

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 19/00 (2006.01)



[12] 发明专利说明书

专利号 ZL 03800936.6

[45] 授权公告日 2007 年 3 月 28 日

[11] 授权公告号 CN 1307611C

[22] 申请日 2003.6.27 [21] 申请号 03800936.6

[30] 优先权

[32] 2002. 7. 1 [33] US [31] 10/186,862

[86] 国际申请 PCT/US2003/020354 2003. 6. 27

[87] 国际公布 WO2004/003499 英 2004. 1. 8

[85] 进入国家阶段日期 2004. 2. 27

[73] 专利权人 朗迅科技公司

地址 美国新泽西州

[72] 发明人 金度锡

[56] 参考文献

US5794188A 1998. 8. 11

US4352182 1982. 9. 28

EP1187100A1 2002. 3. 13

审查员 吴 芸

[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

代理人 吴丽丽

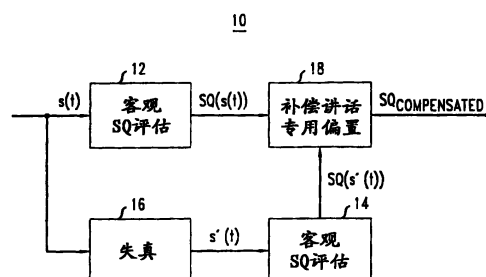
权利要求书 2 页 说明书 6 页 附图 3 页

[54] 发明名称

一种评估语音质量的方法

[57] 摘要

本发明提供了一种客观语音质量评估的方法，这种方法通过在语音质量评估下使语音信号失真来考虑语音内容、说话风格或单个说话者差异。由于使用的是语音信号的失真版本，因此在语音质量评估时可以对不同的语音内容、不同的讲话者和不同的说话风格进行补偿。在客观语音质量评估中由于语音信号失真而导致的退化量对于不同的语音信号基本上是类似的，特别是在语音信号的失真版本的失真比较严重时。将对失真的语音信号的客观语音质量评估与对原来的无失真的语音信号的客观语音质量评估进行比较，以获得与讲话有关的发音得到补偿的语音质量评估。



