



# [12] 发明专利申请公布说明书

[21] 申请号 200580037908.9

[43] 公开日 2007 年 10 月 10 日

[11] 公开号 CN 101053020A

[22] 申请日 2005. 11. 2  
 [21] 申请号 200580037908.9  
 [30] 优先权  
     [32] 2004. 11. 5 [33] EP [31] 04105545.0  
 [86] 国际申请 PCT/IB2005/053570 2005. 11. 2  
 [87] 国际公布 WO2006/048824 英 2006. 5. 11  
 [85] 进入国家阶段日期 2007. 5. 8  
 [71] 申请人 皇家飞利浦电子股份有限公司  
     地址 荷兰艾恩德霍芬  
 [72] 发明人 T·J·F·诺登 S·V·安德森  
     S·H·詹森 W·B·克利恩  
     N·H·范施恩德尔

[74] 专利代理机构 中国专利代理(香港)有限公司  
 代理人 张雪梅 张志醒

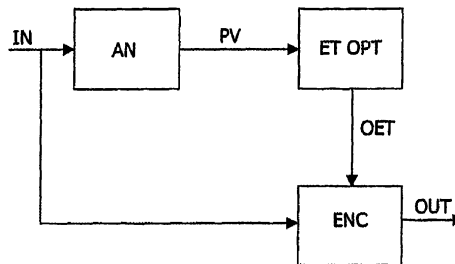
权利要求书 2 页 说明书 14 页 附图 2 页

[54] 发明名称  
 利用信号属性的有效音频编码

### [57] 摘要

一种音频编码器，包括基于例如为属性向量的形式的输入音频信号 IN 的特性 PV，以产生优化的编码模板 OET 的优化装置 ET OPT。该优化的编码模板 OET 针对预定编码效率标准得到优化。编码装置 ENC 随后依据该优化的编码模板 OET 产生编码的音频信号 OUT。该音频编码器可包括基于输入信号 IN 产生输入信号特新集合 PV 的分析装置 AN。在优选实施例中，该优化装置 ET OPT 适于评估与编码模板相关的结果失真。该优化装置 ET OPT 还可以评估与编码模板相关的比特率。在一个实施例中，该优化装置 ET OPT 基于输入信号属性 (PV) 优化大量子编码器的比特率分配。在另一个实施例中，该优化装置 ET OPT 基于输入信号属性 (PV) 预先决定自适应分割。根据本发明的编码器的优点为，由于是基于输入信号属性 (PV) 发现优化

的编码模板 OET，因此可以避免在决定优化编码模板 OET 之前的多个编码的复杂过程。



1. 一种适于根据编码模板编码音频信号 (IN) 的音频编码器, 所述音频编码器包括:

优化装置 (ET OPT), 适于基于所述音频信号 (IN) 的预定属性集合 (PV) 产生优化的编码模板 (OET), 所述优化的编码模板 (OET) 是针对预定编码效率标准被优化的, 以及

编码装置 (ENC), 适于根据所述优化的编码模板 (OET) 产生编码的音频信号 (OUT).

2. 根据权利要求 1 的音频编码器, 还包括分析装置 (AN), 其适于分析所述音频信号 (IN) 并响应于此产生所述音频信号 (IN) 的属性集合 (PV).

3. 根据权利要求 1 的音频编码器, 其中所述优化装置 (ET OPT) 包括适于基于音频信号 (IN) 的预定属性集合 (PV) 来预测与所述编码模板相关的感觉失真的装置。

4. 根据权利要求 1 的音频编码器, 其中所述音频信号 (IN) 的属性集合 (PV) 包括选自包括下述属性的组中的至少一种属性: 音调、噪声、调和性、平稳性、线性预测增益、长时间预测增益、谱平坦度、低频谱平坦度、高频谱平坦度、过零率、响度、调音率、谱矩心、谱带宽、Mel 倒频谱、帧能量、ERB 带 1-10 的谱平坦度、ERB 带 10-20 的谱平坦度、ERB 带 20-30 的谱平坦度、以及 ERB 带 30-37 的谱平坦度。

5. 根据权利要求 1 的音频编码器, 适于为所述音频信号的每段优化所述编码模板。

6. 根据权利要求 1 的音频编码器, 其中所述预测装置 (ET OPT) 还包括适于基于所述音频信号 (IN) 的所述属性集合 (PV) 预测与所述编码模板相关的结果比特率的装置。

7. 根据权利要求 1 的音频编码器, 其中所述优化装置 (ET OPT) 适于基于所述音频信号的所述属性集合 (PV) 优化所述音频信号的分割。

8. 根据权利要求 1 的音频编码器, 其中所述优化装置 (ET OPT) 适于从预定编码模板集合中选择被优化的编码模板 (OET)。

9. 根据权利要求 1 的音频编码器, 其中所述编码装置包括第一

(SENC) 和第二 (TENC) 子编码器, 且其中所述优化装置 (R-OPT) 适于响应于所述音频信号 (IN) 的所述预定属性集合 (PV), 为所述第一 (SENC) 和第二 (TENC) 子编码器产生优化的第一 (R-SE) 和第二 (R-TE) 编码模板。

10. 一种编码音频信号 (IN) 的方法, 所述方法包括步骤:

基于音频信号 (IN) 的预定属性集合 (PV) 产生优化的编码模板 (OET), 所述优化的编码模板 (OET) 是针对预定编码效率标准被优化的, 以及

根据所述优化的编码模板 (OET) 产生编码的音频信号 (OUT)。

11. 一种优化适于编码音频信号 (IN) 的音频编码器的编码模板 (OET) 的方法, 所述方法包括步骤:

