

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G10L 11/00 (2006.01)

G10L 13/08 (2006.01)



[12] 发明专利说明书

专利号 ZL 200410061685.7

[45] 授权公告日 2009年12月23日

[11] 授权公告号 CN 100573662C

[22] 申请日 2004.6.24

[21] 申请号 200410061685.7

[30] 优先权

[32] 2003.6.25 [33] US [31] 10/603,212

[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

WO0243051A1 2002.5.30

CN1192309A 1998.9.2

US5799133A 1998.8.25

US5454375A 1995.10.3

审查员 徐红

[74] 专利代理机构 中科专利商标代理有限责任公司

代理人 王波波

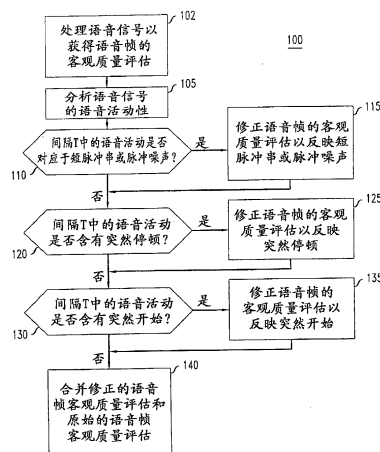
权利要求书2页 说明书10页 附图5页

[54] 发明名称

客观语音质量评估中反映时间和语言失真的方法和系统

[57] 摘要

本发明涉及一种客观语音质量评估中反映时间和语言失真的方法和系统，具体而言，提供了一种客观语音质量评估技术，该技术通过模拟可能对整个语音质量评估起支配作用的失真对主观语音质量评估的影响，反映了这些失真的影响，从而考虑了在客观语音质量评估中语言的影响。



1. 一种客观评估语音质量的方法，包括以下步骤：

使用包络信息检测语音活动区间中的失真；

修正与该语音活动相联系的客观语音质量评估值，以反映所述失真对主观语音质量评估的影响；以及

在所述检测步骤之前，使用该包络信息确定该语音活动区间，其中该客观语音质量评估值基于所检测失真的类型。

2. 根据权利要求1的方法，其中修正的步骤包括确定对语音活动的客观语音质量评估值的步骤。

3. 根据权利要求1的方法，其中所检测的失真是脉冲噪声、突然停顿或突然开始。

4. 根据权利要求1的方法，其中检测的步骤包括检测失真类型的步骤。

5. 根据权利要求4的方法，其中，如果包络信息指示出语音活动能够由听者感知为噪声，如果该客观语音质量评估值与调制噪声参考单位的比值指示出听者将感知令人烦躁的噪声，而且如果所述区间的持续时间足够长，能够让听者感知到，但又没有长到成为一个短脉冲串，那么失真的类型可以确定为脉冲噪声。

6. 根据权利要求4的方法，其中，如果包络信息指示出从一帧到另一帧能量中有足够大的负变化，以致于可以认为是一个突然停顿，而且如果所述区间的持续时间长于短脉冲串，那么失真的类型可以确定为突然停顿。

7. 根据权利要求4的方法，其中，如果包络信息指示出从一帧到另一帧能量中有足够大的正变化，以致于可以认为是一个突然开始，而且如果所述区间的持续时间长于短脉冲串，那么失真的类型可以确定为突然开始。

8. 一种客观语音质量评估系统，包括：

用于使用包络信息检测语音活动区间中的失真的装置；

用于修正与该语音活动相联系的客观语音质量评估值以反映所述失真对主观语音质量评估的影响的装置；以及

用于在执行所述检测之前使用该包络信息确定该语音活动区间的装置，其中该客观语音质量评估值基于所检测失真的类型。

9. 根据权利要求8的客观语音质量评估系统，其中用于修正的装置包括在不考虑失真的情况下确定对语音活动的客观语音质量评估值的装置。

10. 根据权利要求8的客观语音质量评估系统，其中用于检测的装置包括用于确定失真类型的装置。

客观语音质量评估中反映 时间和语言失真的方法和系统

技术领域

本发明通常涉及通信系统，尤其涉及语音质量评估。

背景技术

无线通信系统的性能从语音质量方面来说是可以度量的。在现有技术中，语音质量评估的技术有两种。第一种是主观技术（以下指“主观语音质量评估”）。在主观语音质量评估中，典型的方法是利用听众对经过处理的语音进行质量评估，该语音是经接收器端处理的传输语音信号。该评估技术之所以是主观的，是因为它是建立在个人感知基础上的，而且由本地听众对语音质量的评价通常考虑到语言的影响，本地听众的母语就是提供的或倾听的语料的语言。研究表明主观听音测试的得分受到听众语言知识的影响。当语音中缺少语言信息即静音时，在主观听音测试中本地听众的打分低于非本地听众。在正常的电话交谈中，受话人通常是本地听众。这样，为了仿真典型的条件，在主观语音质量评估中使用本地听众更好一些。主观语音质量评估技术提供了语音质量的良好评价，但可能比较昂贵和费时。

