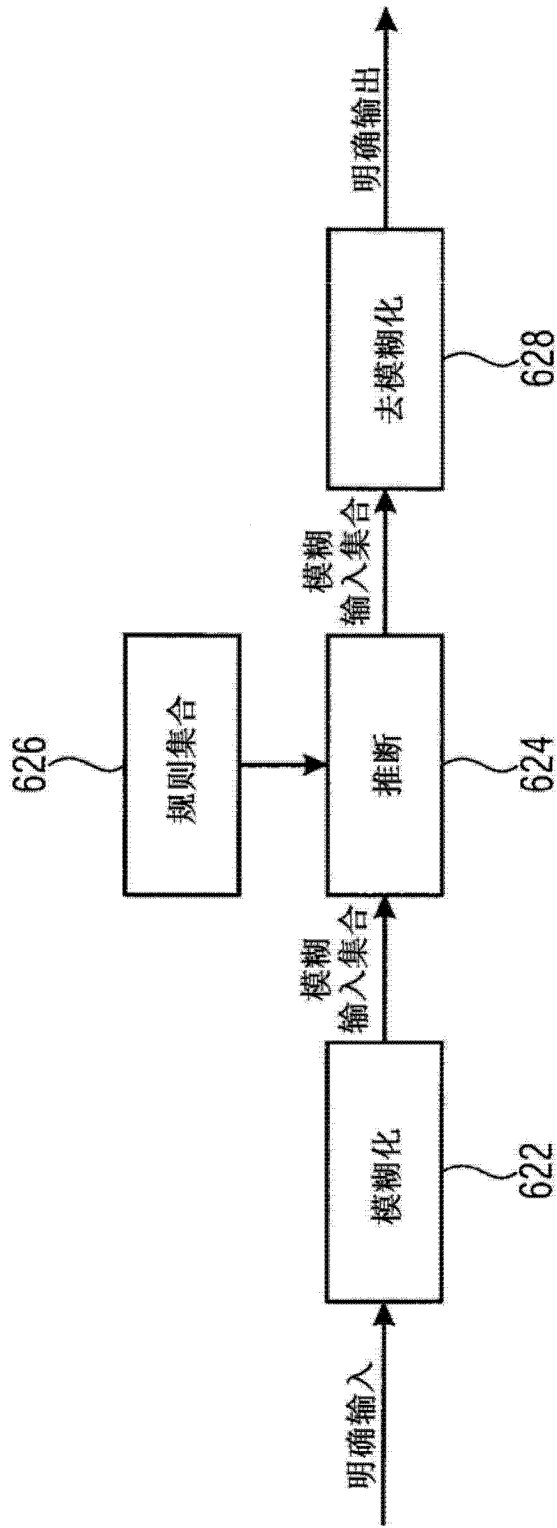


图 6

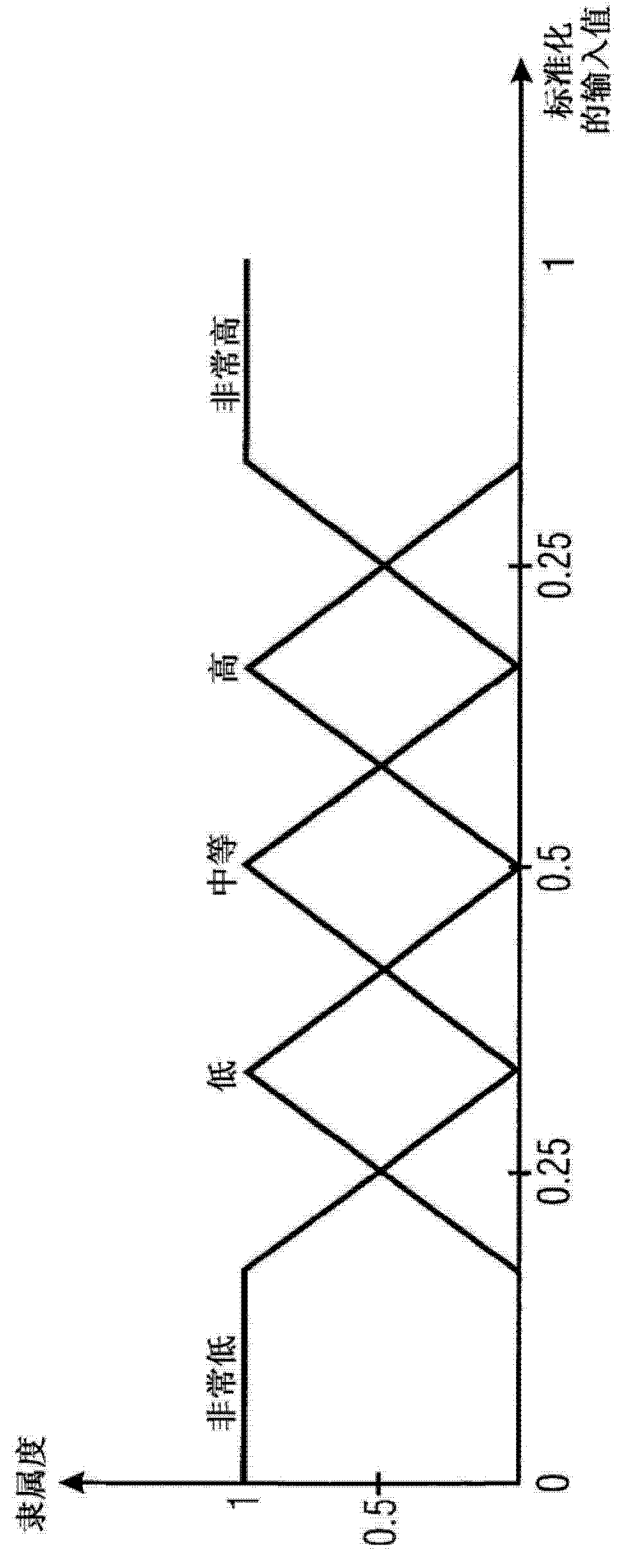


