

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04J 3/06 (2006.01)

H04N 9/475 (2006.01)



[12] 发明专利申请公开说明书

[21] 申请号 03808749.9

[43] 公开日 2006 年 3 月 8 日

[11] 公开号 CN 1745526A

[22] 申请日 2003.4.15 [21] 申请号 03808749.9

[30] 优先权

[32] 2002. 4. 19 [33] US [31] 60/374,269

[32] 2003. 1. 10 [33] US [31] 10/340,477

[86] 国际申请 PCT/US2003/011630 2003. 4. 15

[87] 国际公布 WO2003/090443 英 2003. 10. 30

[85] 进入国家阶段日期 2004. 10. 19

[71] 申请人 汤姆森许可公司

地址 法国布洛涅

[72] 发明人 C·C·王 A·M·奥斯特

S·S·莫蒂

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 程天正 梁 永

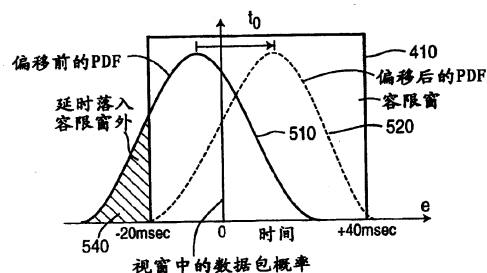
权利要求书 3 页 说明书 9 页 附图 4 页

[54] 发明名称

用于同步音频和视频流的设备和方法

[57] 摘要

公开了一种通过自适应瞬时误差(510)的统计学分布来建立一个新的瞬时误差(520)的统计学分布以用于在相应的音频和视频流中减少音频-视频同步问题(例如“唇同步”问题)的方法和设备。新的瞬时误差(520)的统计学分布完全在一个可接受的较少令观众/听众讨厌的同步容限窗内。



1. 一种方法, 包括:

接收 (610) 视频访问单元和相应的音频访问单元, 用于表示视一听信息的
5 所述视频和相应的音频访问单元有助于显示出由第一概率分布函数(pdf) (510) 描述的视一听瞬时同步误差; 以及

以一个延时量来瞬时延迟 (630) 接收到的音频和视频访问单元中之一, 用于表示视听信息的未延时和相应的延时的访问单元有助于显示由第二 pdf (520) 描述的视一听瞬时同步误差, 所述的第二 pdf 比所述第一 pdf 利用了同步容限窗
10 (410) 的更大份额。

2. 权利要求 1 的方法, 其中:

所述的第一 pdf 已使一个负的瞬时失配值和一个正的瞬时失配值与它相关联, 所述的瞬时失配值具有不同的绝对值; 以及

对所述的延时量进行选择以减小所述的负的和正的瞬时失配值 (630) 的绝对值之间的差。
15

3. 权利要求 2 的方法, 其中所述负瞬时失配值大约为 20 毫秒, 所述正瞬时失配值大约 40 毫秒, 并且所述音频访问单元被延时大约 10 毫秒。

4. 权利要求 1 的方法, 其中所述延时量在对产生所述音频和视频访问单元 (710) 的音频和视频流进行编码之前被传递。

20 5. 权利要求 1 的方法, 其中所述延时量在对产生所述音频和视频访问单元的音频和视频流进行编码期间被传递。

6. 权利要求 1 的方法, 进一步包括:

调整 (810, 820) 所述延时量以便响应在所述 pdf 中的变化。

7. 权利要求 6 的方法, 其中所述变化通过在接收器中检查音频和视频访问单元的显示时间标记来确定。
25

8. 权利要求 1 所述的方法, 进一步包括:

调整所述延时量以便响应在音频-视频瞬时同步误差产生源中的变化。

9. 权利要求 8 所述的方法, 其中所述音频-视频瞬时同步误差产生源包括一个可变延时网络 (140)。

30 10. 一种用于产生编码视频和相应的音频流的方法, 包括:

瞬时编码(120)相应的视频和音频信息以产生包含有相应的视频和音频访问单元的编码视频和音频流; 以及

以与一个同步误差容限模型相应的延迟量来瞬时延迟(710)所述编码视频和音频流之一。

5 11. 权利要求10所述的方法, 其中瞬时延时的所述步骤包括在所述编码之前延时所述视频和音频流中之一。

12. 权利要求10所述的方法, 其中每一个所述编码视频和音频流包括多个各自的视频和音频单元; 以及

10 所述瞬时延时的所述步骤包括通过所述延时量来自适应与所述的视频和音频访问单元中的至少一个相关的时间标记。

13. 权利要求10所述的方法, 其中与所述视频和音频流中的至少一个相关的时间标记通过一个预定的量被舍入。

14. 权利要求13的方法, 其中所述视频流时间标记通过以由如模运算处理的各自视频流时间标记来减少每个视频流时间标记而进行估计。

15 15. 权利要求10的方法, 其中瞬时延时量适应于响应在视一听瞬时同步误差产生源中的变化。

16. 一种设备, 包括:

