(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 101606195 B
(45)授权公告日 2012.05.02

(21)申请号 200880004749.6

(22)申请日 2008.02.12

(30)优先权数据
60/900,821 2007.02.12 US

(85)PCT申请进入国家阶段日
2009.08.12

(86)PCT申请的申请数据
PCT/US2008/001841 2008.02.12

(87)PCT申请的公布数据
WO2008/100503 EN 2008.08.21

(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 H·谬西

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 李玲

(51)Int.Cl.
G10L 19/14(2006.01)
G10L 21/02(2006.01)

(56)对比文件
WO 9953612 A1,1999.10.21, 全文.
CN 1211775 C,2005.07.20, 全文.
CN 1447963 A,2003.10.08, 全文.
US 2003182104 A1,2003.09.25, 全文.

审查员 耿中泽

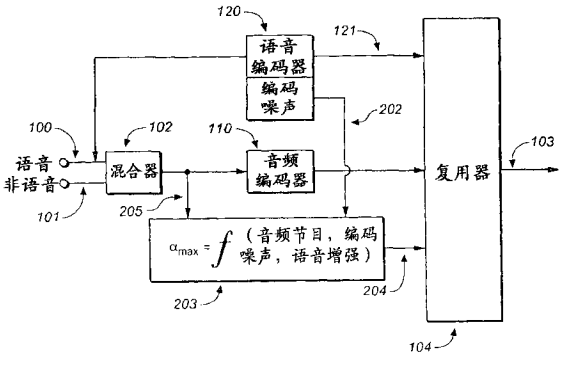
权利要求书 5 页 说明书 10 页 附图 3 页

(54)发明名称

用于年长或听力受损的收听者的改进的语音与非语音音频比值

(57)摘要

本发明涉及音频信号处理和语音增强。根据一个方面，为了产生语音与非语音音频的比值增加的高质量音频节目，以使年长的、听力受损的或是其他的收听者受益，本发明将混合了语音和非语音音频的高质量音频节目与包含在所述音频节目中的语音分量的低质量副本相组合。本发明的这些方面尤其有益于电视和家庭影院音响，但其同样适用于其他音频和音响应用。本发明涉及的是方法，用于执行此类方法的设备，以及保存在计算机可读介质上并且使计算机执行此类方法的软件。



CN 101606195 B

1. 一种用于增强具有语音和非语音分量的音频节目的语音部分的方法，包括：

接收具有语音和非语音分量的音频节目，所述音频节目具有高质量，使得在被独立再现时，所述节目不会具有让收听者觉得讨厌的听觉杂讯，

接收音频节目的语音分量副本，所述副本具有低质量，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，以及

以这样的比例组合语音分量的低质量副本与高质量音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，并且语音分量的低质量副本的听觉杂讯被高质量音频节目所遮蔽。

2. 一种通过具有语音和非语音分量的音频节目的语音分量副本来增强音频节目的语音部分的方法，所述副本具有低质量，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，所述方法包括：

以这样的比例组合语音分量的低质量副本与音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，并且语音分量的低质量副本中的听觉杂讯被音频节目所遮蔽。

3. 根据权利要求 1 或 2 所述的方法，其中组合语音分量副本与音频节目的比例使得在得到的音频节目中的语音分量具有与音频节目中的相应语音分量基本相同的动态特性，以及得到的音频节目中的非语音分量相对于音频节目中的相应非语音分量具有压缩动态范围。

4. 根据权利要求 3 所述的方法，其中得到的音频节目中的语音分量的电平与所述音频节目中的相应语音分量的电平基本相同。

5. 根据权利要求 4 所述的方法，其中得到的音频节目中的非语音分量的电平的增加比所述音频节目中的非语音分量的电平的增加要慢。

6. 根据权利要求 1 或 2 所述的方法，其中所述组合依照分别应用于语音分量副本和音频节目的互补比例因子。

7. 根据权利要求 1 或 2 所述的方法，其中所述组合是语音分量副本与音频节目的累加组合，其中用比例因子 α 来扩缩语音分量副本，以及用互补比例因子 $(1-\alpha)$ 来扩缩音频节目，α 具有 0 ～ 1 的范围。

8. 根据权利要求 7 所述的方法，其中 α 是音频节目的非语音分量的电平的函数。

9. 根据权利要求 8 所述的方法，其中 α 具有固定最大值 α_{max}。

10. 根据权利要求 8 所述的方法，其中 α 具有动态最大值 α_{max}。

11. 根据权利要求 10 所述的方法，其中值 α_{max} 基于主音频节目所导致的听觉遮蔽预测。

12. 根据权利要求 11 所述的方法，还包括接收 α_{max}。

13. 根据权利要求 7 所述的方法，其中 α 具有固定最大值 α_{max}。

14. 根据权利要求 7 所述的方法，其中 α 具有动态最大值 α_{max}。

15. 根据权利要求 14 所述的方法，其中值 α_{max} 基于主音频节目所导致的听觉遮蔽预测。

16. 根据权利要求 14 所述的方法，还包括接收 α_{max}。

17. 根据权利要求 1 或 2 所述的方法，其中组合语音分量副本与音频节目的比例使得在

得到的音频节目中的语音分量相对于音频节目中的相应语音分量具有压缩动态范围，以及得到的音频节目中的非语音分量具有与所述音频节目中的相应非语音分量基本相同的动态特性。

18. 一种用于汇编在增强具有语音和非语音分量的音频节目的语音部分的过程中使用的音频信息的方法，包括：

获取具有语音和非语音分量的音频节目，

对高质量的音频节目进行编码，使得在被解码和独立再现时，所述节目不具有让收听者觉得讨厌的听觉杂讯，

获取所述音频节目的语音分量副本，

对低质量的所述副本进行编码，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，以及

传送或存储编码音频节目以及所述音频节目的编码语音分量副本。

19. 根据权利要求 18 所述的方法，还包括：在传送或存储音频节目以及所述音频节目的语音分量副本之前，对其进行复用。

20. 一种用于汇编在增强具有语音和非语音分量的音频节目的语音部分的过程中使用的音频信息的方法，包括：

获取具有语音和非语音分量的音频节目，

对高质量的音频节目进行编码，使得在被解码和独立再现时，所述节目不具有让收听者觉得讨厌的听觉杂讯，

推导编码音频节目的听觉遮蔽阈值预测，

获取所述音频节目的语音分量副本，

对低质量的所述副本进行编码，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，

推导编码副本的编码噪声的量度，以及

传送或存储编码音频节目、其听觉遮蔽阈值预测、音频节目的编码语音分量副本、以及其编码噪声的量度。

21. 根据权利要求 20 所述的方法，还包括：在传送或存储音频节目、其听觉遮蔽阈值预测、音频节目的语音分量副本以及其编码噪声的量度之前对其进行复用。

22. 一种用于汇编在增强具有语音和非语音分量的音频节目的语音部分的过程中使用的音频信息的方法，包括：

获取具有语音和非语音分量的音频节目，

对高质量的音频节目进行编码，使得在被解码和独立再现时，所述节目不具有让收听者觉得讨厌的听觉杂讯，

推导编码音频节目的听觉遮蔽阈值预测，

获取所述音频节目的语音分量副本，

对低质量的所述副本进行编码，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，

推导编码副本的编码噪声的量度，

推导基于听觉遮蔽阈值预测以及编码噪声的量度的函数的参数，以及

传送或存储编码音频节目、音频节目的编码语音分量副本以及所述参数。

23. 根据权利要求 22 所述的方法，还包括：在传送或存储音频节目、音频节目的语音分量副本以及所述参数之前对其进行复用。

24. 一种用于增强具有语音和非语音分量的音频节目的语音部分的设备，包括：

用于接收具有语音和非语音分量的音频节目的装置，所述音频节目具有高质量，使得在被独立再现时，所述节目不会具有让收听者觉得讨厌的听觉杂讯，

用于接收音频节目的语音分量副本的装置，所述副本具有低质量，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，以及