第二种技术是一种客观技术（以下指“客观语音质量评估”）。客观语音质量评估不是基于个人的感知。一些客观语音质量评估技术基于已知的源语音或从处理过的语音中估计重建的源语音。另一些客观语音质量评估技术则不是基于源语音而是仅仅基于处理过的语音。后一种技术在这里是指“单端客观语音质量评估技术”，通常应用于难以获得源语音或重建的源语音的场合。

但是与主观语音质量评估技术相比，当前的单端客观语音质量评估技术不能提供同样良好的评价。当前的单端客观语音质量评估技术

不如主观语音质量评估技术好的一个原因是前者没有考虑语言的影响。当前的单端客观语音质量评估技术在其语音评估过程中无法考虑语言的影响。

因此，需要一种单端客观语音质量评估技术，能够在语音质量评估过程中考虑到语言的影响。

发明内容

本发明是一种客观语音质量评估技术，该技术通过模拟可能对整个语音质量评估起支配作用的失真对主观语音质量评估的影响，反映了这些失真的影响，从而考虑在客观语音质量评估中语言的影响。在本发明的一种实施方式中，该客观语音质量评估技术包含以下几个步骤：用包络信息检测语音活动区间中的失真；以及修改与该语音活动关联的客观语音质量评估值，以反映这些失真对主观语音质量评估的影响。在本发明的一种实施方式中，该客观语音质量评估技术还区分出失真的类型，如短脉冲串、突然停顿和突然开始等，然后修改客观语音质量评估值，以反映每种失真对主观语音质量评估的不同影响。

附图说明

通过下面的描述、所附权利要求以及附图，本发明的特征、方式以及优越性都可以得到更好的理解，在附图中：

图 1 描述了根据本发明的一个实施例的一种考虑到语言影响的客观语音质量评估技术的流程图；

图 2 描述了根据本发明的一个实施例的一个语音活动检测器的操作流程图，该检测器通过检查与语音信号关联的包络信息检测语音活动；

图 3 是一个语音活动检测器活动的实例图解，该图分别示出了语音和非语音活动的间隔 T 和 G ；

图 4 描述了一个实施例的流程图，该实施例确定语音活动是否是一个短脉冲串或脉冲噪声，而且当确定是短脉冲串或脉冲噪声时，

修正语音帧的客观质量评估 $v_s(m)$;

图 5 描述了一个实施例的流程图, 该实施例确定语音活动是否含有一个突然停顿或静音, 而且当确定含有一个突然停顿或静音时, 修正语音帧的客观质量评估 $v_s(m)$; 以及

图 6 描述了一个实施例的流程图, 该实施例确定语音活动是否含有一个突然开始, 而且当确定含有一个突然开始时, 修正语音帧的客观质量评估 $v_s(m)$ 。

具体实施方式

本发明是一种客观语音质量评估技术, 该技术通过模拟可能对整个语音质量评估起支配作用的失真对主观语音质量评估的影响, 反映了这些失真的影响, 从而考虑了在客观语音质量评估中语言的影响。

图 1 描述了流程图 100, 说明按照本发明的一个实施例的考虑语言影响的客观语音质量评估技术。在步骤 102 中, 处理语音信号 $s(n)$, 以确定语音帧的客观质量评估 $v_s(m)$, 即语音在帧 m 的客观质量。在一个实施例中, 每一帧 m 对应时间间隔 64ms。处理语音信号 $S(n)$ 以获得语音帧的客观质量评估 $v_s(m)$ (不考虑语言的影响) 的方法在本技术领域是众所周知的。在下述共同待审申请中描述了这种处理方法的一个实例: 该申请的序列号是 10/186,862, 题目是“Compensation Of Utterance-Dependent Articulation For Speech Quality Assessment”, 申请日为 2002 年 7 月 1 日, 申请人是 Doh-Suk Kim。

在步骤 105 中, 分析语音信号 $S(n)$ 的语音活动, 例如用语音活动检测器 (VAD)。语音活动检测器在本技术领域是众所周知的。图 2 描述了一个语音活动检测器的操作流程图 200, 该检测器通过检查与语音信号关联的包络信息检测语音活动。在步骤 205 中, 将所有耳蜗通道 k 的包络信号 $\gamma_k(n)$ 累加起来得到总的包络信号 $\gamma(n)$, 见公式 (1) 所示:

$$\gamma(n) = \sum_{k=1}^{N_{cb}} \gamma_k(n) \quad \text{公式 (1)}$$

其中, $\gamma_k(n) = \sqrt{s_k^2(n) + \hat{s}_k^2(n)}$, n 代表时间标号, N_{cb} 代表临界带的总数, $s_k(n)$

代表语音信号 $s(n)$ 通过第 k 个耳蜗通道的输出, 也就是说, $s_k(n) = s(n) * h_k(n)$, $\hat{s}_k(n)$ 是 $s_k(n)$ 的希尔伯特变换。

在步骤 210 中, 按照公式 (2) 将总包络信号 $\gamma(n)$ 和 4ms 的汉明窗 $w(n)$ 相乘, 每 2ms 计算一个帧包络 $e(l)$,

$$e(l) = \log \left[\sum_{n=0}^{31} \gamma^{(l)}(n)w(n) + 1 \right] \quad \text{公式 (2)}$$