1. 一种评估语音质量的方法，所述方法包括下列步骤：

使第二语音信号失真，以产生第一语音信号，所述第一语音信号是第二语音信号的已调噪声参考单元版本；

确定第一和第二语音信号的第一和第二语音质量评估；以及

对第一和第二语音质量评估进行比较，以得到经补偿的语音质量评估。

2. 权利要求1的方法，其中第一和第二语音质量是用相同的客观语音质量评估技术评估的。

3. 权利要求1的方法，其中经补偿的语音质量评估与第一和第二语音质量评估之差相应。

4. 权利要求1的方法，其中经补偿的语音质量评估与第一和第二语音质量评估之比相应。

5. 一种评估语音质量的方法，所述方法包括下列步骤：

使用听觉发音分析确定第一和第二语音信号的第一和第二语音质量评估，第一语音信号是第二语音信号的失真版本；以及

对第一和第二语音质量评估进行比较，以得到经补偿的语音质量评估。

6. 一种评估语音质量的方法，所述方法包括下列步骤：

确定第一和第二语音信号的第一和第二语音质量评估，第一语音信号是第二语音信号的失真版本；以及

对第一和第二语音质量进行比较，以得到经补偿的语音质量评估；

其中确定第一和第二语音质量评估的步骤包括对第一和第二语音信号的发音功率和非发音功率进行比较；以及

根据比较评估第二或第一语音质量。

7. 权利要求6的方法，其中发音频率大致为2~12.5 Hz。

8. 权利要求6的方法，其中发音频率与人类发音的频率相应。

9. 权利要求6的方法，其中非发音频率高于发音频率。

10. 权利要求6的方法，其中比较发音功率与非发音功率包括确定发音功率与非发音功率的比值。

11. 权利要求9的方法，其中所述比值包括一个分母和一个分子，分子包括发音功率加一个小的常数，分母包括非发音功率加上这个小的常数。

12. 权利要求6的方法，其中比较发音功率与非发音功率包括确定发音功率与非发音功率之差。

13. 权利要求6的方法，其中确定第一和第二语音质量评估的步骤包括下列步骤：

用比较确定局部语音质量。

14. 权利要求13的方法，其中还用基于DC分量功率的加权系数确定局部语音质量。

15. 权利要求13的方法，其中用局部语音质量确定第一和第二语音质量。

16. 权利要求6的方法，其中比较发音功率和非发音功率的步骤包括下列步骤：

对从通过对第一和第二语音信号进行滤波获得的多个临界频带信号得到的多个包络中的每个包络执行Fourier变换。

17. 权利要求6的方法，其中比较发音功率和非发音功率的步骤包括下列步骤：

对第一和第二语音信号进行滤波，以得到多个临界频带信号。

18. 权利要求17的方法，其中比较发音功率和非发音功率的步骤包括下列步骤：

对多个临界频带信号进行包络分析，以得到多个调制频谱。

19. 权利要求18的方法，其中比较发音功率和非发音功率的步骤包括下列步骤：

对多个调制频谱中的每个调制频谱进行 Fourier 变换。

一种评估语音质量的方法

技术领域

本发明与通信系统有关，具体地说与语音质量评估（speech quality assessment）有关。

背景技术

可以测量无线通信系统的性能，其中之一是测量语音质量。在当前的技术中，有两种语音质量评估的技术。第一种技术是主观技术(以下称为“主观语音质量评估（subjective speech quality assessment）”)。在主观语音质量评估中，由一些人类收听者来评估经处理的语音的语音质量，经处理的语音是经接收机处理的发送语音信号。这种技术是主观的，因为它是以各人的感觉为基础的，而人对语音质量的评估通常考虑了语音内容、说话风格或单个说话者差异。主观语音质量评估可能是昂贵而费时的。

第二种技术是客观技术(以下称为“客观语音质量评估（objective speech quality assessment）”)。客观语音质量评估不是以各个人的感觉为基础。大多数的客观语音质量评估技术以已知的源语音（source speech）或从经处理的语音估计的重构源语音（reconstructed source speech）为基础。然而，这些客观技术没有考虑到语音内容、说话风格或单个说话者差异。

因此，有必要开发考虑语音内容、说话风格或单个说话者差异的客观评估语音质量的技术。

发明内容

本发明提供了一种客观语音质量评估的方法，这种方法通过在语音质量评估下使语音信号失真来考虑语音内容、说话风格或单个说话者

差异。通过使用语音信号的一个失真版本，在语音质量评估时就能对不同的语音内容、不同的讲话者和不同的说话风格进行补偿。在客观语音质量评估中由于语音信号失真而导致的退化量对于不同的语音信号基本上是类似的，特别是在语音信号的失真版本的失真严重时。将对失真的语音信号的客观语音质量评估与对原来的无失真的语音信号的客观语音质量评估进行比较，以获得与讲话有关的发音(utterance dependent articulation)得到补偿的语音质量评估。在一个实施例中，这种比较与对失真与无失真的语音信号的客观语音质量评估间的差别相应。

附图说明

从以下说明、所附权利要求书和附图可以更好地理解本发明的特征、情况和优点。在这些附图中：

图1示出了按照本发明设计的补偿与讲话有关的发音的客观语音质量评估设备；

图2示出了按照本发明设计的使用发音分析模块的客观语音质量评估模块的实施例；

图3示出了按照本发明的一个实施例在一个发音分析模块内处理多个包络 $a_i(t)$ 的流程图；以及

图4示出了一个例示功率与频率关系的调制频谱 $A_i(m, f)$ 的例子。

具体实施方式

本发明提供了一种客观语音质量评估的方法，这种方法通过使经处理的语音失真来考虑语音内容、说话风格或单个说话者差异。客观语音质量评估对于具有相同主观语音质量评分的不同的语音信号会得到不同的值。这些值所以不同是由于在调制频谱域(modulation spectral domain)内频谱内容分布不同。通过使用经处理的语音信号的失真版本，可以对不同的语音内容、不同的说话者和不同的说话风格进行补偿。在客观语音质量评估中由于语音信号失真而导致的退化量对于不同的语音信号基本上是类似的，特别是在失真严重时。将对失真的语音信

号的客观语音质量评估与对原来的无失真的语音信号的客观语音质量评估进行比较,以获得与讲话有关的发音得到补偿的语音质量评估。

图1示出了按照本发明设计的补偿与讲话有关的发音的客观语音质量评估设备10。客观语音质量评估设备10包括:多个客观语音质量评估模块12、14,一个失真模块16,以及一个补偿讲话专用偏置模块18。语音信号 $s(t)$ 作为输入提供给失真模块16和客观语音质量评估模块12。在失真模块16内,使语音信号 $s(t)$ 失真,产生一个调制噪声参考单位(MNRU)语音信号 $s'(t)$ 。也就是说,失真模块16产生输入信号 $s(t)$ 的一个有噪声的版本。然后,MNRU语音信号 $s'(t)$ 作为输入提供给客观语音质量评估模块14。

