



(12) 发明专利

(10) 授权公告号 CN 102750956 B

(45) 授权公告日 2014. 07. 16

(21) 申请号 201210201879. 7

(22) 申请日 2012. 06. 18

(73) 专利权人 歌尔声学股份有限公司

地址 261031 山东省潍坊市高新技术产业开
发区东方路 268 号

(72) 发明人 楼厦厦 吴晓婕 李波

(74) 专利代理机构 北京市隆安律师事务所
11323

代理人 权鲜枝

(51) Int. Cl.

G10L 21/0208(2013. 01)

(56) 对比文件

WO 2005/122640 A1, 2005. 12. 22,

CN 101040512 A, 2007. 09. 19,

Tomohiro Nakatani et

al. Harmonicity-Based Blind Dereverberation
for Single-Channel Speech Signals. 《IEEE
TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE
PROCESSING》. 2007, 80-95.Kinoshita K et al. Spectral Subtraction
Steered by Multi-Step Forward Linear
Prediction For Single Channel Speech
Dereverberation. 《ICASSP'2006》. 2006,
Jan S. Erkelens and Richard Heusdens.
Correlation-Based and Model-Based Blind
Single-Channel Late-Reverberation
Suppression in Noisy Time-Varying
Acoustical Environments. 《IEEE TRANSACTIONS
ON AUDIO, SPEECH, AND LANGUAGE
PROCESSING》. 2010, 1746-1765.Keisuke Kinoshita et al. Suppression
of Late Reverberation Effect on Speech
Signal Using Long-Term Multiple-step
Linear Prediction. 《IEEE TRANSACTIONS ON
AUDIO, SPEECH, AND LANGUAGE PROCESSING
》. 2009, 534-545.

审查员 游晓梅

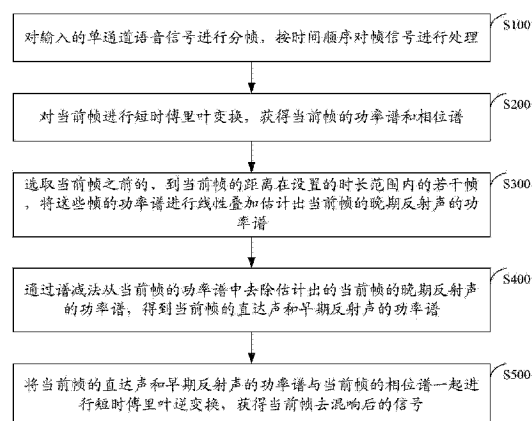
权利要求书2页 说明书9页 附图4页

(54) 发明名称

一种单通道语音去混响的方法和装置

(57) 摘要

本发明公开一种单通道语音去混响的方法和装置,所述方法包括:对输入的单通道语音信号进行分帧,按时间顺序对帧信号进行如下处理:对当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱;选取当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱;通过谱减法从当前帧的功率谱中去除估计出的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱;将当前帧的直达声和早期反射声的功率谱与当前帧的相位谱一起进行短时傅里叶逆变换,获得当前帧去混响后的信号。本发明能够解决单通道语音去混响中估计混响环境的传递函数或估计混响时间困难的问题。



1. 一种单通道语音去混响的方法,其特征在于,所述方法包括:

对输入的单通道语音信号进行分帧,按时间顺序对帧信号进行如下处理:

对当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱;

选取当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱;

通过谱减法从当前帧的功率谱中去除估计出的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱;

将当前帧的直达声和早期反射声的功率谱与当前帧的相位谱一起进行短时傅里叶逆变换,获得当前帧去混响后的信号。

2. 根据权利要求1所述的方法,其特征在于,

依据晚期反射声的衰减特性,设置所述时长范围的上限值;

和/或,

依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域,设置所述时长范围的下限值。

3. 根据权利要求1所述的方法,其特征在于,

所述时长范围的上限值选择在0.3秒~0.5秒之间的值。

4. 根据权利要求1所述的方法,其特征在于,

所述时长范围的下限值选择在50毫秒~80毫秒之间的值。

5. 根据权利要求1-4任一项所述的方法,其特征在于,

所述将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱具体包括:

应用自回归AR模型将这些帧的功率谱中全部成分进行线性叠加估计出当前帧的晚期反射声的功率谱;

或者,

应用滑动平均MA模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加估计出当前帧的晚期反射声的功率谱;

或者,

应用自回归AR模型将这些帧的功率谱中全部成分进行线性叠加,并且应用滑动平均MA模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加,估计出当前帧的晚期反射声的功率谱。

6. 一种单通道语音去混响的装置,其特征在于,所述装置包括:

分帧单元,用于对输入的单通道语音信号进行分帧,按时间顺序向傅里叶变换单元输出帧信号;

傅里叶变换单元,用于对接收的当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱,向谱减单元和谱估计单元输出当前帧的功率谱,向傅里叶逆变换单元输出相位谱;

谱估计单元,用于将当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧的功率谱进行线性叠加,估计出当前帧的晚期反射声的功率谱,向谱减单元输出估计的当前帧的晚期反射声的功率谱;

谱减单元,用于通过谱减法从傅里叶变换单元获得的当前帧的功率谱中去除从谱估计单元获得的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱,向傅里叶逆变换单元输出当前帧的直达声和早期反射声的功率谱;

傅里叶逆变换单元,用于将从谱减单元获得的当前帧的直达声和早期反射声的功率谱与从傅里叶变换单元获得的当前帧的相位谱一起进行短时傅里叶逆变换,输出当前帧去混响后的信号。

7. 根据权利要求6所述的装置,其特征在于,

所述谱估计单元具体用于,依据晚期反射声的衰减特性设置所述时长范围的上限值;和/或,依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域设置所述时长范围的下限值。

8. 根据权利要求6所述的装置,其特征在于,

所述谱估计单元具体用于,选择时长范围的上限值为0.3秒~0.5秒之间的值。

9. 根据权利要求6所述的装置,其特征在于,

所述谱估计单元具体用于,选择时长范围的下限值为50毫秒~80毫秒之间的值。

10. 根据权利要求6-9任一项所述的装置,其特征在于,

所述谱估计单元具体用于:

对于当前帧之前的、到当前帧的距离在所述设置的时长范围内的若干帧,应用自回归AR模型将这些帧的功率谱中全部成分进行线性叠加估计出当前帧的晚期反射声的功率谱;

或者,

对于当前帧之前的、到当前帧的距离在所述设置的时长范围内的若干帧,应用滑动平均MA模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加估计出当前帧的晚期反射声的功率谱;