其中, $\gamma^{(l)}(n)$ 是总包络信号 $\gamma(n)$ 第 l 个 2ms 帧信号。应该知道这里帧包络 $e(l)$ 和汉明窗 $w(n)$ 的持续时间仅仅用来举个例子, 其他的持续时间是可能的。在步骤 215 中, 按照公式 (3) 对帧包络 $e(l)$ 进行向下取整 (flooring) 运算:

$$e(l) = \begin{cases} e(l) & \text{如果 } e(l) > 5 \\ 5 & \text{其它} \end{cases} \quad \text{公式 (3)}$$

在步骤 220 中, 按照公式 (4) 得到向下取整的帧包络 $e(l)$ 的时间导数 $\Delta e(l)$:

$$\Delta e(l) = \frac{\sum_{j=-3}^3 je(l-j)}{\sum_{j=-3}^3 j^2} \quad \text{公式 (4)}$$

其中 $-3 \leq j \leq 3$ 。

在步骤 225 中, 按照公式 (5) 进行语音活动检测。

$$vad(l) = \begin{cases} 1 & \text{如果 } e(l) > 5 \\ 0 & \text{其它} \end{cases} \quad \text{公式 (5)}$$

在步骤 230 中, 公式 (5) 的结果即 $vad(l)$ 可基于输出中 0 和 1 的持续时间进行改进。例如, 如果在 $vad(l)$ 中 0 的持续时间少于 8ms, 则 $vad(l)$ 在该段时间内应改为 1。与此类似, 如果在 $vad(l)$ 中 1 的持续时间少于 8ms, 则 $vad(l)$ 在该段时间内应改为 0。图 3 描述了一个语音活动检测器活动的实例图解, 该图分别示出了语音和非语音活动的间隔 T 和 G。应该知道间隔 T 相关的语音活动可以包含, 比如说, 真实语音、数据或噪音。

回到图 1 的流程图 100, 在对语音信号 $s(n)$ 的语音活动分析后, 在步骤 110 中检查间隔 T 以确定相关的语音活动是否对应于短脉冲串或脉冲噪声。如果确定间隔 T 中的语音活动是短脉冲串或脉冲噪声,

则在步骤 115 中修正语音帧的客观质量评估 $v_s(m)$ ，得到修正的语音帧客观质量评估 $\chi_s(m)$ 。修正的语音帧客观质量评估 $\chi_s(m)$ 通过模拟或仿真短脉冲串或脉冲噪声对主观语音质量评估的影响，考虑了短脉冲串或脉冲噪声的影响。

从步骤 115 或者如果在步骤 110 中确定间隔 T 中的语音活动不是短脉冲串或脉冲噪声，流程图 100 进到步骤 120，该步骤检查以确定间隔 T 中的语音活动是否包含突然停顿或静音。如果确定间隔 T 中的语音活动包含突然停顿或静音，则在步骤 125 中修正语音帧的客观质量评估 $v_s(m)$ ，得到修正的语音帧客观质量评估 $\chi_s(m)$ 。修正的语音帧客观质量评估 $\chi_s(m)$ 通过模拟或仿真突然停顿或静音及随后的释放对主观语音质量评估的影响，考虑了突然停顿或静音的影响。

从步骤 125 或者如果在步骤 120 中确定间隔 T 中的语音活动不包含突然停顿或静音，流程图 100 进到步骤 130，该步骤检查以确定间隔 T 中的语音活动是否包含突然开始。如果确定间隔 T 中的语音活动包含突然开始，则在步骤 135 中修正语音帧的客观质量评估 $v_s(m)$ ，得到修正的语音帧客观质量评估 $\chi_s(m)$ 。修正的语音帧客观质量评估 $\chi_s(m)$ 通过模拟或仿真突然开始对主观语音质量评估的影响，考虑了突然开始的影响。

从步骤 135 或者如果在步骤 130 中确定间隔 T 中的语音活动不包含突然开始，流程图 100 进到步骤 140，在该步骤中，如果对语音帧客观质量评估 $v_s(m)$ 进行了修正，则将修正的结果和步骤 102 中原始的语音帧客观质量评估 $v_s(m)$ 进行合并。

下面将描述根据本发明一个实施例的确定语音活动是否为短脉冲串或脉冲噪声、是否包含突然停顿（或静音）或者突然开始的技术，也就是步骤 110、120 和 130，以及修正语音帧客观质量评估 $v_s(m)$ 的技术，也就是步骤 115、125 和 135。图 4 描述了流程图 400，其中按照一个实施例确定语音活动是否是一个短脉冲串或脉冲噪声，而且当确定是短脉冲串或脉冲噪声时，修正语音帧的客观质量评估 $v_s(m)$ 。在步骤 405 中，通过在间隔 T_i 中寻找帧包络 $e(l)$ 是最大值的帧 l 来确定脉

