

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2013 年 12 月 27 日 (27.12.2013)



(10) 国际公布号
WO 2013/189199 A1

- (51) 国际专利分类号:
G10L 21/0208 (2013.01)
- (21) 国际申请号: PCT/CN2013/073584
- (22) 国际申请日: 2013 年 4 月 1 日 (01.04.2013)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201210201879.7 2012 年 6 月 18 日 (18.06.2012) CN
- (71) 申请人: 歌尔声学股份有限公司 (GOERTEK INC) [CN/CN]; 中国山东省潍坊市高新技术产业开发区东方路 268 号, Shandong 261031 (CN)。
- (72) 发明人: 楼厦厦 (LOU, Shasha); 中国山东省潍坊市高新技术产业开发区东方路 268 号, Shandong 261031 (CN)。 吴晓婕 (WU, Xiaojie); 中国山东省潍坊市高新技术产业开发区东方路 268 号, Shandong 261031 (CN)。 李波 (LI, Bo); 中国山东省潍坊市高新技术产业开发区东方路 268 号, Shandong 261031 (CN)。
- (74) 代理人: 北京市隆安律师事务所 (BEIJING LONGAN LAW FIRM); 中国北京市朝阳区建国门

外大街 21 号北京国际俱乐部 188 室, Beijing 100020 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第 21 条(3))。

(54) Title: METHOD AND DEVICE FOR DEREVERBERATION OF SINGLE-CHANNEL SPEECH

(54) 发明名称: 一种单通道语音去混响的方法和装置

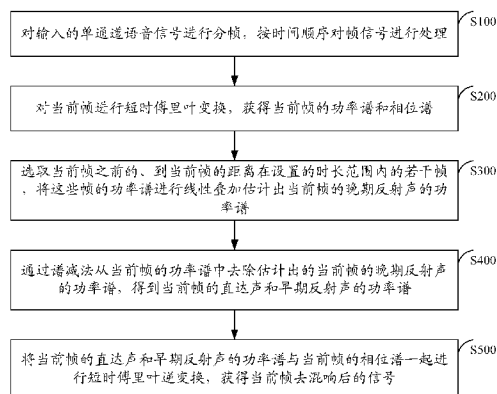


图 1 / Fig. 1

S100 FRAMING AN INPUT SINGLE-CHANNEL SPEECH SIGNAL, AND PROCESSING THE FRAME SIGNALS ACCORDING TO A TIME SEQUENCE
S200 PERFORMING SHORT-TIME FOURIER TRANSFORM ON A CURRENT FRAME TO OBTAIN A POWER SPECTRUM AND A PHASE SPECTRUM OF THE CURRENT FRAME
S300 SELECTING SEVERAL FRAMES PREVIOUS TO THE CURRENT FRAME AND HAVING A DISTANCE FROM THE CURRENT FRAME WITHIN A SET DURATION RANGE, AND PERFORMING LINEAR SUPERPOSITION ON THE POWER SPECTRA OF THESE FRAMES TO ESTIMATE THE POWER SPECTRUM OF A LATE REFLECTION SOUND OF THE CURRENT FRAME
S400 REMOVING THE ESTIMATED POWER SPECTRUM OF THE LATE REFLECTION SOUND OF THE CURRENT FRAME FROM THE POWER SPECTRUM OF THE CURRENT FRAME BY MEANS OF A SPECTRAL SUBTRACTION METHOD TO OBTAIN THE POWER SPECTRA OF A DIRECT SOUND AND AN EARLY REFLECTION SOUND OF THE CURRENT FRAME
S500 PERFORMING INVERSE SHORT-TIME FOURIER TRANSFORM ON THE POWER SPECTRA OF THE DIRECT SOUND AND THE EARLY REFLECTION SOUND OF THE CURRENT FRAME AND THE PHASE SPECTRUM OF THE CURRENT FRAME TOGETHER TO OBTAIN A SIGNAL OF THE CURRENT FRAME AFTER DEREVERBERATION

