

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04S 3/00 (2006.01)

G10L 19/00 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200780001908.2

[43] 公开日 2009 年 5 月 13 日

[11] 公开号 CN 101433099A

[22] 申请日 2007.1.5

[21] 申请号 200780001908.2

[30] 优先权

[32] 2006.1.5 [33] US [31] 60/743,096

[86] 国际申请 PCT/SE2007/000006 2007.1.5

[87] 国际公布 WO2007/078254 英 2007.7.12

[85] 进入国家阶段日期 2008.7.4

[71] 申请人 艾利森电话股份有限公司

地址 瑞典斯德哥尔摩

[72] 发明人 A·塔尔布 E·卡尔森

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 胡莉莉 张志醒

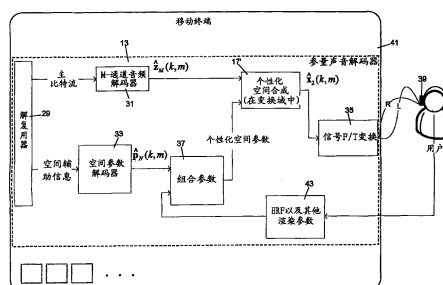
权利要求书 3 页 说明书 11 页 附图 9 页

[54] 发明名称

多声道环绕声的个性化解码

[57] 摘要

在多通道解码器(13)中接收参量多声道环绕音频比特流。接收到的空间参数在组合单元(37)中被变换为一组新的空间参数,该组新的空间参数被使用,以便获得对多声道环绕声的解码,该多声道环绕声并不是原始输入多声道环绕信号的简单等效物,而是例如可以通过基于从单元(43)获得的与用户头相关滤波器的表示进行变换来个性化。这样的个性化后的空间参数也可以通过将接收到的空间参数和与用户头相关滤波器的表示与一组附加渲染系数相组合而获得,所述附加渲染系数例如由用户交互式地来确定并且因此是与时间有关的。



1. 一种解码由参量多通道解码器接收到的参量多声道环绕音频比特流的方法，该方法包含以下步骤：

- 解复用所述比特流，以形成主比特流和空间辅助信息，
- 解码该空间辅助信息，以形成第一组空间参数，
- 修改该第一组空间参数，以形成第二组空间参数，
- 基于或者使用该第二组空间参数，由所述主比特流合成要被提供给收听设备的环绕音频信号。

2. 根据权利要求 1 所述的方法，其特征在于，在修改的步骤中，第二组空间参数是通过将第一组空间参数和与用户头相关滤波器的表示相组合来获得的，使得新参数被个性化并且环绕音频信号也被个性化。

3. 根据权利要求 2 所述的方法，其特征在于，在组合的步骤中，所接收到的空间参数以及与用户头相关滤波器的表示也与用户确定的附加渲染参数相组合。

4. 根据权利要求 3 所述的方法，其特征在于，所述附加渲染参数是响应于用户选择的交互式参数集。

5. 根据权利要求 3 所述的方法，其特征在于，所述附加渲染参数是与时间有关的。

6. 一种将表示声音的数字数据发送到移动单元的方法，该数字数据包括第一数目(N)的第一通道，每个第一通道特别是表示具有特殊特征的声音，诸如从特定方向和在特定频带中接收到的声音，该方法包含以下步骤：

- 分析所述数字数据，以确定声音的参数特征，所述参数特别是被确定来表示由每个第一通道中的数字数据表示的声音之间的空间关系，
- 将第一通道的数字数据彼此降混，以产生第二数目(M)的第二通道中的数字数据，该第二数目小于第一数目 ($M < N$)，
- 将第二通道中的数字数据和将参数以无线方式发送到移动单元，
- 在移动单元中接收第二通道中的数字数据和接收参数，
- 基于接收到的参数，变换接收到的第二通道中的数字数据，以产生适于被渲染到移动单元的声音发射器的变换后的数字数据，并且

将变换后的数字数据渲染到移动单元的声音发射器，

其特征在于在变换的步骤之前修改接收到的参数以形成用于变换步骤中的新参数的附加步骤。

7. 一种用于解码参量多声道环绕音频比特流的参量环绕解码器，该比特流包括指示了在由解码器接收到的比特流的通道中所表示的声音特性的空间参数，其特征在于修改单元，用于修改所述空间参数，以形成在合成中使用的新的空间参数，使得获得对原始的多声道环绕声的不同解码。

8. 根据权利要求 7 所述的参量环绕解码器，其特征在于，在修改空间参数时，修改单元被设置来使用与用户头相关滤波器的表示，使得新参数被个性化并且最后所得到的环绕音频信号也被个性化。

9. 根据权利要求 8 所述的参量环绕解码器，其特征在于，在修改空间参数时，修改单元被设置来还使用由用户确定的附加渲染参数。

10. 根据权利要求 7 所述的参量环绕解码器，其特征在于，修改单元被设置来以与时间有关的方式修改空间参数。

11. 一种包含参量环绕解码器的移动终端，该参量环绕解码器用于解码由移动单元接收到的参量多声道环绕音频比特流，该比特流包括指示在所接收到的比特流解码器的通道中所表示的声音特性的空间参数，其特征在于，该参量环绕解码器包括用于修改所述空间参数的修改单元，以形成在合成中使用的新的空间参数，使得获得对原始的多声道环绕声的不同解码。

12. 根据权利要求 11 所述的移动终端，其特征在于，在修改空间参数时，修改单元被设置来使用与用户头相关滤波器的表示，使得新参数被个性化并且最后所得到的环绕音频信号也被个性化。

13. 根据权利要求 12 所述的移动终端，其特征在于，在修改空间参数时，修改单元被设置来还使用由用户确定的或者用户输入的附加渲染参数，诸如通过压下移动单元的一个或多个键来输入。

14. 根据权利要求 12 所述的移动终端，其特征在于，修改单元被设置来根据来自用户的输入交互式地修改所述空间参数。

15. 根据权利要求 7 所述的移动终端，其特征在于，修改单元被设置来以与时间有关的方式修改空间参数。

16. 一种解码参量多声道环绕音频比特流的方法，该参量多声道环绕音

频比特流包括第一数目(N)的音频通道,所述比特流由参量多通道解码器接收,该方法包括以下步骤:

- 解复用所述比特流,以形成主比特流和空间辅助信息,
- 解码该主比特流,以形成针对所述多个音频通道的单独比特流,
- 解码该空间辅助信息,以形成第一组空间参数,
- 基于或者使用该第一组空间参数,由所述单独的比特流合成第二数目(M)的音频通道中的适于被提供给收听设备的环绕音频信号,其中第二数目(M)小于第一数目(N)。

17. 根据权利要求16所述的方法,其特征在于,第二数目(M)等于2。

18. 根据权利要求16所述的方法,其特征在于,第一数目(N)等于5或6。

多声道环绕声的个性化解码

相关申请

本申请要求于 2006 年 1 月 5 日提交的序列号为 No.60/743,096 的美国临时专利申请的优先权和利益，该美国临时专利申请的全部教导通过引用结合于此。

技术领域

本发明涉及对多声道环绕音频比特流的解码。

背景技术

长期以来，在全世界的电影院中，多声道环绕音频系统将电影观众置于电影场景的音频空间的中心，这些电影场景正在观众的面前被播放并且正给观众真实而令人信服的“身临其境”的感觉。这种音频技术已经作为家庭环绕影院系统进入普通人的家中，并且现在正在他们自己的客厅中给他们提供“身临其境”的感受。

该音频技术将被使用的下一领域包括移动无线单元或者终端，特别是包括诸如蜂窝电话和 PDA 这样的小单元。由于这些单元具有小尺寸的显示器，所以环绕声的融入式特性（immersive nature）甚至更重要。然而，将这种技术迁移到移动单元并不是小事。主要障碍包括：

- 1、在无线移动通道中，可用的比特率在很多情况下低。
- 2、移动终端的处理能力常常有限。
- 3、小的移动终端通常只有两个微型扬声器和耳塞或耳机。

特别是对于诸如蜂窝电话这样的移动终端而言，这意味着用于移动终端的环绕声解决方案不得不采用比在杜比数字 5.1 系统中所使用的 384kbits/s 低很多的比特率。由于处理能力有限，移动终端的解码器必须在计算方面被优化，并且由于移动终端的扬声器配置，该环绕声必须通过耳塞或耳机来输送。