或者,

对于当前帧之前的、到当前帧的距离在所述设置的时长范围内的若干帧,应用自回归AR模型将这些帧的功率谱中全部成分进行线性叠加,并且应用滑动平均MA模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加,估计出当前帧的晚期反射声的功率谱。

一种单通道语音去混响的方法和装置

技术领域

[0001] 本发明涉及语音增强领域,特别涉及单通道语音去混响的方法和装置。

背景技术

[0002] 在远距离语音通讯中,麦克风端接收的信号容易受到环境混响的影响。比如,在房间内,语音经过墙面、地板和家具等多次放射,麦克风端接收到的信号是直达声和反射声的混合信号。这部分反射声就是混响信号。当说话人距离麦克风比较远,且通话环境是一个相对封闭的空间时,就很容易产生混响。混响严重时,会导致语音不清楚,影响通话质量。另外,混响带来的干扰,还会导致声学接收系统性能变差,语音识别系统性能显著下降等。

[0003] 早期的去混响方法主要是利用反卷积来进行的。这类方法需要提前知道准确的混响环境(房间或办公室等)的冲激响应或传递函数。混响环境的冲激响应可以通过某种特别的方法或装置提前测量得到,也可以通过其它方法单独估计得到。然后利用这个已知的混响环境冲激响应,估计逆滤波器,实现对混响信号的反卷积,从而实现去混响。这类方法的问题是,混响环境的冲激响应往往很难提前获得,且求取逆滤波器的过程本身可能引入新的不稳定因素。

[0004] 另一类去混响方法,不需要估计混响环境的冲激响应,因此不需要计算逆滤波器和进行逆滤波运算,也被称为盲去混响方法。这类方法通常基于语音模型假设,比如:混响导致接收的语音激励脉冲发生变化,使得周期性变得不那么明显,从而影响语音清晰度。这类方法一般基于 LPC (Linear Prediction Coding, 线性预测编码) 模型,假定产生语音的模型是一个全极点模型,而混响或其它加性噪声在整个系统中引入了新的零点,从而干扰了语音激励脉冲,但并不影响全极点滤波器。去混响方法是:估计信号的 LPC 残差,然后按照基因同步猝发准则(pitch-synchronous clustering criterion)或峰度(Kurtosis)最大化准则等,来估计干净的脉冲激励序列,从而实现去混响。这类方法的问题是计算复杂度往往非常高,且对于混响只影响全零点滤波器的假设,与实验分析存在不相符的情况。

[0005] 利用谱减法去混响是一个较佳的方案,语音信号包括直达声、早期反射声和晚期反射声,采用谱减法将晚期反射声的功率谱从整个语音的功率谱中除去能够提高语音质量。但其中的关键问题在于晚期反射声的谱的估计,即如何获得比较准确的晚期反射声的功率谱,从而在将晚期反射声的成份有效去除的同时又不损伤语音。在单通道语音去混响中,因为只有一路麦克风信息可用,因此估计混响环境的传递函数或估计混响时间(RT60)非常困难。

发明内容

[0006] 本发明提供了一种单通道语音去混响的方法和装置,以解决单通道语音去混响中估计混响环境的传递函数或估计混响时间困难的问题。

[0007] 本发明公开了一种单通道语音去混响的方法,所述方法包括:

[0008] 对输入的单通道语音信号进行分帧,按时间顺序对帧信号进行如下处理:

- [0009] 对当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱;
- [0010] 选取当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱;
- [0011] 通过谱减法从当前帧的功率谱中去除估计出的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱;
- [0012] 将当前帧的直达声和早期反射声的功率谱与当前帧的相位谱一起进行短时傅里叶逆变换,获得当前帧去混响后的信号。
- [0013] 较佳地,依据晚期反射声的衰减特性,设置所述时长范围的上限值;
- [0014] 和/或,
- [0015] 依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域,设置所述时长范围的下限值。
- [0016] 较佳地,所述时长范围的上限值选择在 0.3 秒~0.5 秒之间的值。
- [0017] 较佳地,所述时长范围的下限值选择在 50 毫秒~80 毫秒之间的值。
- [0018] 较佳地,所述将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱具体包括:
- [0019] 应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加估计出当前帧的晚期反射声的功率谱;
- [0020] 或者,
- [0021] 应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加估计出当前帧的晚期反射声的功率谱;
- [0022] 或者,
- [0023] 应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加,并且应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加,估计出当前帧的晚期反射声的功率谱。
- [0024] 本发明还公开了一种单通道语音去混响的装置,所述装置包括:
- [0025] 分帧单元,用于对输入的单通道语音信号进行分帧,按时间顺序向傅里叶变换单元输出帧信号;
- [0026] 傅里叶变换单元,用于对接收的当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱,向谱减单元和谱估计单元输出当前帧的功率谱,向傅里叶逆变换单元输出相位谱;
- [0027] 谱估计单元,用于将当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧的功率谱进行线性叠加,估计出当前帧的晚期反射声的功率谱,向谱减单元输出估计的当前帧的晚期反射声的功率谱;
- [0028] 谱减单元,用于通过谱减法从傅里叶变换单元获得的当前帧的功率谱中去除从谱估计单元获得的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱,向傅里叶逆变换单元输出当前帧的直达声和早期反射声的功率谱;
- [0029] 傅里叶逆变换单元,用于将从谱减单元获得的当前帧的直达声和早期反射声的功率谱与从傅里叶变换单元获得的当前帧的相位谱一起进行短时傅里叶逆变换,输出当前帧去混响后的信号。

[0030] 较佳地,所述谱估计单元具体用于,依据晚期反射声的衰减特性设置所述时长范围的上限值;和/或,依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域设置所述时长范围的下限值。

[0031] 较佳地,所述谱估计单元具体用于,选择时长范围的上限值为 0.3 秒~0.5 秒之间的值。

[0032] 较佳地,所述谱估计单元具体用于,选择时长范围的下限值为 50 毫秒~80 毫秒之间的值。

[0033] 较佳地,所述谱估计单元具体用于:

[0034] 对于当前帧之前的、到当前帧的距离在所述设置的时长范围内的若干帧,应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加估计出当前帧的晚期反射声的功率谱;

[0035] 或者,

[0036] 对于当前帧之前的、到当前帧的距离在所述设置的时长范围内的若干帧,应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加估计出当前帧的晚期反射声的功率谱;

[0037] 或者,

[0038] 对于当前帧之前的、到当前帧的距离在所述设置的时长范围内的若干帧,应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加,并且应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加,估计出当前帧的晚期反射声的功率谱。