(57) Abstract: A method and device for dereverberation of single-channel speech. The method includes: framing an input single-channel speech signal, and processing the frame signals according to a time sequence (S100); performing short-time Fourier transform on a current frame to obtain a power spectrum and a phase spectrum of the current frame (S200); selecting several frames previous to the current frame and having a distance from the current frame within a set duration range, and performing linear superposition on the power spectra of these frames to estimate the power spectrum of a late reflection sound of the current frame (S300); removing the estimated power spectrum of the late reflection sound of the current frame from the power spectrum of the current frame by means of a spectral subtraction method to obtain the power spectra of a direct sound and an early reflection sound of the current frame (S400); and performing inverse short-time Fourier transform on the power spectra of the direct sound and the early reflection sound of the current frame together to obtain a signal of the current frame after dereverberation (S500). The dereverberation method and device can solve the problem that the estimation of a transfer function of a reverberation environment or the estimation of reverberation time is difficult in the dereverberation of single-channel speech.

(57) 摘要:

[见续页]



WO 2013/189199 A1



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处理（S100）；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（S200）；选取当前帧之前的、到当前帧的距离在设定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（S300）；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（S400）；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（S500）。该去混响方法和装置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。

一种单通道语音去混响的方法和装置

技术领域

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

在电话会议，智能电视网络电话等语音通讯中，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，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、地板和家具等多次反射，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。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。

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。

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，然后按照基音同步猝发准则（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极点滤波器的假设，与实验分析存在不相符的情况。

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

本发明提供的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。

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：

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。

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；

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。

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。

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。

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：

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。

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:

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。

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。

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。

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。

较佳地, 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。

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；

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；

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；

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；

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，便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。

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

图 1 为本发明单通道语音去混响的方法的流程图;

图 2 为真实房间的冲激响应的示意图;

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) 为混响信号和去混响信号的能量包络曲线;

图 4 为本发明单通道语音去混响装置的结构图;

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

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。

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。

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。

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。

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。

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。

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。

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。

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:

$$x(t) = h * s(t) + n(t)$$

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。

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 hd 、早期反射 he 和晚期反射 hl 。 hd 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 hd 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 hd 和 he 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。

因此，混响模型也可表示为：

$$x(t) = (hd + he) * s(t) + hl * s(t) + n(t)$$

hl 部分符合指数衰减模型，可用如下方程近似：

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

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。

下面详细描述如何进行晚期反射声的功率谱估计。

从功率谱分析角度来看，信号功率谱 $X(t, f)$ 可以表示为：

$$X(t, f) = Y(t, f) + R(t, f)$$

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。

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。

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。

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 hl 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。

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。

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，提高语音质量。

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。

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，便可获得较好语音质量。

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。

理论上，上限值的设置与应用方法的具体环境相关。在本发明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 $h(t)$ 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地去除。

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。

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：

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

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阶数， 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。

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。

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：

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

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阶数， 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； $Y(t - j \cdot \Delta t, f)$ 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。

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。

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：

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

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阶数, 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, $Y(t-j \cdot \Delta t, f)$ 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。

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。

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。

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:

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。

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:

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

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。

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。按本发明中提出的 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, 混响拖尾明显衰减, 语音质量得到显著提

升。

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。

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。

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。

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。

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。

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。

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。

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。

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。

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。

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。

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:

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

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。

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。

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:

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

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; $Y(t-j \cdot \Delta t, f)$ 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。

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。

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:

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

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, $Y(t-j \cdot \Delta t, f)$ 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。

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。

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。

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:

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

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。

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。

权利要求书

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 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。