图 7

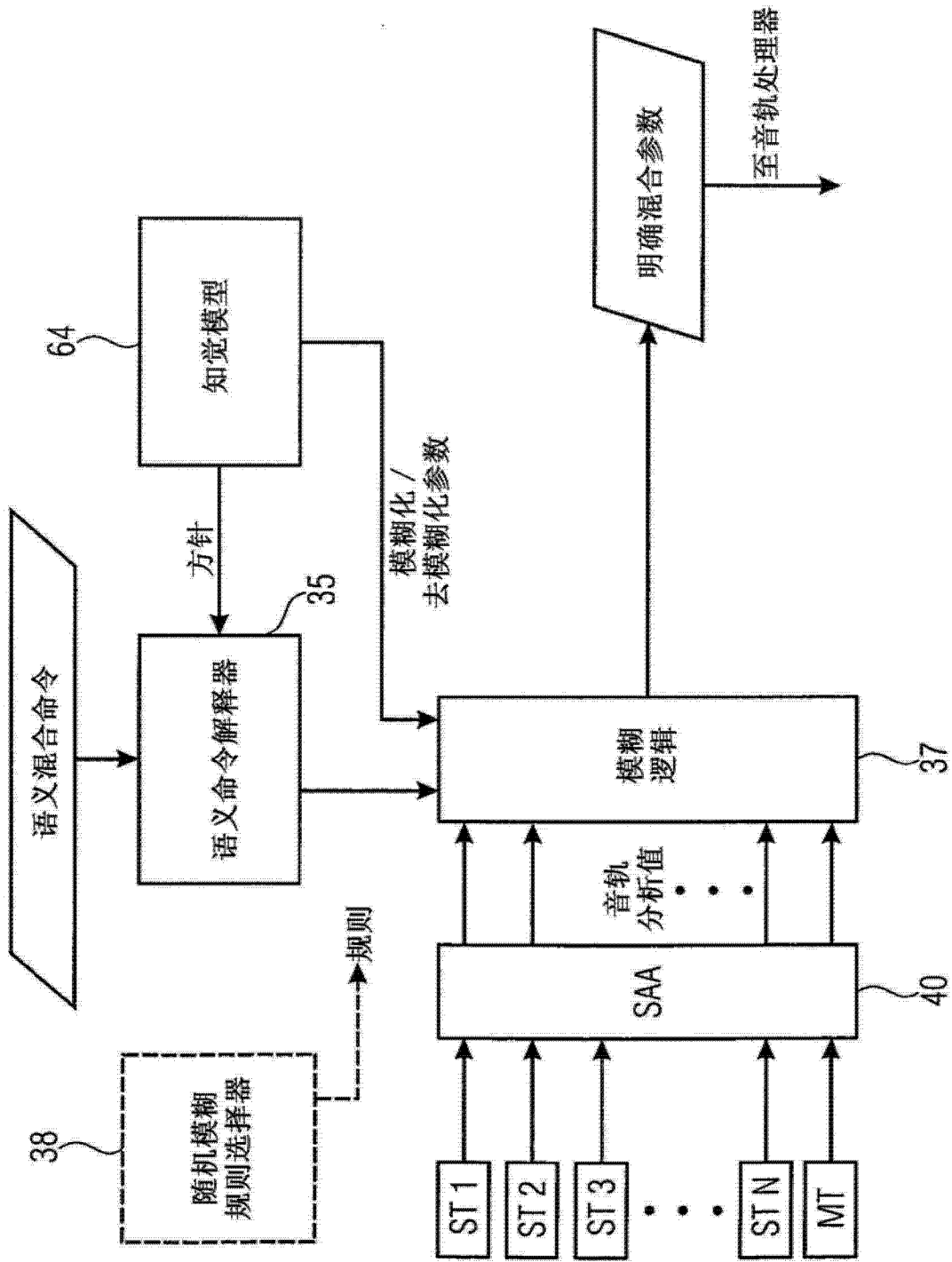


图 8