接收音频信号 (IN) 的预定属性集合 (PV),

基于所述音频信号 (IN) 的所述预定属性集合 (PV), 针对预定编码效率标准优化所述编码模板 (OET)。

12. 一种包括根据权利要求 1 的音频编码器的装置。

13. 一种计算机可读取程序代码, 适于根据权利要求 10 的方法编码音频信号。

## 利用信号属性的有效音频编码

### 技术领域

本发明涉及高效、高质量的音频信号编码。更具体而言，本发明涉及这样的音频编解码器种类，其适应于输入信号，即，具有为了获得就速率失真(rate-distortion)标准而言最佳的编码信号而待优化的多个编码设置。本发明提供一种音频编码器以及优化音频编码器设置的方法。

### 背景技术

编码中的关键问题是发现各个输入信号的最有效的表示。由于音频信号可以呈现范围宽的特性，且对于不同信号特性，不同的编码方法是最有效的，因此期望使用灵活的编解码器，即结合了不同编码方法的编解码器。例如，将音频信号分割，并编码成正弦部分和其余部分。通常，使用针对由正弦波组成的信号的特定编码方法编码音调信号，而采用波形或噪声编码器编码其余信号。因此，在这些编解码器中，需要决定使用哪种设置（或者哪种编码模板），例如，使用哪种编码方法编码信号的哪部分。这种决定可以基于完整的输入信号，即输入信号本身，且在尝试许多编码可能性之后，针对各种可能性计算得到的（感觉的）失真。然而，在采用组合许多不同编码方法并因此具有许多可能设置的所出现的灵活自适应编解码器的情况下，就复杂度而言，关于编码设置的决定变成问题。

在仅使用一种编码方法的大多数编解码器中，也需要进行例如有关于编码器设置的决定，例如对于输入信号的不同部分编码器设置可能不相同。例如对于采用自适应时间分割的编解码器，情况就是如此。可以通过速率失真优化使分割适应，但是这显著提高了复杂度。另一个示例可在参数正弦编码中找到。这种情况下需要决定将多少正弦曲线分配到特定分段，最佳数目取决于输入信号。此外，在转换和子带编解码器中，必须决定量化水平和比例因子带（使用相同量化水平编码的一组频带）。这些决定是基于完整的输入信号，考虑了不同频带中

相应的编码误差。

专利申请 US 2004/0006644 描述了一种对输入信号进行代码转换的方法。根据将要进行代码转换的输入信号，可以选择不同的代码转换方法。在 US 2004/0006644 中提出了基于将要进行代码转换的输入信号的先前确定的属性在不同方法之间进行选择。然而 US 2004/0006644 没有披露用于优化编码器设置的任何方法。

总之，现有技术没有令人满意地解决如何确定最佳编码器设置或者哪种编码方法可以最佳地编码输入信号的哪部分。因此，在高质量音频编码领域中，需要一种有效地优化编码模板（或者编码器设置）从而使编码适应于输入信号的方法。

### 发明内容

因此，本发明的目标是提供一种能够低复杂度地优化编码器模板的音频编码器和音频编码方法，并提供就速率失真标准而言是有效的编码信号。

根据第一方面，本发明提供了一种适于根据编码模板编码音频信号的音频编码器，该音频编码器包括：

优化装置，其适于基于音频信号的预定属性集合产生优化的编码模板，该优化的编码模板针对预定的编码效率标准得到优化，以及

编码装置，其适于依据该优化的编码模板产生编码音频信号。

术语“编码模板”理解为必须为特定编码器选择的参数集合，即设置。“优化的编码模板”理解为一种编码模板，其中部分或所有参数响应于音频信号的预定属性集合被选择或调整，以便得到就所述预定编码效率标准而言更优化的编码的输出信号。“音频信号的预定属性集合”理解为对音频信号参数描述，包括描述该音频信号的信号属性的一个或多个参数。音频信号的预定属性集合例如可以是属性向量的形式，其中纯量值代表各个参数。

通过使用音频信号的预定属性集合，例如借助于属性向量，该音频编码器能够通过利用待编码音频信号的相关属性的现有知识，优化将用于编码过程的编码模板。因此，该音频编码器优选基于该音频信号的预定属性集合评估速率与/或失真度量，由此提供优化的编码模板而实际上未对音频信号编码。换言之，使用例如输入信号属性向量，

可以进行有关最佳编码器设置的决定而无需尝试大量可能的设置，且在最终决定优化编码模板之前，就速率和失真方面监测得到的编码的输出信号。

与传统编码器相比，这实现了一种用于编码模板优化的具有低复杂度的编码器。对于编码模板包括大的待优化参数集合的编码方案而言，为了获得最佳的速率失真效率，这一点尤为有利。一个示例为包括两个或更多个子编码器的编码器类别，其中至少一个任务是决定于编码器之间的比特率分配以便获得最佳的速率-失真效率。尽管使用完整输入信号在所有可能编码模板中进行彻底搜索和（感觉的）失真度量将是最佳的，但是这可能是低效率的，且可能太复杂而无法利用有限量的可用处理能力来实现。

应该理解，代表音频信号的属性集合的数据可以以任意方便的方式排列，例如属性向量或者属性矩阵。

音频编码器可包括适于分析音频信号并响应于其产生音频信号的属性集合的分析装置。然而，可以在音频编码器外部建立音频信号的属性集合。该音频编码器于是适于接收音频信号以及该音频信号的预定属性集合作为输入。