[0039] 本发明实施例的有益效果是:通过选取当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱,能够不需估计混响环境的传递函数或混响时间,便可以估计出当前帧的晚期反射声的功率谱,进而利用谱减法进行去混响,简化了去混响的操作复杂度,使得实现更为简单;

[0040] 依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域设置时长范围的下限值,能够在去除混响的同时更好保留有用的直达声和早期反射声,提高话音质量;

[0041] 依据晚期反射声的衰减特性设置时长范围的上限值,能够在保证估计的晚期反射声的功率谱的准确性的同时,减少叠加运算量;

[0042] 本发明实施例将上限值选择为 0.3 秒~0.5 秒之间的值,该上限值为通过实验获得的门限值,在混响环境发生变化时,无需调整该上限值,都能够获得较好的去混响效果;

[0043] 本发明实施例将下限值设置在 50 毫秒~80 毫秒之间,在混响环境变化时,无需改变下限值,能够有效避开直达声和早期反射声进行叠加,使得叠加结果中基本不包含直达声和早期反射声,从而在去混响的同时保留有用的直达声和早期反射声,取得较好的话音质量。

[0044] 上述混响环境的变化包括:从无混响的消声室到混响非常严重的大礼堂。

附图说明

[0045] 图 1 为本发明单通道语音去混响的方法的流程图;

[0046] 图 2 为真实房间的冲激响应的示意图；

[0047] 图 3 为本发明实施效果示意图,图 3 (a)为混响信号时域示意图,图 3 (b)为去混响后的信号的时域示意图,图 3 (c)为混响信号频域示意图,图 3 (d)为去混响信号频域示意图；

[0048] 图 4 为本发明单通道语音去混响装置的结构图；

[0049] 图 5 为本发明单通道语音去混响装置具体实施方式的结构图。

具体实施方式

[0050] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明实施方式作进一步地详细描述。

[0051] 参见图 1,为本发明提供的单通道语音去混响的方法的流程图。

[0052] 步骤 S100,对输入的单通道语音信号进行分帧,按时间顺序对帧信号进行如下处理。

[0053] 步骤 S200,对当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱。

[0054] 步骤 S300,选取当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱。

[0055] 所述若干帧为一个预设数量的帧,可以为时长范围内的所有帧或该时长范围内的一部分帧。

[0056] 步骤 S400,通过谱减法从当前帧的功率谱中去除估计的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱。

[0057] 步骤 S500,将当前帧的直达声和早期反射声的功率谱与当前帧的相位谱一起进行短时傅里叶逆变换,获得当前帧去混响后的信号。

[0058] 在混响环境中,麦克风采集到的信号 $x(t)$,即单通道语音信号,是直达声和反射声的混合,可用如下混响模型表示：

$$[0059] \quad x(t) = h*s(t)+n(t)$$

[0060] 其中, $s(t)$ 是从声源发出的信号, h 是从声源位置到麦克风位置两点之间的房间冲激响应, $*$ 表示卷积运算, $n(t)$ 表示混响环境内的其它加性噪声。

[0061] 一个真实房间的冲激响应,如图 2 所示。可以将它划分为 3 个部分,直达峰 h_d 、早期反射 h_e 和晚期反射 h_l 。 h_d 和 $s(t)$ 的卷积可以简单地认为是声源发出的信号经过一定的延迟后在麦克风端的再现,对应于 $x(t)$ 中的直达声部分。早期反射部分的冲击响应对应于 h_d 之后一段时长的部分,该时长的结束时间点为 50ms 至 80ms 中的某个时间点。一般认为这一部分和 $s(t)$ 卷积所产生的早期反射声对直达声有加强和改善音质的作用。晚期反射声部分的冲击响应是去除 h_d 和 h_e 后房间冲激响应余下的长长的拖尾部分,这一部分与信号 $s(t)$ 卷积所产生的反射声,就是会对听感造成影响的混响成份。去混响算法主要是去除这一部分的影响。

[0062] 因此,混响模型也可表示为：

$$[0063] \quad x(t) = (h_d+h_e)*s(t)+h_l*s(t)+n(t)$$

[0064] h_l 部分符合指数衰减模型,可用如下方程近似：

[0065]
$$hl(t) = b(t)e^{-\frac{3\ln 10}{T_r}t}$$

[0066] 其中, T_r 是混响环境的混响时间(RT60), $b(t)$ 是零均值高斯分布随机变量。

[0067] 下面详细描述如何进行晚期反射声的功率谱估计。

[0068] 从功率谱分析角度来看, 信号功率谱 $X(t, f)$ 可以表示为:

[0069]
$$X(t, f) = Y(t, f) + R(t, f)$$

[0070] 其中 $R(t, f)$ 为晚期反射声的功率谱, 而 $Y(t, f)$ 是直达声和早期反射声的功率谱, 应予以保留。估计出晚期反射声的功率谱 $R(t, f)$ 后, 可以利用谱减法把 $Y(t, f)$ 从 $X(t, f)$ 中估计出来, 从而实现去混响。

[0071] 根据混响产生模型分析, 晚期反射声的功率谱与在它之前的信号功率谱或信号功率谱中的某些成份成线性关系, 而直达声和早期反射声的功率谱由于人的语音特性, 恰恰和过去的信号功率谱或信号功率谱中的某些成份不构成线性关系。因此, 通过对当前帧之前的特定时长的帧的功率谱中成分进行线性叠加, 能够估计出当前帧的晚期反射声的功率谱。接着再通过谱减法将晚期反射声的功率谱从功率谱中去除掉, 能够实现单通道语音去混响。

[0072] 较佳地, 依据晚期反射声的衰减特性设置所述时长范围的上限值。

[0073] 进行谱估计所用的帧越多, 估计越准确, 但是过多的帧造成运算量的增加。通过图 2 和 h1 部分的指数衰减模型可知距离当前帧越远的反射声能量越小, 在某一时刻之后的反射声能量可以被忽略。因此, 依据晚期反射声的衰减特性获得该反射声能量可以被忽略的时刻, 设置上限值为该时刻距离当前帧时刻的时长。由此, 能够在保证估计的晚期反射声的功率谱的准确性的同时, 减少叠加运算量。

[0074] 较佳地, 依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域, 来设置所述时长范围的下限值。

[0075] 通过图 2 可知直达声和早期反射声能量集中在距离当前帧较近的短时间内。依据直达声和早期反射声在混响环境下的冲击响应分布区域设置下限值, 使得在线性叠加时避开直达声和早期反射声能量集中的时间段, 能够在去除混响的同时更好保留有用的直达声和早期反射声, 提高话音质量。