1/4

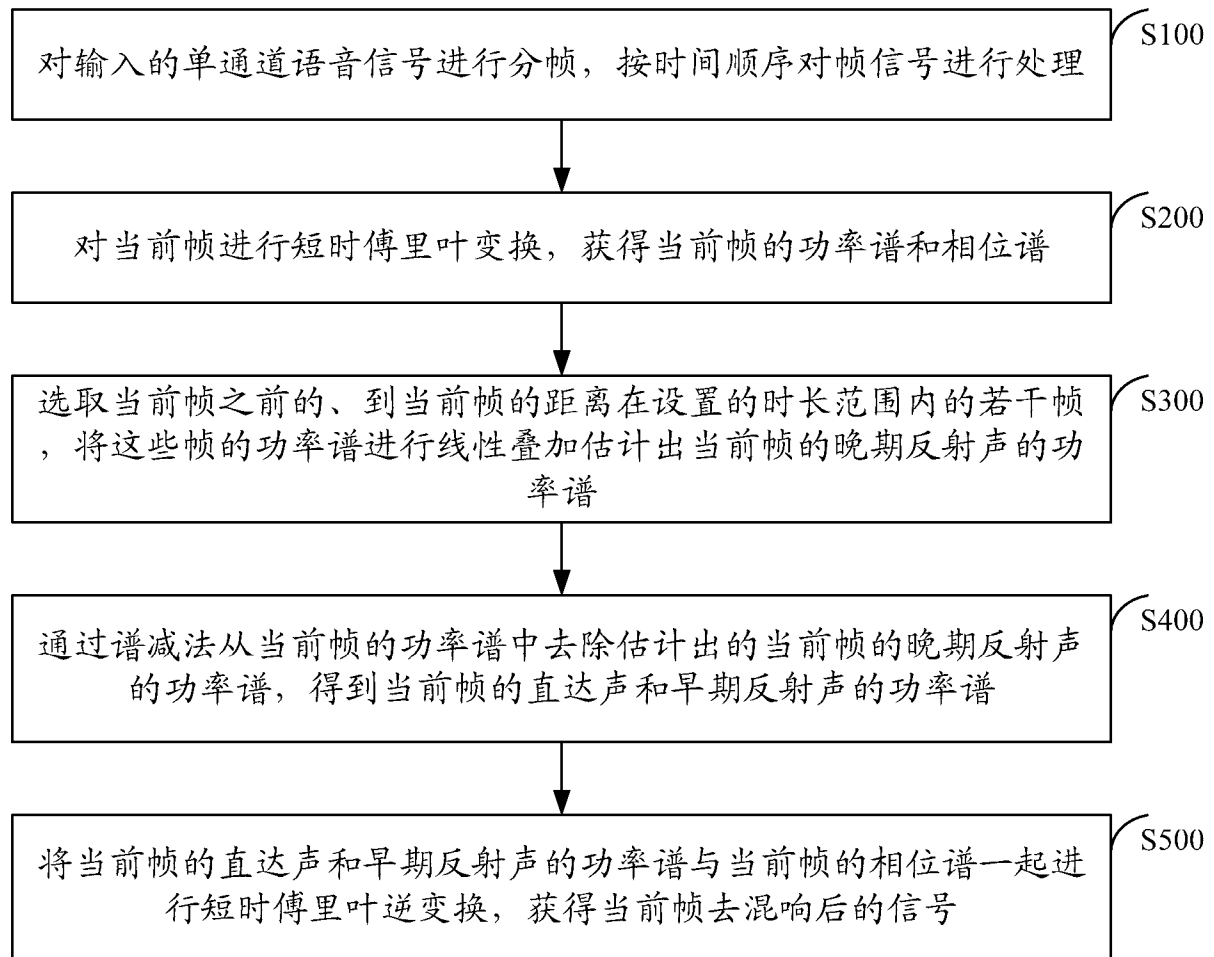


图 1

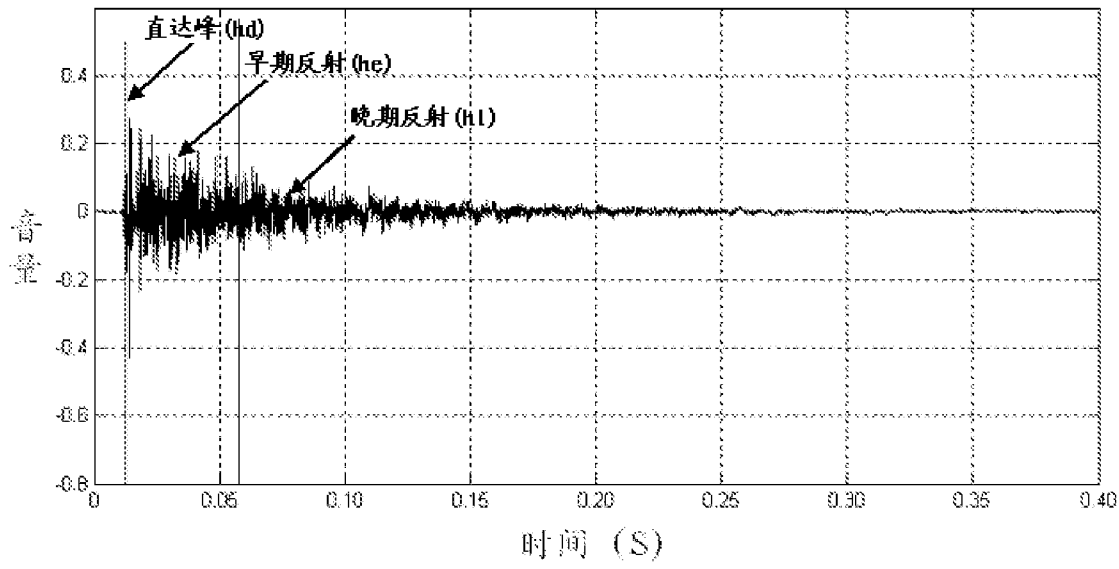