一个延时元件(710), 用于传递一个瞬时延时到音频信号和相应的视频信号中的至少一个, 以响应一个误差容限模型; 以及

20 一个编码器(120), 用于编码音频和视频信号以产生编码音频和视频流。

17. 权利要求16的设备, 其中

所述误差容限模型定义了一个同步容限窗(410); 以及

所述瞬时延时导致了一个描述在所述音频信号和相应的视频信号之间的同步误差的概率分布函数(pdf)(310), 该函数被向着与相应的所述同步容限窗

25 (410)更有益的方向偏移。

18. 权利要求17的设备, 其中所述同步容限窗(410)已使一个负瞬时失配值和一个正瞬时失配值与它相关联, 所述瞬时失配值具有不同的绝对值, 所述的pdf具有相应的负的和正的瞬时失配值, 所述这些瞬时失配值被以向着与所述同步容限窗瞬时失配值对齐的方向偏移。

30 19. 一种具有用于执行步骤的计算机可执行指令的计算机可读介质, 包括:

接收(610)视频访问单元和相应的音频访问单元,所述用于表示视听信息的所述视频和相应的音频访问单元有助于显示出由第一概率分布函数(pdf)(510)描述的瞬时同步误差;以及

- 5 以一个延时量来瞬时延时(630)所述接收到的音频和视频访问单元中之一,用于表示视听信息的未延时和相应的延时单元有助于显示由第二 pdf(520)描述的同步误差,所述第二 pdf 比所述第一 pdf(510)利用了同步容限窗(410)的更大部分。

20. 权利要求 19 的计算机可读介质,其中所述步骤进一步包括:
在对产生所述音频和视频访问单元的音频和视频流编码之前传递所述延时
10 量。

用于同步音频和视频流的设备和方法

5 相关申请的交叉引用

本专利申请要求 2002 年 4 月 19 日提交的、序列号为 60/374269 的美国临时申请的权益，该申请的全部内容在此被引用参考。

发明背景

1. 发明领域

- 10 本发明涉及多媒体通信系统领域，并且更具体地涉及使由可变延时传输网络导致的唇 (lip) 同步误差最小化。

背景技术

- “唇同步” (或 lip sync) 问题是广为人知的。简要地，在通过呈现装置呈现音频和视频流中的瞬时误差可能导致一种音频信息在相应的视频信息呈现之前 (超前) 或之后 (滞后) 显示的情况，导致例如在说话者声音的音频表示和说话者嘴唇的视频表示之间的同步很差。
- 15

用于解决所谓的唇同步问题的现有技术都比较复杂，而且有时会引起音频和/或视频信息的降级。例如，大家知道，丢失视频帧以致引起视频图像的时间提前，因此要对超前音频信号作以校正。

- 20 很多原因可以引起唇同步误差。与之特别相关的是使用可变延时网络，例如因特网和它的分组交换网络。在这样的网络中，音频和视频信息作为分离和独立流进行传输。在引入这些流到可变延时网络之前的传输处理期间，一个包含时间标记和其他元数据 (例如编码器抽样率、数据包次序等) 的传输层报头被加入到一些或全部传输数据包。音频和视频信息的时间标记典型地来自于公共的源，
- 25 例如实时时钟。不幸的是，当音频和视频数据包通过可变延时网络时，由于网络状况，所以出现瞬时异常、丢失数据包、数据包的次序没有被保存以及数据包的延时时间有变化。最终的结果是在接收到的通过可变延时网络传输的音频和视频流中的唇同步误差。

发明内容

- 30 本发明包括一种用于通过把瞬时误差的统计分布调整到低于听众能够注意

和厌恶的误差范围以下来减少相应的音频和视频流的唇同步问题的方法和设备。

特别地，根据本发明的一个实施例的方法包括：接收视频访问单元和相应的音频访问单元，该视频和相应的音频访问单元表示视-听信息，从而有助于显示由第一概率分布函数(pdf)描述的唇同步误差；以及通过一个定时因子来瞬时延迟接收到的音频和视频访问单元之一，未延时和相应的已延时访问单元表示视-听信息，从而显示由第二pdf描述的唇同步误差，第二pdf比第一pdf利用了更大份额的唇同步容限窗。

在另一个实施例中，一种用于产生适合于在可变延时网络使用的编码视频和音频流的方法包括：对相应的视频和音频信息进行瞬时编码来产生编码视频和音频流，每一个编码视频和音频流包括多个分别具有被时间标记的视频和音频数据包的视频和音频数据包；以及通过一个定时因子来自适应视频时间标记的数据包和音频时间标记的数据包中的至少一个，以便减小滞后视频唇同步误差的可能性。

在另一个实施例中，在一个接收器中实现唇同步误差pdf估计器以便动态地估计pdf。基于估计的pdf，根据目标函数计算一个最佳的音频延时时间。然后计算出的延时被引入到接收器端。