通过耳机或耳塞输送多声道环绕声的标准方式是对每个扬声器信号执行

3D 音频或双声道立体声渲染 (binaural rendering)。

通常, 在 3D 音频渲染中, 使用音频场景的模型, 并且每个引入的单声道信号通过一组滤波器进行滤波, 这些滤波器模型化由人的头部、躯干和耳朵所产生的变换。这些滤波器被称为具有与头相关传递函数 (HRTF) 的与头相关滤波器 (HRF, head related filter), 并且如果进行适当的设计, 这些滤波器给出良好的 3D 音频场景感知。

图 1 的图解出了一种根据杜比数字 5.1 系统对音频信号进行完整的 3D 音频渲染的方法。根据杜比数字 5.1 系统的六个多声道信号是:

- 右环绕 (SR),
- 右 (R),
- 中 (C),
- 低频 (LFE),
- 左 (L)
- 左环绕 (SL)

在图 1 所示的实例中, 中信号和低频信号被组合成一个信号。然后, 需要五个不同的滤波器 H_I^B 、 H_C^B 、 H^C 、 H_I^F 和 H_C^F , 以便实现与头相关滤波的这种方 法。SR 信号被输入到滤波器 H_I^B 和 H_C^B , R 信号被输入到滤波器 H_I^F 和 H_C^F , C 和 LFE 信号被共同地输入到滤波器 H^C , L 信号被输入到滤波器 H_C^F 和 H_I^F , 并且 SL 信号被输入到滤波器 H_C^B 和 H_I^B 。从滤波器 H_I^B 、 H_I^F 、 H^C 、 H_C^F 和 H_C^B 输出的信号在右求和元件 1R 中被求和, 以给出意图被提供给右侧耳机 (未示出) 的信号。从滤波器 H_C^B 、 H_C^F 、 H^C 、 H_I^F 和 H_I^B 输出的信号在左求和元件 1L 中被求和, 以给出意图被提供给左侧耳机 (未示出) 的信号。

这样的渲染的 3D 感知方面的质量取决于 HRF 有多接近地模型化或表示在听众正在收听时与她/他自己的头相关的滤波。因此, 有利的是, 如果希望得到好的或者非常好的质量, 则 HRF 能够针对每个听众被适配和被个性化。该适配和个性化步骤可以包括模型化、测量和通常取决于用户的调谐, 以便改善所感知到的 3D 音频场景的质量。

目前的现有技术的标准化多通道音频编解码器要求大量的带宽或高比特率, 以便达到可接受的质量, 并且因此, 这些要求阻止将这样的编解码器用于诸如无线移动流的业务中。

例如,即使杜比数字 5.1 系统(AC-3 编解码器)与 AAC 多通道编解码器相比具有非常低的复杂度,对于类似的质量,该杜比数字 5.1 系统仍然要求更高的比特率。AAC 多通道编解码器和 AC-3 编解码器这两种编解码器直到今天仍然不可用于无线移动领域中,因为这两种编解码器对计算复杂度和比特率的要求都很高。

已经开发出新的基于双耳线索编码(binaural cue coding)原理的参量多通道编解码器(parametric multi-channel codec)。最近标准化的参量立体声工具是用于编码立体声的低复杂度/高质量参量技术的好实例。在 MPEG 中,参量立体声到多声道编码的扩展目前正以空间音频编码(Spatial Audio coding)为名而被标准化,并且其也被称为 MPEG-环绕。

可以根据图 2 的框图对参量多声道编码的原理进行解释和理解,图 2 示出了一般情况。参量环绕编码器 3(也被称为多通道参量环绕编码器)接收到多声道复合音频信号,该多声道复合音频信号包括各个信号 $x_1(n)$ 至 $x_N(n)$,其中 N 是输入通道的数目。如上所述,对于杜比数字 5.1 环绕系统而言, $N=6$ 。然后,该编码器 3 在降混单元(down-mixing unit)5 中形成复合的降混后的信号,该复合的降混后的信号包含各个降混后的信号 $z_1(n)$ 至 $z_M(n)$ 。降混后的通道数目 M ($M < N$) 取决于所要求的或所允许的最大比特率、所要求的质量以及 M -通道音频编码器 7 的可用性。编码过程的一个关键方面是降混后的复合信号得自多声道输入信号,该降混后的复合信号通常是立体声信号但是也可以是单声道信号,并且在音频编码器 7 中被压缩用于通过无线通道 9 传送的信号是这个降混后的复合信号而不是原始的多声道信号。参量编码器 3 并且特别是其降混单元 5 能够执行降混过程,以致其创建单声道的或立体声的降混中的多声道信号的或多或少真正的等效物。该参量环绕编码器还包括空间参数估计单元 9,该空间参数估计单元 9 由输入信号 $x_1(n)$ 至 $x_N(n)$ 计算出线索或空间参数,这些线索或空间参数以某种方式可以被说成描述降混过程或其中所做的假设。压缩后的音频信号从 M -通道音频编码器被输出并且也是主信号,该压缩后的音频信号与构成辅助信息(side information)的空间参数一起通过诸如无线接口的接口 11 被传送到接收侧,该接收侧在这里所考虑的情况下通常是移动终端。

可替换地,该降混可以由某个外部单元来供给,诸如由使用艺术家风格

降混 (Artistic Downmix) 的单元来供给。

在接收侧, 互补的参量环绕解码器 13 包括音频解码器 15, 并且应该被构造来能够基于在发送侧使用的降混算法的知识以及与压缩后的多声道信号并行被接收到的编码后的空间参量或线索来产生最好的可能的多声道解码。该音频解码器 15 产生信号 $\hat{z}_1(n)$ 至 $\hat{z}_M(n)$, 这些信号 $\hat{z}_1(n)$ 至 $\hat{z}_M(n)$ 应该尽可能与发送侧的信号 $z_1(n)$ 至 $z_M(n)$ 类似。这些信号和空间参数一起被输入到空间合成单元 17, 该空间合成单元 17 产生输出信号 $\hat{x}_1(n)$ 至 $\hat{x}_N(n)$, 这些输出信号应该尽可能地与发送侧的原始输入信号 $x_1(n)$ 至 $x_N(n)$ 类似。输出信号 $\hat{x}_1(n)$ 至 $\hat{x}_N(n)$ 可以被输入到如图 1 中所示的双声道立体声渲染系统。

很明显, 根据接口 11 上通常相对低的发送通道带宽, 将会出现信息的丢失, 因此在接收侧的信号 $\hat{z}_1(n)$ 至 $\hat{z}_M(n)$ 以及 $\hat{x}_1(n)$ 至 $\hat{x}_N(n)$ 不能与他们在发送侧的配对信号相同。即使接收侧的信号并不是其配对信号的非常真实的等效信号, 这些接收侧的信号也可以是足够好的等效物。

通常, 这样的环绕编码过程独立于用于在图 2 中的音频编码器 7 和音频解码器 15 这些单元中所使用的所发送的通道的压缩算法。该编码过程可以使用多种高性能压缩算法中的任何算法, 这些高性能压缩算法诸如 AMR-WB+、MPEG-1 第三层 (Layer III)、MPEG-4 AAC 或者是 MPEG-4 高效 AAC, 并且该编码过程甚至可以使用 PCM。

通常, 上述操作是在变换过的信号域中完成的, 变换诸如是傅立叶变换或 MDCT。如果在单元 9 和 17 中的空间参数估计和合成使用与在音频编码器 7 (也被称为核心编解码器) 中所使用的相同变换类型, 那么这是特别有益的。

图 3 是有效的参量音频编码器的详细框图。N-通道离散时间输入信号 (用向量形式 $\mathbf{x}_N(n)$ 表示) 首先在变换单元 21 中被变换到频域, 并且通常被变换到给出信号 $\mathbf{X}_N(k, m)$ 的变换域。索引 k 是变换系数的索引, 或者如果选择频域变换, 那么索引 k 就是子带的索引。索引 m 表示十中抽一的 (decimated) 时域索引, 该十中抽一的时域索引也与可能通过重叠帧的输入信号相关。