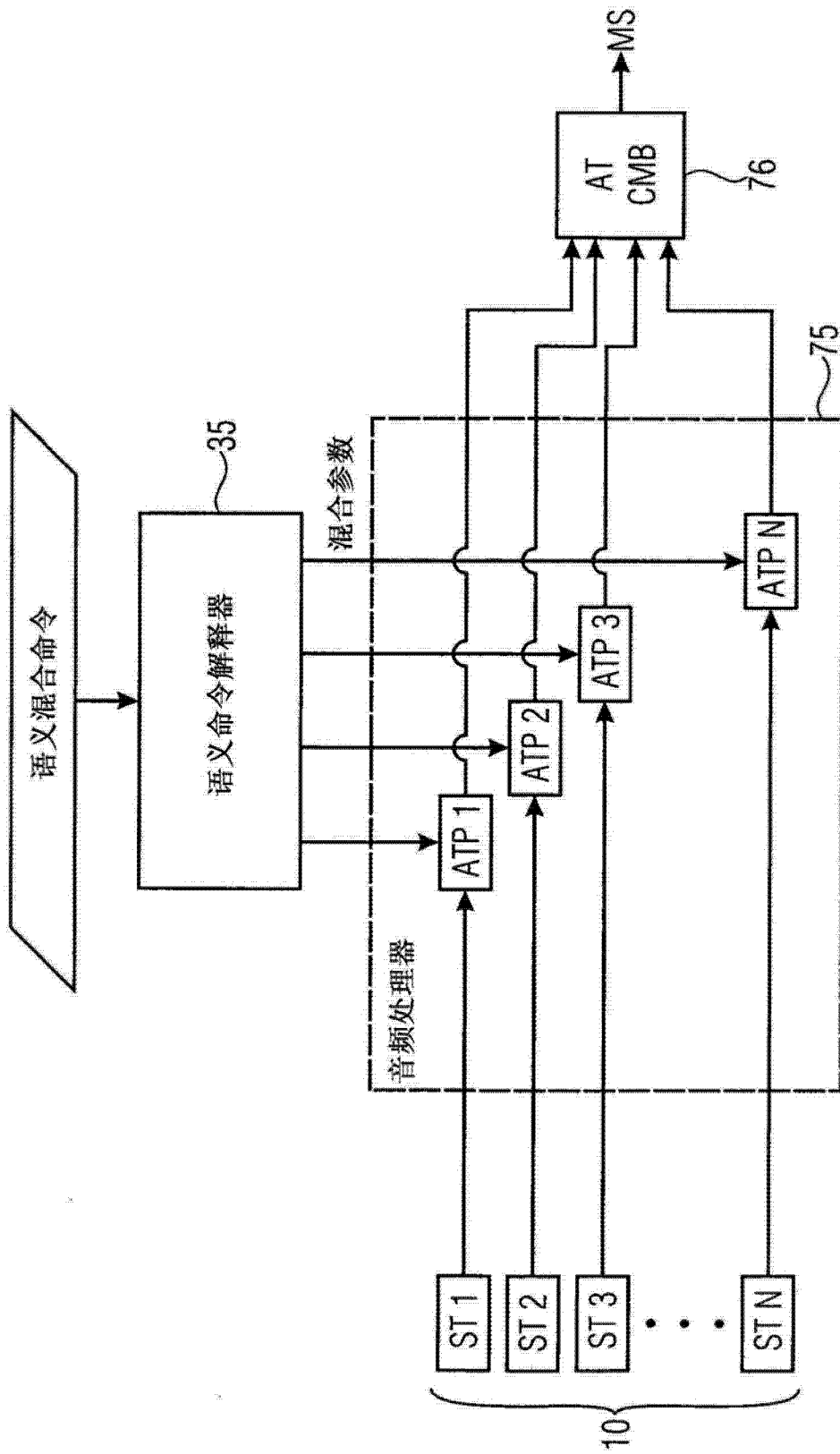


图 9

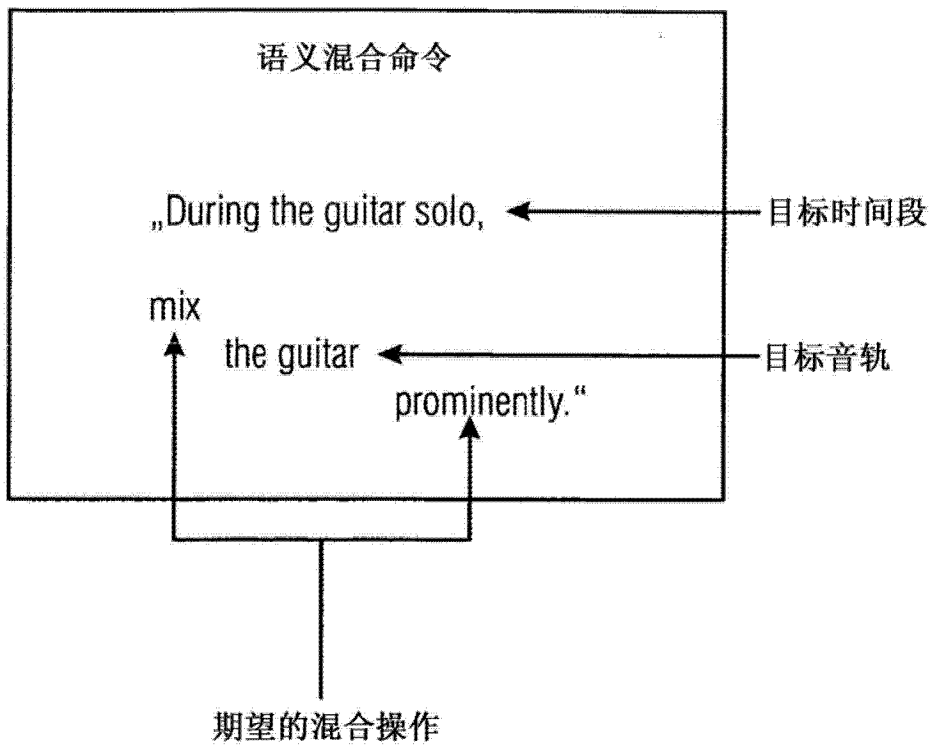