用于以这样的比例组合语音分量的低质量副本与高质量音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，并且语音分量的低质量副本的听觉杂讯被高质量音频节目所遮蔽的装置。

25. 一种通过具有语音和非语音分量的音频节目的语音分量副本来增强音频节目的语音部分的设备，所述副本具有低质量，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，所述方法包括：

用于以这样的比例组合语音分量的低质量副本与音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，并且语音分量的低质量副本中的听觉杂讯被音频节目所遮蔽的装置。

26. 根据权利要求 24 或 25 所述的设备，其中组合语音分量副本与音频节目的比例使得在得到的音频节目中的语音分量具有与音频节目中的相应语音分量基本相同的动态特性，以及得到的音频节目中的非语音分量相对于音频节目中的相应非语音分量具有压缩动态范围。

27. 根据权利要求 26 所述的设备，其中得到的音频节目中的语音分量的电平与所述音频节目中的相应语音分量的电平基本相同。

28. 根据权利要求 27 所述的设备，其中得到的音频节目中的非语音分量的电平的增加比所述音频节目中的非语音分量的电平的增加要慢。

29. 根据权利要求 24 或 25 所述的设备，其中所述组合依照分别应用于语音分量副本和音频节目的互补比例因子。

30. 根据权利要求 24 或 25 所述的设备，其中所述组合是语音分量副本与音频节目的累加组合，其中用比例因子 α 来扩缩语音分量副本，以及用互补比例因子 $(1-\alpha)$ 来扩缩音频节目，α 具有 0 ～ 1 的范围。

31. 根据权利要求 30 所述的设备，其中 α 是音频节目的非语音分量的电平的函数。

32. 根据权利要求 31 所述的设备，其中 α 具有固定最大值 α_{max}。

33. 根据权利要求 31 所述的设备，其中 α 具有动态最大值 α_{max}。

34. 根据权利要求 33 所述的设备，其中值 α_{max} 基于主音频节目所导致的听觉遮蔽预测。

35. 根据权利要求 34 所述的设备，还包括用于接收 α_{max} 的装置。

36. 根据权利要求 30 所述的设备，其中 α 具有固定最大值 α_{max}。

37. 根据权利要求 30 所述的设备，其中 α 具有动态最大值 α_{max}。

38. 根据权利要求 37 所述的设备，其中值 α_{max} 基于主音频节目所导致的听觉遮蔽预

测。

39. 根据权利要求 37 所述的设备，还包括用于接收 α_{max} 的装置。

40. 根据权利要求 24 或 25 所述的设备，其中组合语音分量副本与音频节目的比例使得在得到的音频节目中的语音分量相对于音频节目中的相应语音分量具有压缩动态范围，以及得到的音频节目中的非语音分量具有与所述音频节目中的相应非语音分量基本相同的动态特性。

41. 一种用于汇编在增强具有语音和非语音分量的音频节目的语音部分的过程中使用的音频信息的设备，包括：

用于获取具有语音和非语音分量的音频节目的装置，

用于对高质量的音频节目进行编码，使得在被解码和独立再现时，所述节目不具有让收听者觉得讨厌的听觉杂讯的装置，

用于获取所述音频节目的语音分量副本的装置，

用于对低质量的所述副本进行编码，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯的装置，以及

用于传送或存储编码音频节目以及所述音频节目的编码语音分量副本的装置。

42. 根据权利要求 41 所述的设备，还包括：用于在传送或存储音频节目以及所述音频节目的语音分量副本之前，对其进行复用的装置。

43. 一种用于汇编在增强具有语音和非语音分量的音频节目的语音部分的过程中使用的音频信息的设备，包括：

用于获取具有语音和非语音分量的音频节目的装置，

用于对高质量的音频节目进行编码，使得在被解码和独立再现时，所述节目不具有让收听者觉得讨厌的听觉杂讯的装置，

用于推导编码音频节目的听觉遮蔽阈值预测的装置，

用于获取所述音频节目的语音分量副本的装置，

用于对低质量的所述副本进行编码，使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯的装置，

用于推导编码副本的编码噪声的量度的装置，以及

用于传送或存储编码音频节目、其听觉遮蔽阈值预测、音频节目的编码语音分量副本、以及其编码噪声的量度的装置。

44. 根据权利要求 43 所述的设备，还包括：用于在传送或存储音频节目、其听觉遮蔽阈值预测、音频节目的语音分量副本以及其编码噪声的量度之前对其进行复用的装置。

45. 一种用于汇编在增强具有语音和非语音分量的音频节目的语音部分的过程中使用的音频信息的设备，包括：

用于获取具有语音和非语音分量的音频节目的装置，

用于对高质量的音频节目进行编码，使得在被解码和独立再现时，所述节目不具有让收听者觉得讨厌的听觉杂讯的装置，

用于推导编码音频节目的听觉遮蔽阈值预测的装置，

用于获取所述音频节目的语音分量副本的装置，

用于对低质量的所述副本进行编码，使得在被独立再现时，所述副本具有让收听者觉

得讨厌的听觉杂讯的装置，

用于推导编码副本的编码噪声的量度的装置，

用于推导基于听觉遮蔽阈值预测以及编码噪声的量度的函数的参数的装置，以及

用于传送或存储编码音频节目、音频节目的编码语音分量副本以及所述参数的装置。

46. 根据权利要求 45 所述的设备，还包括：用于在传送或存储音频节目、音频节目的语音分量副本以及所述参数之前对其进行复用的装置。

用于年长或听力受损的收听者的改进的语音与非语音音频比值

技术领域

[0001] 本发明涉及音频信号处理和语音增强。根据一个方面，为了产生语音与非语音音频比值增加的高质量音频节目，以使年长的、听力受损的或是其他的收听者受益，本发明将混合了语音和非语音音频的高质量音频节目与包含在所述音频节目中的语音分量的低质量副本组合。本发明的这些方面尤其有益于电视和家庭影院音响，但是它们同样适用于其他的音频和音响应用。本发明涉及的是方法、用于执行此类方法的设备、以及保存在计算机可读介质上使计算机执行此类方法的软件。

背景技术

[0002] 在电影或电视中，对话和叙述通常是连同音乐、广告词、效果以及周围环境之类的其他非语音声响一起呈现的。在很多情况中，语音声响和非语音声响是被单独记录的，并且是在录音师的控制下混合在一起的。在混合语音和非语音声响时，非语音声响有可能局部遮蔽语音，由此导致一部分语音无法被听到。结果，收听者必须根据剩余的部分信息来理解该语音。少量遮蔽是很容易被耳朵健康的年轻收听者容忍的。但是，随着遮蔽的增加，理解起来将会逐渐变得困难，直至最终无法理解语音（相关示例参见 ANSI S3.5 1997“Methods forCalculation of the Speech Intelligibility Index”）。录音师凭直觉知道这种关系，并且会以那些通常为大多数观众提供足够可懂度的相对水平来混合语音与背景。

[0003] 当背景声响妨碍了所有观众的可懂度时，对于年长者和听力受损的人来说，背景声响的有害效果要更大一些（比照 Killion，M. 在 2002 年发表于 Thieme Medical Publishers，New York，NY 出版的 Seminarsin Hearing 第 23 卷第 1 号第 57 ～ 75 页的“New thinking on hearing innoise：A generalized Articulation Index”）。录音师通常具有正常的听力并且至少要比他的一部分听众年轻，该录音师根据其自身的内部标准来选择语音与非语音音频的比值。有时，这会使相当一部分听众花费很大力气才能跟得上对话或叙述。