此后, 信号在降混单元 5 中被降混, 以产生 M-通道降混信号 $z_M(k, m)$, 其中 $M < N$ 。空间模型参数向量 $\mathbf{p}_N(k, m)$ 序列在估计单元 9 中被估计。这可以以开环形式或者是闭环形式来完成。

空间参数包括代表环绕声感受的心理学声学线索。例如, 在 MPEG 环绕编码器中, 这些参数包括在等效于 ILD、ITD 和 IC 线索的电平、相位和相干性的通道间差异, 以捕获多声道音频信号相对于所发送的降混后的信号 $z_M(k, m)$ (或者如果在闭环中, 则为解码后的信号 $\hat{z}_M(k, m)$) 的空间图像。线索

$p_N(k,m)$ 可以以非常紧凑的形式诸如在空间参数量化单元 23 中被编码,从而产生信号 $\tilde{p}_N(k,m)$,其后跟随的是空间参数编码器 25。M-通道音频编码器 7 产生主比特流,该主比特流在复用器 27 中与由参数编码器产生的空间辅助信息一起被复用。来自复用器的复用后的信号被发送到接收侧的解复用器 29,在解复用器 29 中,辅助信息和主比特流如在图 4 的框图中所看到的那样被恢复。

在接收侧,主比特流被解码来使用接收到的空间参数合成高质量的多通道表示。主比特流首先在 M-通道音频解码器 31 中被解码,解码后的信号 $\hat{z}_M(k,m)$ 从该 M-通道音频解码器 31 被输入到空间合成单元 17。含有空间参数的空间辅助信息由解复用器 29 提取并被提供给空间参数解码器 33,该空间参数解码器 33 产生解码后的参数 $\hat{p}_N(k,m)$ 并将这些参数发送给合成单元 17。空间合成单元产生信号 $\hat{x}_N(k,m)$,该信号 $\hat{x}_N(k,m)$ 被提供给信号 F/T 变换单元 35,从而变换到时域,以产生信号 $\hat{x}_N(n)$,也就是多声道解码信号。

多声道环绕声的 3D 音频渲染可以通过如下方式被输送给移动终端用户:使用有效的参量环绕解码器来首先获得多个环绕声声道,例如使用如上参考图 4 所述的多通道解码器。因此,图 1 中所示的系统被用于合成双声道立体声的 3D 音频渲染过的多声道信号。该操作在图 5 的示意图中被示出。

也已完成其中已在子带域中执行的空间或 3D 音频滤波的工作。在 C.A. Lanciani 和 R.W. Schafer 在 1999 年 3 月 21 日至 23 日于美国亚拉巴马州的奥本举行的第 31 届系统理论研讨会上发表的“Application of Head-related Transfer Functions to MPEG Audio Signals”中,公开了 MPEG 编码后的单声道信号如何通过子带域中执行 HR 滤波操作而被空间化。在 A.B. Touimi、M. Emerit 和 J.M. Pernaux 在 2004 年 10 月 27 日至 29 日于美国马里兰州学院公园举行的第三届移动和普适多媒体国际会议论文集第 229-235 页发表的“Efficient Method for Multiple Compressed Audio Streams Spatialization”中,公开了多个单独进行 MPEG 编码的单声道信号可以通过子带域中进行 HR 滤波操作而被空间化。该解决方案是基于 HR 滤波器的特殊实施方案,其中所有的 HR 滤波器被模型化为数个预定的基本滤波器的线性组合。

3D 音频渲染的应用有很多并且包括游戏、移动 TV 显示、使用诸如 3GPP MBMS 或 DVB-H 的标准、听音乐会、观看电影以及普通的多媒体业务,这些应用包含多声道音频分量。

上述的渲染多声道环绕声的方法尽管由于其允许要被提供给无线移动单元的一整套新业务而变得引人注目,但是这些方法还是有很多缺陷:

首先, 这样的渲染的计算需求受到抑制, 因为解码和 3D 渲染都必须并行且实时地执行。即使参量多通道解码器的复杂度与全波形多通道解码器相比是低的, 但参量多通道解码器的复杂度仍然十分高并且至少要高于简单的立体声解码器。空间解码的合成阶段具有与所编码的通道的数目至少成比例的复杂度。此外, 3D 渲染的滤波操作同样与通道的数目成比例。

第二个缺点包括为了存储中间的解码过的通道而需要的临时存储器。由于在 3D 渲染的第二阶段需要这些中间的解码过的通道, 所以这些中间的解码过的通道实际上被缓存了。

最后, 通常作为语音和音频编解码器的部分的可能的后处理步骤可能会影响这样的 3D 音频渲染的质量。对于在扩音器的环境中进行收听而言, 这些后处理是有益的。然而, 这些后处理引入了严重的非线性的相位失真, 该相位失真不均衡地分布在多个通道上并且可能影响 3D 音频渲染质量。

发明内容

本发明的目的是提供一种解码参量多声道环绕音频比特流的有效且通用的方法。

本发明的另一目的是提供一种移动终端, 其中参量多声道环绕音频比特流能够有效地被解码来产生适于被提供给移动终端中的或者连接到移动终端的收听设备的一个或多个信号。

在解码参量多声道环绕音频比特流的方法中, 使用了诸如解码多声道环绕声和特别是双声道立体声解码多声道环绕声的概念。

在这样的方法中, 由参量多通道解码器所接收到的空间参数可以被变换为一组新的空间参数, 使用这组新的空间参数以便获得对多声道环绕声的不同解码。

变换后的参数也可以是个性化的空间参数, 并且可以通过将接收到的空间参数和与用户头相关滤波器的表示进行组合而获得。

该个性化的空间参数也可以通过将接收到的空间参数、与用户头相关滤波器的表示以及由用户确定的一组附加渲染参数进行组合而获得。

该组附加渲染参数的子集可以是交互式参数, 这些交互式参数响应于在收听过程期间可以被改变的用户选择来设置。

这组附加渲染参数可以是与时间相关的参数。

这里所描述的方法可以允许一种简单而有效的方式来渲染环绕声，其通过移动设备上的参量编码器进行编码。主要的优点包括降低了复杂度并且在使用移动设备通过耳机进行收听时增加了交互性。

本发明的另外的目的和优点将在随后的说明中被阐述，并且从这些说明中，一部分是显而易见的，或者可以通过本发明的实践而认识到。本发明的目的和优点可以通过特别是在随附的权利要求中指出的方法、过程、手段以及组合来实现和获得。

附图说明

虽然本发明的新颖性特征已经在随附的权利要求书中详细描述，对上述的本发明及其他特征的全面理解（无论是组织结构上和还是内容上）可以通过考虑下面参考附图呈现的非限制性实施例的具体描述来获得，并且通过考虑该具体描述将更好地被理解，其中：

图1是框图，该框图示出了对5.1音频信号的可能的3D音频或双声道立体声渲染，

图2是对参量多声道编码和解码系统的原理的高级说明，

图3是对参量多通道音频编码器的详细说明，

图4是对参量多通道音频解码器的详细说明，

图5是解码后的多声道信号的3D音频渲染（现有技术），

图6是多声道环绕声的个性化的双声道立体声解码，

图7是MPEG-环绕解码器中的空间音频处理的广义图，

图8是用于个性化双声道立体声解码的本发明的实施例，

图9是示出组合参数的示意图，并且

图10是示出收听测试结果的图解。

具体实施方式

图6的框图示出了如在参量声音解码器13中执行的对参量多声道环绕音频比特流进行解码的方法中的主要步骤。在解复用器29中，主比特流和空间辅助信息被恢复。该主比特流首先在M-通道音频解码器31中被解码，解码