图 10

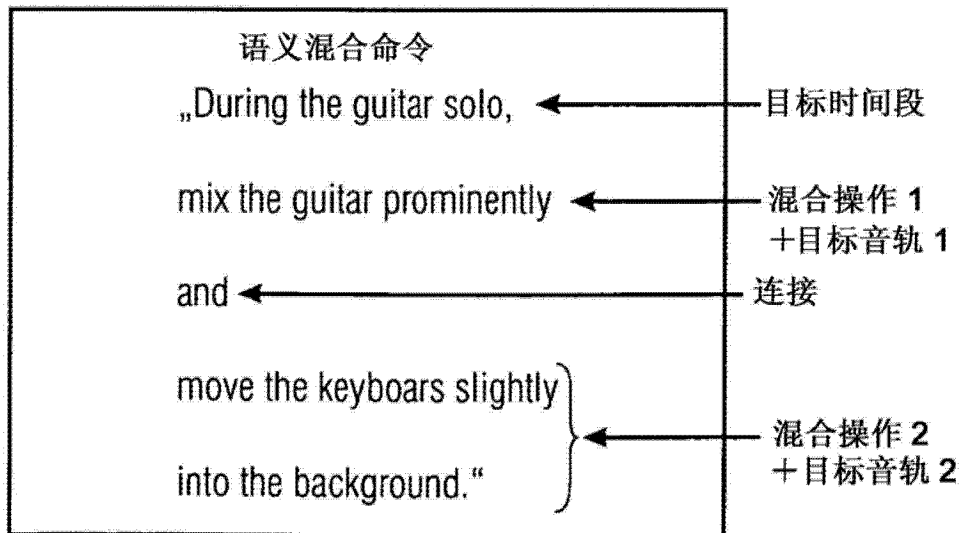


图 11

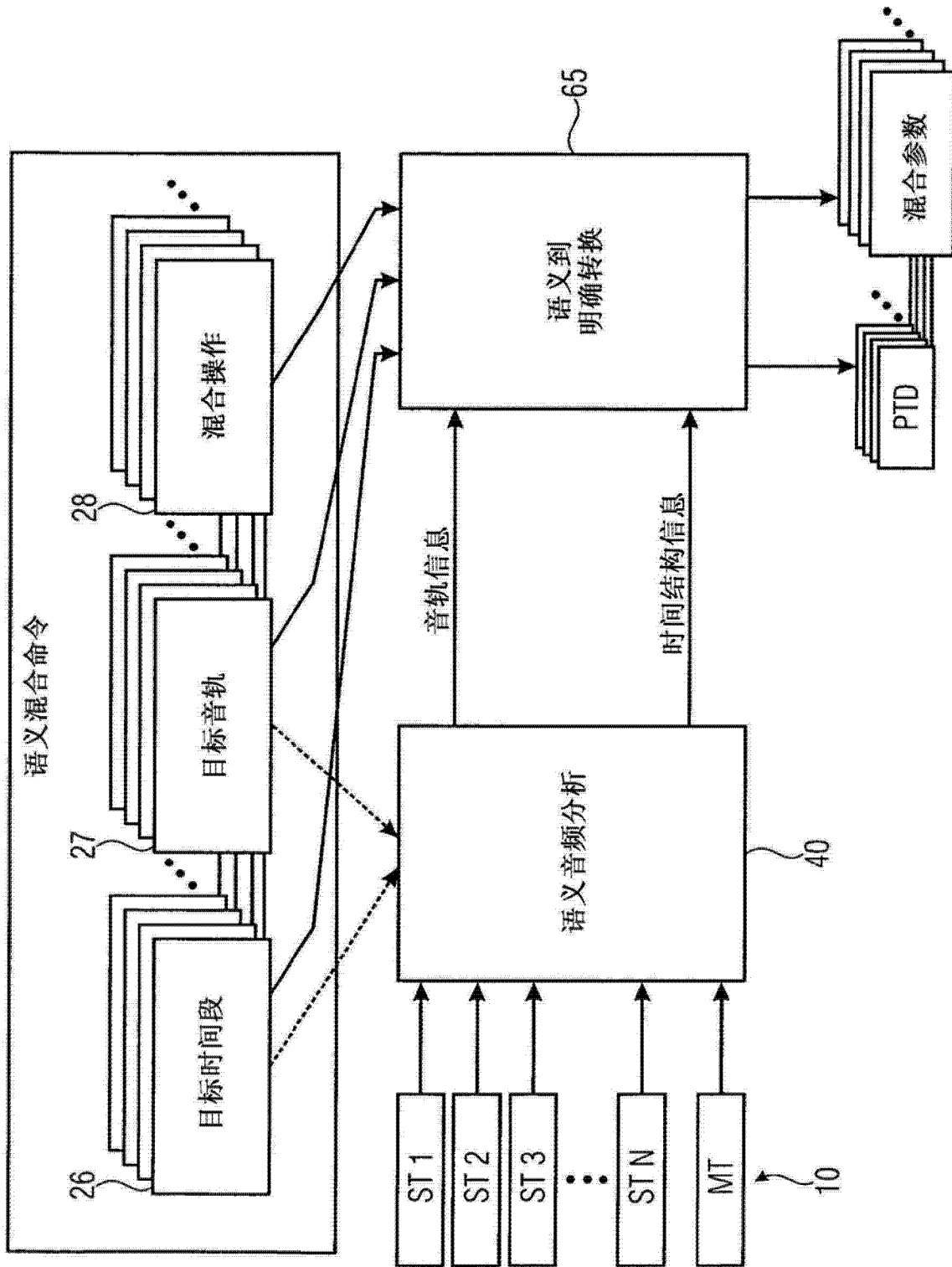