图 2

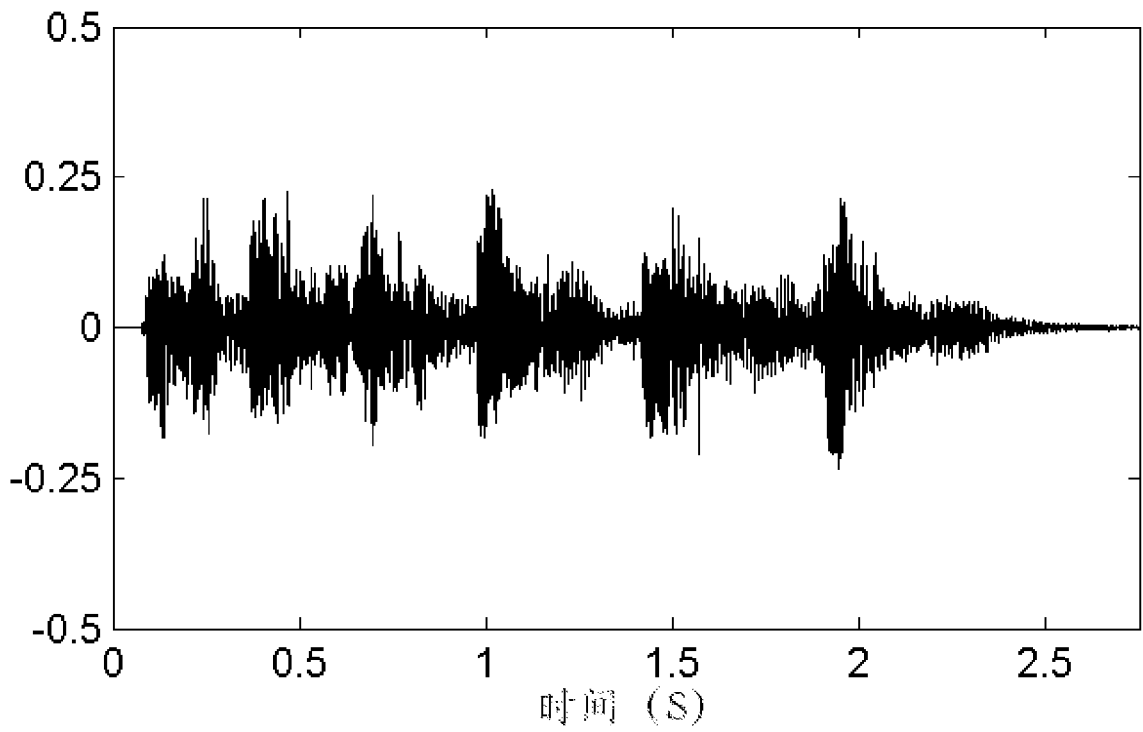


图 3 (a)