后的信号 $\hat{x}_s(k, m)$ 从该 M-通道音频解码器 31 被输入到个性化的空间合成单元 17'。含有空间参数的空间辅助信息从解复用器 29 被提供给产生解码后的参数 $\hat{p}_s(k, m)$ 的空间参数解码器 33。解码后的空间参数被输入到参数组合单元 37, 该参数组合单元 37 还可接收其他参数信息, 特别是接收个性化的参数和 HRF 信息。该组合单元产生新的参数, 这些新的参数特别是可以是个性化的空间参数并被输入到合成单元 17'。该空间合成单元产生被提供给信号 F/T 变换单元 35 的信号 $\hat{x}_2(k, m)$, 从而变换回时域。时域信号被提供给例如其中正在运行参量环绕解码器的移动终端 41 的耳机 39。由组合单元 37 所接收到的附加信息和参数可以由参数单元 43 获得, 该参数单元 43 例如可以被构造来在收听时间期间诸如通过压下移动终端或单元 41 的某个适当的键来交互式地接收用户输入。

现在将描述如在 MPEG 环绕多通道解码器中实施的方法, 参照 ISO/IEC 14496-3:200X/PDAM 4 (MPEG Surround, N7530, 2005 年 10 月, 尼斯, 法国) 的文本。然而, 很明显该方法同样能够在其他环境下很好地被使用。

MPEG 环绕解码器中的处理可以通过如在图 7 的图解中所示出的两个矩阵乘法来限定, 这些乘法被示为包括相应的信号被输入到其中的矩阵单元 M1 和 M2 (也被分别称为预去相关器(preddecorrelator)矩阵单元和混合矩阵单元)。第一矩阵乘法形成去相关单元或去相关器 D₁、D₂、... 的输入信号, 而第二矩阵乘法基于降混输入和去相关器的输出形成输出信号。上述操作针对每个混合子带完成, 混合子带由混合子带索引 k 进行索引。

在下文中, 索引 n 被用于多个时隙, k 被用于对混合子带进行索引, 而 l 被用于对参数集进行索引。然后, 对输入通道处理以形成输出通道可以被描述为:

$$\mathbf{v}^{n,k} = \mathbf{M}_1^{n,k} \mathbf{x}^{n,k}$$

$$\mathbf{y}^{n,k} = \mathbf{M}_2^{n,k} \mathbf{w}^{n,k}$$

其中, $\mathbf{M}_1^{n,k}$ 是将一定数目的输入通道映射到一定数目的进入去相关器的通道的二维矩阵, 并且针对每个时隙 n 和每个混合子带 k 被定义, 而 $\mathbf{M}_2^{n,k}$ 是将一定数目的预处理过的通道映射到一定数目的输出通道的二维矩阵, 并且针对每个时隙 n 和每个混合子带 k 被定义。根据是使用去相关过的信号的时域时间整形 (TP) 还是使用去相关过的信号的时间包络整形 (TES), 矩阵 $\mathbf{M}_2^{n,k}$ 成

为两个版本，这两个版本被表示为 $\mathbf{M}_{2_wv}^{n,k}$ 和 $\mathbf{M}_{2_dy}^{n,k}$ 。

第一矩阵单元 M1 的输入向量 $x^{n,k}$ 对应于从音频解码器 31 获得的图 6 的解码过的信号 $\hat{z}_M(k, m)$ 。被输入到混合矩阵单元 M2 的向量 $w^{n,k}$ 是去相关器 D₁、D₂、... 的输出 d₁、d₂、...、第一矩阵乘法（也就是预去相关器矩阵单元 M₁）的输出和残留信号 res₁、res₂、... 的组合，并且该向量 $w^{n,k}$ 针对每个时隙 n 和每个混合子带 k 被定义。输出向量 $y^{n,k}$ 具有基本上对应于如上所述的信号 L、SL、R、SR、C 和 LFE 的分量 l_f、l_s、r_f、r_s、cf 和 lfe。这些分量必须被变换到时域并且以某种方式被渲染，以被提供给所使用的耳机，也就是说它们不能直接被使用。

一种用于 3D 音频渲染并且特别是个性化解码的方法使用包含“由模型重构”块的解码器，该“由模型重构”块在混合滤波组（filter-bank）域中取诸如个人 3D 音频滤波器的表示的外部输入并且利用该外部输入将模型参数的衍生参数变换为其他模型参数，这允许直接在变换域中产生两个双声道立体声信号，使得与图 6 中的变换单元 35 相比只有双声道立体声的 2 声道信号必须被变换到离散的时域。

基于 MPEG 环绕的个性化双声道立体声解码的实施例在图 8 的图解中被示出。

第三矩阵 $\mathbf{M}_3^{n,k}$ （象征性地被示为参数修改矩阵 M3）在该实例中是从 6 个声道到两个声道的线性映射，这两个声道被用作通过变换单元 35 到用户耳机 39 的输入。该矩阵乘法可以被写为：

$$(\mathbf{r}) = \mathbf{M}_3^{n,k} \mathbf{y}^{n,k}$$

也可以完成附加的双声道立体声后处理并且在如在此所述的方法的范围之外。这可以进一步包括左声道和右声道的后处理。

通过线性化（结合律），清楚的是：矩阵 $\mathbf{M}_2^{n,k}$ 和 $\mathbf{M}_3^{n,k}$ 可以被一起组合来形成存储在新的混合矩阵 $\mathbf{M}_4^{n,k} = \mathbf{M}_3^{n,k} \mathbf{M}_2^{n,k}$ 中的一组新参数。这个组合操作在图 9 中被示出，其中对应于新矩阵的乘法单元被示为混合矩阵单元 M4 并且两个矩阵的乘法在乘法单元 45 中进行。

新的混合矩阵 $\mathbf{M}_4^{n,k}$ 具有的取决于比特流参数和用户预定义的与头相关滤波器 HRF 以及取决于其他动态渲染参数（如果期望的话）的参数。

对于只有与头相关滤波器的情况，矩阵 $\mathbf{M}_3^{n,k}$ 可以被写为：

$$\mathbf{M}_3^{n,k} = \begin{bmatrix} H_C^F(k) & H_C^B(k) & H_I^F(k) & H_I^B(k) & H^C(k) & H^C(k) \\ H_I^F(k) & H_I^B(k) & H_C^F(k) & H_C^B(k) & H^C(k) & H^C(k) \end{bmatrix}$$

该矩阵元素是五个不同的滤波器, 这些滤波器被用来实现与头相关滤波, 并且如上所述被表示为 H_I^B 、 H_C^B 、 H^C 、 H_I^F 和 H_C^F 。在这种情况下, 滤波器被表示在混合域中。表示滤波器从时域到频域或变换域的这样的操作在信号处理文献中是公知的。这里, 形成矩阵 $\mathbf{M}_3^{n,k}$ 的滤波器是混合子带索引 k 的函数并且类似于图 1 中所示的那些滤波器。

应该注意到, 对于这种情况, 矩阵 $\mathbf{M}_3^{n,k}$ 和时隙索引 n 无关。如果用户想要通过耳机 39 体验另一虚拟扬声器配置, 那么与头相关滤波器也可能动态地改变。

在另一实施例中, 用户可能想要交互地改变他的空间位置。到这时意味着: 则用户可能想要体验如何接近音乐会场景 (如果例如在播放现场音乐会) 或者远离音乐会场景。通过为参数修改矩阵 $\mathbf{M}_3^{n,k}$ 增加延迟线可以容易地实现这一点。用户动作可能是动态的, 并且在那种情况下, 矩阵 $\mathbf{M}_3^{n,k}$ 取决于时隙索引 n 。

在又一实施例中, 用户可能想要体验不同的空间感受。在这种情况下, 回响和其他音效可以有效地被引入矩阵 $\mathbf{M}_3^{n,k}$ 。

与用户交互性有关的、矩阵 $\mathbf{M}_3^{n,k}$ 的动态特性可以得益于两个用户动作之间的插值。参数插值的方法是众所周知的, 并且在这里不被说明。

如已阐述的那样, 参数修改矩阵 $\mathbf{M}_3^{n,k}$ 可以包含附加渲染参数, 这些附加渲染参数是可交互的并且响应于用户输入而改变。

上述本发明的特定实施例已经被实施和被测试为 MPEG 标准化成果的部分, 用于对 MPEG 环绕解码器的双声道立体声扩展。在图 10 的图解中示出了由独立组执行的多个收听测试的测试结果。可以清楚地看到, 根据本发明的特定实施例进行的双声道立体声渲染的感知到的质量对于大部分测试信号而言好于根据如图 5 中所示的标准 3D 音频后处理方法所获得的质量。