附图说明

通过考虑此后的详细描述和附图，本发明的教导可以被容易地理解，其中：

附图1描述了一个通信系统的高层框图；

附图2描述了一个控制器的高层框图；

附图3描述了一个有助于理解本发明的唇同步误差 e 的概率密度函数 $p(e)$ 的图解表示；

附图4描述了一个有助于理解本发明的唇同步误差容限(LSET)窗的图解表示；

附图5描述了一个在容限窗中pdf偏移的图解表示；

附图6描述了根据本发明处理音频和/或视频数据包的方法；

附图7描述了根据本发明的一个备用实施例的通信系统的高层框图；以及

附图8描述了一个本发明实施例的高层框图，其中在一个接收机端执行了一个pdf估计器。

为了便于理解，只要有可能，相同的参考数字已经被使用来标明在附图中相

同的元件。

发明的详细描述

本发明将在诸如因特网之类的可变延时网络的背景下讨论，其中可变延时网络往往会在视频和/或音频数据包通过时产生瞬时误差，以致可能产生唇同步误差。然而，本发明的方法可以容易地适合各种瞬时误差源。本发明是在视频和/或音频呈现单元（例如视频和音频帧）上运行的，所述呈现单元可以被打包以适合于通过例如可变延时网络的网络进行传输。

而且，尽管“唇同步”的标准通信定义使语音或歌唱的同步（或同步处理）与视频相关，以便视频中嘴唇的动作看起来和声音自然地一致；但对本发明来说，该定义的解释不能受到这样的限制。相反地，“唇同步”涉及到视频中表示的任何动作和相应的音频轨道或比特流的同步，以便由该动作特意地产生的声音与特意地产生该声音的视频适当地匹配。换句话说，对于本发明来说，“唇同步”涉及由音频信息信号表示的声音和相应的由视频信息信号表示的视频之间的同步；而不管相应的音频和视频的内容。因此，对“唇同步误差”的引用实质上是很全面的，并且可以被解释为各种“视听时间同步误差”。

附图1描述了包括本发明的一个通信系统的高层框图。特别地，通信系统100包括一个视听源110，例如一个大容量存储装置、摄象机、话筒、网络输入或者其它的视听信息源。视听源110分别向一个视频编码器120V提供视频流V和向音频编码器120A提供相应的音频流A。所述的编码器120V和120A（举例说，它们构成MPEG或其它的压缩编码器）对视频流V和音频流A分别进行编码以便产生编码视频流VE和编码音频流AE。编码视频VE和音频AE流根据适合于可变延时网络140（举例地说，是以太网、ATM、或其它的传输流编码器，它根据适合于可变延时网络140的特定传输格式对视频VE和音频AE流进行编码）的特定传输格式通过传输处理器130进行处理。

传输流T通过可变延时网络140传播到目的地，可变延时网络140例如因特网、内联网，ATM、以太网、LAN、WAN、公共交换电话网络（PSTN）、卫星、或其它的网络，在该目的地它被接收为传输流T'。传输流T'包括具有一些由可变延时网络140传输引入的延时或其它误差的原始传输流T。

合成的传输流T'被传输处理器150（举例来说，是一个以太网、ATM、或其它的传输流解码器）接收，该处理器从接收到的传输流T'中提取出一个编码视频

流 VE' 和相应的编码音频流 AE'。编码视频 VE' 和音频 AE' 流包括原始的具有一些例如由传输处理器 130、可变延时网络 140 和/或传输处理器 150 引起的瞬时误差的编码视频 VE 和音频 AE 流。接收到的编码视频 VE' 和音频 AE' 流通过解码器 160 被解码来产生合成的视频 V' 和音频 A' 流。产生的视频 V' 和音频 A' 流被一个
 5 呈现装置 170 (例如电视或其它显示装置) 呈现, 170V 已经使一个音频呈现装置 (例如扬声器 170A) 与它相关联。

附图 2 描述了适合于使用在根据本发明的原理的系统和设备中的控制器的高层框图。特别地, 附图 2 中的控制器 200 可以用于执行一个或多个在上述参考附图 1 的描述中的功能元件, 以及后面参考附图 7 和 8 的描述中的各种功能元件。

10 附图 2 中示例性的控制器 200 包括一个处理器 230 和用于存储各种程序 245 的存储器 240。处理器 230 和常规的支持电路 220 例如电源、时钟电路、高速缓冲存储器等以及协助执行存储在存储器 240 中的软件程序的电路配合运行。同样地, 可以预见在此讨论的处理步骤中的一些例如软件处理可以在硬件中执行, 例如, 像与处理器 230 协作的电路来执行各种步骤。控制器 200 还包括输入/输出
 15 (I/O) 电路 210, 它形成在与具有控制器 200 的功能元件通信的各种功能元件的接口。

尽管附图 2 中的控制器 200 被描述为一个用于进行对根据本发明的音频和/或视频流的各种瞬时校正的通用计算机, 但是本发明可以在硬件中执行, 例如一个专用集成电路 (ASIC)。同样地, 在此描述的处理步骤可以被广泛地解释为被
 20 同等地通过软件、硬件、或其组合来实现。

唇同步误差 (LSE) 可以根据公式 1 定义如下:

$$e = (t_a^v - t_v^v) - (t_a^a - t_v^a) \quad (\text{公式 1})$$

25 在公式 1 中, t_a^v 和 t_v^v 是相关的音频和视频帧分别到达接收器端的再现装置 170 的时间; 以及 t_a^a 和 t_v^a 是音频和视频帧分别到达音频和视频编码器的时间。

附图 3 描述了有助于理解本发明的唇同步误差 e 的概率密度函数 $p(e)$ 的图解表示。由于例如由可变延时网络 140 引入的随机延时, 所以在接收器端的音频数据包和它相应的视频数据包之间的延时是随机变量。所述的随机变量被它的概
 30 率密度函数 (pdf) $p(e)$ 所限制, 如附图 3 中的实线 310。特别地, 附图 3 的图解