[0004] 本领域中已知的一种解决方案利用语音与非语音音频单独存在于生产线 (production chain) 上的某些点的事实来为观众提供两个独立音频流。一个流传送主内容音频（主要是语音），另一个流传送次要内容音频（排除语音的剩余音频节目）。用户被给予混合处理的控制权。不幸的是，由于该方案并不是构建在传送完全混合的音频节目的现行实践上的，因此，该方案是不切实际的。相反，它会用两个现今未使用的音频流来替换主音频节目。该方法的另一个缺点还在于：由于必须向用户递送两个独立音频流，并且每一个音频流都具有广播质量，因此，其需要的带宽大约是当前广播实践的两倍。

[0005] 成功的音频编码标准 AC-3 允许同时传递主音频节目和其他相关音频流。所有的流都具有广播质量。这些相关音频流之一是用于听力受损的人的。根据可以在 http://www.dobly.com/assets/pdf/tech_library/46_DDEncodingGuidelines.pdf 得到的“Dolby DigitalProfessional Encoding Guidelines”第 5.4.4 节，这个音频流通常只包括对

话，并且是以固定比率添加到已经包含该对话副本的主音频节目的中心声道的（如果主音频是双声道立体声，则添加至左和右声道）。相关情况也可以参见 ATSC 标准：Digital Television Standard(A/53)，revision D，Including Amendment No.1，Section 6.5 Hearing Impaired(HI)。关于 AC-3 的更多细节可以在标题“引入的参考文献”下方的 AC-3 引文中找到。

[0006] 从以上论述中可以清楚了解，当前需要但却无法实现的是：以利用语音与非语音音频是被单独记录的事实的方式来增加语音与非语音音频的比值，同时还要构建在传送完全混合的音频节目的现行实践上以及还需要最小的附加带宽。因此，本发明的目的是提供一种用于在电视广播中可选地增加语音与非语音音频的比值的方法，所述方法只需要少量附加带宽，利用语音与非语音音频是被单独记录的事实，并且是现有广播实践的扩展而不是替换。

发明内容

[0007] 根据本发明用于增强具有语音和非语音分量的音频节目中的语音部分的第一个方面，接收具有语音和非语音分量的音频节目，所述音频节目具有高质量使得在独立再现所述节目时，所述节目不会具有让收听者觉得讨厌的听觉杂讯 (audible artifact)。接收音频节目的语音分量副本，所述副本具有低质量使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，以及以这样的比例组合语音分量的低质量副本与高质量音频节目，使得在得到的音频节目中，语音与非语音分量的比值被增加，并且语音分量的低质量副本的听觉杂讯被高质量音频节目所遮蔽。

[0008] 根据本发明的一个方面，其中具有语音和非语音分量的音频节目的语音部分是用所述音频节目的语音分量副本来增强的，所述副本具有低质量使得在被独立再现时，所述副本具有让收听者觉得讨厌的听觉杂讯，所述语音分量的低质量副本与音频节目以这样的比例组合，使得在得到的音频节目中语音与非语音分量的比值被增加，并且语音分量的低质量副本的听觉杂讯被音频节目所遮蔽。

[0009] 在上述任一方面中，组合语音分量副本与音频节目的比例使得在得到的音频节目中的语音分量具有与音频节目中的相应语音分量基本相同的动态特性，以及得到的音频节目中的非语音分量相对于音频节目中的相应非语音分量具有压缩的动态范围。

[0010] 作为替换，在上述任一方面中，组合语音分量副本与音频节目的比例使得在得到的音频节目中的语音分量相对于音频节目中的相应语音分量具有压缩的动态范围，以及得到的音频节目中的非语音分量与所述音频节目中的相应非语音分量具有基本相同的动态特性。

[0011] 根据本发明的另一个方面，用于增强具有语音和非语音分量的音频节目的语音部分的处理包括：接收具有语音和非语音分量的音频节目，接收音频节目的语音分量副本，以及以这样的比例组合语音分量副本与音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，得到的音频节目中的语音分量具有与音频节目中的相应语音分量基本相同的动态特性，以及得到的音频节目中的非语音分量相对于音频节目中的相应非语音分量具有压缩的动态范围。

[0012] 根据本发明的另一个方面，使用具有语音和非语音分量的音频节目的语音分量副

本来增强音频节目的语音部分的处理，包括：以这样的比例组合语音分量副本与音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，得到的音频节目中的语音分量具有与音频节目中的相应语音分量基本相同的动态特性，以及得到的音频节目中的非语音分量相对于音频节目中的相应非语音分量具有压缩的动态范围。

[0013] 根据本发明用于增强具有语音和非语音分量的音频节目的语音部分的另一个方面，接收具有语音和非语音分量的音频节目，接收音频节目的语音分量副本，以及以这样的比例组合语音分量副本与音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，得到的音频节目中的语音分量相对于音频节目中的相应语音分量具有压缩的动态范围，以及得到的音频节目中的非语音分量具有与音频节目中的相应非语音分量基本相同的动态特性。

[0014] 根据本发明使用具有语音和非语音分量的音频节目的语音分量副本来增强音频节目的语音部分的另一个方面，以这样的比例组合语音分量副本与音频节目，使得在得到的音频节目中语音与非语音分量的比值被增加，得到的音频节目中的语音分量相对于音频节目中的相应语音分量具有压缩的动态范围，以及得到的音频节目中的非语音分量具有与音频节目中的相应非语音分量基本相同的动态范围特性。

[0015] 虽然用于实施本发明的示例处于电视或家庭影院音响的环境中，但是本领域普通技术人员将会理解，本发明同样可以在其他的音频和音响应用中使用。

[0016] 如果电视或家庭影院观众可以使用主音频节目以及只包含语音分量的单独音频流，则可以通过适当地扩缩和混合这两个分量来实现任何的语音与非语音音频比值。举个例子，如果希望完全抑制非语音音频而只听语音，那么所播放的仅仅是包含语音声响的流。在另一种极端情况中，如果希望完全抑制语音而只听非语音音频，则简单地从主音频节目中扣除语音音频。在这两种极端情况之间，语音与非语音音频的任何中间比值都是可以实现的。

[0017] 为使次要语音声道具有商业可行性，分配给主音频节目的带宽是不允许增加太多的。为了满足这个约束条件，次要语音必须用一个极大减小数据速率的编码器来进行编码。这种数据速率降低是以导致语音信号失真为代价的。因为低比特率编码而发生失真的语音可以被描述成是原始语音与失真分量（编码噪声）的组合。当这种失真可以被听到时，它会降低觉察到的语音的声音质量。虽然编码噪声对信号的声音质量具有严重影响，但其水平通常远远低于所编码的信号。

[0018] 在实践中，主音频节目具有“广播质量”，并且与之关联的编码噪声几乎是觉察不到的。换句话说，在被独立再现时，所述节目具有不让收听者觉得讨厌的听觉杂讯。另一方面，根据本发明的一些方面，如果独立收听次要语音，那么由于其数据速率严重受限，因此，所述次要语音有可能具有让收听者觉得讨厌的听觉杂讯。如果是被独立听到的，那么所述次要语音的质量是不适合广播应用的。

[0019] 在与主音频节目混合之后，是否能够听到与次要语音相关联的编码噪声取决于主音频节目是否遮蔽所述编码噪声。这种遮蔽很有可能是在主节目除了语音音频之外还包含很强的非语音音频的时候发生。相比之下，当语音在主节目中占首要并且非语音音频很弱或是没有非语音音频的时候，编码噪声不太可能会被遮蔽。从使用次要语音来提高语音在主音频节目中的相对电平的角度看，这些关系是非常有利的。最有可能从添加次要语音中