优选地，该优化装置包括适于基于音频信号的预定属性集合来预测与编码模板相关的感觉失真。“与编码模板相关的失真”理解为通过根据该编码模板编码该音频信号引起的经编码的音频信号和该音频信号本身之间的差异。“感觉失真”理解为人听觉系统感觉的失真相关的失真的度量，即反映感觉的音质的失真度量。优选地，该感觉失真度量基于感觉模型，例如人遮蔽曲线（masking curve）等的表示。

优选地，该优化装置包括适于基于音频信号的预定属性集合来预测与编码模板相关的比特率的装置。

最优选地，该优化装置适于基于音频信号的预定属性集合来预测与编码模板相关的感觉失真和比特率。由此，编码器能够根据如下标准对编码模板进行优化，所述标准为在给定最大目标比特率下的最佳音质或者在预定的就感觉失真而言最低的音质下的最低可能比特率。

优选地，音频信号的属性集合包括选自包括下述属性的组中的至少一种属性：音调、噪声、调和性、平稳性、线性预测增益、长时间预测增益、谱平坦度、低频谱平坦度、高频谱平坦度、过零率（zero

crossing rate)、响度、调音率(voicing ratio)、谱矩心(spectral centroid)、谱带宽、Mel倒频谱、帧能量、ERB带1-10的谱平坦度、ERB带10-20的谱平坦度、ERB带20-30的谱平坦度、以及ERB带30-37的谱平坦度。优选地,音频信号的预定属性集合包括属性向量,其具有代表一个或多个所述参数的标量。然而应该理解,可以使用许多其他类型的参数。原则上可以选择任何描述信号的参数。然而,优选地音频信号的预定属性集合包括与感觉相关的属性,即与人听觉系统感觉到的有关的特性。

音频信号的预定属性集合可包括可由现有技术中已知的标准定义确定的属性。

优选地,音频信号的属性集合专门设计成为所关心的特定编码器考虑相关属性。例如,对于具有正弦编码器部分和噪声编码器部分的组合编码器的情形,可以包括音调和噪声。由此,比特率分配任务变得简单,且容易由音调和噪声参数确定。例如,一种非常简单的决定标准为,在音调参数超过某一数值的情况下选择正弦编码器部分,否则选择噪声编码器部分。然而应该理解,基于所关心特定编码器的现有知识,即使仅使用一个、两个或者几个参数描述音频信号,也可能精确地预测编码性能。

优选地,音频解码器适于为音频信号的每段优化编码模板。因此,编码器能够跟踪音频信号的快速变化,例如瞬变,并相应地调整其编码模板。

该优化装置适于基于音频信号的所述属性集合来优化音频信号的分割。除了编码模板之外,已经证明使用自适应分割可以有效地编码。使用基于音频信号的信号属性的预先自适应分割,这种自适应分割变得更为有效,因为在现有技术编码器中,除了优化编码模板之外,自适应编码只是添加额外的复杂的优化任务。

该优化装置适于从一组预定义的编码模板中选择优化的编码模板。为了进一步促进该编码模板优化过程,优选地该组预定编码模板覆盖整个编码器参数空间的大部分。该优化任务于是可以评价该组预定编码参数,并根据预定编码效率标准选择其中最佳的一个。

在优选实施例中,所述编码装置包括第一和第二子编码器,而所述优化装置适于响应于音频信号的预定属性集合为第一和第二子编码

器优化第一和第二编码模板。如果是优选的，音频编码器可包括3、4、5、10或者更多个单独的子编码器，并适于基于音频信号的预定属性集合为所有子编码器优化编码模板。因此，该实施例覆盖组合的编解码器。

在第二方面，本发明提供了一种编码音频信号的方法，该方法包括步骤：

基于音频信号的预定属性集合产生优化的编码模板，所述优化的编码模板是针对预定编码效率标准被优化的，以及  
依据优化的编码模板产生编码的音频信号。

上述针对本发明第一方面的解释和优选变型也适用于第二方面。

在第三方面，本发明提供了一种对适于编码音频信号的音频编码器的编码模板进行优化的方法，该方法包括步骤：

接收音频信号的预定属性集合，

基于音频信号的预定属性集合，针对预定编码效率标准优化编码模板。

基于音频信号的预定属性集合（例如使用属性向量），为编码器优化编码模板，使得与优化编码模板的现有技术方法相比，优化过程的复杂度显著降低。原因在于，优化编码效率的现有技术方法基于实际编码音频信号获得的必要的比特率以及结果失真。因此，这些现有技术方法涉及编码过程。通过基于音频信号的预定属性集合的优化方法，消除了优化方法中的编码过程。这对于具有大量设置需要优化的编码器而言尤为有利。相反，该优化可基于感觉失真度量的预测以及给定编码模板的比特率的预测。

尽管没有根据编码模板实际编码信号那样精确，但是通过仔细考虑例如哪些数据要包含在音频信号的预定属性集合中并建立所关心（多个）编码器的精确模型，预测精度可以得到改进。对于每个编码器具有大量可能设置的组合编码器的复杂集合，现有技术方法可能提供不良结果，因为可能不能实际测试整个参数空间，而只能非常粗糙地覆盖参数空间。相反，只要可获得给定的计算能力，预测可以证明足够快速以覆盖整个参数空间，因此得到更接近理论上最优的编码模板。