[0076] 较佳地, 所述时长范围的下限值选择为 50 毫秒 ~ 80 毫秒之间的值。

[0077] 通过实验发现, 在各种环境下, 只要保证下限值取值为 50ms ~ 80ms 之间的数值, 就可以有效地绕过直达声和早期反射声部分, 更好地估计出有效的晚期反射声的功率谱。当环境发生变化后, 无需调整下限值设置, 便可获得较好话音质量。

[0078] 较佳地, 所述时长范围的上限值选择为在 0.3 秒 ~ 0.5 秒之间的值。

[0079] 理论上, 上限值的设置与应用方法的具体环境相关。在本专利所涉及的晚期反射声的功率谱估计中, 上限值理论上对应于房间冲激响应的长度, 但结合混响产生模型以及真实环境的冲激响应 h1 部分按指数模型衰减, 距离当前时刻越远的反射声能量越小, 超过 0.5s 后反射声的能量几乎可以忽略不计。因此, 实际中只需要使用一个非常粗略的上限值就可以适用于绝大多数混响环境。经验证, 上限值取在 0.3 秒 ~ 0.5 秒之间的值时, 对消声室环境(混响时间非常短)、一般办公室环境(混响时间 0.3 ~ 0.5s)、甚或大礼堂(混响时间 >1s)的多种混响环境都具有很好的适应性。在消声室环境下, 几乎没有晚期反射声。本发

明的方法只估计线性成份,且绕过了直达声和早期反射声的能量集中时间段,因此即便上限值的取值比消声室的混响时间长很多,但有效的语音成份并不会被去除。而在大礼堂环境中,虽然上限值的取值可能远小于真实的混响时间,但由于冲激响应按指数衰减得非常快,前 0.3s 内的晚期反射声成份占据了总体晚期反射声成份的绝大部分能量,因为也可以把混响很好的去除。

[0080] 在一具体实施方式中,所述将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱具体包括:应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加估计出当前帧的晚期反射声的功率谱。

[0081] 例如,按如下公式使用 AR 模型估计当前帧的晚期反射声的功率谱:

$$[0082] \quad R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t-j \cdot \Delta t, f)$$

[0083] 其中, $R(t, f)$ 为估计的晚期反射声的功率谱, J_0 为由设置的时长范围的下限值得出的起始级数, J_{AR} 为由设置的时长范围的上限值得出的 AR 模型的阶数, $\alpha_{j,f}$ 为 AR 模型估计参数; $X(t-j \cdot \Delta t, f)$ 为当前帧之前 j 帧的功率谱, Δt 为帧间距。

[0084] 在一具体实施方式中,所述将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱具体包括:应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加估计出当前帧的晚期反射声的功率谱。

[0085] 例如,按如下公式使用 MA 模型估计当前帧的晚期反射声的功率谱:

$$[0086] \quad R(t, f) = \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t-j \cdot \Delta t, f)$$

[0087] 其中, $R(t, f)$ 为估计的晚期反射声的功率谱, J_0 为由设置的时长范围的下限值得出的起始级数, J_{MA} 为由设置的时长范围的上限值得出的 MA 模型的阶数, $\beta_{j,f}$ 为 MA 模型估计参数; $Y(t-j, f)$ 为当前帧之前 j 帧的直达声和早期反射声的功率谱, Δt 为帧间距。

[0088] 在一具体实施方式中,所述将这些帧的功率谱进行线性叠加估计出当前帧的晚期反射声的功率谱具体包括:应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加,并且应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加,估计出当前帧的晚期反射声的功率谱。

[0089] 例如,按如下公式使用 ARMA 模型估计当前帧的晚期反射声的功率谱:

$$[0090] \quad R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t-j \cdot \Delta t, f) + \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t-j \cdot \Delta t, f)$$

[0091] 其中, $R(t, f)$ 为估计的晚期反射声的功率谱, J_0 为由设置的时长范围的下限值得出的起始级数, J_{AR} 为由设置的时长范围的上限值得出的 AR 模型的阶数, $\alpha_{j,f}$ 为 AR 模型估计参数, J_{MA} 为由设置的上限值得出的 MA 模型的阶数, $\beta_{j,f}$ 为 MA 模型估计参数, $Y(t-j, f)$ 为当前帧之前 j 帧的直达声和早期反射声的功率谱, $X(t-j \cdot \Delta t, f)$ 为当前帧之前 j 帧的功率谱, Δt 为帧间距。

[0092] AR 模型、MA 模型、ARMA 模型的具体求解,现有技术中存在公知算法,比如,利用 Yule-Walker (尤利-沃克) 方程求解或 Burg (伯格) 算法。

[0093] 利用谱减法来去混响,估计晚期反射声的功率谱最为关键。现有技术中提到的晚期反射声的功率谱估计往往是上述提出的 AR 或 MA 或 ARMA 模型的某种特例,此外,其它晚期反射声的功率谱估计方法往往需要在语音间歇阶段估计混响环境的混响时间(RT60),作为晚期反射声的功率谱估计中的一个重要参数。在本专利中,不需要估计混响时间或对各种环境估计冲激响应,便可以适应多种不同的混响环境,以及说话人在混响环境中由于运动等造成的混响冲激响应或混响时间发生改变的情况。

[0094] 在一具体实施方式中,通过谱减法从所述帧的功率谱中去除混响成分具体包括:

[0095] 依据晚期反射声的功率谱通过谱减法求得增益函数;

[0096] 将增益函数与当前帧的功率谱相乘得当前帧的直达声和早期反射声的功率谱。

[0097] 晚期反射声的功率谱 $R(t, f)$ 估计完成后,去除混响的语音信号 $Y(t, f)$ 可以通过谱减法得到:

[0098] $Y(t, f) = G(t, f) \cdot X(t, f)$

[0099] 其中, $G(t, f) = \frac{X(t, f) - R(t, f)}{X(t, f)}$ 为谱减法求得的 Gain (增益) 函数。

[0100] 本专利的实施效果如图 3 所示。混响信号(单通道语音信号)采集自会议室,声源和麦克风距离 2m,混响时间(RT60)约 0.45s。按本专利中提出的 AR 模型估计晚期反射声的功率谱,下限值设置为 80ms,上限值设置为 0.5s。依图示可知,应用本发明方法去混响后,语音质量得到显著提升。

[0101] 本发明的装置如图 4 所示,单通道语音去混响的装置包括如下单元。

[0102] 分帧单元 100,用于对输入的单通道语音信号进行分帧,按时间顺序向傅里叶变换单元 200 输出帧信号。

[0103] 傅里叶变换单元 200,用于对接收的当前帧进行短时傅里叶变换,获得当前帧的功率谱和相位谱,向谱减单元 400 和谱估计单元 300 输出当前帧的功率谱,向傅里叶逆变换单元 500 输出相位谱。