受益的节目部分（也就是具有很强的非语音音频的部分）同样最有可能遮蔽编码噪声。相反，最容易被编码噪声降级的节目部分（例如缺少背景声响的语音）也最不可能需要增强的对话。

[0020] 这些观察表明，如果利用信号自适应混合处理，则有可能将听觉上失真的次要语音与高质量的主音频节目相组合，以创建没有听觉失真且语音与非语音音频比值被增加的音频节目。优选地，自适应混合器限制了相对混合等级，使得编码噪声保持在主音频节目所引起的遮蔽阈值以下。所述处理可以通过初始时只在音频节目中那些具有低的语音与非语音音频比值的部分中添加低质量次要语音来实现。以下描述这种原理的示例性实施方式。

附图说明

[0021] 图 1 是实现本发明各方面的编码器或编码功能的示例。

[0022] 图 2 是实现本发明各方面并且包括自适应交叉渐变器 (crossfader) 的解码器或解码功能的示例。

[0023] 图 3 是可以在图 2 的示例中使用的函数 α = f(P) 的示例。

[0024] 图 4 是在函数 α = f(P) 具有图 3 所示的特性时，将得到的音频节目中的非语音音频功率 P' 与图 2 示例中的得到的音频节目中的非语音音频功率 P 相对比的图表。

[0025] 图 5 是实现本发明各方面并且包含某些非语音分量的动态范围压缩的解码器或解码功能的示例。

[0026] 图 6 是在理解图 5 的过程中使用的压缩器输入功率与输出功率特性的对比图。

[0027] 图 7 是实现本发明各方面的编码器或编码功能的示例，其中所述编码器或编码功能可选地包括产生可在解码过程中使用的一个或多个参数。

具体实施方式

[0028] 图 1 和 2 分别显示的是实现本发明各方面的编码和解码方案。图 5 显示的是实现本发明各方面的备选解码方案。参考图 1 中用于实现本发明各方面的编码器或编码功能的示例，作为音频节目生成处理器或进程的一部分，在混合控制台或混合功能（“混合器”）102 中混合电视音频节目的两个分量，其中一个分量主要包含的是语音 100，另一个则主要包含的是非语音 101。包含语音和非语音信号二者的得到的音频节目是用 AC-3 或 AAC 之类的高比特率、高质量的音频编码器或编码功能（“音频编码器”）100 编码的。关于 AAC 的更多细节可以在标题“引入的参考文献”下方的 AAC 引文中找到。主要包含语音 100 的节目分量是用编码器或编码功能（“语音编码器”）120 同时编码的，所述编码器或编码功能以低于音频编码器 110 产生的比特率的比特率来产生编码音频。语音编码器 120 实现的音频质量远远不如音频编码器 110 实现的音频质量。语音编码器 120 可以通过优化来编码语音，但是还应该尝试保持信号相位。满足这个准则的编码器本身是已知的。一个示例是码激励线性预测 (CELP) 编码器。与其他那些所谓的“混合编码器”相似，CELP 编码器使用语音生成源过滤器模型来为语音信号建模，以便实现高编码增益，此外它还尝试保持要编码的波形，从而限制相位失真。

[0029] 在关于本发明各方面的实验性实施方式中，发现以 8K 比特 / 秒运行的 CELP 声码器实现的语音编码器是非常合适的，并且它提供的可感知当量大约是 10-dB 的语音与非语

音音频电平增量。

[0030] 如果这两个编码器的编码延迟不同，则应该在时间上移位至少一个信号，以便保持信号之间的时间对准（未显示）。随后，高质量音频编码器 110 和低质量语音编码器 120 二者的输出都可以由复用器或复用功能（“复用器”）组合成单个比特流，并被封装到适合广播或存储的比特流 103 中。

[0031] 现在参考图 2 中用于实现本发明各方面的解码器或解码功能的示例，接收比特流 103。其中举例来说，所述比特流是从广播接口接收或是从存储介质中检索得到的，并且所述比特流被施加给解复用器或解复用功能（“解复用器”）105，在那里它会被拆包和解复用，以便产生编码主音频节目 111 以及编码语音信号 121。编码主音频节目由音频解码器或解码功能（“音频解码器”）130 解码，以便产生解码主音频信号 131，并且解码语音信号由语音解码器或解码功能（“语音解码器”）140 解码，以便产生解码语音信号 141。在本示例中，这两个信号在交叉渐变器或交叉渐变功能（“交叉渐变器”）160 中组合，以便产生输出信号 180。这个信号同样会被传递到用于测量非语音音频 151 的功率电平 P 的设备或功能（“非语音音频电平”）150，其中所述测量是通过从解码主音频节目的功率中减去解码语音信号的功率来执行的。交叉渐变由加权或比例因子 α 控制。所述加权因子 α 转而是通过变换 170 而从非语音音频 150 的功率电平 P 中得到的。换句话说，α 是 P 的函数（即 $\alpha = f(P)$）。最终得到的是信号自适应混合器。这种变换或函数通常会使局限于非负数的 α 值随着功率电平 P 的增加而增加。比例因子 α 应该被限制成不超出最大值 α_{max}，其中 $\alpha_{max} < 1$，但是如下文进一步说明的那样，所述比例因子无论如何也不会大到无法遮蔽编码噪声。如下文中进一步说明的那样，非语音音频 150 的电平、变换 170 以及交叉渐变 160 构成了信号自适应交叉渐变器或交叉渐变功能（“信号自适应交叉渐变器”）181。

[0032] 在交叉渐变器 160 中，在累加组合解码次要语音和解码主音频节目之前，信号自适应交叉渐变器 181 将解码次要语音扩缩 α 倍，并且将解码主音频节目扩缩 (1-α) 倍。扩缩处理的对称性使得到的信号中的语音分量的电平和动态特性与比例因子 α 无关——扩缩处理既不会影响得到的信号中的语音分量的电平，也不会对语音分量的动态范围施加任何动态范围压缩或其他修改。相比之下，得到的信号中的非语音音频的电平会受到扩缩处理的影响。特别地，由于 α 的值会随着非语音音频的功率电平 P 的增加而增加，因此，扩缩处理往往会抵消该电平的任何变化，由此有效压缩非语音音频信号的动态范围。动态范围压缩形式是由变换 170 确定的。举个例子，如果函数 $\alpha = f(P)$ 采用的是图 3 所示的形式，那么如图 4 所示，得到的音频节目中的非语音音频的功率 P’与非语音音频的功率 P 相对比的图表示出了一个压缩特性——高于最小非语音功率电平，与非语音功率电平相比，得到的非语音功率增加的较慢。

[0033] 自适应交叉渐变器 181 的功能可以概括如下：当非语音音频分量电平很低时，比例因子 α 为零或者很小，并且自适应交叉渐变器输出一个与解码主音频节目相等或几乎相等的信号。当非语音音频的电平增加时，α 的值也会增加。这导致解码次要语音为最终的音频节目 180 做出较大贡献，并且更大地抑制解码主音频节目，包括其非语音音频分量。次要语音对增强信号贡献的增加是通过语音在主音频节目中的贡献的减小来平衡的。结果，增强信号中的语音的电平保持不受自适应交叉渐变操作的影响——增强信号中的语音的电平与解码语音音频信号 141 的电平基本上是相同的电平，并且非语音音频分量的动态