在客观语音质量评估模块12、14中,对语音信号 $s(t)$ 和MNRU语音信号 $s'(t)$ 进行处理,得到客观语音质量评估 $SQ(s(t))$ 和 $SQ(s'(t))$ 。客观语音质量评估模块12、14就对任何输入语音信号执行的处理而言实质上是相同的。也就是说,如果客观语音质量评估模块12、14接收同样的输入语音信号,那末这两个模块12、14的输出信号将大致相同。注意,在其他一些实施例中,客观语音质量评估模块12、14可以以相互不同的方式处理语音信号 $s(t)$ 和 $s'(t)$ 。客观语音质量评估模块在该技术领域内是众所周知的。在这里,稍后将对这种模块的一个例子进行说明。

然后,对客观语音质量评估 $SQ(s(t))$ 和 $SQ(s'(t))$ 进行比较,得到补偿了与讲话有关的发音的语音质量评估 $SQ_{compensated}$ 。在一个实施例中,用客观语音质量评估 $SQ(s(t))$ 与 $SQ(s'(t))$ 之差确定语音质量评估 $SQ_{compensated}$ 。例如, $SQ_{compensated}$ 等于 $SQ(s(t))$ 减去 $SQ(s'(t))$, 或 $SQ_{compensated}$ 等于 $SQ(s'(t))$ 减去 $SQ(s(t))$ 。在另一实施例中,根据客观语音质量评估 $SQ(s(t))$ 与 $SQ(s'(t))$ 之比确定语音质量评估 $SQ_{compensated}$ 。例如,

$$SQ_{compensated} = \frac{SQ(s(t)) + \mu}{SQ(s'(t)) + \mu}$$

或

$$SQ_{compensated} = \frac{SQ(s'(t)) + \mu}{SQ(s(t)) + \mu}$$

其中， μ 为一个小数。

正如前面提到的那样，客观语音质量评估模块12、14在这个技术领域内是众所周知的。图2示出了按照本发明设计的使用听觉发音分析模块的客观语音质量评估模块12、14的实施例20。如图2所示，客观质量评估模块20包括耳蜗滤波器组（cochlear filterbank）22、包络分析模块24和发音分析模块26。在客观质量评估模块20中，语音信号 $s(t)$ 提供给耳蜗滤波器组22作为输入。耳蜗滤波器组22包括多个耳蜗滤波器 $h_i(t)$ ，用来按照第一级外周听觉系统处理语音信号 $s(t)$ ，其中 $i=1,2,\dots,N_c$ 表示具体的耳蜗滤波器信道，而 N_c 表示耳蜗滤波器信道的总数。具体地说，耳蜗滤波器组22对语音信号 $s(t)$ 进行滤波，产生多个临界频带信号（critical band signal） $s_i(t)$ ，其中临界频带信号 $s_i(t)$ 等于 $s(t)*h_i(t)$ 。

这些临界频带信号 $s_i(t)$ 提供给包络分析模块24作为输入。在包络分析模块24内，对这些临界频带信号 $s_i(t)$ 进行处理，得到多个包络 $a_i(t)$ ，其中 $a_i(t)=\sqrt{s_i^2(t)+\hat{s}_i^2(t)}$ ，而 $\hat{s}_i(t)$ 是 $s_i(t)$ 的Hilbert变换。

然后，这些包络 $a_i(t)$ 提供给发音分析模块26作为输入。在发音分析模块26内，对这些包络 $a_i(t)$ 进行处理，得到对语音信号 $s(t)$ 的语音质量评估。具体地说，发音分析模块26对与人的发音系统产生的信号关联的功率（以下称为“发音功率 $P_A(m,i)$ ”）和与不是人的发音系统产生的信号关联的功率（以下称为“非发音功率 $P_{NA}(m,i)$ ”）进行比较。然后，用这种比较作出语音质量评估。

图3示出了按照本发明的一个实施例在发音分析模块26内处理多个包络 $a_i(t)$ 的流程图300。在步骤310，对这些包络 $a_i(t)$ 中的每个包络的帧 m 执行Fourier变换，产生调制频谱 $A_i(m,f)$ ，其中 f 为频率。