根据第三方面的方法可包括分析音频信号的初始集合，并据此产

生该音频信号的预定属性集合。

优选地，该优化步骤包括预测感觉失真度量（见上文定义）。

优选地，该优化步骤包括预测比特率。优选地，该优化步骤包括预测感觉失真和比特率，以便能够根据下述标准对编码模板进行优化，所述标准为给定最大目标比特率下的最佳音质或者在预定的就感觉失真而言最低的音质下的最低可能比特率。

优选地，对音频信号的每段执行该优化方法。

优选地，该优化方法包括基于音频信号的预定属性集合来优化该音频信号的分割。

在第四方面，本发明提供了一种包括根据第一方面的音频编码器的装置。该装置优选为例如下述的音频装置：固态音频装置、CD 播放器、CD 记录器、DVD 播放器、DVD 记录器、硬盘记录器、移动通信装置、（便携）计算机等。然而，该装置还可以是音频装置以外的装置。

在第五方面，本发明提供了一种适于根据第二方面的方法编码音频信号的计算机可读取程序代码。

在第六方面，本发明提供了一种适于根据第三方面的方法优化编码模板的计算机可读取程序代码。

根据第五和第六方面的计算机可读取程序代码可包括适于信号处理器、个人计算机等的软件算法。该计算机可读取程序代码可以呈现在例如磁盘或存储卡或存储棒的便携介质上，或者可以呈现在 ROM 芯片中或以其他方式存储于装置内。

#### 附图说明

在下文中将参照附图更详细地描述本发明，附图中：

图 1 示出了现有技术编码器，其中编码设置固定或者基于编码信号的结果失真反复调整；

图 2 示出了根据本发明的编码器，其中编码器设置的决定基于对输入信号的先前分析；

图 3 示出了优选的用于评估编码失真的基于高斯混合的最小均方误差（MMSE）评估器；

图 4 示出了现有技术的组合编码器，其中通过评估编码信号的失真决定两个子编码器之间的比特率分配；

图 5 示出了根据本发明的组合编码器，其中基于输入信号的属性决定两个子编码器之间的比特率分配；

图 6 示出了根据本发明的编码器，其中基于输入信号的属性决定输入信号的自适应分割。

尽管本发明可以采取各种改进以及备选形式，但是在图示中示范性地示出了特定实施例并在此详细描述这些实施例。然而，应该理解，本发明不限于所披露的具体形式。相反，本发明覆盖落在由所附权利要求界定的本发明的精神和范围内的所有改进、等价物和备选。

### 具体实施方式

图 1 示出了现有技术编码器 ENC，其接收输入信号 IN 并响应该输入信号产生被编码的输出信号 OUT。在现有技术编码器 ENC 中，编码器设置或编码模板是固定的或者基于涉及对输入信号编码的优化算法。尝试不同的编码模板，各一个均涉及输入音频信号 IN 的编码，且对于每个编码模板，监测例如与每个编码模板相关的失真和比特率，最终选择最有效的编码模板用于产生输出信号 OUT。

图 2 通过优选的音频编码器实施例示出了本发明的原理。输入音频信号 IN 被信号分析装置 AN 接收并分析。作为响应，分析装置 AN 产生包括该音频信号 IN 的属性集合的属性向量 PV。该属性向量 PV 随后由编码模板优化单元 ET OPT 接收，该编码模板优化单元基于接收的属性向量 PV 产生优化的编码模板 OET。编码器装置 ENC 随后使用该优化的编码模板 OET 和输入音频信号 IN 产生编码的输出信号 OUT，该输出信号 OUT 为输入音频信号 IN 的被编码的版本。

因此，在图 2 的音频编码器中，属性向量 PV 以及不同编码配置的数学模型，例如其率失真性能，被用于产生优化的编码模板 OET。于是无需尝试所有可能的编码模板，因为属性向量 PV 已经表明编码模板的与输入类型有关的性能。与图 1 的现有技术编码器相比，根据本发明的音频编码器能够为编码器装置优化编码模板而无需对输入音频信号 IN 进行编码，而仅利用输入音频信号 IN 的属性即可决定最佳的编码模板。

应该理解，图 2 的图示中所示的分析装置 AN 是可选的。因此，根据本发明的音频编码器可适于接收输入音频信号 IN 和属性向量 PV 作

为输入。

属性向量PV的应用是有效的，并降低了优化过程的复杂度。使用属性向量PV的缺点为，编码变得（略微）亚于最优。然而，目前用于音频编码的专门方法极可能更偏离最佳解决方案。

可以按照多种方式使用输入音频信号的预定属性集合的应用，这些应用可以同时使用。将在下文中进一步描述这些应用。出于简化的原因，在下文中用属性向量表示输入音频信号的预定属性集合。

在第一实施例中，属性向量用于评估不同编码模板的失真，例如感觉失真。例如，不同编码方法的组合或者一种编码方法中不同设置的组合。在复杂度方面，这具有两个优点：1) 不需要实际编码，2) 不需要计算（感觉）失真。换言之，属性向量用于获得（感觉）失真，而无需实际编码和计算相应的失真。

在第二实施例中，属性向量用于直接确定在混合编码器中，即在包括若干编码方法或子编码器的组合的编码器中，输入信号哪部分使用哪种编码方法进行编码。这比先前的方案更进一步：这种情况下，属性向量不仅指示编码方法的与输入类型有关的性能，还指示使用哪个（些）编码方法。