虽然在这里描述的实施例涉及用于双声道立体声耳机收听的解码, 但是对于本领域中的技术人员显而易见的是, 在不背离参数映射和组合的基本思想的情况下, 将这些实施例也应用于扬声器收听或其他空间配置。

虽然在这里对本发明的特定实施例进行了说明和描述, 但是应该认识到在不背离本发明的精神和范围的情况下, 本领域技术人员容易想到可被设想

的大量其他实施例和大量的附加优点、修改和变化。因此，本发明在其较宽的方面并不限于在此示出和描述的特定细节、代表性装置和所示的实例。因此，在不背离如随附的权利要求及其等效物所限定的一般发明概念的精神和范围的情况下，可以进行各种修改。因此，应理解的是，随附的权利要求意图覆盖诸如落在本发明的真正精神和范围之内内的所有修改和变化。在不背离本发明的精神和范围的情况下，应该想象到大量的其他实施例。

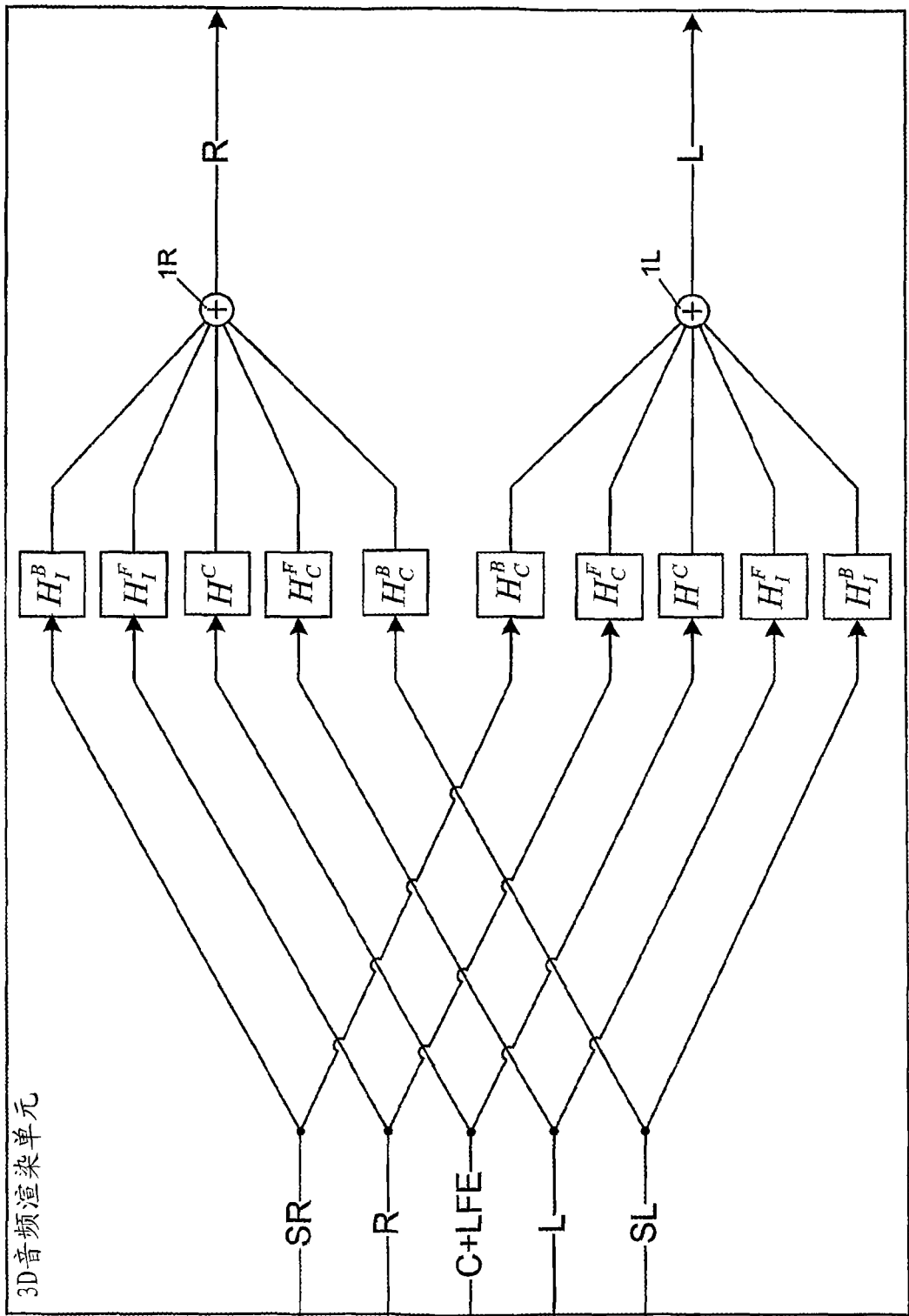


图 1
现有技术

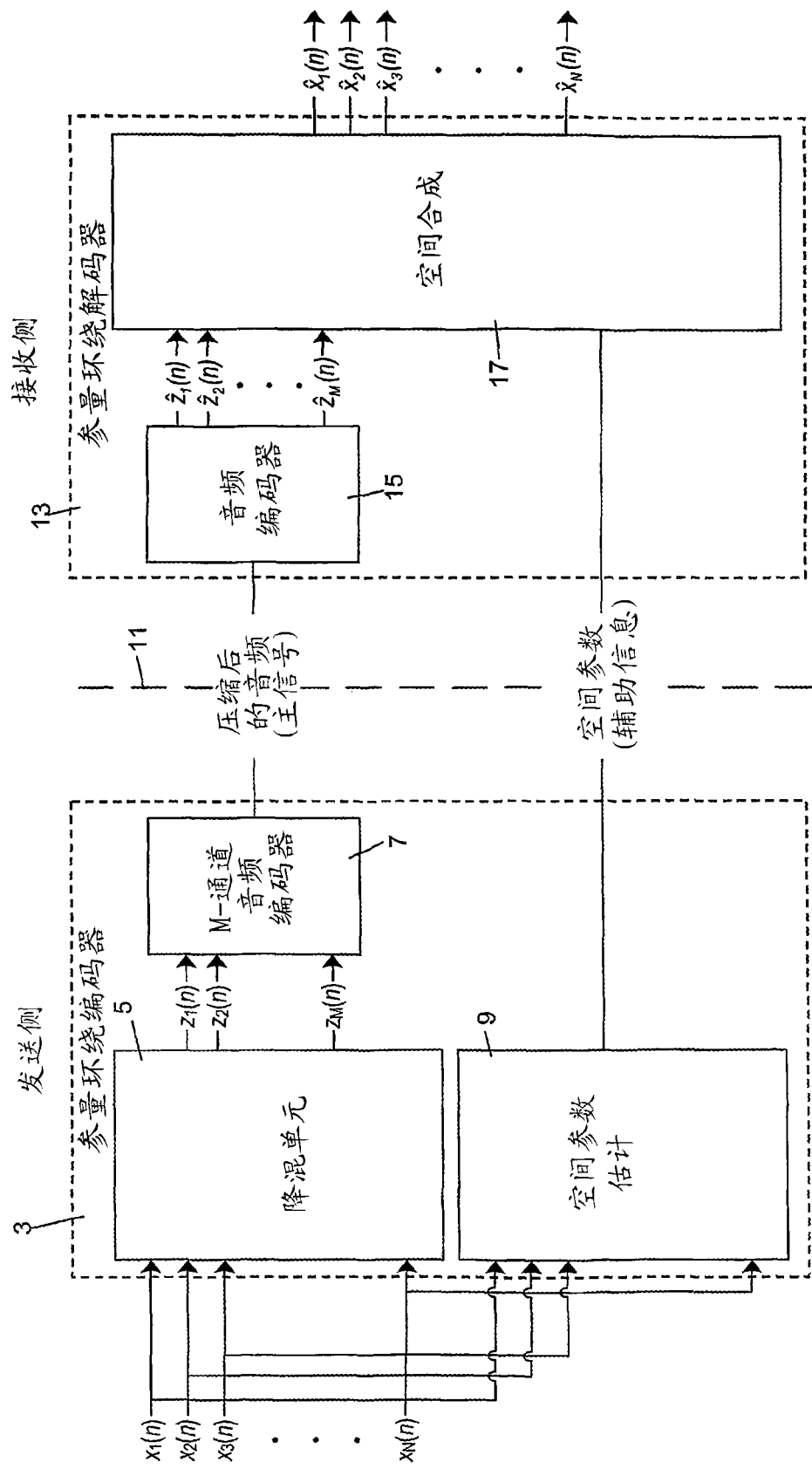


图 2
现有技术

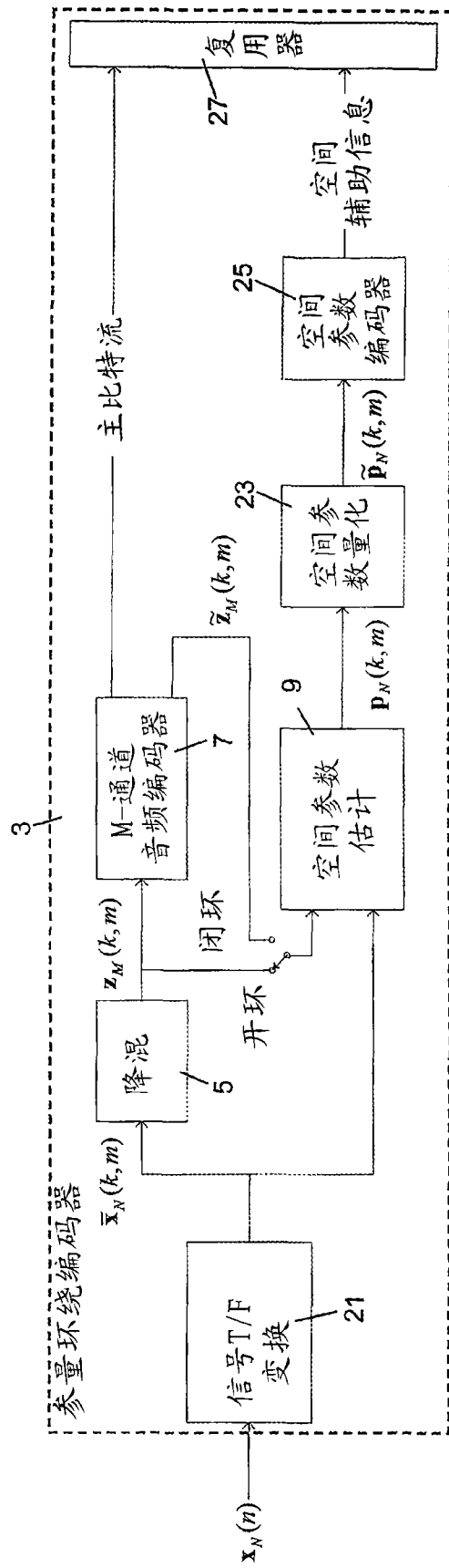


图 3
现有技术

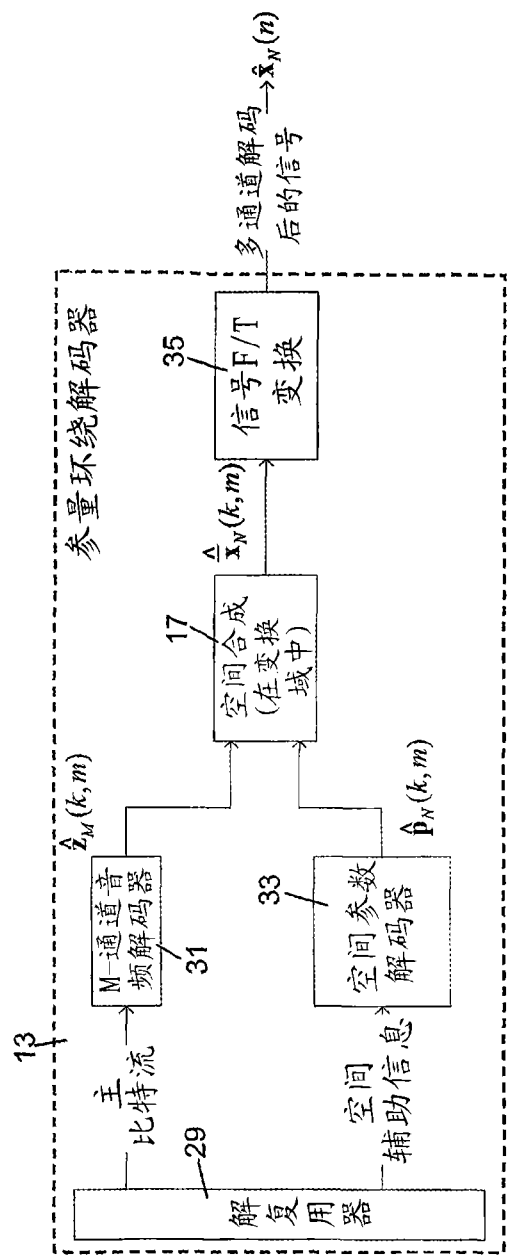


图 4
现有技术

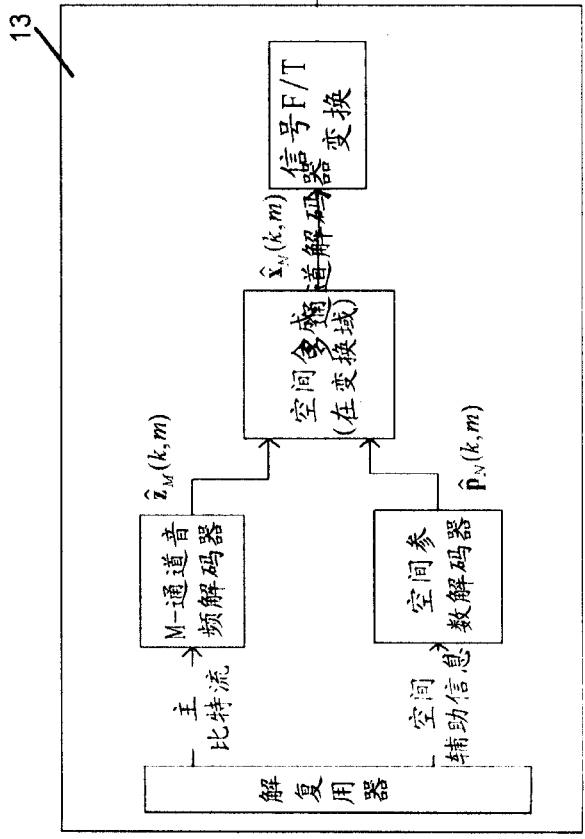
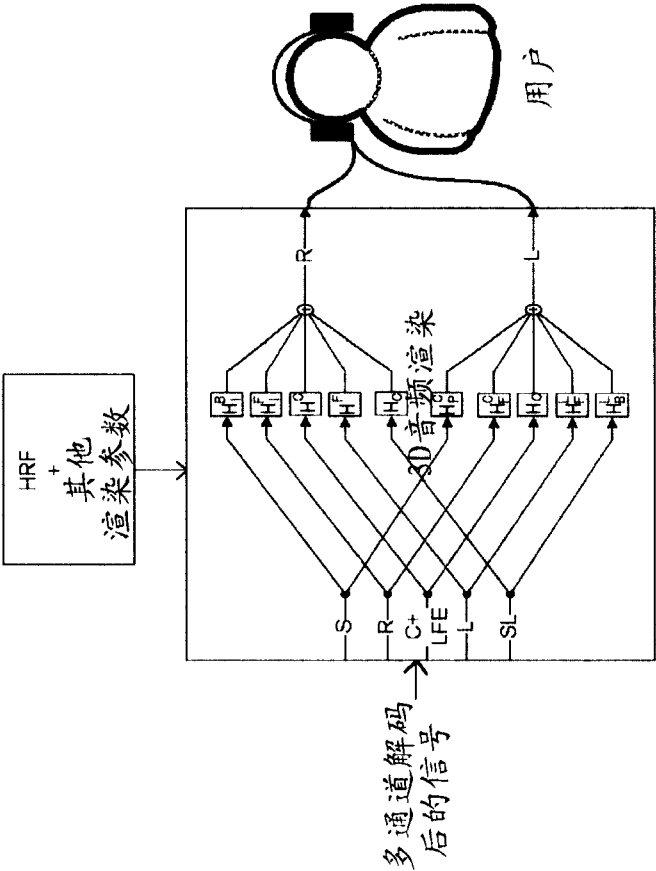