表示描述了一个定义视频数据和相应的音频数据的瞬时关系的水平轴。时间零被选择作为表示已经结合了其同步的音频数据的视频数据内容的时间。当这个分布被描述为高斯分布时，可以利用其它对称的或不对称的 pdf 曲线，这取决于特定的模拟的误差源以及模拟的误差源的数目（也就是可以使用一个用于复合视频和音频源的复合对称的或不对称的 pdf 曲线）。

随着时间从零向正方向的增加，音频数据更加落后于视频数据（也就是，音频数据包愈加延迟于相应的视频数据包）。随着时间从零向负方向的增加，音频数据更加超前于视频数据（也就是，视频数据包愈加延迟于相应的音频数据包）。

附图 4 描述了有助于理解本发明的唇同步误差容限 (LSET) 窗 410 的图解表示。特别地，该 LSET 窗由函数公式 (2) 定义如下，其中 a 和 b 是 LSET 窗容限的下限和上限。

$$W(e) = \begin{cases} 1(a \leq e \leq b) \\ 0(\text{其它}) \end{cases} \quad (\text{公式 2})$$

发明者注意到音频和视频数据包的不对称误差容限以及在音频数据包在相应的视频数据包之前被接收时的情况下产生的大量问题。典型的数值变化范围，例如 $[a, b] = [-20ms, 40ms]$ 。

附图 5 描述了在容限窗中 pdf 偏移的图解表示。特别地，附图 5 的图解表示描述了一个定义视频数据和相应的音频数据之间的瞬时关系的水平轴。时间零以 20 上述参考附图 3 和 4 的描述的方式被选择。延时容限窗 410 表示结合了不被观众讨厌的唇同步的延时容限或瞬时误差。可以注意到，附图 5 中的延时容限窗 410 从 -20 毫秒（即音频数据包超前视频数据包达到 20 毫秒）延伸到 +40 毫秒（即音频数据包滞后视频数据包达到 40 毫秒）。可以注意到，音频信息超前视频信息的唇同步误差比音频信息滞后视频信息的唇误差往往更令人讨厌（例如更容易使观众留意到和/或分散注意），因此附图 5 中的延时容限窗 410 不对称。

参考附图 5，pdf 曲线 510 的左边的“尾部”部分落入到超出了较低的延时容限窗范围的区域 540。可以注意到，pdf 曲线 510 的右边的“尾部”部分在延时窗容限范围上限之前被充分地归零。误差窗容限范围被定义为在其中的瞬时误差例如唇同步误差被视为较少令人讨厌的范围。因此，或者正的或者负的超过

延时容限范围的延时包括这些被视为令一般观众讨厌或高度讨厌的延时。

偏移的 pdf 曲线 520 表示在时间上被偏移的原始的概率分布曲线 510, 以便一个在 pdf 曲线之下的更大的区域在误差容限窗 410 内。因此, 原始的或第一 pdf 已经在时间上被偏移, 使得在最终或第二 pdf 之下的增加的区域 (优选地最大区域) 被包括在误差容限窗 410 中。在 pdf 上的所述偏移由结合了视频和/或音频信息的自适应定时参量引起, 例如视频和/或音频访问单元的显示时间标记。因此, 如果音频和/或视频瞬时信息适合于引起在相应的 pdf 上这样的偏移, 那么令人讨厌的唇同步误差的可能性会最小, 或者至少减少了一个与由偏移引起的在曲线误差之下的 pdf 中的减少相应的量。因此, 最大化在唇 LSET 中的唇 LSE 曲线之下的区域的最佳解决方法是使由公式 3 给出的目标函数最大化如下:

$$\begin{aligned} J &= \int_{-\infty}^{+\infty} p(e - t_0) W(e) de \\ &= \int_a^b p(e - t_0) de \\ &= P(b - t_0) - P(a - t_0) \end{aligned} \quad (\text{公式 3})$$

15

在公式 3 中, $p(e)$ 是 LSE 的 pdf, $P(e)$ 是累积分布函数以及 $W(e)$ 是在 [2] 中定义的 LSET 窗函数。最优化处理是用来使被 pdf 曲线在 $[a, b]$ 的界限内包围的区域最大化。这相当于超出窗的“尾部”区域的最小化处理。该最优化问题可以通过求 J 对 t_0 的导数并解如下公式 4 得到 t_0 来解决:

20

$$\frac{\partial J}{\partial t_0} = 0 \quad (\text{公式 4})$$

可以证明, 如附图 2 所示的对称高斯 LSE pdf 的 t_0 的最佳解是 LSET 窗的下限和上限的平均值:

25

$$t_0 = \frac{a + b}{2} \quad (\text{公式 5})$$

对于其它的 LSE pdf, 最佳的 t_0 可能是一个正的或负的值, 取决于 pdf 和误差容限窗之间的相对位置关系。一个正的 t_0 意味着在音频帧中延时, 以及一个负的 t_0