范围会减小。由于没有不必要的语音信号调制，因此，这是一个非常期望的结果。

[0034] 为了保持语音电平不变，为动态范围压缩的主音频信号所添加的次要语音的量应该是施加给主音频信号的压缩的量的函数。所添加的次要语音补偿了因为压缩而导致产生的电平减小。这种减小是将比例因子 α 应用于次要语音信号以及将互补比例因子 (1-α) 应用于主音频而自动得到的，其中 α 是应用于主音频的动态范围压缩的函数。作用于主音频的效果与 AC-3 中的“夜间模式”所提供的效果相似，其中随着主音频电平的增加，输出会根据压缩特性而被调低。

[0035] 为了确保编码噪声不会暴露，自适应交叉渐变器 160 应该防止对于主音频节目的抑制作用超出一个临界值。这可以通过将 α 限制成小于或等于 α_{max} 来实现。虽然在 α_{max} 是固定值时可以实现令人满意的性能，但如果 α_{max} 是用心理声学遮蔽模型得到的，则有可能获得更好的性能，其中所述心理声学遮蔽模型将关联于低质量语音信号 141 的编码噪声频谱与主音频节目信号 131 引起的预测听觉遮蔽阈值相比较。

[0036] 参考图 5 中用于实现本发明各方面的解码器或解码功能的替换示例，其中举例来说，比特流 103 是从广播接口接收或是从存储介质检索得到的，并且所述比特流被应用于解复用器或解复用功能 (“解复用器”)105，以便产生编码主音频节目以及编码语音信号 121。编码主音频节目由音频解码器或解码功能 (“音频解码器”)130 解码，以便产生解码主音频信号 131，并且解码语音信号由语音解码器或解码功能 (“语音解码器”)140 解码，以便产生解码语音信号 141。信号 131 和 141 被传递到用于测量非语音音频 151 的功率电平 P 的设备或功能 (“非语音音频电平”)150，其中举例来说，所述测量是通过从解码主音频节目的功率中减去解码语音信号的功率来执行的。到目前为止的描述中，图 5 的示例与图 2 的示例是相同的。但是，图 5 解码器示例的剩余部分是不同的。在图 5 的示例中，解码语音信号 141 会进行动态范围压缩器或压缩功能 (“动态范围压缩器”)301。压缩器 301 是图 6 所示的输入 / 输出功能的示例，它不但会传递未修改的语音信号的高电平部分，而且还会随着应用于压缩器 301 的语音信号电平的减小而逐渐施加更大的增益。在压缩之后，解码语音副本会在用乘法器符号 302 显示的复用器 (或扩缩器 (scalar)) 或乘法 (或扩缩) 功能中扩缩 α 倍，并且会在用加法符号 304 显示的累加组合器或组合功能中被添加给解码主音频节目。压缩器 301 和乘法器 302 的顺序可以是颠倒的。

[0037] 图 5 示例的功能可以概括如下 :当非语音音频分量的电平很低时，比例因子 α 为零或者很小，并且为主音频节目添加的语音的量为零或是可以忽略。由此，所产生的信号与解码主音频节目相等或近似相等。当非语音音频分量电平增加时，α 的值也会增加。这会导致压缩语音为最终的音频节目做出更大的贡献，由此导致最终的音频节目中的语音与非语音分量的比值增加。当语音电平低时，次要语音的动态范围压缩处理会允许大的语音电平增加，而语音电平高时，所述处理只会少量增大语音电平。由于所述处理确保了语音的峰值音量不会大为增加，同时极大增加了软语音部分的音量，因此所述处理是一个非常重要的属性。这样一来，得到的音频节目中的语音与非语音分量的比值会增加，得到的音频节目中的语音分量相对于音频节目中的相应语音分量具有压缩的动态范围，而得到的音频节目中的非语音分量与音频节目中的相应的非语音分量具有基本相同的动态范围特性。

[0038] 图 2 和 5 的解码示例都具有增加语音与非语音比值并且由此使语音更易于理解的属性。在图 2 的示例中，语音分量的动态特性原则上是不会改变的，而非语音分量的动态特

性则会改变（其动态范围被压缩）。而在图 5 的示例中，情况则正好相反——语音分量的动态特性被改变（其动态范围被压缩），而非语音动态特性原则上是不会改变的。

[0039] 在图 5 的示例中，解码语音副本信号会进行动态范围压缩处理，并且会按照比例因子 α 而被扩缩（无论哪一种顺序）。以下的说明可以用于理解其组合效果。设想这样一种情况，其中非语音音频具有高电平，由此 α 很大（例如 α ＝ 1）。此外，在这里还设想语音电平来自压缩器 301 ：

[0040] (a) 当语音电平高时（语音峰值），压缩器不会提供增益，并且会在不做修改的情况下传递所述信号（如图 6 的输入 / 输出功能所示，在高电平，响应特性与虚对角线相重合，其中所述虚对角线标记的是输出与输入相同时的关系）。由此，在语音峰值期间，压缩器输出端的语音电平与主音频中的语音峰值的电平是相同的。一旦在主音频中添加了解码语音副本音频，那么相加得到的语音峰值的电平比原始语音峰值高 6dB。非语音音频的电平不会改变，由此语音与非语音音频的比值增加 6dB ；以及

[0041] (b) 当语音电平低时（例如软辅音），压缩器提供相当大的增益量（输入 / 输出曲线高出图 6 的虚对角线很多）。出于论述目的，假设压缩器施加了 20dB 的增益。由于所述语音主要是来自解码语音副本信号的语音，因此，一旦在压缩器输出中添加了主音频，则语音与非语音音频的比值会增加大约 20dB。当非语音音频的电平减小时，α 会减小并且会添加逐渐消弱的解码语音副本。

[0042] 虽然压缩器 301 的增益并不重要，但是我们发现，可以接受的增益约为 15 ～ 20dB。

[0043] 通过考虑图 5 示例在没有压缩器 301 时的操作，可以更好地理解其用途。在所述情况中，语音与非语音音频比值的增加与 α 成正比。如果 α 被限制成不超过 1，则语音与非语音改进的最大量是 6dB，这是一个合理的改进，但其小于可能的期望值。如果允许 α 大于 1，则语音与非语音改进同样有可能变得更大，但是，如果假设语音电平高于非语音音频的电平，则总的电平同样也会增加，并且有可能产生诸如过载或音量过大之类的问题。

[0044] 诸如过载或音量过大之类的问题可以通过包含压缩器 301 以及在主音频中添加压缩语音来克服。再次假设 α ＝ 1。当瞬时语音电平很高时，压缩器是没有效果的 (0dB 增益），并且总和信号的语音电平的增加量是很少的 (6dB)。这与没有压缩器 301 的情形是相同的。但是，当瞬时语音电平很低时（假设比峰值电平低 30dB)，压缩器将会施加高增益（假设是 15dB)。在被添加给主音频时，得到的音频中的瞬时语音电平实际上是受压缩次要音频支配的，也就是说，瞬时语音电平被提高大约 15dB。这相当于 6dB 的语音峰值提升。由此，即使 α 是恒定的（例如由于非语音音频分量的功率电平 P 是恒定的），也还是存在时变的语音与非语音改进，并且这种改进在语音低谷中是最大的，而在语音峰值处则是最小的。

[0045] 随着非语音音频电平的减小以及 α 的减小，总和音频中的语音峰值几乎保持不变。这是因为解码语音副本信号的电平远远低于主音频中的语音的电平（由于 α ＜ 1 引入的衰减），并且将这二者加在一起也不会显著影响得到的语音信号的电平。对低电平语音部分来说，情况是不同的。它们接收来自压缩器的增益以及由于 α 所导致的衰减。最终结果是次要语音的电平可以相当于（甚至会大于，取决于压缩器设置）主音频中的语音电平。在将其加在一起时，它们不会影响（增加）总和信号中的语音分量的电平。