冲噪声帧 l_i ，例如见公式 (6)：

$$l_i = \arg \max_{u_i \leq l \leq d_i} e(l) \quad (6)$$

其中， u_i 和 d_i 分别表示在间隔 T_i 中位于开始和末端的帧 l 的序号。在步骤 410 中，将帧包络 $e(l_i)$ 与听者的听阈值相比较，该听阈值指示一个听者是否会将相应的帧 l_i 听为令人烦躁的短脉冲串。在一个实施例中，听者的听阈值是 8，也就是说，在步骤 410，检查以确定 $e(l_i)$ 是否大于 8。如果帧包络 $e(l_i)$ 没有超过听者的听阈值，则在步骤 415 中确定语音活动不是短脉冲串或脉冲噪声。

如果帧包络 $e(l_i)$ 超过听者的听阈值，则在步骤 420 中检查以确定间隔 T_i 的持续时间是否既满足一短脉冲串阈值又满足一感知阈值。也就是说，检查以确定间隔 T_i 是否太短以致于不能被人听见，以及是否太长以致于不能分类为短脉冲串。在一个实施例中，如果间隔 T_i 的持续时间大于等于 28ms 且小于等于 60ms，即 $28 \leq T_i \leq 60$ ，则步骤 420 的两个阈值都得到了满足。否则就不满足。如果步骤 420 的阈值没有得到满足，那么在步骤 425 中确定语音活动不是短脉冲串或脉冲噪声。

如果步骤 420 的两个阈值都得到了满足，则在步骤 430 中，从间隔 T_i 的开始处前面的一帧或几帧到间隔 T_i 的第一帧或前几个帧中的包络 $e(l)$ 确定差分帧包络 $\Delta e(l)$ 的最大值，然后与一个突变阈值如 0.25 比较。突变阈值表征了一个识别帧包络中突变的标准。在一个实施例中，从帧包络 $e(u_i - 1)$ 即间隔 T_i 之前一帧的帧包络到帧包络 $e(u_i + 5)$ 即间隔 T_i 中的第 5 个帧包络，确定差分帧包络 $\Delta e(l)$ 的最大值，然后与一个突变阈值 0.25 相比较，也就是在步骤 430 中检查以确定是否满足公式 (7)：

$$\max_{u_i - 1 \leq l \leq u_i + 5} \Delta e(l) > 0.25 \quad \text{公式 (7)}$$

如果差分帧包络 $\Delta e(l)$ 的最大值没有超过阈值，那么在步骤 435 中确定语音活动不是短脉冲串或脉冲噪声。

如果差分帧包络 $\Delta e(l)$ 的最大值超过阈值，那么在步骤 440 中确定帧 m_i 是否足够引起听者烦恼，其中 m_i 对应于帧 m ，该帧受到脉冲噪声帧 l_i 的影响最大。在一个实施例中，通过判断语音帧的客观质量评

估 $v_s(m_l)$ 与调制噪声参考单位 $v_q(m_l)$ 的比值是否超过一噪声阈值，完成步骤 440 的处理过程。例如，采用噪声阈值 1.1，步骤 440 可以表示为公式 (8)：

$$\frac{v_s(m_l)}{v_q(m_l)} < 1.1 \quad \text{公式 (8)}$$

在这里如果满足公式 (8)，就可以确定帧 m_l 足够引起听者的烦恼。如果确定语音帧的客观质量评估 $v_s(m_l)$ 足够引起听者的烦恼，那么在步骤 445 中确定语音活动不是短脉冲串或脉冲噪声。

如果确定语音帧的客观质量评估 $v_s(m_l)$ 不足以引起听者的烦恼，那么在步骤 450 中检查与间隔 $G_{i-1,i}$ ， $G_{i,i+1}$ ， T_{i-1} 和/或 T_{i+1} 的持续时间相关的条件是否满足某种最大或最小持续时间的阈值，以验证其是否属于人的语音。在一个实施例中，步骤 450 中的条件表示为公式(9)和(10)。

$$G_{i-1,i} < 180ms, \quad G_{i,i+1} > 40ms \text{ 和 } T_{i-1} > 50ms \quad \text{公式 (9)}$$

$$G_{i-1,i} > 40ms, \quad G_{i,i+1} < 100ms \text{ 和 } T_{i+1} > 60ms \quad \text{公式 (10)}$$

如果这些公式或条件中任一个得到了满足，那么在步骤 455 中确定语音活动不是短脉冲串或脉冲噪声。而确定该语音活动为自然语音。应该理解，在公式 (9) 和 (10) 中使用的最大和最小持续时间的阈值仅仅是用来举例说明，是可以取其它值的。

如果在步骤 450 中没有条件得到满足，那么在步骤 460 中将语音帧的客观质量评估 $v_s(m)$ 按公式 (11) 进行修正：

$$\chi_s(m) = \frac{v_s(m)}{1 + \exp[-8.2(m - m_l)/e(l_l) - 10]} \quad \text{公式 (11)}$$