图 5
现有技术

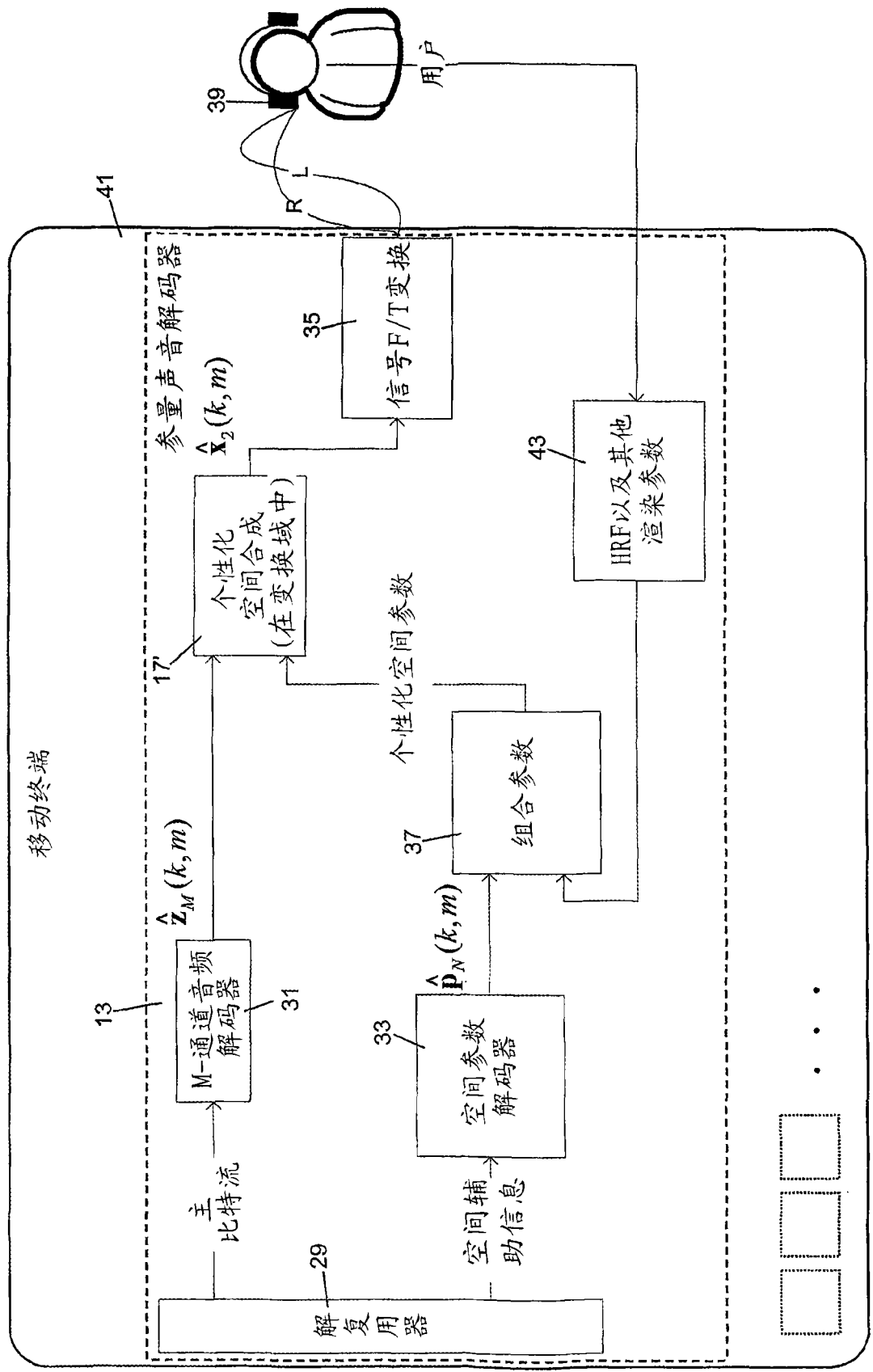
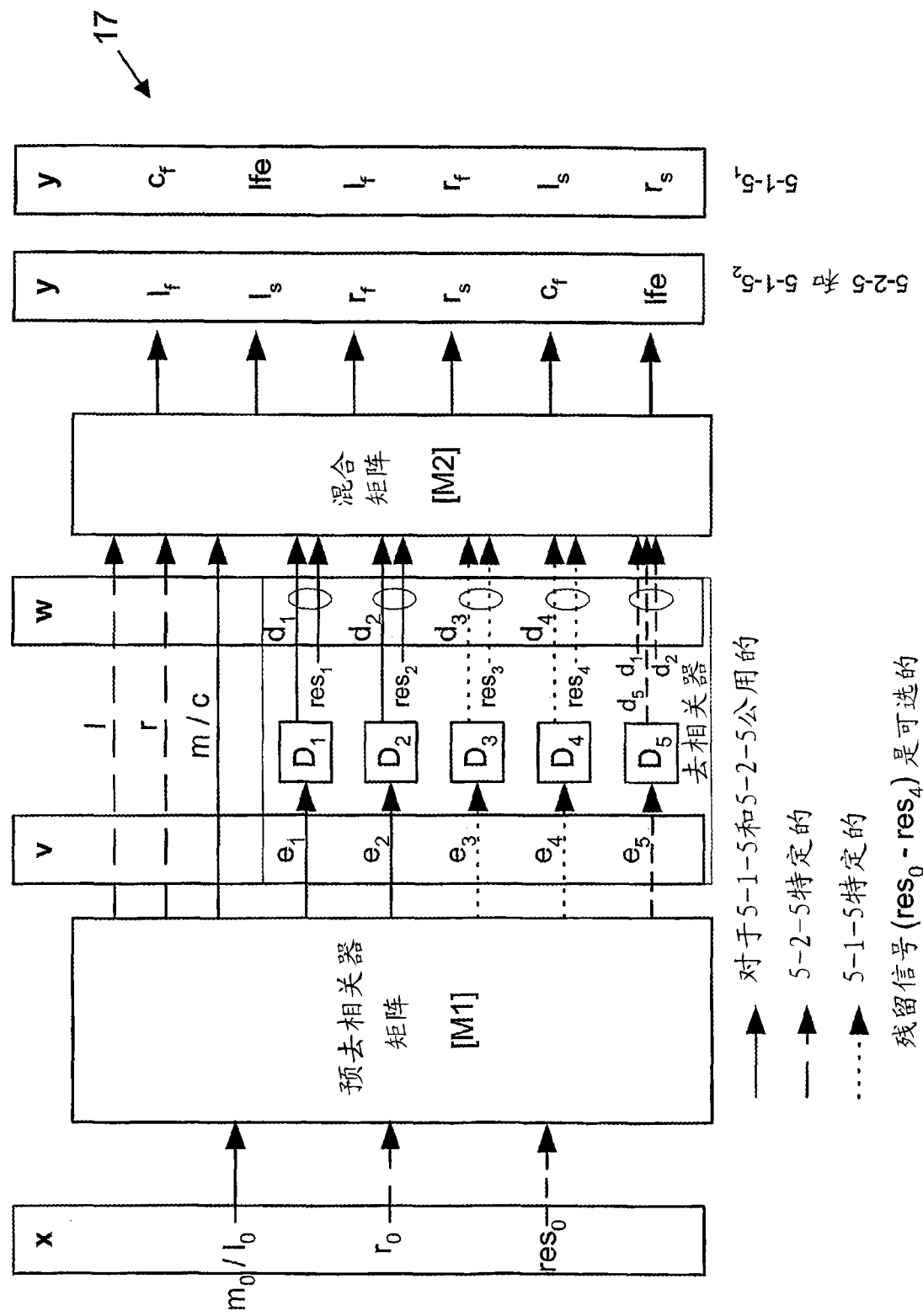