例如，如果输入信号具有主要的正弦曲线，则无需使用所有编码方法编码该信号并选择最有效的一种编码方法。相反，属性向量表明该信号主要包含正弦曲线，因此检查哪种编码方法（例如正弦编码器）可有效地编码正弦曲线就足够了，随后使用其开始编码。因此，察看属性向量，无需实际编码即可立即清楚哪种编码方法可以最有效地编码（部分）输入信号。该属性向量也可以用于评估编码方法之间的潜在的相互作用。有关这些相互作用的知识对于编解码器的有效配置也是重要的。

在第三实施例中，属性向量评估编解码器的最佳时变自适应分割。通过利用属性向量，可以基于输入信号的随时间变化的特性预先设置该自适应分割，与探索多种分割可能性的效果的方法相比，这使得复杂度降低。

现在将更详细地描述所述三个实施例。

第一实施例是用于瞬时失真评估的基于属性向量的方案。框架基于从待编码的帧提取的属性向量，据此执行失真评估。更详细地，对

编码器 $Q(\cdot)$ 所发生的编码失真 $\theta$ 进行评估的任务得以解决。对于给定帧 $x$ ，所发生的失真表达为：

$$\theta = \delta(x, \tilde{x}) = \delta(x, Q(x)), \quad (1)$$

其中 $\delta(\cdot)$ 为恰当的失真度量。

该评估被分成属性提取 $f(\cdot)$ 和评估 $g(\cdot)$ 。随机输入向量 $X$ 处理成维数减小的随机向量 $P$ ，由此获得编码失真 $\Theta$ 的评估 $\hat{\Theta}$ 。该方案的目标是执行无偏评估，并最小化评估误差方差：

$$\sigma_z^2 = E[(Z)^2] = E[(\Theta - \hat{\Theta})^2] = E[(\Theta - g(P))^2]. \quad (2)$$

这个方案的性能高度依赖于属性向量的选择。因此，属性向量提取器 $f(\cdot)$ 的基本任务是提取这样的属性 $P$ ，其对于所需提取器精度 $\sigma_z^2$ 包含有关 $\Theta$ 的充分信息，即足够高的交互信息， $I(\Theta; P)$ ，例如在 T. M. Cover 和 J. A. Thomas 的 *Elements of Information Theory* (John Wiley & Sons, New York, NY, 1991) 中所发现的。

提取器 $g(\cdot)$ 的目标是基于对属性向量 $P=p$ 的观察，发现所发生的失真 $\theta$ 的评估 $\hat{\theta}$ 。该任务（即最小化 $\sigma_z^2$ 的任务）的最小均方误差评估器（MMSE）为条件平均评估器：

$$\hat{\theta}_{\text{mmse}} = E[\Theta | P = p] = \int \theta f_{\Theta|P}(\theta | P = p) d\theta \quad (3)$$

图 3 示出了使用 J. Lindblom、J. Samuelsson 和 P. Hedelin，在 "Model based spectrum prediction," (Proc. IEEE Workshop Speech Coding, (Delawan, WI, USA), 2000, pp.117-119) 中所述的基于模型的方法的选择实施。在图 3 中，T 0-L 表示离线训练联合 pdf  $f_{\Theta,P}^{(M)}(\theta, p)$ 。

对联合 pdf  $f_{\Theta,P}^{(M)}(\theta, p)$  采用高斯混合模型 (GMM)，则每个编码时刻的 MMSE 近似为：

$$\hat{\theta} = g(p) = \int \theta f_{\Theta|P}^{(M)}(\theta | P = p) d\theta, \quad (4)$$

其中  $f_{\theta,p}^{(M)}(\theta|P=p)$  为条件模型 pdf, 可以示为高斯密度的混合, 且可以从联合模型 pdf  $f_{\theta,p}^{(M)}(\theta,p)$  容易地导出。实践中, 该评估器计算条件均值的加权和:

$$\hat{\theta} = \sum_{i=1}^M \rho_i' m_{i,\theta|P=p}, \quad (5)$$

其中  $M$  为混合分量的数目,  $\{\rho_i'\}$  和  $\{m_{i,\theta|P=p}\}$  分别代表条件模型 pdf  $f_{\theta,p}^{(M)}(\theta|P=p)$  的权重和平均值。参考方程 (3), 当模型 pdf 接近真实 pdf 时, 评估器输出将接近真实条件平均值。

通过失真评估而非编码和失真计算获得的复杂度减小, 依赖于 3 个因素: 使用属性向量的失真评估的复杂度、编码方法的复杂度、以及失真计算的复杂度。

失真评估的复杂度明显依赖于所使用的模型。对于上述实施例, 假设独立地评估每个 RD 点, 复杂度可以写为:  $N_{RD} \cdot N_{mixt} \cdot (C_{product} + C_{pdf})$ , 其中  $N_{RD}$  为 RD 点的数目,  $N_{mixt}$  为混合的数目,  $C_{product}$  为矩阵向量乘积的复杂度,  $C_{pdf}$  为高斯 pdf 评估的复杂度。矩阵向量乘积具有所采用的属性向量的“维数”, 但该矩阵是对称的, 由此复杂度可以降低到大约其一半。