图4示出了一个例示功率与频率关系的调制频谱 $A_i(m,f)$ 的例子40。在例子40中，发音功率 $P_A(m,i)$ 是与频率为2~12.5 Hz关联的功率，而非发音功率 $P_{NA}(m,i)$ 是与频率高于12.5 Hz关联的功率。与频率低于2 Hz关联的功率 $P_{No}(m,i)$ 是临界频带信号 $a_i(t)$ 的帧 m 的DC分量。在这个例子中，根据人发音的速度为2~12.5 Hz而与发音功率 $P_A(m,i)$ 和非发音功率 $P_{NA}(m,i)$ 关联的频率范围（以下分别称为“发音频率范围”和“非发音频率范

围”)是相邻而不相叠的频率范围这样的事实,选择发音功率 $P_A(m,i)$ 用作与频率2~12.5 Hz关联的功率。可以理解,对这个申请来说,所谓“发音功率 $P_A(m,i)$ ”不应该局限于人发音的频率范围或者说上述的频率范围2~12.5 Hz。同样,所谓“非发音功率 $P_{NA}(m,i)$ ”也不应该将频率范围局限于高于与发音功率 $P_A(m,i)$ 关联的频率范围。非发音频率范围可以与也可以不与发音频率范围交叠或邻接。非发音频率范围也可以包括低于发音频率范围内最低频率的频率,诸如与临界频带信号 $a_i(t)$ 的帧 m 的DC分量关联的那些频率。

在步骤320,对于每个调制频谱 $A_i(m,f)$,发音分析模块26将发音功率 $P_A(m,i)$ 与非发音功率 $P_{NA}(m,i)$ 进行比较。在发音分析模块26的这个实施例中,发音功率 $P_A(m,i)$ 与非发音功率 $P_{NA}(m,i)$ 之间的比较为发音与非发音之比 $ANR(m,i)$ 。 ANR 由下式定义

$$ANR(m,i) = \frac{P_A(m,i) + \varepsilon}{P_{NA}(m,i) + \varepsilon} \quad (1)$$

其中 ε 为某个小常数。发音功率 $P_A(m,i)$ 与非发音功率 $P_{NA}(m,i)$ 之间的其他比较方式也是可行的。例如,比较可以是式(1)的倒数,也可以是发音功率 $P_A(m,i)$ 与非发音功率 $P_{NA}(m,i)$ 之差。为了便于讨论,流程图300所示的发音分析模块26的实施例将就用式(1)的 $ANR(m,i)$ 进行比较予以讨论。然而,不应将此以任何方式解释为对本发明的限制。

在步骤330,用 $ANR(m,i)$ 确定帧 m 的局部语音质量 $LSQ(m)$ 。局部语音质量 $LSQ(m)$ 用所有信道 i 的发音与非发音比 $ANR(m,i)$ 与一个基于DC分量功率 $P_{No}(m,i)$ 的加权系数 $R(m,i)$ 的总和来确定。具体地说,局部语音质量 $LSQ(m)$ 用下式确定:

$$LSQ(m) = \log \left[\sum_{i=1}^{N_c} ANR(m,i) R(m,i) \right] \quad (2)$$

其中

$$R(m,i) = \frac{\log(1 + P_{No}(m,i))}{\sum_{k=1}^{N_c} \log(1 + P_{No}(m,k))} \quad (3)$$

而 k 为频率附标。

在步骤340, 语音信号 $s(t)$ 的总体语音质量 SQ 用帧 m 的局部语音质量 $LSQ(m)$ 和对数功率 $P_s(m)$ 确定。具体地说, 语音质量 SQ 用下式确定:

$$SQ = L\{P_s(m)LSQ(m)\}_{m=1}^T = \left[\sum_{m=1, P_s > P_{th}}^T P_s^\lambda(m) LSQ^\lambda(m) \right]^{1/\lambda} \quad (4)$$

其中: $P_s(m) = \log\left[\sum_{t \in m} s^2(t)\right]$, L 为 L_p 范数, T 为语音信号 $s(t)$ 内的帧的总数, λ 为任何值, 而 P_{th} 为可听信号与无声的区别门限。在一个实施例中, λ 可取地为一个奇整数值。