[0104] 谱估计单元 300,用于将当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧的功率谱进行线性叠加,估计出当前帧的晚期反射声的功率谱,向谱减单元 400 输出估计的当前帧的晚期反射声的功率谱。

[0105] 谱减单元 400,用于通过谱减法从傅里叶变换单元 200 获得的当前帧的功率谱中去除从谱估计单元 300 获得的当前帧的晚期反射声的功率谱,得到当前帧的直达声和早期反射声的功率谱,向傅里叶逆变换单元 500 输出当前帧的直达声和早期反射声的功率谱。

[0106] 傅里叶逆变换单元 500,用于将从谱减单元 400 获得的当前帧的直达声和早期反射声的功率谱与从傅里叶变换单元 200 获得的当前帧的相位谱一起进行短时傅里叶逆变换,输出当前帧去混响后的信号。

[0107] 较佳地,所述谱估计单元 300 具体用于,依据晚期反射声的衰减特性设置所述时长范围的上限值。

[0108] 较佳地,谱估计单元 300 具体用于,依据语音相关特性及直达声和早期反射声在混响环境下的冲击响应分布区域设置所述时长范围的下限值。

[0109] 较佳地,谱估计单元 300 具体用于,选择时长范围的上限值为 0.3 秒~0.5 秒之间的值。

[0110] 较佳地,谱估计单元 300 具体用于,选择时长范围的下限值为 50 毫秒~80 毫秒之间的值。

[0111] 具体实施方式的装置如图 5 所示,所述谱估计单元 300 具体用于:对于当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加估计出当前帧的晚期反射声的功率谱。

[0112] 例如,按如下公式使用 AR 模型估计当前帧的晚期反射声的功率谱:

$$[0113] \quad R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t-j \cdot \Delta t, f)$$

[0114] 其中, $R(t, f)$ 为估计的晚期反射声的功率谱, J_0 为由设置的下限值得出的起始级数, J_{AR} 为由设置的上限值得出的 AR 模型的阶数, $\alpha_{j,f}$ 为 AR 模型估计参数; $X(t-j \cdot \Delta t, f)$ 为当前帧之前 j 帧的功率谱, Δt 为帧间距。

[0115] 在另一具体实施方式中,所述谱估计单元 300 具体用于:对于当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加估计出当前帧的晚期反射声的功率谱。

[0116] 例如,按如下公式使用 MA 模型估计当前帧的晚期反射声的功率谱:

$$[0117] \quad R(t, f) = \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t-j \cdot \Delta t, f)$$

[0118] 其中, $R(t, f)$ 为估计的晚期反射声的功率谱, J_0 为由设置的下限值得出的起始级数, J_{MA} 为由设置的上限值得出的 MA 模型的阶数, $\beta_{j,f}$ 为 MA 模型估计参数; $Y(t-j, f)$ 为当前帧之前 j 帧的直达声和早期反射声的功率谱, Δt 为帧间距。

[0119] 在另一具体实施方式中,所述谱估计单元 300 具体用于:对于当前帧之前的、到当前帧的距离在设置的时长范围内的若干帧,应用自回归 AR 模型将这些帧的功率谱中全部成分进行线性叠加,并且应用滑动平均 MA 模型将这些帧的功率谱中直达声和早期反射声成分进行线性叠加,估计出当前帧的晚期反射声的功率谱。

[0120] 例如,按如下公式使用 ARMA 模型估计当前帧的晚期反射声的功率谱:

$$[0121] \quad R(t, f) = \sum_{j=J_0}^{J_{AR}} \alpha_{j,f} \cdot X(t-j \cdot \Delta t, f) + \sum_{j=J_0}^{J_{MA}} \beta_{j,f} \cdot Y(t-j \cdot \Delta t, f)$$

[0122] 其中, $R(t, f)$ 为估计的晚期反射声的功率谱, J_0 为由设置的下限值得出的起始级数, J_{AR} 为由设置的上限值得出的 AR 模型的阶数, $\alpha_{j,f}$ 为 AR 模型估计参数, J_{MA} 为由设置的上限值得出的 MA 模型的阶数, $\beta_{j,f}$ 为 MA 模型估计参数, $Y(t-j, f)$ 为当前帧之前 j 帧的直达声和早期反射声的功率谱, $X(t-j \cdot \Delta t, f)$ 为当前帧之前 j 帧的功率谱, Δt 为帧间距。

[0123] AR 模型、MA 模型、ARMA 模型的具体求解,现有技术中存在公知算法,比如,利用 Yule-Walker (尤利-沃克) 方程求解或 Burg (伯格) 算法。

[0124] 所述谱减单元 400 具体用于:依据晚期反射声的功率谱通过谱减法求得增益函数,将增益函数与当前帧的功率谱相乘得当前帧的直达声和早期反射声的功率谱。

[0125] 晚期反射声的功率谱 $R(t, f)$ 估计完成后,去除混响的语音信号 $Y(t, f)$ 可以通过谱减法得到:

[0126] $Y(t, f) = G(t, f) \cdot X(t, f)$

[0127] 其中, $G(t, f) = \frac{X(t, f) - R(t, f)}{X(t, f)}$ 为谱减法求得的 Gain (增益) 函数。

[0128] 以上所述仅为本发明的较佳实施例而已, 并非用于限定本发明的保护范围。凡在本发明的精神和原则之内所作的任何修改、等同替换、改进等, 均包含在本发明的保护范围内。