该编码方法的复杂度明显依赖于所使用的方法, 且在不同编解码器之间大不相同。然而, 该复杂度预计高于失真评估的复杂度。

已经使用所发生的信噪比 (SNR) 作为待评估的失真<sup>①</sup>, 为诸如码激励线性预测 (CELP) 的编码器  $Q(\cdot)$ , 评价了所实施的评估方案。已经测试了六种不同的属性向量: 10 阶线性预测增益 ( $G_{LPC}$ )、长时间预测增益 ( $G_{LTP}$ )、谱平坦度 ( $G$ )、低频谱平坦度 ( $G_{low}$ )、高频谱平坦度  $G_{high}$ 、以及 LPC 和 LTP 增益的组合 ( $G_{LPC} G_{LTP}$ )。所有评估器都基于 32 元混合 (32-mixture) 模型, 且基于 Timit 语音数据库、使用单独的评估和训练集合评估结果。

结果为, 当所采用的属性向量 $P$ 中交互信息 $I(\Theta;P)$ 增大时, 评估误差方差 $\sigma_z^2$ 减小。因此, 与真实失真的接近程度随着所采用的属性向量的交互信息 $I(\Theta;P)$ 而增大。结果表明, 只要属性向量具有足够高的交互信息 $I(\Theta;P)$ , 则高精度评估可得以执行。结果证实了使用属性向量指示编码配置的与输入类型有关的性能(由此减小复杂度)的可行性。

还使用每帧 30 个正弦曲线, 为正弦编码器评估了属性向量方案。该编码器是基于 R. Heusdens 和 S. van de Par 的 "Rate-distortion optimal sinusoidal modeling of audio and speech using psychoacoustical matching pursuits," (Proc. IEEE Int. Conf. Acoust, Speech, and Signal Proc, (Orlando, FL, USA), 2002, vol. 2, pp. 1809-1812) 中发现的心理声学匹配跟踪, 使用 S. van de Par, S. Kohlrausch, A. Charestan 和 R. Heusdens 的 "A new psychoacoustical masking model for audio coding applications," (Proc. Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc, (Orlando, FL, USA), 2002, vol. 2, pp. 1805-1808) 中发现的感觉谱失真度量作为待评估的失真 $\Theta$ 。

针对 8 个不同的属性向量进行了测试: 过零率 (ZCR)、响度 (L)、调音率 (V)、谱矩心 (SC)、谱带宽 (BW)、谱平坦度 (SF)、12 阶 Mel 倒频谱 (MFCC)、以及基于 L+SF+SC+BW 组合的 4 维属性向量。所有评估器都是基于 16 元混合模型, 且基于分成评估和训练集合的包含 900.000 帧 35ms 的音频数据库评估这些结果。对于该实施, 结果也表明, 只要属性向量具有足够高的交互信息 $I(\Theta;P)$ , 则可以高精度地评估失真。

在下文中将描述第二实施例, 其中属性向量用于确定输入信号的哪部分由混合编码器中哪种编码方法进行编码。

本实施例的混合编码器包括两种编码方法: 正弦编码器, 之后是转换编码器。该正弦编码器类似于结合第一实施例所述的编码器。该转换编码器是基于例如 R. D. Koilpillai 和 P. P. Vaidyanathan 的 "Cosine-modulated fir filter banks satisfying perfect reconstruction," (IEEE Trans. Signal Processing, vol. 40, no. 4, pp. 770-783, April 1992) 中所发现的 MDCT 滤波器组, 并编码该正弦编码器的剩余部分。关键的问题为, 哪些信号分量由该正弦编

码器编码, 哪些分量由该转换编码器编码。在本实施例中该问题转变为, 可用比特预算的哪些部分由该正弦编码器支配, 哪些部分由转换编码器支配。

图 4 示出了现有技术方法。输入信号 IN 施加到正弦编码器 SENC, 该正弦编码器 SENC 将其余信号 res 传送到转换编码器 TENC, 该转换编码器 TENC 因此旨在编码正弦编码器 SENC 所无法编码的信号。速率失真优化单元 R-D OPT 分配分别用于这两个编码器 SENC、TENC 的比特率 R-SE 和 R-TE。作为响应, 优化单元 R-D OPT 从最后一个编码器 TENC 接收结果失真 D。尝试了若干不同比特分配 R-SE 和 R-TE, 速率失真优化单元 R-D OPT 于是选择最优的分配, 即导致最低失真 D 的分配, 且该分配 R-SE 和 R-TE 随后用于产生编码的输出信号 OUT。

在选定的示例中, 尝试了下述比特分配: 100% 分给正弦编码器 (SENC), 0% 分配给转换编码器 (TENC); 75% 分配给 SENC, 25% 分配给 TENC; 50% 分配给 SENC, 50% 分配给 TENC; 25% 分配给 SENC, 75% 分配给 TENC; 0% 分配给 SENC, 100% 分配给 TENC。使用不同比特分配编码该信号, 根据得到的参数合成信号以确定相应的感觉失真。为此, 使用 S. van de Par, A. Kohlrausch, G. Charestan 和 R. Heusdens 的 "A new psychoacoustical masking model for audio coding applications," (Proc. Proc. IEEE Int. Conf. Acoust, Speech, and Signal Proc, (Orlando, Florida, USA), 2002, vol. 2, pp. 1805-1808) 中发现的感觉相关的失真度量, 其利用了输入信号的谱听觉遮蔽属性。该优选算法选择导致最低感觉失真的比特分配。

图 5 示出了根据本发明的方法。与图 4 的现有技术方法不同之处在于, 前述的属性向量 PV 输入到比特率优化单元 R-OPT, 该单元决定两个编码器 SENC 和 TENC 的最佳比特分配 R-SE 和 R-TE。在所示实施例中, 分析单元 AN 分析输入信号 IN, 并响应于该输入信号产生属性向量 PV。使用该属性向量 PV 评估最佳比特分配 R-SE 和 R-TE, 而不尝试不同的比特分配。