[0046] 最终结果是 ：与语音谷底的语音电平相比，语音峰值的电平更加“稳定”（也就是不会发生大于 6dB 的变化）。语音与非语音比值会在最需要增加的时候增加最多，而语音峰

值电平的变化则相对较小。

[0047] 由于心理声学模型的计算成本很高，因此，从成本的角度来看，较为期望的是在编码端而不是解码端推导 α 的最大可允许值，并且将这个值或是易于计算出该值的分量作为一个或多个参数来进行传送。例如，所述值可以作为一系列的 α_{max} 值而被传送到解码端。在图 7 中显示了关于这种方案的示例。所述方案的关键要素是用于推导满足约束条件的 α 的最大值的功能或设备（“$\alpha_{max} = f$（音频节目，编码噪声，语音增强）”）203，其中所述约束条件是因为解码器的得到的音频输出中的音频信号分量所导致的预测听觉遮蔽阈值比解码器的得到的音频输出中的次要语音分量的编码噪声超出指定的安全裕量。为此目的，功能或设备 203 接收作为输入的主音频节目 205 以及与次要语音 100 的编码处理相关联的编码噪声 202。编码噪声的表示可以采用若干种方式来获取。例如，编码语音 121 可以再次解码，并被从输入语音 100 中减去（未示出）。包括 CELP 编码器之类的混合编码器的很多编码器是根据“合成 - 分析”准则工作的。作为正常操作的一部分，根据合成 - 分析准则工作的编码器执行的是从原始语音中减去解码语音，以便获取编码噪声量度的步骤。如果使用这种编码器，则可以在不需要附加计算的情况下直接得到编码噪声 202 的表示。

[0048] 根据使用 α_{max} 的解码器配置，功能或设备 203 还知道解码器执行的处理及其操作细节。适当的解码器配置可以采用图 2 示例或图 5 示例的形式。

[0049] 如果功能或设备 203 产生的关于 α_{max} 值的信息流将要供如图 2 所示的解码器使用，那么功能或设备 203 可以执行以下操作：

[0050] a) 将主音频节目 205 扩缩 $1-\alpha_i$ 倍，其中 α_i 是期望结果 α_{max} 的初始猜测值。

[0051] b) 使用听觉遮蔽模型来预测经过扩缩的主音频节目所导致的听觉遮蔽阈值。对本领域普通技术人员来说，听觉遮蔽模型是众所周知的。

[0052] c) 将关联于次要语音的编码噪声 202 扩缩 α_i 倍。

[0053] d) 将经过扩缩的编码噪声与预测听觉遮蔽阈值相比较。如果预测听觉遮蔽阈值比经过扩缩的编码噪声超出期望安全裕量以上，则增加 α_i 的值并重复步骤 (a) 到 (d)。相反，如果关于 α_i 的初始猜测值产生的是比经过扩缩的编码噪声加上安全裕量还要小的预测听觉遮蔽阈值，则减小 α_i 的值。所述迭代处理会继续进行，直至找到期望的 α_{max} 值。

[0054] 如果功能或设备 203 产生的关于 α_{max} 值的信息流被如图 5 所示的解码器使用，那么功能或设备 203 可以执行以下操作：

[0055] a) 按照某个增益以及比例因子 α_i 来扩缩与次要语音相关联的编码噪声 202，其中所述增益与图 5 压缩器 301 施加的增益相等，并且 α_i 是期望结果 α_{max} 的初始猜测值。

[0056] b) 使用听觉遮蔽模型来预测主音频节目所导致的听觉遮蔽阈值。如果音频编码器 110 引入了听觉遮蔽模型，则可以使用关于所述模型的预测，由此极大节约了计算成本。

[0057] c) 将经过扩缩的编码噪声与预测听觉遮蔽阈值相比较。如果预测听觉遮蔽阈值比经过扩缩的编码噪声超出期望安全裕量以上，则增加 α_i 的值，并且重复步骤 (a) 到 (c)。相反，如果关于 α_i 的初始猜测值产生的是比经过扩缩的编码噪声加上安全裕量还要小的预测听觉遮蔽阈值，则减小 α_i 的值。这种迭代处理会继续进行，直至找到期望的 α_{max} 值。

[0058] α_{max} 的值应该以一个足够高的速率来更新，以便充分反映预测遮蔽阈值和编码噪声 202 的变化。最后，编码次要语音 121、编码主音频节目 111 以及关于 α_{max} 值的信息流可以依次通过复用器或复用功能（“复用器”）104 而被组合为单个比特流，并且随后被封装到

适合广播或存储的单个数据流 103 中。本领域技术人员会了解，在不同的例示实施例中，用于比特流的复用、解复用、封装以及拆包之类的细节对本发明并不重要。

[0059] 本发明的各方面包括上述示例的修改和扩展。例如，语音信号和主信号中的每一个都可以分成相应的频率子波段，其中在一个或多个这种子波段中应用上述处理，并且得到的子波段信号被重新组合，以便产生输出信号，这与解码器或解码处理时一样的。

[0060] 本发明的各方面还允许用户控制对话增强度。所述处理可以通过使用附加用户可控比例因子 β 来扩缩比例因子 α 来获取经过修改的比例因子 α’来实现，也就是说，α’＝ β ＊ α，其中 0 ≤ β ≤ 1。如果选择 β 为零，则会始终听到未经修改的主音频节目。如果选择 β 等于 1，则应用最大量的对话增强。由于 α_{max} 确保了永远都会遮蔽编码噪声，并且由于用户只能相对于最大增强度来减小对话增强度，所述调整并不会带来让编码失真可能被听到的风险。

[0061] 在刚刚描述的实施例中，对话增强是在解码音频信号上执行的。这一点并不是本发明的固有限制。在一些状况中，例如在音频编码器和语音编码器使用相同的编码准则时，至少某些操作是可以在编码域中执行的（也就是在完全或部分解码之前）。

[0062] 引入的参考文献

[0063] 作为参考，在这里全面引入以下的专利、专利申请和公开。

[0064] AC-3

[0065] ATSC Standard A52/A :Digital A udio Compression Standard(AC-3, E-AC-3), Revision B, Advanced Television Systems Committee, 2005 年 6 月 14 日。A/52B 文档可以在万维网上的地址 http://www.atsc.org/standards.html 得到。

[0066] Steve Vernon 于 1995 年 8 月发表于 IEEE Trans.ConsumerElectronics 第 41 卷第 3 号的“Design and Implementation of AC-3Coders”。

[0067] Mark Davis 于 1993 年 10 月发表于 Audio Engineering SocietyPreprint 3774, 95th AES Convention 的“The AC-3 MultichannelCoder”。

[0068] Bosi 等人于 1992 年 10 月发表于 Audio Engineering SocietyPreprint 3365, 93rd AES Convention 的“High Quality, Low-RateAudio Transform Coding for Transmission and MultimediaApplications”。

[0069] 美国专利 5,583,962 ;5,632,005 ;5,633,981 ;5,727,119 以及 6,021,386。

[0070] AAC

[0071] ISO/IEC JTC1/SC29,“Information technology-very lowbitrate audio-visual coding,”ISO/IEC IS-14496(Part 3,Audio),19961)ISO/IEC 13818-7.“MPEG-2 advanced audio coding,AAC”.International Standard,1997 ;

[0072] M.Bosi, K.Brandenburg, S.Quackenbush, L.Fielder, K.Akagiri, H.Fuchs, M.Dietz, J.Herre, G.Davidson 以及 Y.Oikawa 于 1996 年发表于 Proc.of the 101st AES-Convention 的“ISO/IECMPEG-2 Advanced Audio Coding”;