3/4

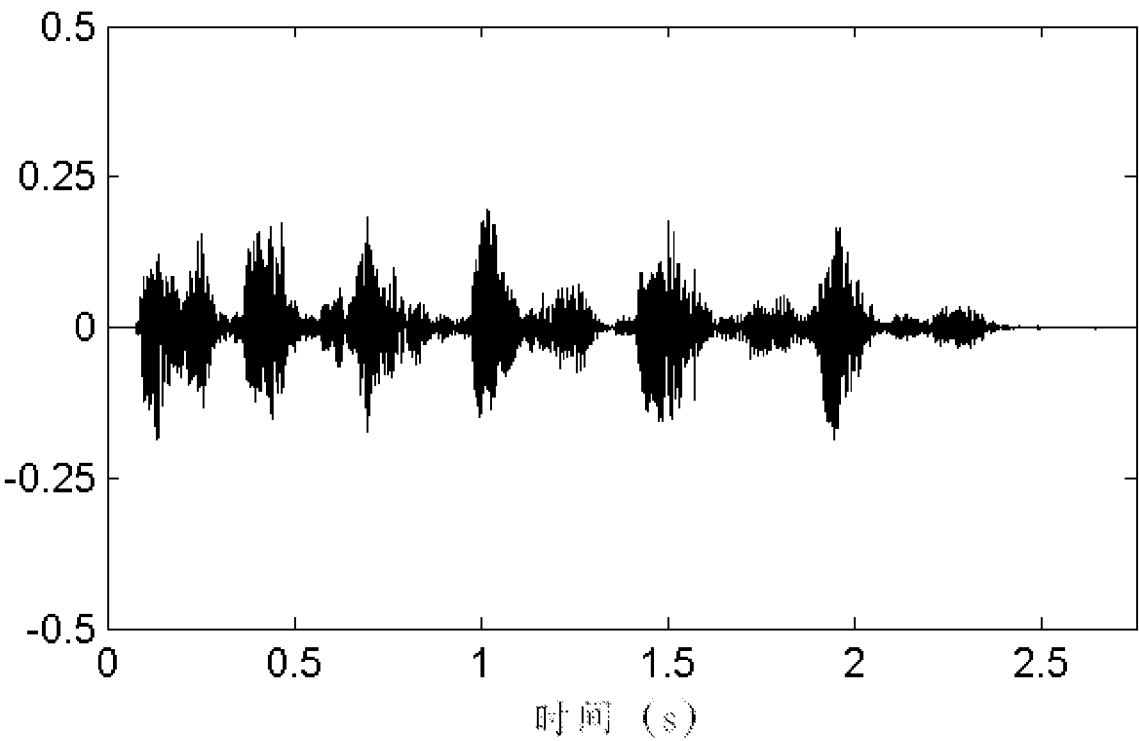


图 3 (b)