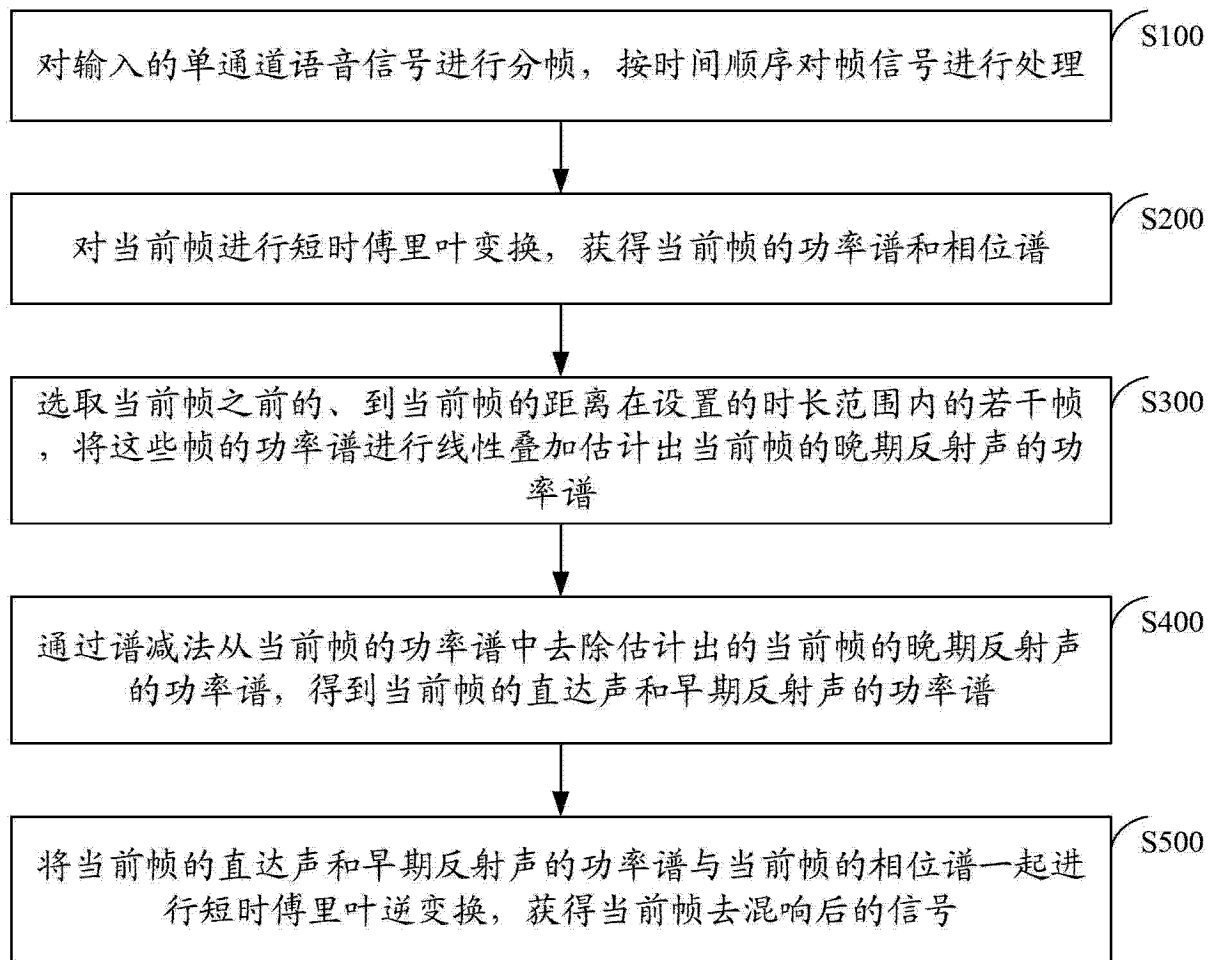


图 1

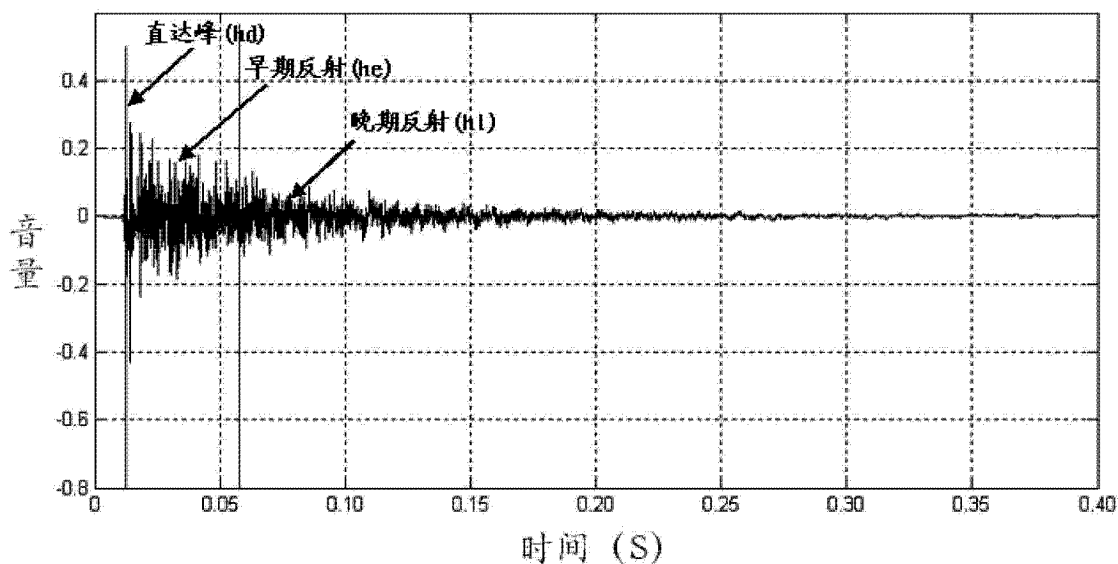


图 2

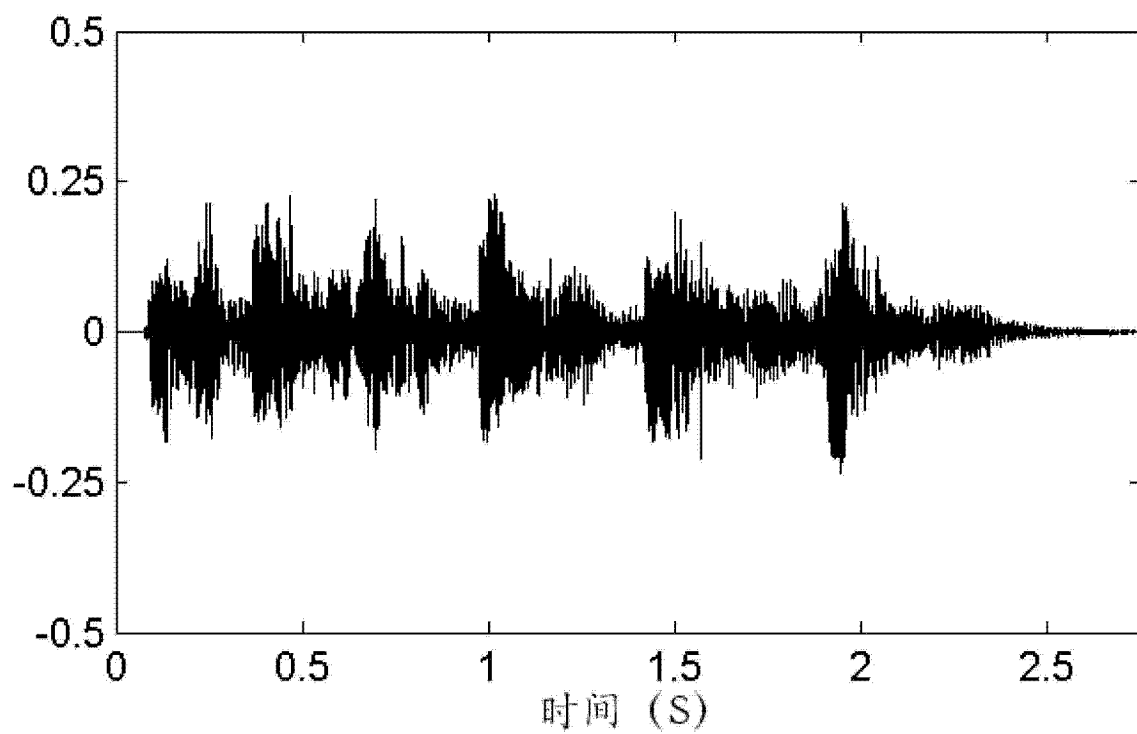