[0073] M.Bosi, K.Brandenburg, S.Quackenbush, L.Fielder, K.Akagiri, H.Fuchs, M.Dietz, J.Herre, G.Davidson, Y.Oikawa 于 1997 年 10 月发表于 Journal of the AES 第 45 卷第 10 号第 789-814 页的“ISO/IEC MPEG-2 Advanced Audio Coding”;

[0074] Karlheinz Brandenburg 于 1999 年发表于 Proc.of the AES 17thInternational

Conference on High Quality Audio Coding, Florence, Italy 的“MP3 and AAC explained”;以及

[0075] G.A.Soulodre 等人于 1998 年 3 月发表于 J.Audio Eng.Soc, 第 46 卷第 3 号第 164-177 页 的“Subjective Evaluation of State-of-the-ArtTwo-Channel Audio Codecs”。

[0076] 实施方式

[0077] 本发明可以用硬件、软件或是软硬件组合（例如可编程逻辑阵列）来实现。除非以别的方式加以规定，否则，作为本发明的一部分而被包含的算法并不是固有地涉及任何特定的计算机或其他设备的。特别地，各种通用机器都可以与依照这里的教导所编写的程序结合使用，或者更为便利的可以是构造更为专用的设备（例如集成电路），以执行所需要的方法步骤。由此，本发明可以以一个或多个可编程计算机系统上运行的一个或多个计算机程序实施，其中每一个计算机系统都包括至少一个处理器，至少一个数据存储系统（包括易失和非易失存储器和 / 或存储部件），至少一个输入设备或端口，以及至少一个输出设备或端口。程序代码被应用于输入数据，以便执行这里描述的功能并产生输出信息。所述输出信息则以已知的方式而被应用于一个或多个输出设备。

[0078] 每一个这种程序都可以用任何期望的计算机语言来实现（包括机器、汇编或高级程序、逻辑或面向对象的编程语言），以便与计算机系统进行通信。在任何情况中，所述语言都可以是编译或解释性语言。

[0079] 优选地，每一个这种计算机程序都被保存或下载至可供通用或专用可编程计算机读取的存储介质或设备（例如固态存储器或媒体，或是磁性或光学媒体），以便在计算机系统读取存储媒体或设备的时候配置和操作计算机，从而执行这里描述的过程。本发明的系统还可以被认为是作为用计算机程序配置的计算机可读存储介质来实现的，其中所述存储介质被配置成使计算机系统以规定和预定方式执行操作，以便执行这里描述的功能。

[0080] 在这里业已描述了本发明的众多实施例。但是应该理解，在不脱离本发明的实质和范围的情况下，各种修改都是可行的。例如，这里描述的某些步骤可以与顺序无关，由此可以按照与所描述的顺序不同的顺序来执行。

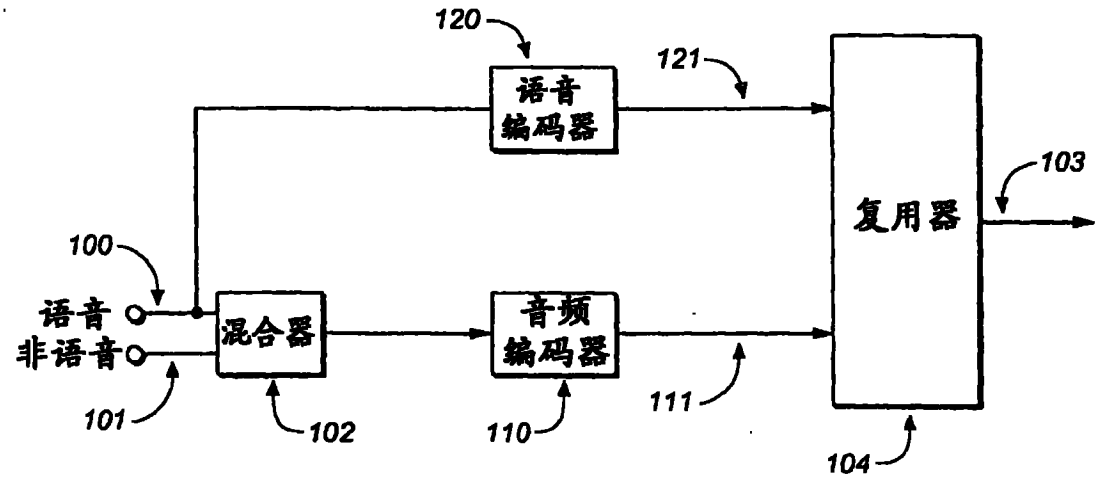

120
语音
编码器
121
复用器
103
100
语音
非语音
混合器
音频
编码器
111
101
102
110
104

图 1

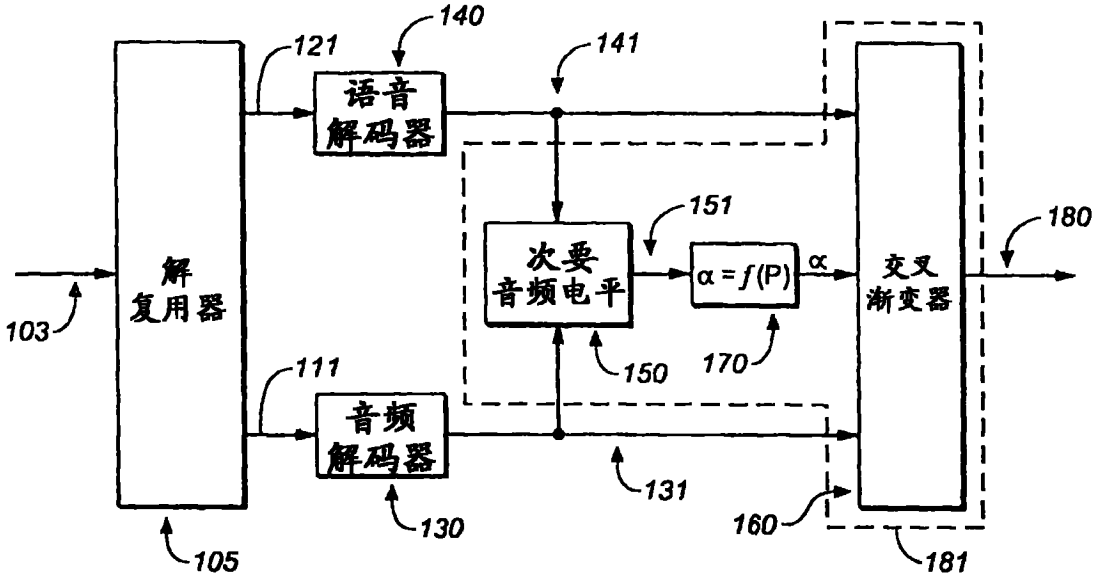

140
121
141
语音
解码器
151
次要
音频电平
α = f(P)
α
交叉
渐变器
180
解
复用器
103
150
170
111
音频
解码器
131
160
105
130
181