为了确定该任务可用的属性, 已经检查了十二个属性向量: 八个 1 维向量 (过零率、响度 (L)、调音率、谱矩心、谱带宽 (BW)、谱平坦度、帧能量、LPC 平坦度), 两个 4 维向量 (L+BW 和 SFERB: ERB 带 1-10,

10-20, 20-30, 30-37 的谱平坦度), 基于该两个 4 维属性向量的组合的一个 8 维向量, 以及一个 12 维向量 (12 阶 Mel 倒频谱)。如上所述, 高斯混合模型用于评估比特分配。所有评估器都是基于 32 元混合模型, 使用包含 6.000 帧 43ms 的音频数据库训练该模型。通过使用多维属性向量获得最佳的结果。因此, 采用与用于训练的数据库不同的数据库, 将该 4 维属性向量用于所述评估。

对图 4 和 5 的两种方法进行了比较。使用在 S. van de Par, A. Kohlrausch, G. Charestan 和 R. Heusdens, "A new psychoacoustical masking model for audio coding applications," (Proc. Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc, (Orlando, Florida, USA), 2002, vol. 2, pp. 1805-1808) 中所发现的失真度量, 确定了每帧的结果感觉失真。这两种方法得到相似的失真, 表明使用属性向量确定比特分配的可行性。

然而, 图 5 所示实施例可以按多个方式改进, 例如使用更佳的属性或者改进图 3 所示的高斯混合模型。后者的示例为: 使用更多的混合, 将评估器的可能结果限制在 0 至 100% 之间 (目前评估器是基于高斯, 高斯可以采取任意值), 改变模型的任务 (可以将帧分类成以下类别: 0、25、50、75、100%, 而不是评估 0-100% 之间的百分比)。除了高斯混合模型之外, 可以使用另外的模型。

与通过速率失真优化确定不同编解码器策略 SENC 和 TENC 之间的比特分布 R-SE 和 R-TE 的编解码器相比, 使用属性向量 PV 评估所述分布显著降低了计算复杂度。在上述实施例中, 复杂度减小到与该优化中检查的比特分配的数目相等的因数分之一。因此, 在所述示例中复杂度减小到原来的 1/5。

图 6 示出了第三实施例, 即基于属性向量 PV 的、用于确定自适应于输入信号 IN 的预先优化的分割 OSEG 的方案。

分割优化单元 SEG OPT 中针对该自适应分割 OSEG 的决定基于属性向量 PV, 并基于不同分割的模型, 例如其速率失真性能。优化后的分割 OSEG 与输入信号 IN 一起随后应用于编码器 ENC, 可编码的输出信号 OUT 产生。于是无需编码所有不同的分割可能性, 因为属性向量 PV 已经表明分割的、与输入类型有关的性能。

实际上, 对预先分割使用属性向量与速率失真评估的属性向量使

用相似。按照与第一实施例所述相同的方式，属性向量可以用于评估不同分割可能性的速率失真性能，选择具有最佳性能的分割。

与通过完全速率失真优化的速率失真相比，对预先自适应时间分割使用属性向量，显著降低了计算复杂度。复杂度降低了约等于所允许的不同分割长度的因数分之一（忽略了由属性向量引入的额外复杂度）。例如，假设在具有自适应分割的正弦编码器中允许4种不同的分割长度：10.7、16.0、21.3和26.8ms。于是通过预先分割将复杂度降低为原来的1/4。

将会理解，根据本发明的编码原理可以应用于范围广泛的应用，例如固态音频装置、CD播放器/记录器、DVD播放器/记录器、移动通信装置、（便携）计算机、例如因特网上的音频多媒体流等。