意味着在视频帧中延时，以偏移 LSE 并使公式 4 最大化。

附图 6 描述了根据本发明的用于处理音频和/或视频数据包的方法。特别地，附图 6 描述了一种用于自适应相应的视频和/或音频帧或访问单元数据包使得唇同步误差和尤其是超前音频型唇同步误差被最小化的方法。在附图 6 的方法的背景下，唇同步误差由按照框 605 的包括可变延时网络、编码器、传输处理器或其它误差源中的一个或多个的误差源引起。

在步骤 610，可能由误差源产生的瞬时误差被表示为概率密度函数 (pdf)。例如，和上述参考附图 5 的描述一样，一个结合了可能由可变延时网络产生的瞬时误差的 pdf 被示出。该 pdf 包括例如一个具有高斯形状的随机数分布（可能是或不是以零点为中心），这里的零点表示没有唇同步误差（也就是，视频和音频数据的瞬时校正）。

在步骤 620，定义了一个与 pdf 相关联的误差容限窗。如同在框 615 中所示的，误差容限窗可以相关于唇同步误差或其它误差来定义。如附图 5 中所示，一个关联着唇同步误差的延时容限窗被定义为例如在 -20 毫秒到 +40 毫秒之间的延时。也就是说，一个具有超前相应的视频访问单元达到 20 毫秒或滞后相应的视频数据包达到 40 毫秒的音频访问单元的不对称音频延时容限值（相对于零时间点）被认为是可以接受的。其它的容限窗可以根据与利用本发明的通信系统的相关的因素来定义。

在步骤 630，该方法自适应与组成内容流的视频和音频帧的至少一个相关的定时参数，例如时间标记。可选地，未压缩的音频和视频流中的一个或全部都在编码之前延时。这样的修改以一种力图引起与误差源相关的 pdf 从初始的位置（例如以零为中心）向最大利用延时容限窗的位置偏移的方式进行。可以注意到，在框 625 中上述自适应可以在编码处理、传输处理或其它的处理期间发生。返回头参考附图 5，一个适当的 pdf 偏移被示出为增加了在由延时容限窗确定的范围内的概率分布曲线以下区域的数量。

附图 7 描述了根据本发明的一个备用实施例的通信系统的高层框图。特别地，附图 7 中的通信系统 700 和附图 1 中的通信系统 100 实质上相同。主要的区别是延时元件 710A 被用于在音频编码器 120A 对原始音频流编码之前延时该音频流 A。依照前面讨论的唇同步误差容限 (LSET) 模型，该延时元件 710A 向音频流产生一个延时时间 t_0 来偏移相应的 pdf。可以注意到，附图 7 中的通信系统 700

可以被修改以便包括一个相应的视频延时元件 710V (未示出), 用于在视频源信号 V 被视频编码器 120V 编码之前延时该信号。音频 710A 和视频 710V 延时元件中的一个或全部都可以被利用。

在本发明的这个实施例中, 利用了如附图 5 中示出的误差容限窗 410, 例如
5 每一个音频帧或访问单元在编码之前相对于每个视频帧被延时大约 t_0 毫秒。通过将每一个音频帧在时间上向回偏移 t_0 毫秒, 与可变延时网络产生的误差相关的 pdf 被以关于附图 5 的上面描述的方式偏移。也就是说, pdf 向前或向后偏移取决于标记 t_0 , 在时间上从具有超前音频数据包的唇同步误差趋势到没有唇同步误差或滞后音频数据包唇同步误差的趋势 (它比超前音频数据包唇同步误差更少令人讨
10 厌)。因此, 任何音频数据包延时将仍然保持在由误差容限窗 410 建立的误差容限中的概率增加了。

在本发明的一个实施例当中, 其中假定了一个如附图 2 所示的对称高斯 pdf, 音频或视频帧的时间标记被修改以使定时失配最小化。对于一个恒定比特率的音频编码器, 视频时间标记可选地以一种用于增加保持在解码器端的 LSET 中的定
15 时失配概率的方式进行修改。在该实施例中, 例如视频标记舍入到较低的数十毫秒, 如公式 6 所指出的, 如下 (其中 t_v 和 $\hat{t}_v$ 是用于视频帧的以毫秒表示的原始和舍入的时间标记):

$$\hat{t}_v = t_v - (t_v \bmod 10) \quad (\text{公式 6})$$

20

上述技术介绍了音频数据包中在 0 到 9 毫秒的范围内的均匀分布式延时。其它的范围也可以被选择 (例如模 15、模 20 等等), 并且音频数据包同样可以以这种方式进行处理。

在先前描述的实施例中, pdf 的 LSE 都是已知的并且认为较为稳定。结果,
25 预定的时间偏移在所有的音频 (或视频) 访问单元上进行。在更先进的一个实施例中的 LSE pdf 可能不是已知的或是不稳定, LSE pdf 被监测和估计, 并且时间偏移不是预定的。

附图 8 描述了实施例中的 LSE, 其中的 pdf 估计器在接收端执行。特别地, 例如在上面附图 1 和 7 中描述的接收端设备被修改以包括有 LSE pdf 估计器 810
30 和音频延时元件 820A。虽然没有示出, 一个视频延时元件 820V 也可以被使用。