图 2

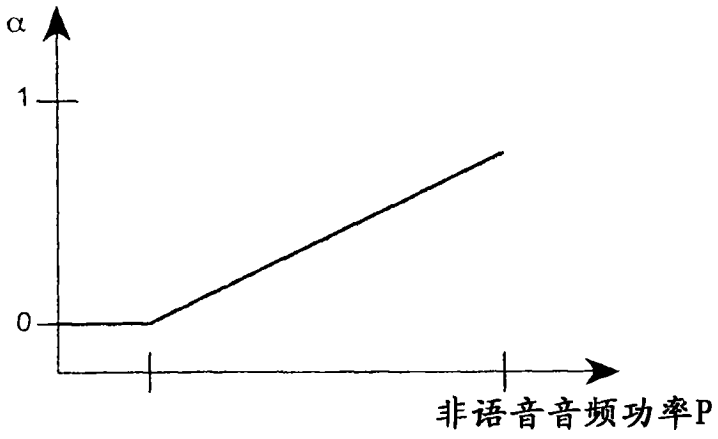


图 3

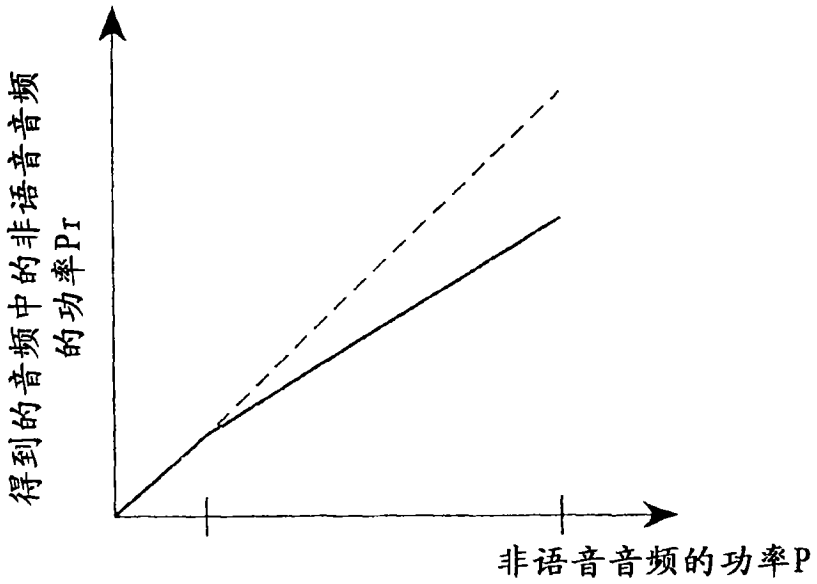


图 4

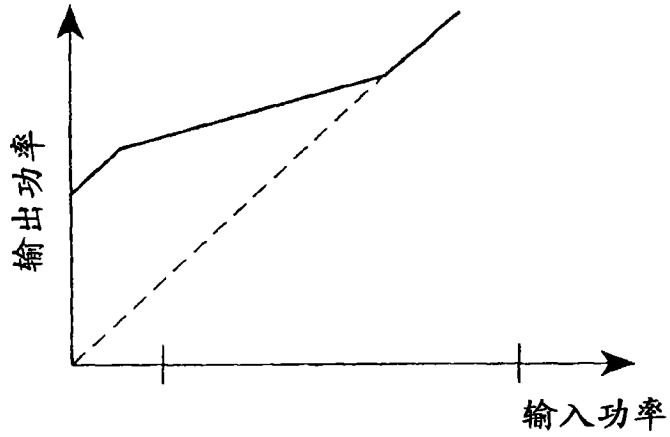


图 6

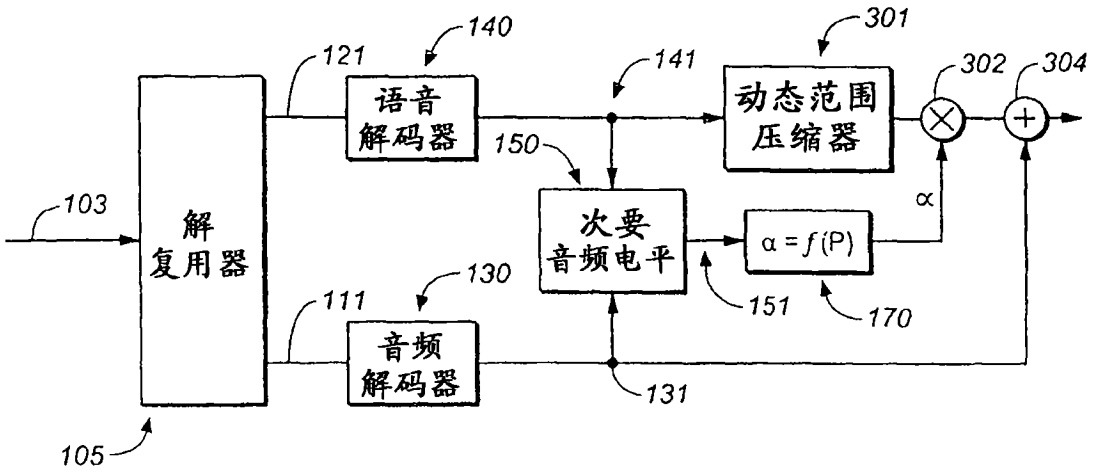

140
121
301
141
302
304
语音
解码器
动态范围
压缩器
150
103
解
复用器
次要
音频电平
α = f (P)
α
130
111
151
170
音频
解码器
131
105

图 5

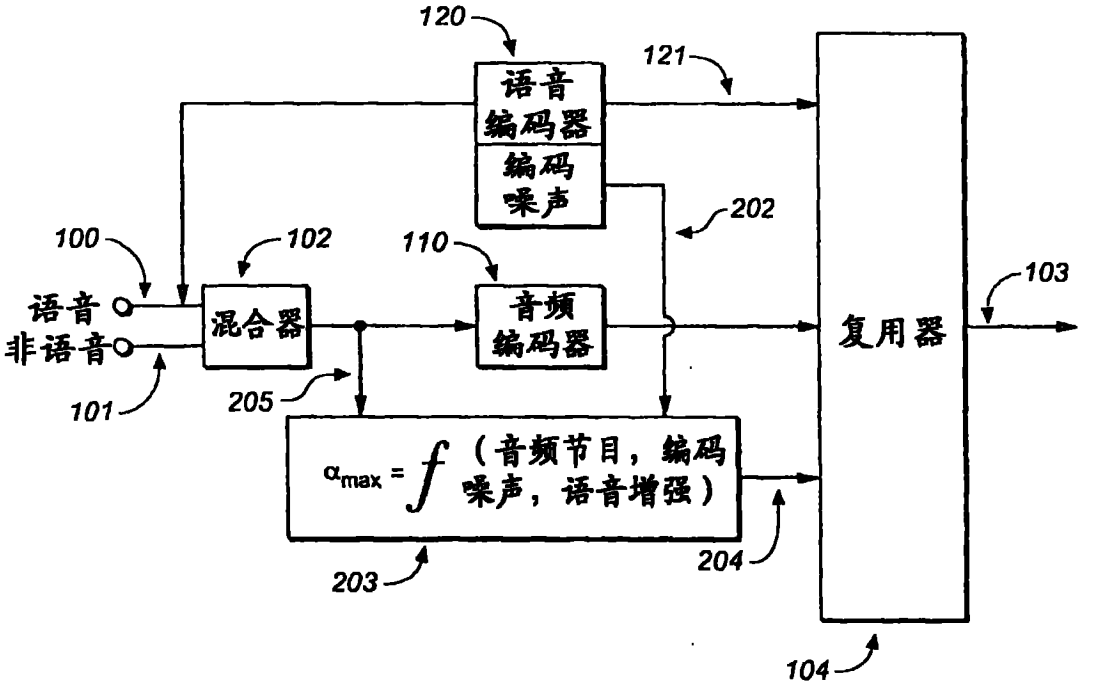

120
121
语音
编码器
编码
噪声
202
100
102
110
103
语音
非语音
混合器
音频
编码器
复用器
101
205
$\alpha_{max} = f$（音频节目，编码
噪声，语音增强）
204
203
104

图 7