图 5 描述了一个实施例的流程图 500，该实施例确定语音活动是否含有一个突然停顿或静音，以及当确定含有突然停顿或静音时，修正语音帧的客观质量评估 $v_s(m)$ 。在步骤 505 确定突然停顿帧 l_M 。确定突然停顿帧 l_M 时，首先利用间隔 T_i 中的所有帧 l 寻找语音活动中差分帧包络 $\Delta e(l)$ 的负峰值。如果当 $-3 \leq j \leq 3$ 时满足 $\Delta e(l) < \Delta e(l+j)$ ，则说明差分帧包络 $\Delta e(l)$ 拥有一个负峰值。找到负峰值以后，将差分帧包络 $\Delta e(l)$ 的负峰值的最小值所在帧确定为突然停顿帧 l_M 。在步骤 510 中检查以确定差分帧包络 $\Delta e(l_M)$ 是否满足突然停顿阈值。突然停顿阈值代表一个标准，该标准确定从帧 l 到帧 $l+1$ 的帧包络中是否含有足够的负变化以

被认为是突然停顿。在一个实施例中，突然停顿阈值是-0.56，步骤 510 可以用公式 (12) 来表示：

$$\Delta e(l_M) < -0.56 \quad \text{公式 (12)}$$

如果差分帧包络 $\Delta e(l_M)$ 不满足突然停顿阈值，那么在步骤 515 确定语音活动没有突然停顿或静音。

如果差分帧包络 $\Delta e(l_M)$ 满足突然停顿阈值，那么在步骤 520 中检查以确定间隔 T_i 的语音活动是否具有足够的持续时间，例如长于一个短脉冲串。在一个实施例中，检查以确定间隔 T_i 的持续时间是否超过持续时间阈值，例如 60ms。也就是说，如果 $T_i < 60\text{ms}$ ，则间隔 T_i 关联的语音活动没有足够的持续时间。如果认为语音活动没有足够的持续时间，那么在步骤 525 确定语音活动没有突然停顿或静音。

如果认为语音活动具有足够的持续时间，那么在步骤 530 中，对帧 l_M 前一帧或几帧到帧 l_M 或更往后确定帧包络 $e(l)$ 的最大值，然后与停顿能量阈值相比较。停顿能量阈值表示确定帧包络在静音之前是否具有足够能量的一个标准。在一个实施例中，对帧 $l_M - 7$ 到帧 l_M 确定帧包络 $e(l)$ 的最大值，然后与停顿能量阈值 9.5 相比较，即 $\max_{l_M-7 \leq l \leq l_M} e(l) > 9.5$ 。如果最大帧包络 $e(l)$ 不满足停顿能量阈值，那么步骤 535 确定语音活动没有突然停顿或静音。

如果最大帧包络 $e(l)$ 满足停顿能量阈值，那么在步骤 540 按照公式 (13) 对若干语音帧如 $m_M, \dots, m_M + 6$ 的客观质量评估 $v_s(m)$ 进行修正：

$$\chi_s(m) = |\Delta e(l_M)| \left[\frac{6}{1 + \exp[-2(m - m_M - 3)]} - 6 \right] \quad \text{公式 (13)}$$

其中 m_M 对应于受突然停顿帧 l_M 影响最大的帧 m 。

图 6 描述了一个实施例的流程图 600，该实施例确定语音活动是否含有一个突然开始，以及当确定含有一个突然开始时，修正语音帧的客观质量评估 $v_s(m)$ 。在步骤 605 确定突然开始帧 l_s 。确定突然开始帧 l_s 时，首先利用间隔 T_i 中的所有帧 l 寻找语音活动中差分帧包络 $\Delta e(l)$ 的正峰值。如果当 $-3 \leq j \leq 3$ 时满足 $\Delta e(l) > \Delta e(l+j)$ ，则说明差分帧包络 $\Delta e(l)$ 拥有一个正峰值。找到正峰值以后，将差分帧包络 $\Delta e(l)$ 的正峰值的最大值所在帧确定为突然开始帧 l_s 。在步骤 610 中检查以确定差分帧包

络 $\Delta e(l_s)$ 是否满足突然开始阈值。突然开始阈值代表一个标准, 该标准确定从帧 l 到帧 $l+1$ 的帧包络中是否含有足够的正变化以被认为是突然开始。在一个实施例中, 突然开始阈值是 0.9, 步骤 610 可以用公式 (14) 来表示:

$$\Delta e(l_s) > 0.9 \quad \text{公式 (14)}$$

如果差分帧包络 $\Delta e(l_s)$ 不满足突然开始阈值, 那么在步骤 615 确定语音活动没有突然开始。

如果差分帧包络 $\Delta e(l_s)$ 满足突然开始阈值, 那么在步骤 620 中检查以确定间隔 T_i 的语音活动是否具有足够的持续时间, 例如长于一个短脉冲串。在一个实施例中, 检查以确定间隔 T_i 的持续时间是否超过持续时间阈值, 例如 60ms。也就是说, 如果 $T_i < 60\text{ms}$, 则间隔 T_i 关联的语音活动没有足够的持续时间。如果语音活动没有足够的持续时间, 那么在步骤 625 确定语音活动没有突然开始。