图 12

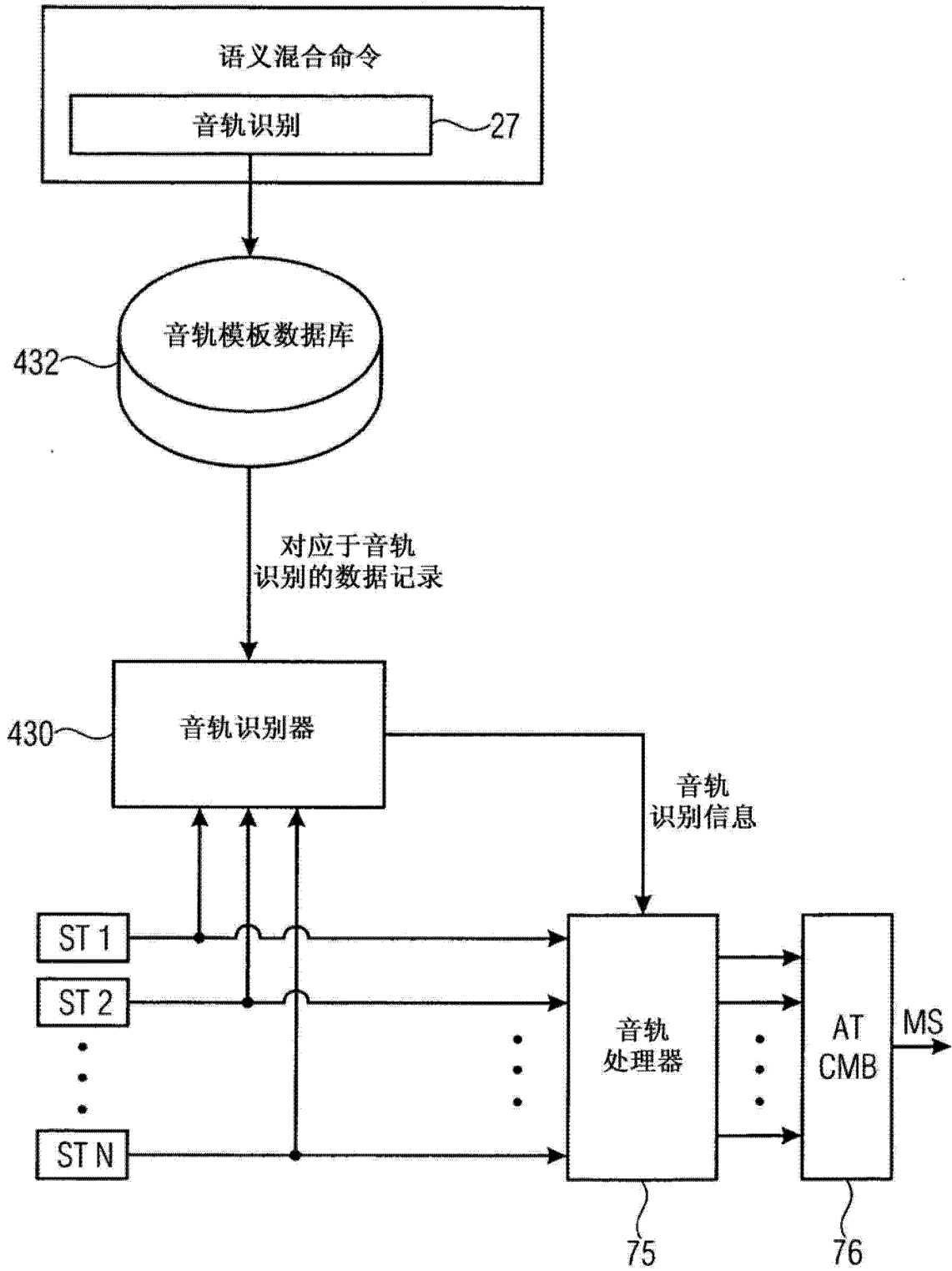


图 13