在权利要求中，仅仅为了清楚而引用附图参考符号。不应将对附图中示范性实施例的参考理解为限制本发明的范围。

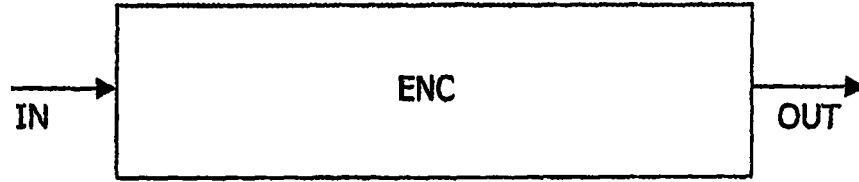


图 1  
现有技术

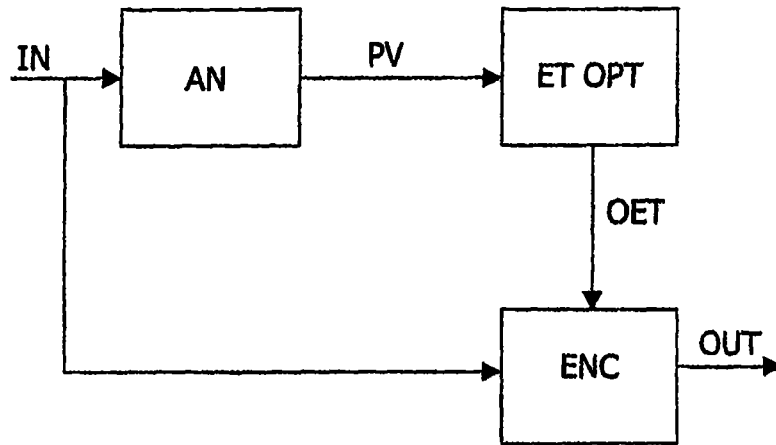


图 2

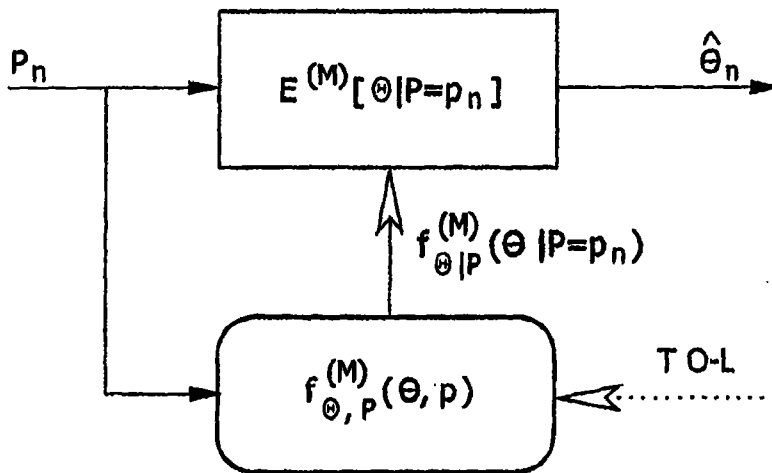


图 3

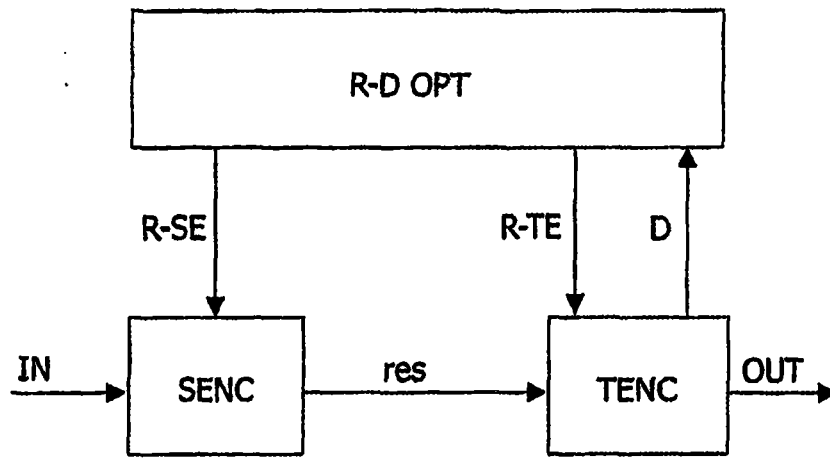


图 4  
现有技术

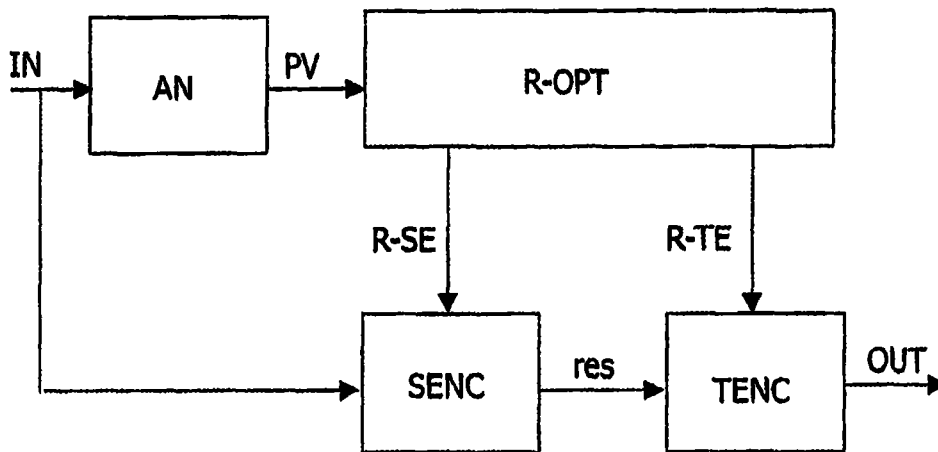


图 5

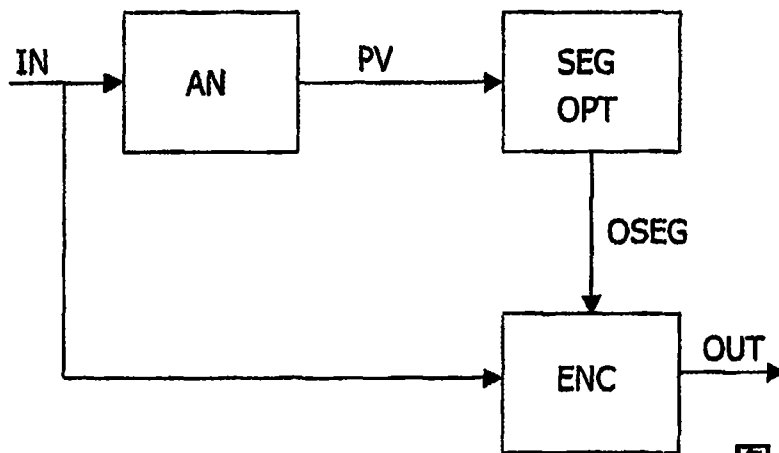


图 6