如果语音活动具有足够的持续时间, 那么在步骤 630 中, 对帧 l_s 或更往前到帧 l_s 后面一帧或几帧确定帧包络 $e(l)$ 的最大值, 然后与开始能量阈值相比较。开始能量阈值表示确定帧包络是否具有足够能量的一个标准。在一个实施例中, 对帧 l_s 到帧 $l_s + 7$ 确定帧包络 $e(l)$ 的最大值, 然后与开始能量阈值 12 相比较, 即 $\max_{l_s \leq l \leq l_s + 7} e(l) < 12$ 。如果最大帧包络 $e(l)$ 不满足开始能量阈值, 那么在步骤 635 确定语音活动不含有突然开始。

如果最大帧包络 $e(l)$ 满足开始能量阈值, 那么在步骤 640 按照公式 (16) 对若干语音帧如 $m_M, \dots, m_M + 6$ 的客观质量评估 $v_s(m)$ 进行修正:

$$\chi_s(m) = \frac{v_s(m)}{1 + \exp[-0.4(m - m_s) / \Delta e(l_s) - 10]} \quad \text{公式 (16)}$$

其中 m_s 对应于受突然开始帧 l_s 影响最大的帧 m 。应当理解, 公式 (11)、(13) 和 (16) 中使用的值都是通过经验推导出来的, 取其它值也是可能的。因此, 本发明不局限于这些特定的值。

注意在确定了修正的语音帧客观质量评估 $\chi_s(m)$ 之后, 在步骤 140 可按照公式 (17) 进行合并处理:

$$v_s(m) = \min(v_{s,I}(m), v_{s,M}(m), v_{s,S}(m)) \quad \text{公式 (17)}$$

其中 $v_{s,I}(m), v_{s,M}(m), v_{s,S}(m)$ 分别对应于公式 (11)、(13) 和 (16) 中修正

的语音帧客观质量评估 $\chi_s(m)$ 。

虽然本发明的一定实施例的描述已经相当地详尽，但是其它的方式也是可能的。例如，流程图中步骤次序可以重新安排，其中一些步骤（或标准）可以从流程图中删除或向其中增加。因此，本发明的实质和范围不应局限于这里的实施例的描述。还应当理解，对于那些熟练的技术人员来说，本发明既可以用硬件实现，也可以在某些类型的处理器上用软件实现。