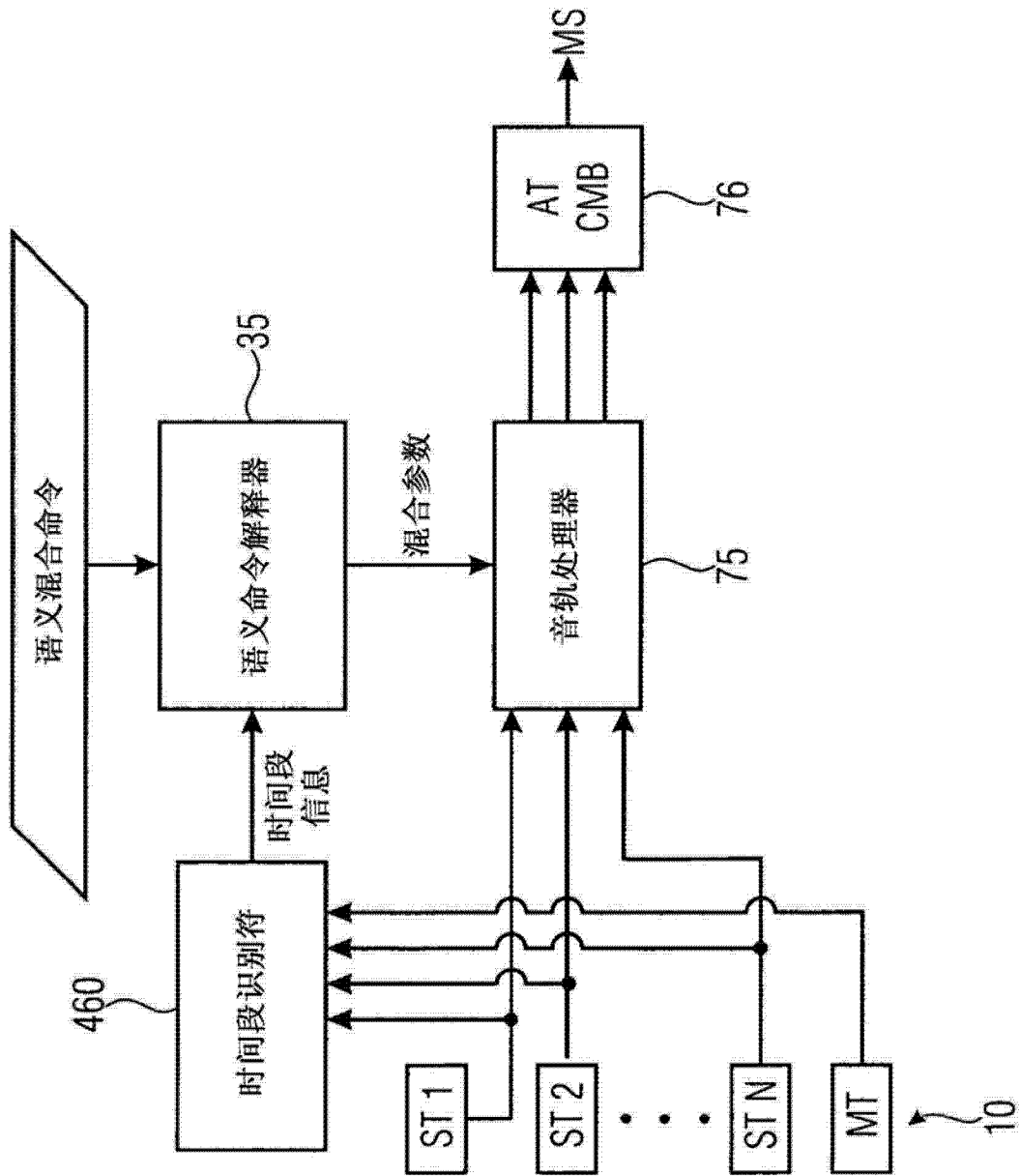


图 14

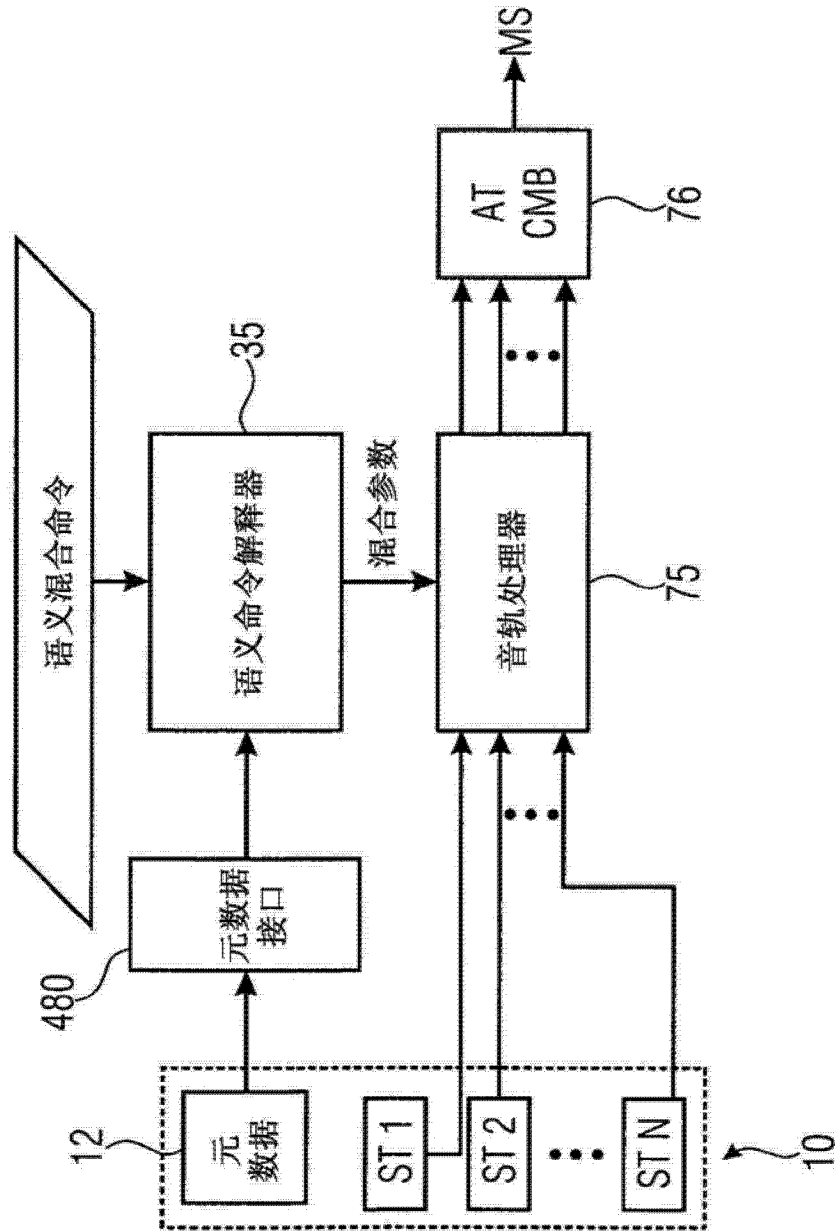