图 3(a)

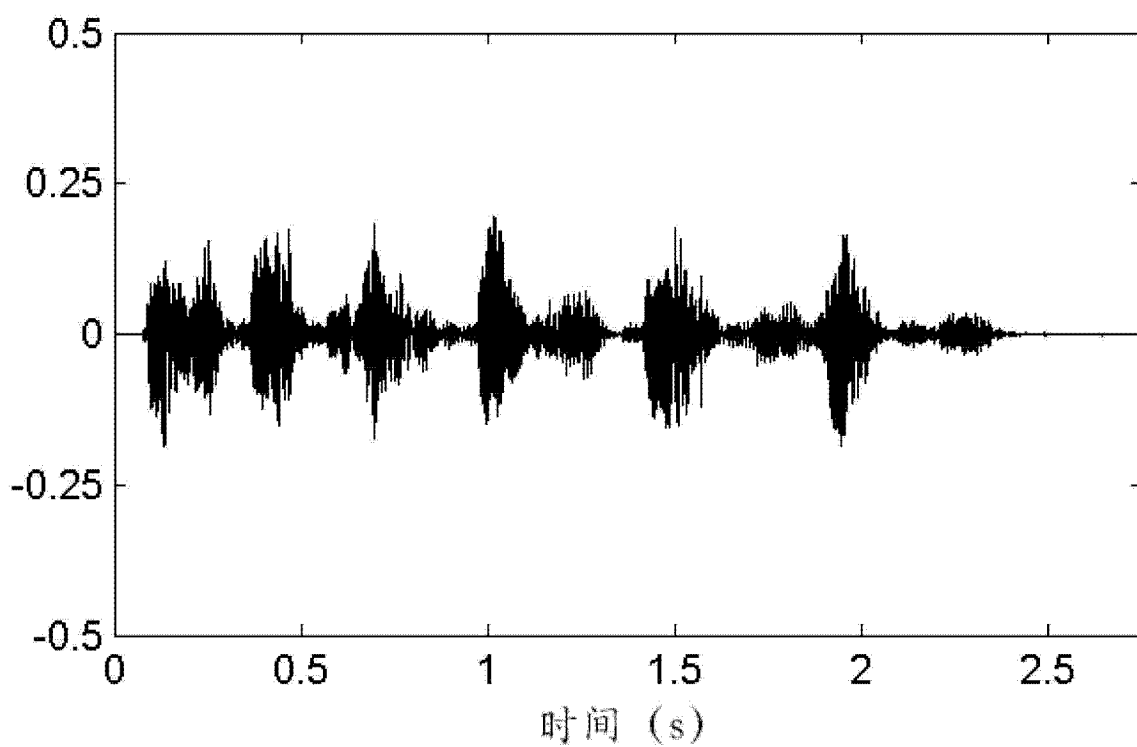


图 3(b)

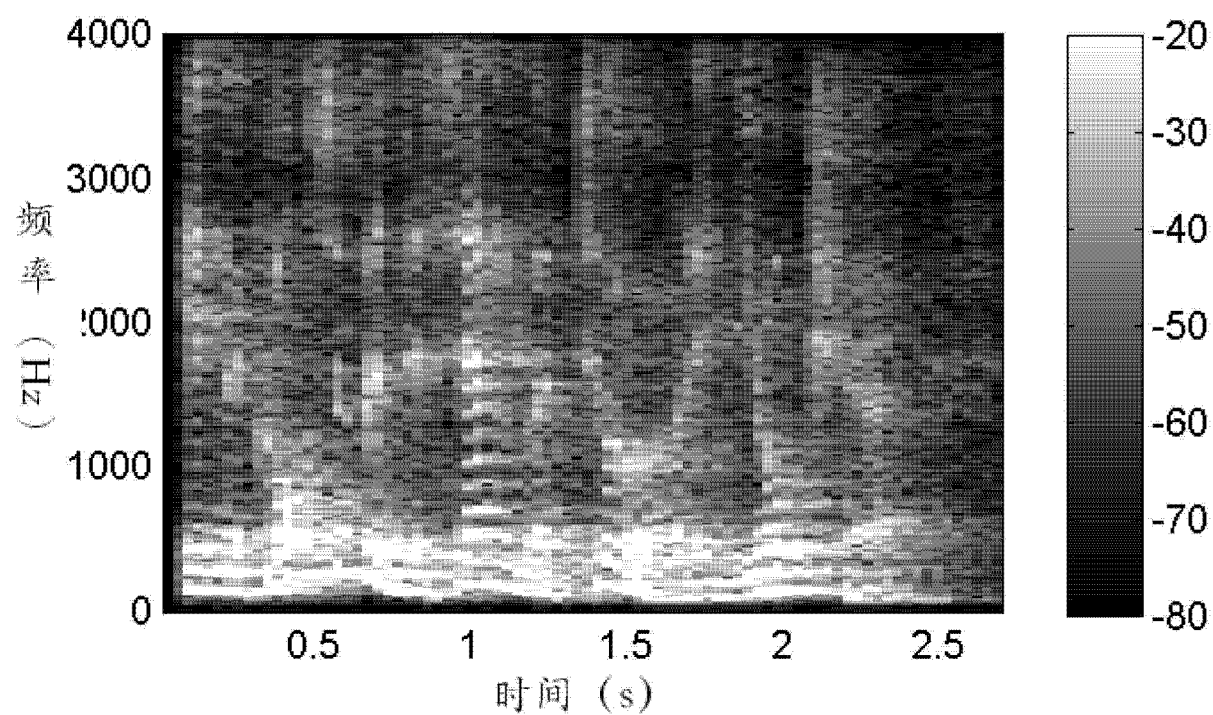


图 3(c)