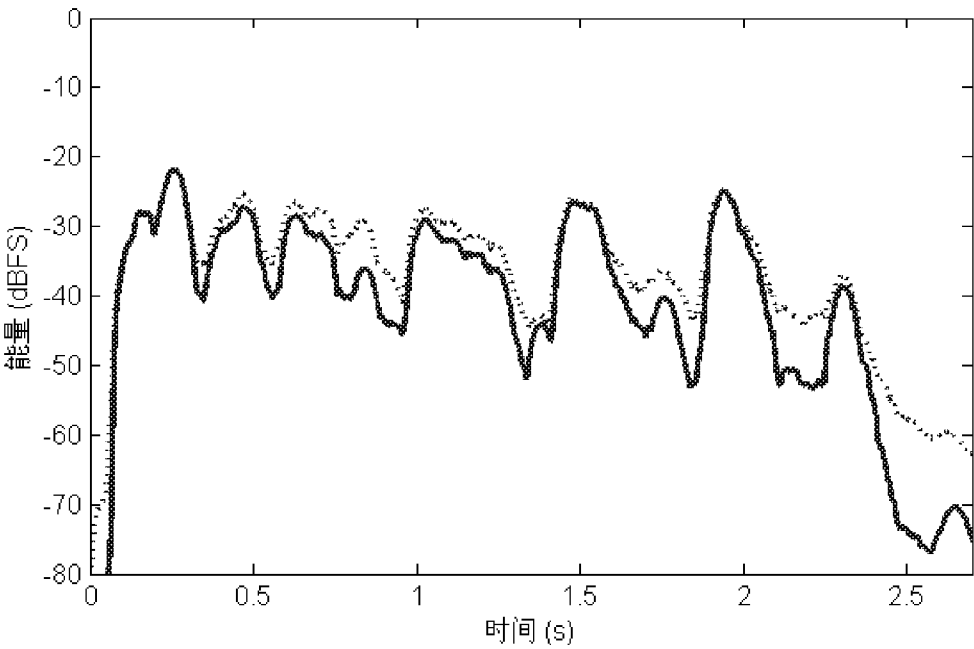


图 3 (c)