图 15

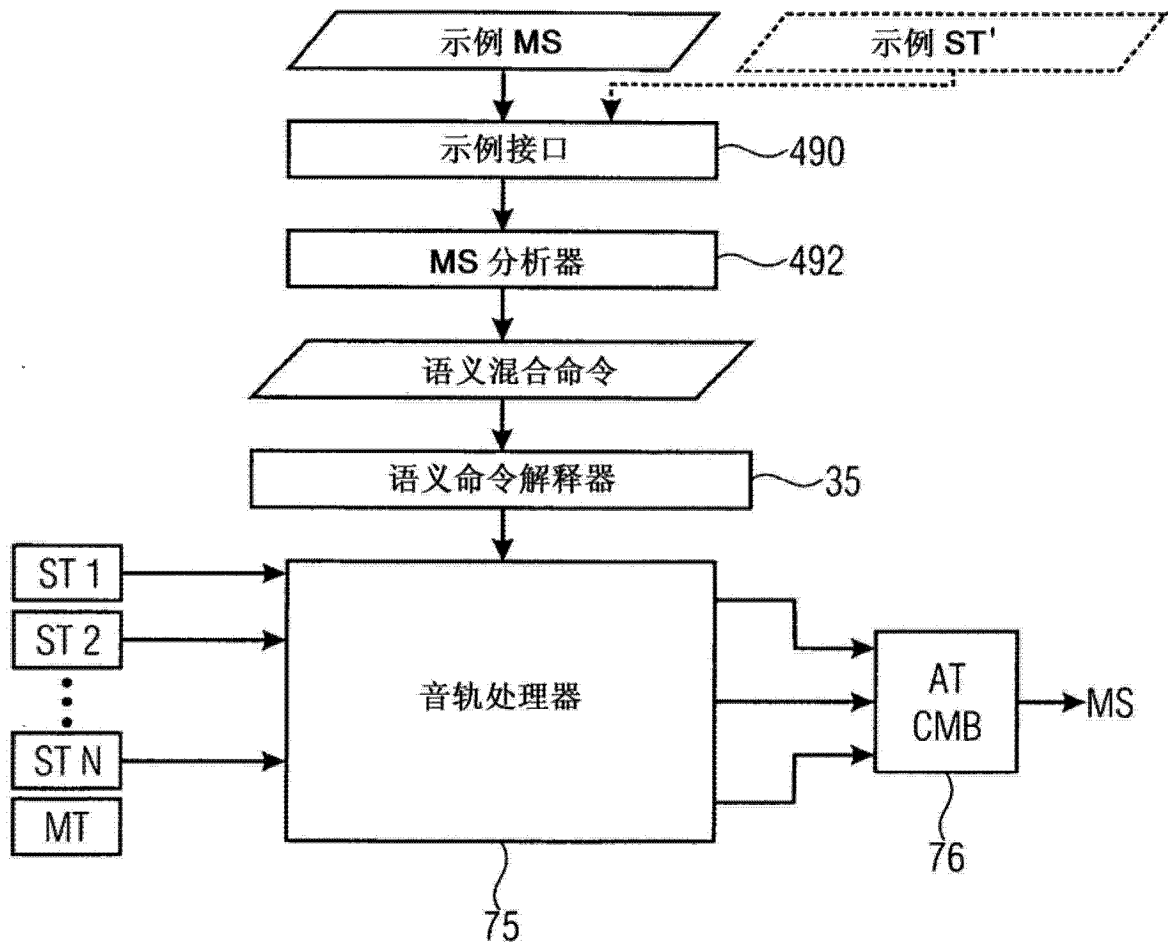


图 16