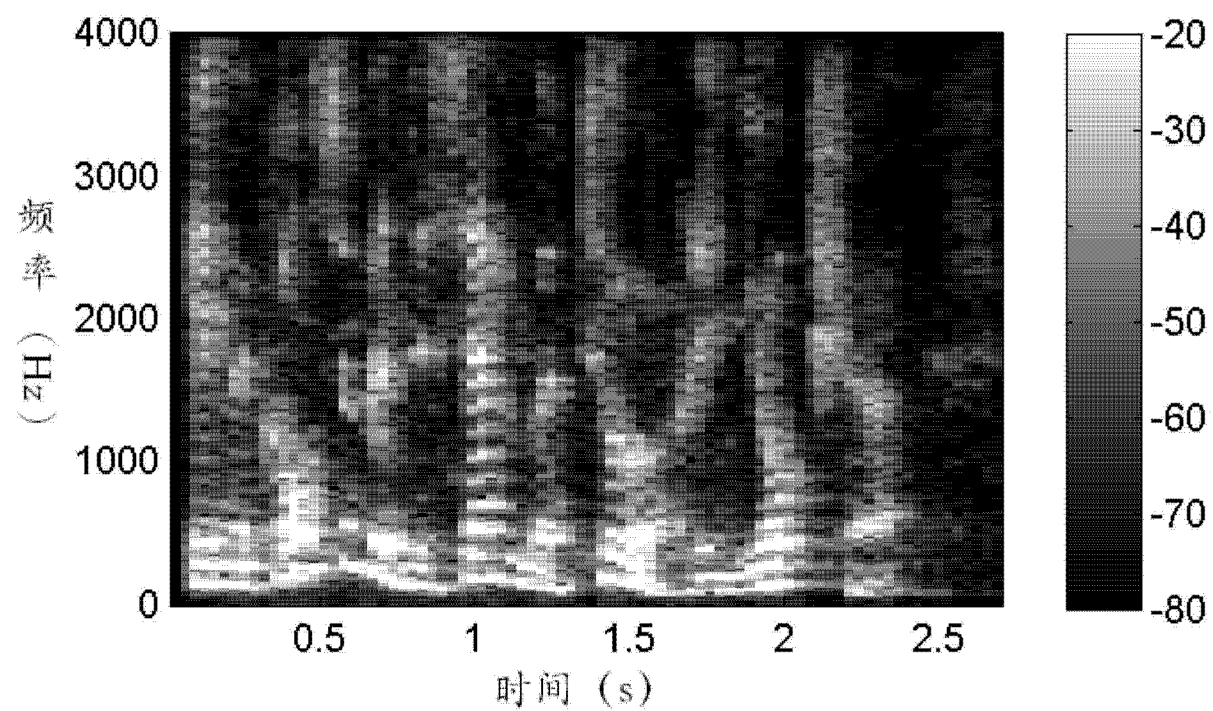


图 3(d)

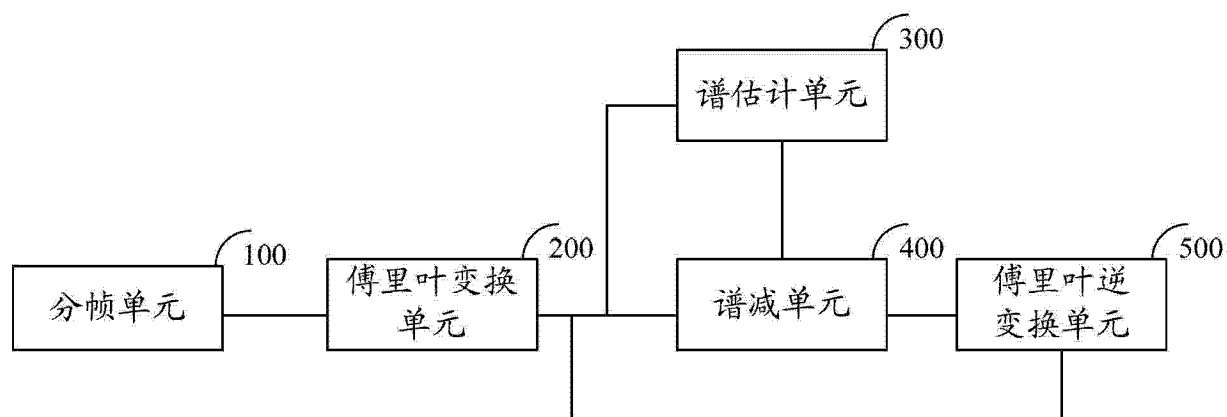


图 4

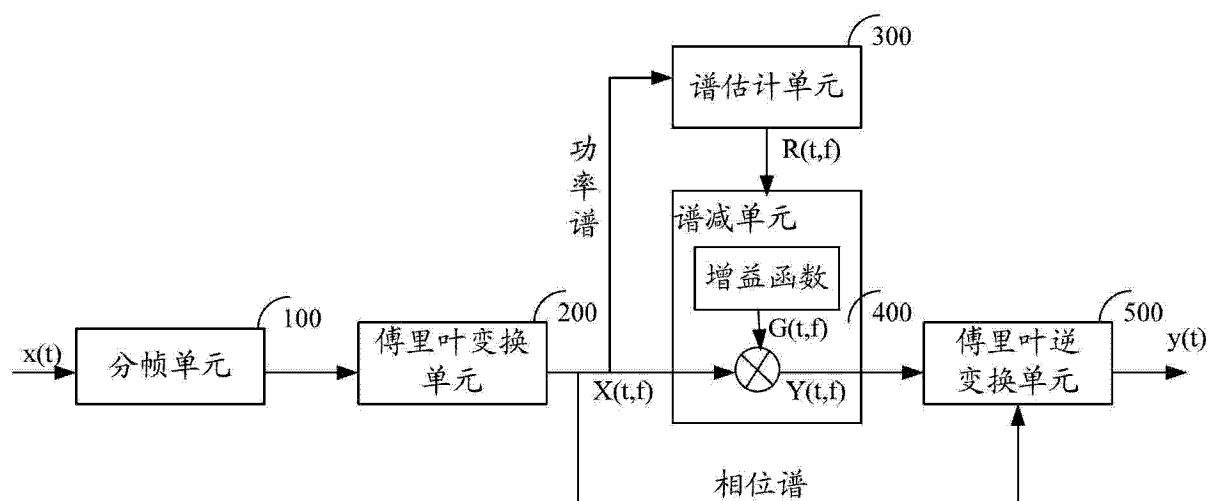


图 5