图 6



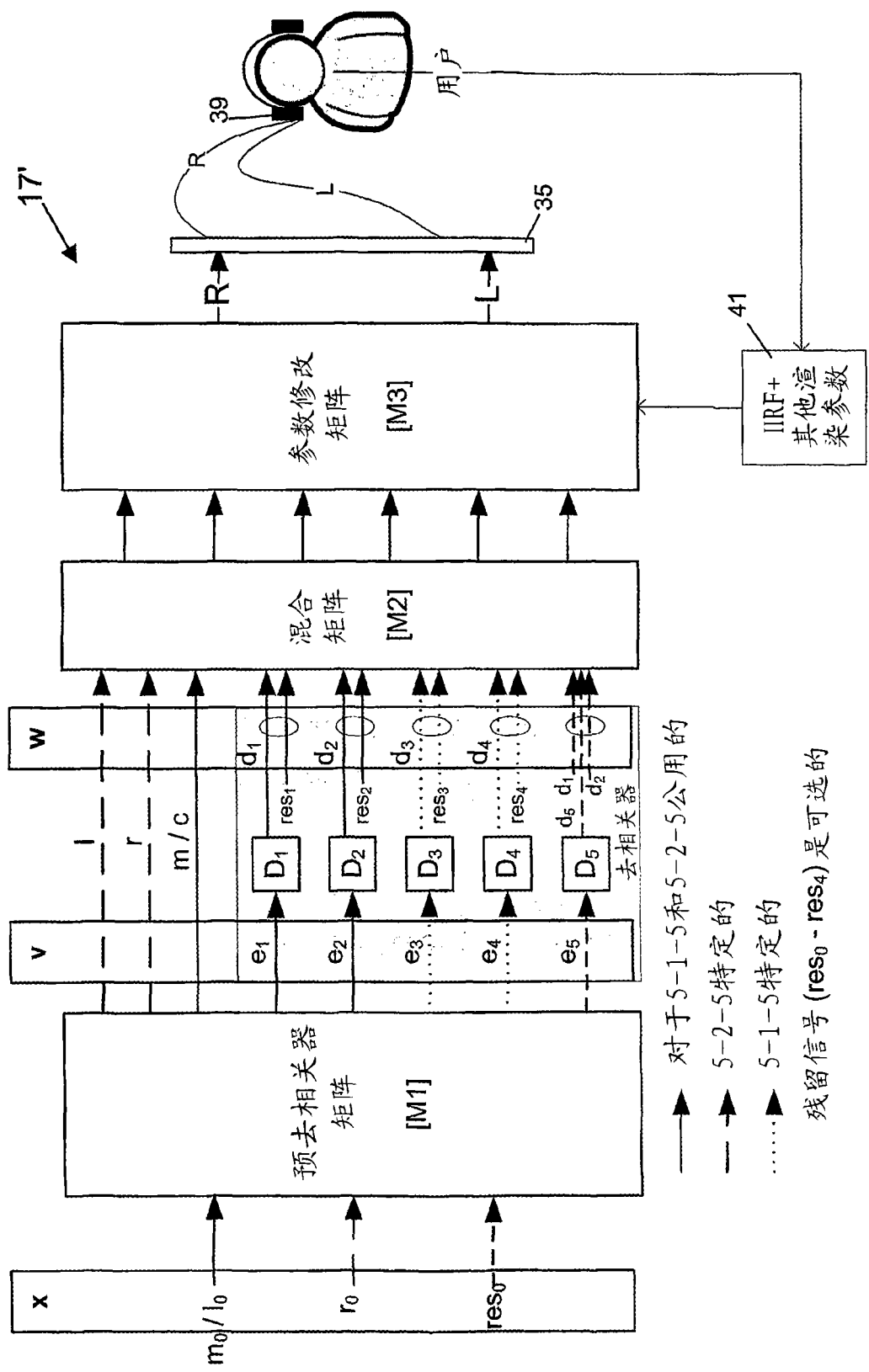
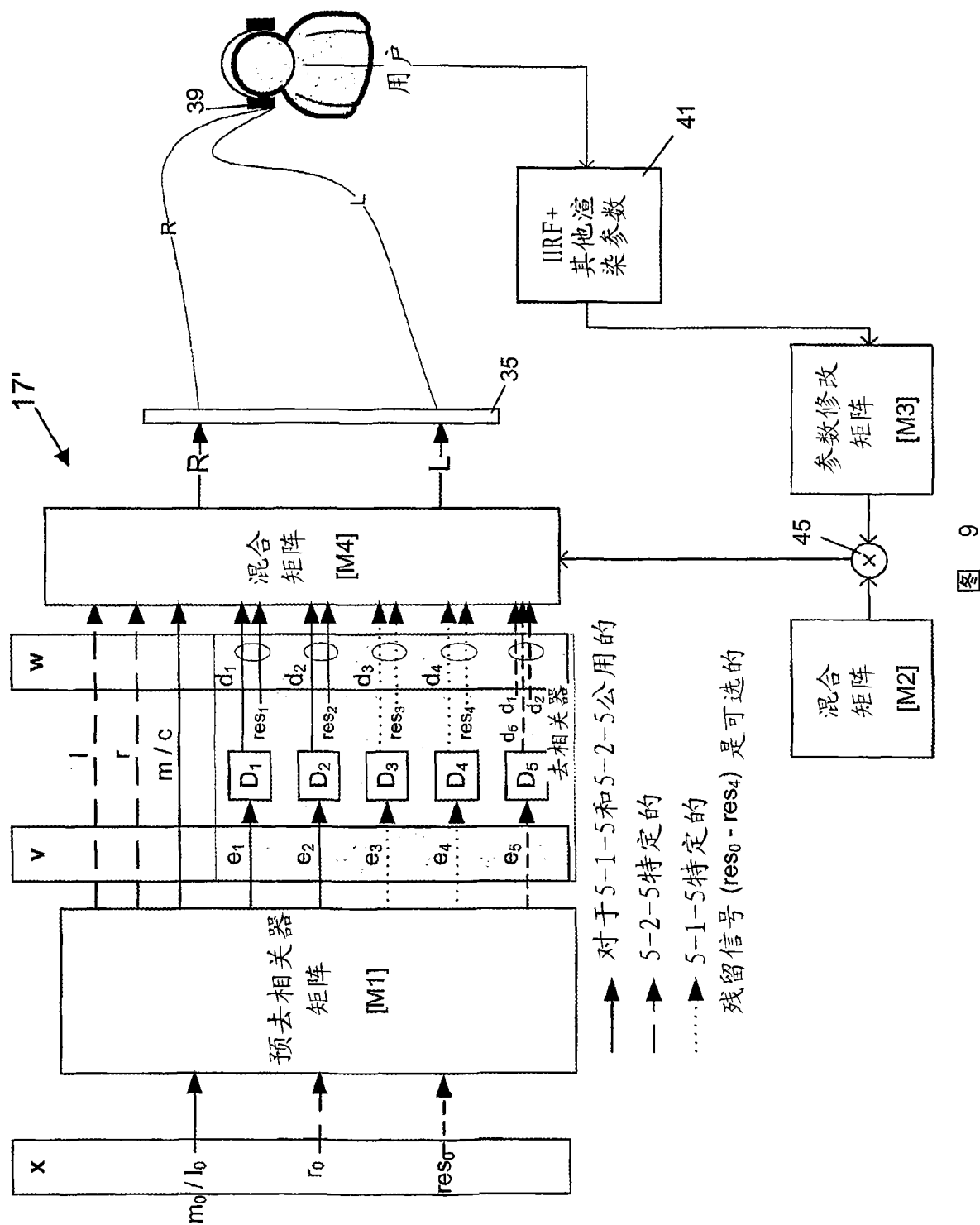


图 8

—— 对于5-1-5和5-2-5公用的
- - - 5-2-5特定的
..... 5-1-5特定的
残留信号 ($res_0 - res_4$) 是可选的



9