图1

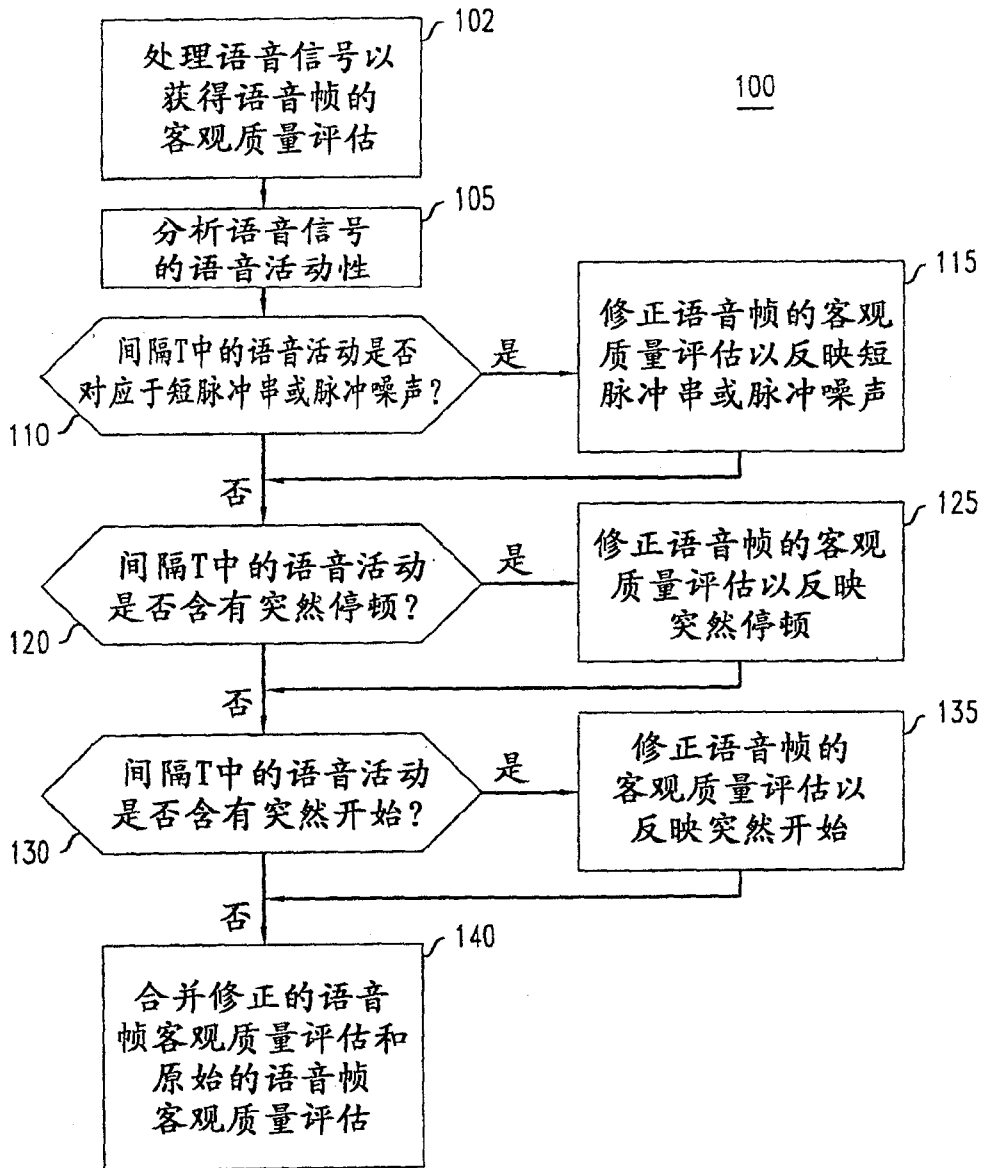


图 2

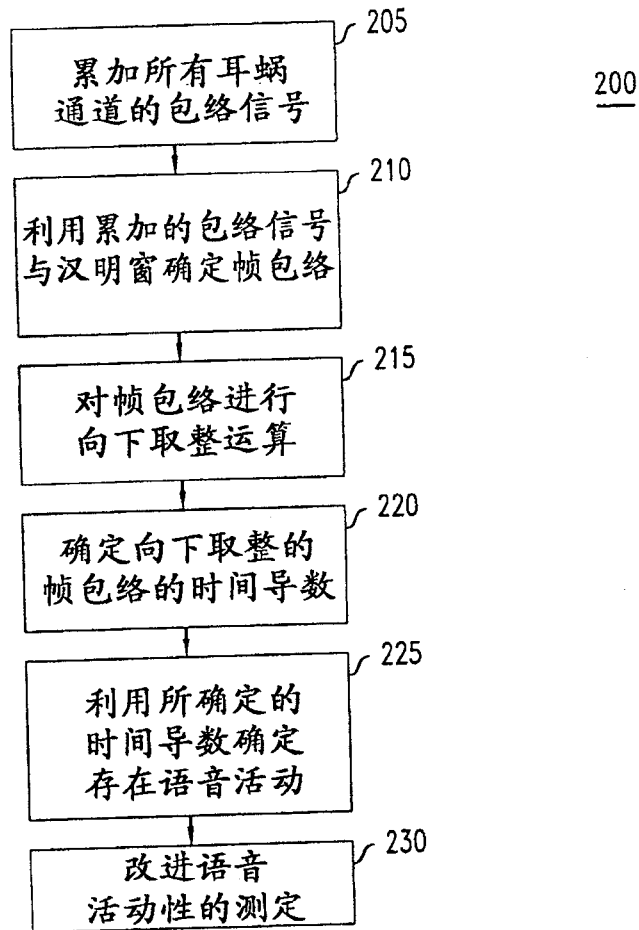


图 3

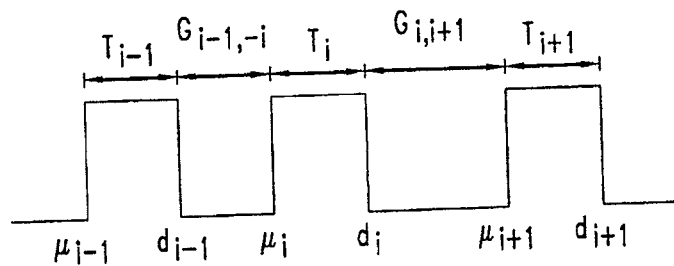