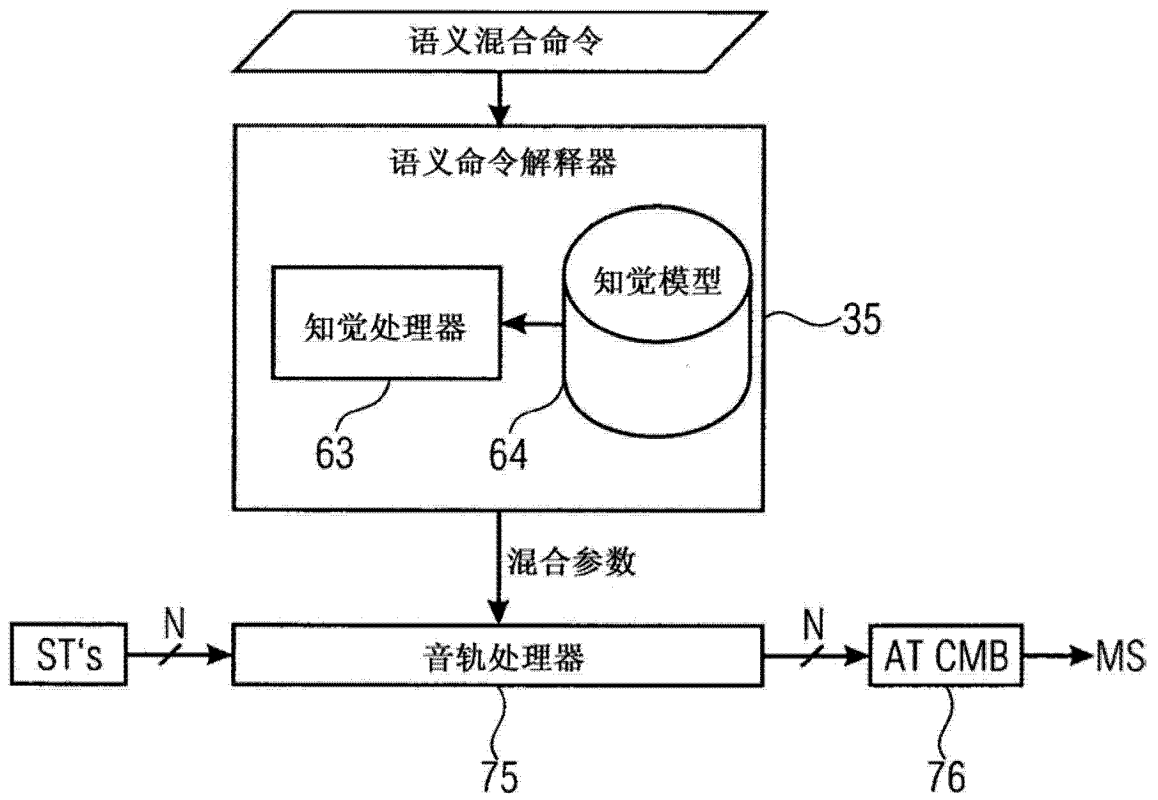


图 17

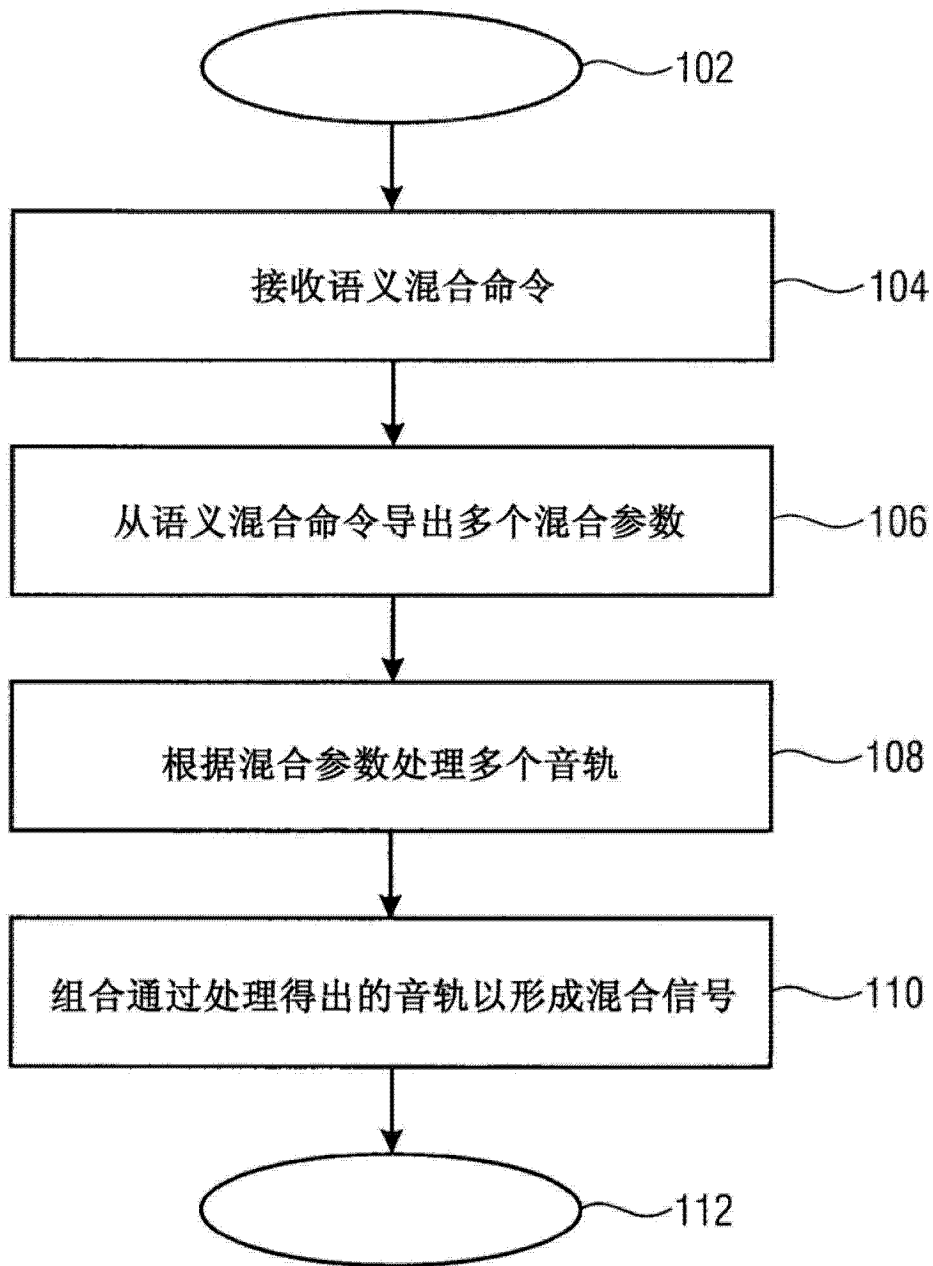


图 18