LSE pdf 估计器 810 接收解码的音频 A' 和视频 V' 信号以及响应于 LSET 模型信息而产生一个延时指示信号 t_0 。在附图 8 的实施例中，延时指示信号 t_0 被音频延时元件 820A 处理向已解码音频流 A' 传送相应的延时量，从而产生一个延时的音频流 A''。估计器 810 不断地收集音频和视频访问单元的显示时间标记。每一个 LSE 5 e 都使用公式 1 进行计算。所有的 LSE 被用于构成 LSE 的 pdf。通过使用 LSET 模型，最佳的时间偏移 t_0 可以通过解公式 4 求出时间偏移 t_0 公式来得到。在音频帧 ($t_0 > 0$) 或视频帧中 ($t_0 < 0$) 的延时被加入以便偏移 LSEpdf。

在一个实施例中，一个确定的光学时移被从接收器传输到编码器，使得要被编码和传输的音频和视频流中的至少一个在编码之前、传输处理之前和/或传输到 10 接收器之前被延时。

尽管各种包括本发明的教导的实施例在此已被详细地描述和示出，但是本领域熟练的技术人员可以容易地设计许多其它还包含这些教导的各种各样的实施例。

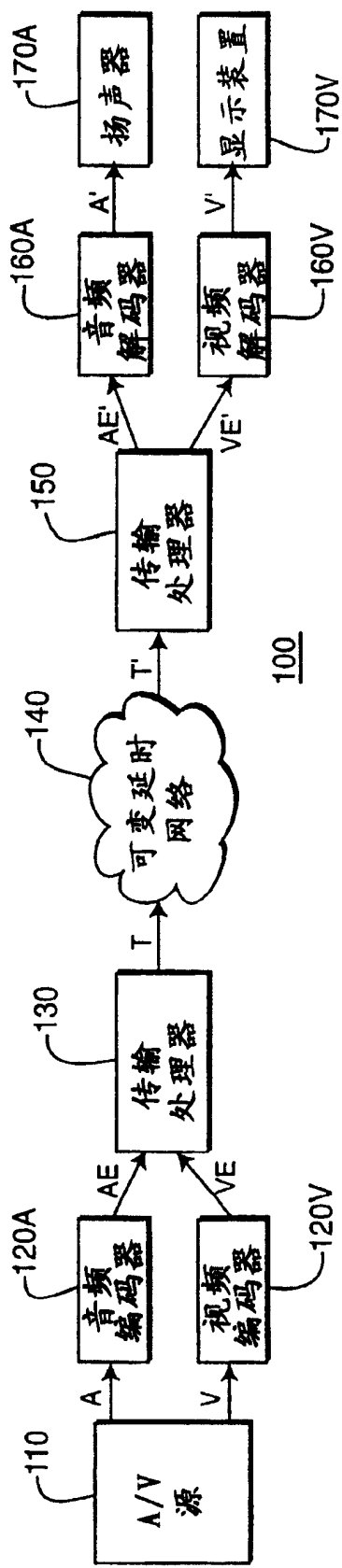


图 1

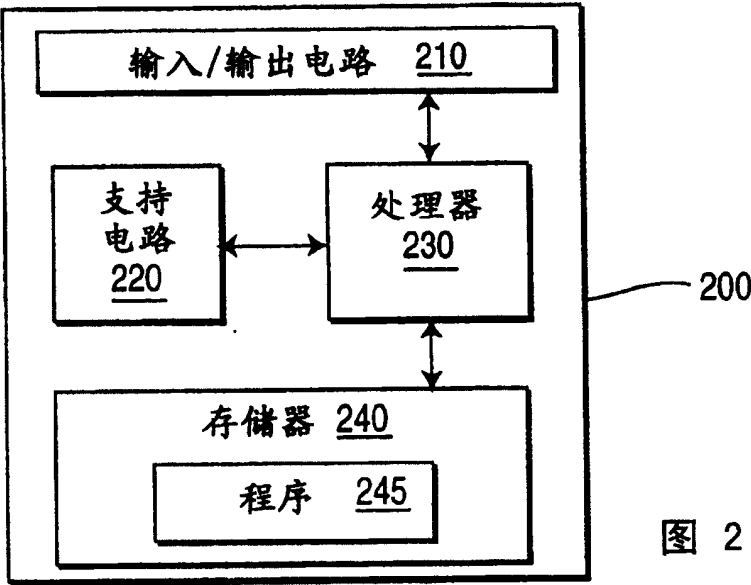


图 2

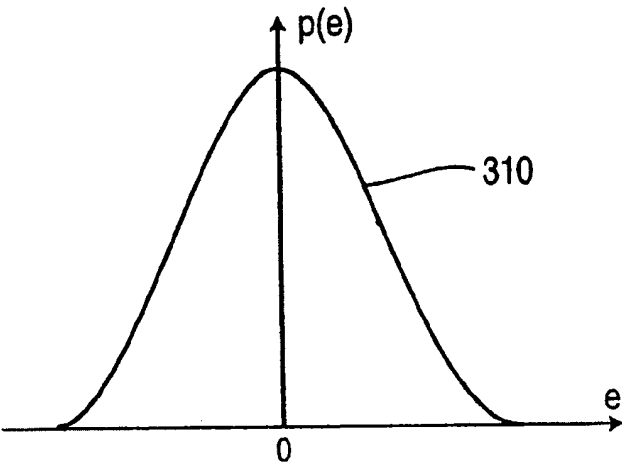