图 4

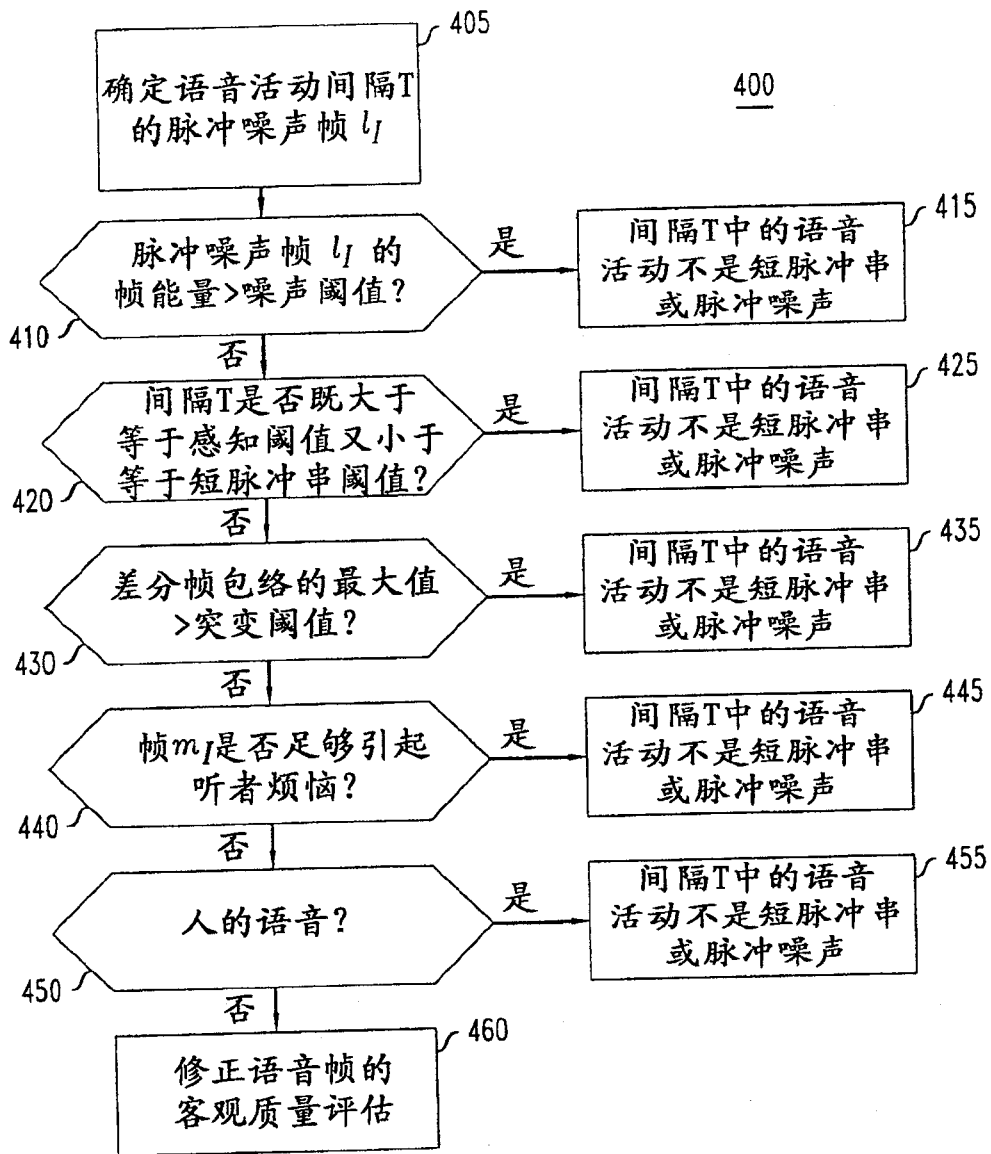


图5

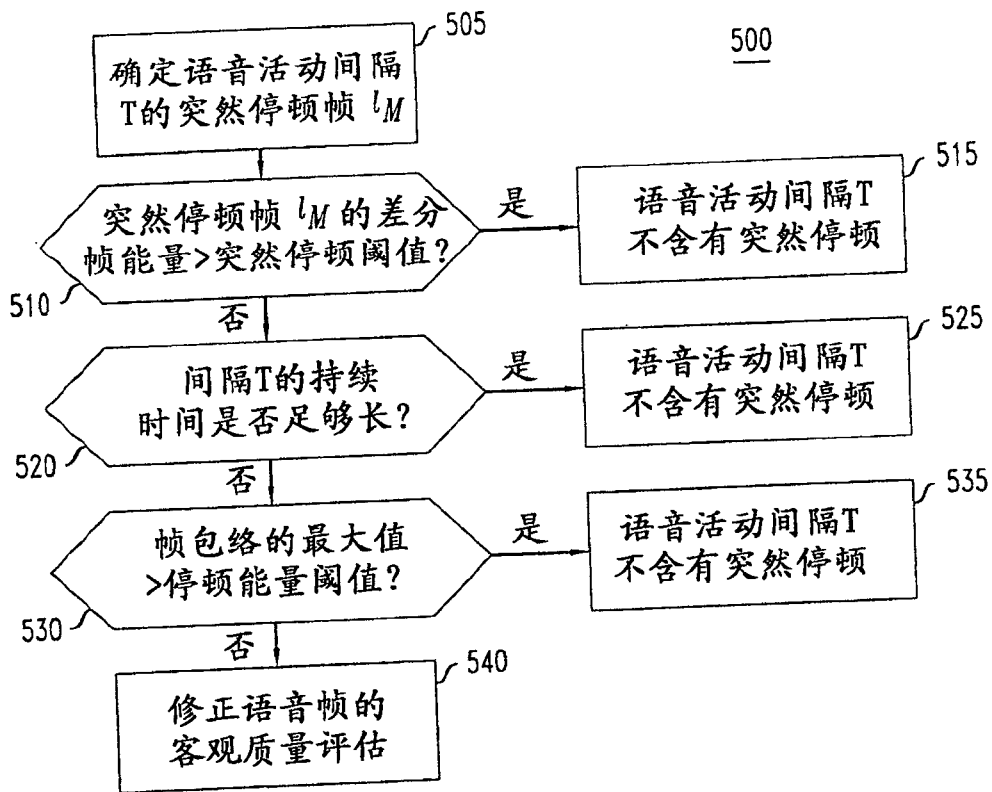


图6