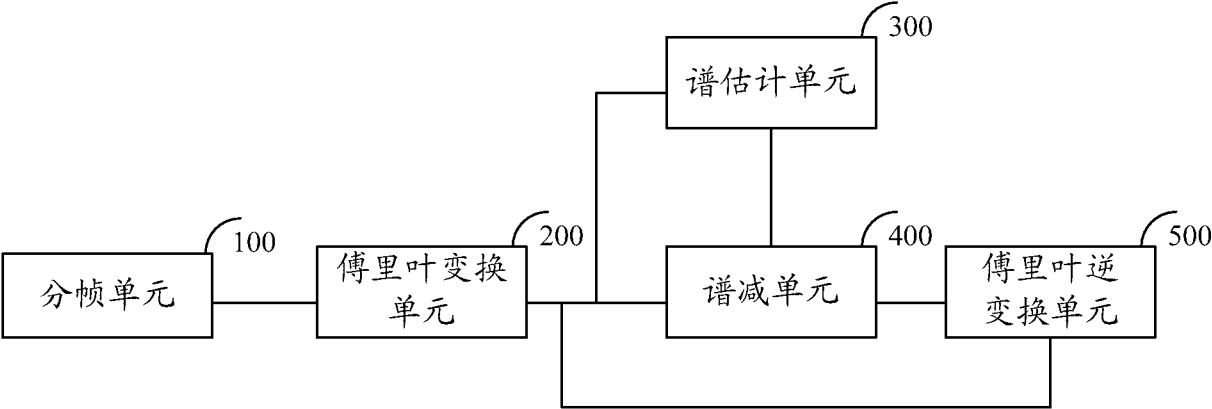


图 4

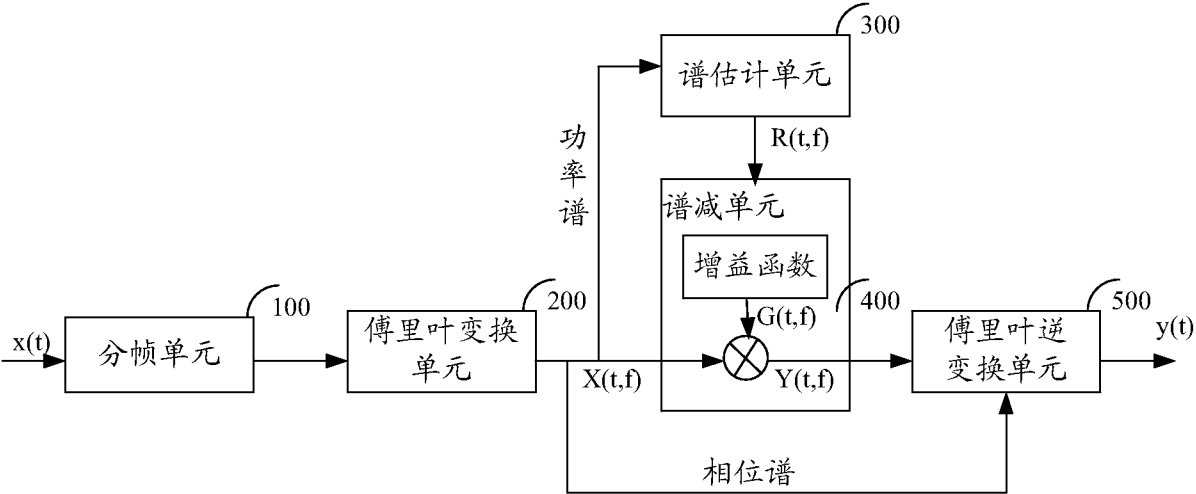


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/073584

A. CLASSIFICATION OF SUBJECT MATTER

G10L 21/0208 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G10L; G10; H04B; H04M; H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNKI, CNTXT, VEN, USTXT, EPTXT, WOTXT, TWTXT: reverberat+, reverberant+, dereverberat+, de, cancel+, remov+,
restrain+, suppress+, reduc+, lessen+, decreas+, attenuat+, eliminat+, echo??. late, latter, reflect+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 102750956 A (GOERTEK INC.) 24 October 2012 (24.10.2012) claims 1 to 10, description, paragraphs [0007] to [0127] and figures 1 to 5	1-10
X	CN 101385386 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 11 March 2009 (11.03.2009) description, page 8 to page 42 and figures 1-23C	1-10
X	US 8160262 B2 (NUANCE COMMUNICATIONS INC.) 17 April 2012 (17.04.2012) description, column 3, line 42 to column 13, line 41 and figures 3 to 7	1-10
A	US 2008292108 A1 (BUCK M. et al.) 27 November 2008 (27.11.2008) the whole document	1-10
A	US 2008059157 A1 (FUKUDA T. et al.) 06 March 2008 (06.03.2008) the whole document	1-10

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents: “A” document defining the general state of the art which is not considered to be of particular relevance “E” earlier application or patent but published on or after the international filing date “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) “O” document referring to an oral disclosure, use, exhibition or other means “P” document published prior to the international filing date but later than the priority date claimed		“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art “&” document member of the same patent family
Date of the actual completion of the international search 28 June 2013 (28.06.2013)		Date of mailing of the international search report 18 July 2013 (18.07.2013)
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451		Authorized officer YANG, Shilin Telephone No. (86-10) 62085717

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/073584

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1989550 A (KONINK PHILIPS ELECTRONICS N.V.) 27 June 2007 (27.06.2007) the whole document	1-10
A	CN 101454825 A (HARMAN INTERNATIONAL INDUSTRIES INC.) 10 June 2009 (10.06.2009) the whole document	1-10
A	CN 101315772 A (SHANGHAI JIAOTONG UNIVERSITY) 03 December 2008 (03.12.2008) the whole document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/073584

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102750956 A	24.10.2012	None	
CN 101385386 A	11.03.2009	WO 2007100137 A1	07.09.2007
		JPWO 2007100137 SX	23.07.2009
		US 2009248403 A1	01.10.2009
		JP 4774100 B2	14.09.2011
		CN 101385386 B	09.05.2012
		US 8271277 B2	18.09.2012
		EP 1993320 A1	19.11.2008
US 8160262 B2	17.04.2012	EP 2058804 A1	13.05.2009
		US 2009117948 A1	07.05.2009
US 2008292108 A1	27.11.2008	EP 1885154 A1	06.02.2008
US 2008059157 A1	06.03.2008	JP 2008058900 A	13.03.2008
		US 7590526 B2	15.09.2009
		JP 4107613 B2	25.06.2008
CN 1989550 A	27.06.2007	KR 20070036777 A	03.04.2007
		US 2008300869 A1	04.12.2008
		JP 2008507720 A	13.03.2008
		KR 1149591 B1	29.05.2012
		JP 5042823 B2	03.10.2012
		US 8116471 B2	14.02.2012
		WO 2006011104 A1	02.02.2006
		IN 200700280 P4	24.08.2007
		CN 1989550 B	13.10.2010
		EP 1774517 A1	18.04.2007
CN 101454825 A	10.06.2009	WO 2008034221 A1	27.03.2008
		JP 4964943 B2	04.07.2012
		JP 2009531722 A	03.09.2009
		EP 2064699 A1	03.06.2009
CN 101315772 A	03.12.2008	None	