图 3

唇同步误差 e 的概率密度函数 $p(e)$

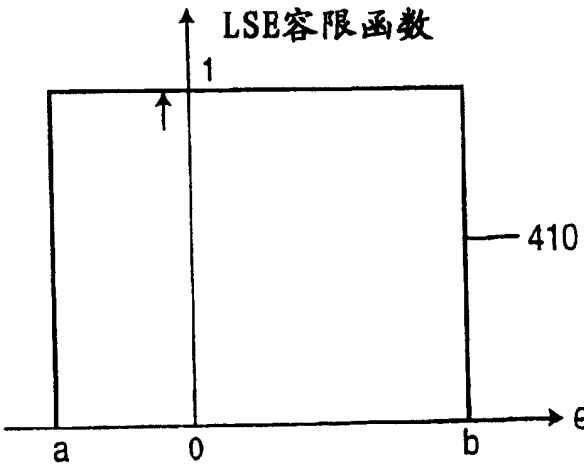


图 4

唇同步误差容限(LSET)窗

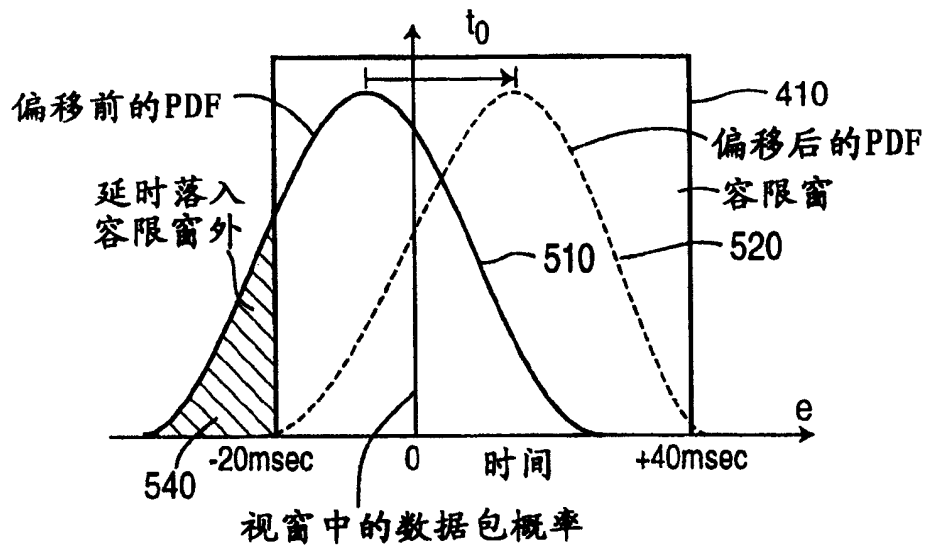


图 5

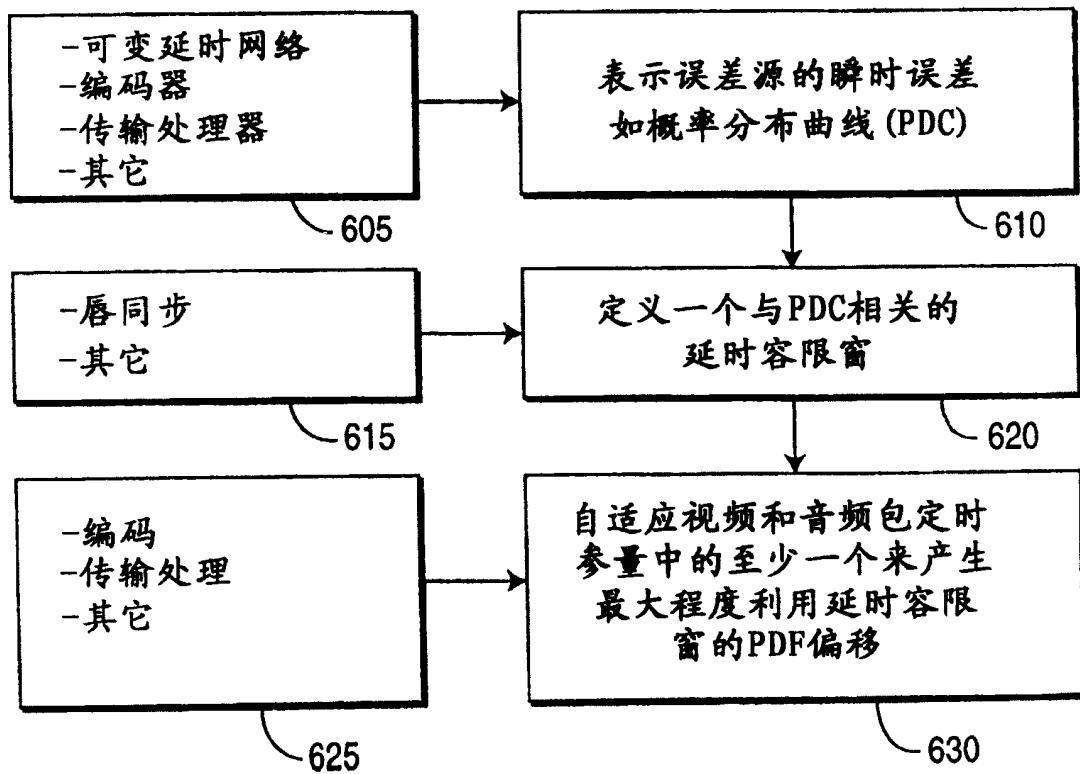


图 6

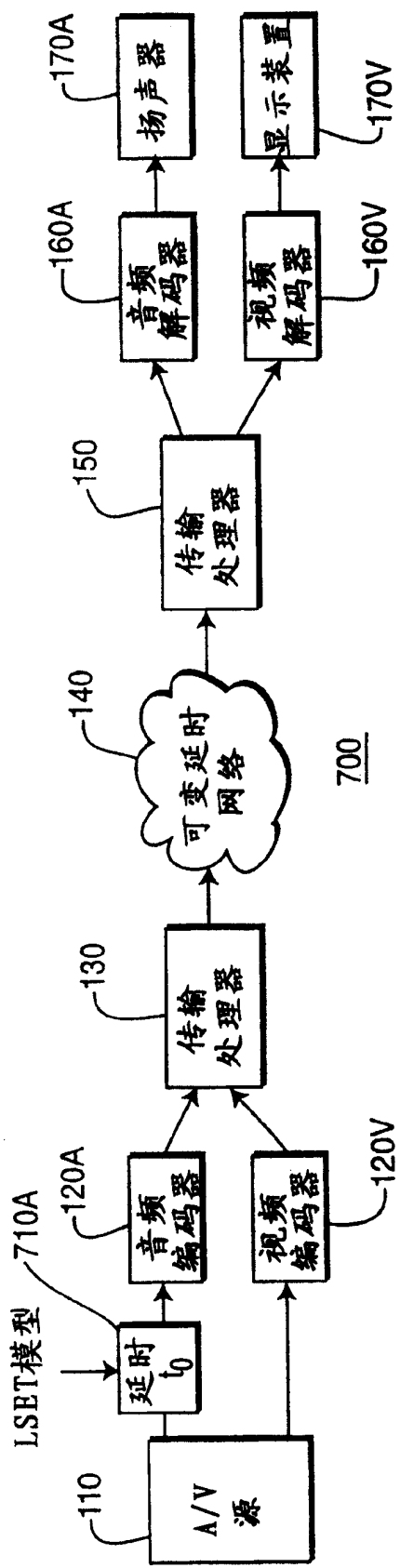


图 7

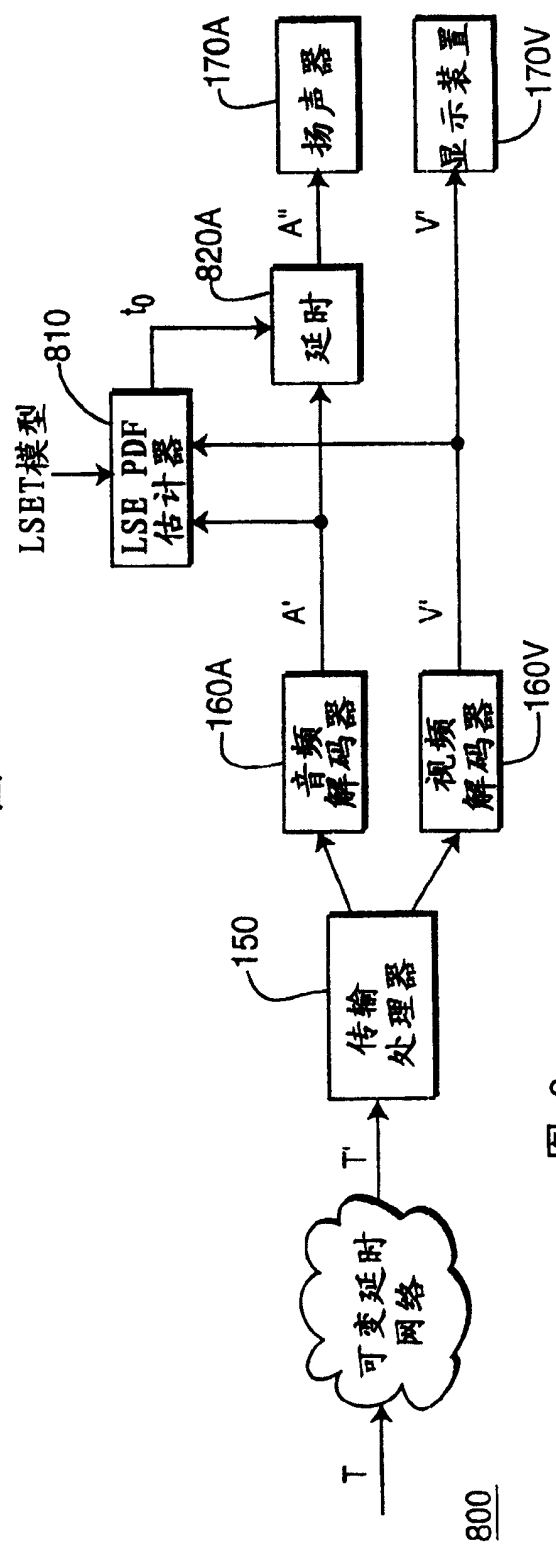


图 8