发音分析模块26的输出是对全部的帧 m 的语音质量 SQ 的评估。也就是说, 语音质量 SQ 是对语音信号 $s(t)$ 的语音质量评估。

虽然以上结合一定的实施例对本发明作了相当详细的说明, 但其他实施型式也是可行的。因此, 本发明的精神实质和专利保护范围不应局限于在本说明中所说明的这些实施例。

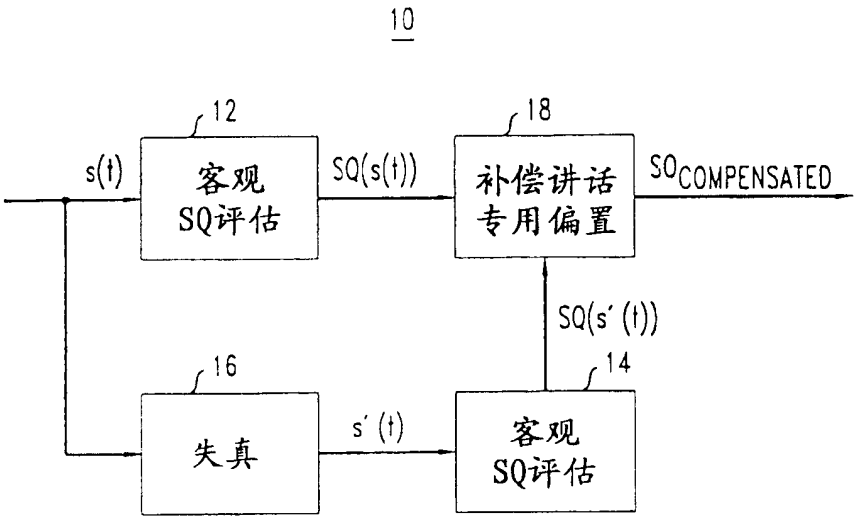


图1

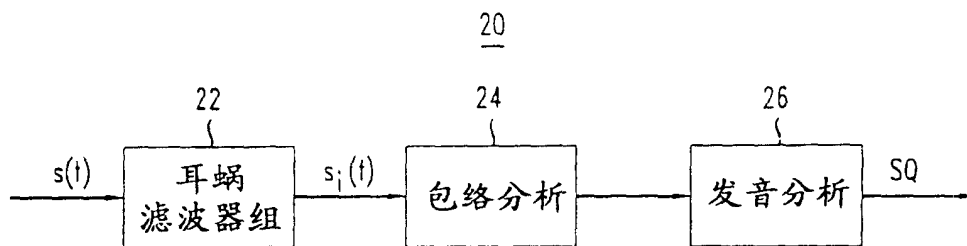


图 2

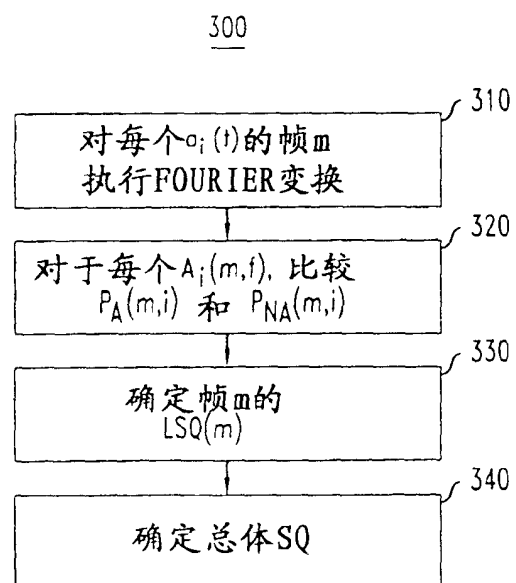


图 3

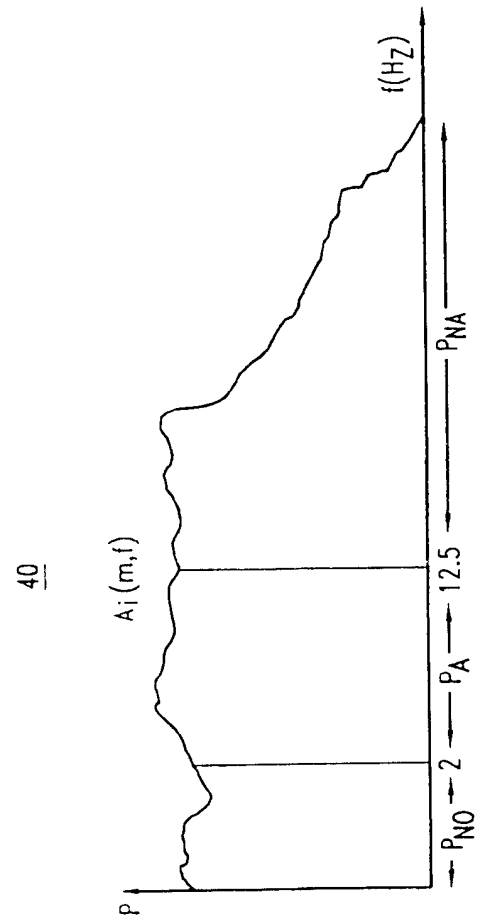


图 4