国际检索报告

国际申请号

PCT/CN2013/073584

A. 主题的分类

G10L 21/0208 (2013.01) i

按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: G10L, G10, H04B, H04M, H04R

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, CNKI, CNTXT: 混响, 回响, 回声, 反射, 漫射, 后期, 晚期, 后, 除, 去, 消, 减, 抑制

VEN, USTXT, EPTXT, WOTXT, TWTXT: reverberat+, reverberant+, dereverberat+, de, cancel+, remov+, restrain+,

suppress+, reduc+, lessen+, decreas+, attenuat+, eliminat+, echo??, late, latter, reflect+,

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 102750956 A (歌尔声学股份有限公司) 24.10 月 2012 (24.10.2012), 权利要求 1-10, 说明书第[0007]段-第[0127]段, 附图 1-5	1-10
X	CN 101385386 A (日本电信电话株式会社) 11.3 月 2009 (11.03.2009), 说明书第 8 页-第 42 页, 附图 1-23C	1-10
X	US 8160262 B2 (NUANCE COMMUNICATIONS INC.) 17.4 月 2012 (17.04.2012), 说明书第 3 栏第 42 行-第 13 栏第 41 行, 附图 3-7	1-10
A	US 2008292108 A1 (BUCK M.等) 27.11 月 2008 (27.11.2008), 全文	1-10
A	US 2008059157 A1 (FUKUDA T.等) 06.3 月 2008 (06.03.2008), 全文	1-10
A	CN 1989550 A (皇家飞利浦电子股份有限公司) 27.6 月 2007 (27.06.2007), 全文	1-10

☒ 其余文件在 C 栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

28.6 月 2013 (28.06.2013)

国际检索报告邮寄日期

18.7 月 2013 (18.07.2013)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

杨士林

电话号码: (86-10) 62085717

国际检索报告

国际申请号
PCT/CN2013/073584

C(续). 相关文件

类 型	引用文件，必要时，指明相关段落	相关的权利要求
A	CN 101454825 A（哈曼国际工业有限公司）10.6 月 2009（10.06.2009），全文	1-10
A	CN 101315772 A（上海交通大学）03.12 月 2008（03.12.2008），全文	1-10

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2013/073584

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN 102750956 A	24.10.2012	无	
CN 101385386 A	11.03.2009	WO 2007100137 A1	07.09.2007
		JP WO2007100137S X	23.07.2009
		US 2009248403 A1	01.10.2009
		JP 4774100 B2	14.09.2011
		CN 101385386 B	09.05.2012
		US 8271277 B2	18.09.2012
		EP 1993320 A1	19.11.2008
US 8160262 B2	17.04.2012	EP 2058804 A1	13.05.2009
		US 2009117948 A1	07.05.2009
US 2008292108 A1	27.11.2008	EP 1885154 A1	06.02.2008
US 2008059157 A1	06.03.2008	JP 2008058900 A	13.03.2008
		US 7590526 B2	15.09.2009
		JP 4107613 B2	25.06.2008
CN 1989550 A	27.06.2007	KR 20070036777 A	03.04.2007
		US 2008300869 A1	04.12.2008
		JP 2008507720 A	13.03.2008
		KR 1149591 B1	29.05.2012
		JP 5042823 B2	03.10.2012
		US 8116471 B2	14.02.2012
		WO 2006011104 A1	02.02.2006
		IN 200700280 P4	24.08.2007
		CN 1989550 B	13.10.2010
		EP 1774517 A1	18.04.2007
CN 101454825 A	10.06.2009	WO 2008034221 A1	27.03.2008
		JP 4964943 B2	04.07.2012
		JP 2009531722 A	03.09.2009
		EP 2064699 A1	03.06.2009
CN 101315772 A	03.12.2008	无	