



(12)发明专利申请

(10)申请公布号 CN 106960672 A

(43)申请公布日 2017.07.18

(21)申请号 201710203054.1

H04S 7/00(2006.01)

(22)申请日 2017.03.30

(71)申请人 国家计算机网络与信息安全管理中心

地址 100029 北京市朝阳区裕民路甲3号

申请人 中国科学院声学研究所

(72)发明人 高昕 颜永红 邹潇湘 白海钊
舒敏 云晓春 王锟 张震 计哲
董琳 金暉 王中华 李海灵
李佳

(74)专利代理机构 工业和信息化部电子专利中心 11010

代理人 于金平

(51)Int.Cl.

G10L 21/0388(2013.01)

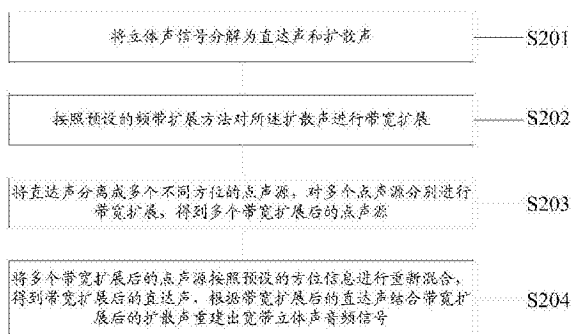
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种立体声音频的带宽扩展方法与装置

(57)摘要

本发明公开了一种立体声音频的带宽扩展方法与装置。该方法包括：将立体声信号分解为直达声和扩散声；按照预设的频带扩展方法对扩散声进行带宽扩展；将直达声分离成多个不同方位的点声源，对多个点声源分别进行带宽扩展，得到带宽扩展后的多个点声源；将带宽扩展后的多个点声源按照预先估计的方位信息进行重新混合，得到带宽扩展后的直达声；根据带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号。借助于本发明的技术方案，解决了现有技术中仅根据单个声道重建信号的主观质量实现对信号带宽的扩展，没有考虑到两个声道中信号能量和相位的相关性，其重建立体声信号严重影响了听者对声源位置和距离的判定的问题。



1. 一种立体声音频的带宽扩展方法,其特征在于,包括:
将立体声信号分解为直达声和扩散声;
按照预设的频带扩展方法对所述扩散声进行带宽扩展;
将所述直达声分离成多个不同方位的点声源,对多个点声源分别进行带宽扩展,得到带宽扩展后的多个点声源;
将所述带宽扩展后的多个点声源按照预先估计的方位信息进行重新混合,得到带宽扩展后的直达声,根据所述带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号。
2. 如权利要求1所述的立体声音频的带宽扩展方法,其特征在于,所述将立体声信号分解为直达声和扩散声,包括:
将所述立体声信号分解为左声道和右声道;
分别将分帧处理后的左声道和右声道进行时频变换,得到立体声信号的左声道短时频谱成分和右声道短时频谱成分;
分别根据所述左声道短时频谱成分和右声道短时频谱成分,得到左右声道信号能量谱之间的和 P_{sum} 、左右声道信号能量谱之间的差 P_{diff} 、左右声道信号能量谱之间的互相关 P_{cc} ;
利用所述 P_{sum} 、 P_{diff} 及 P_{cc} ,通过最小二乘法得到直达声矩阵;
利用所述直达声矩阵从所述立体声信号中分离出直达声;
在所述立体声信号中剔除所述直达声,得到扩散声。
3. 如权利要求1所述的立体声音频的带宽扩展方法,其特征在于,
所述将所述直达声分离成多个不同方位的点声源,包括:
计算每一个时频点上直达声的方向信息,对全部时频点的方向信息进行聚类,得到方向信息的聚类中心,所述聚类中心分别对应各个点声源的方向信息;
根据某一时频点上直达声的方向信息和所述方向信息的聚类中心,得到掩蔽矩阵;
利用所述掩蔽矩阵对直达声进行分离,得到多个不同方位的点声源。
4. 如权利要求1所述的立体声音频的带宽扩展方法,其特征在于,
所述对多个点声源分别进行带宽扩展,包括:
将多个点声源分别输入到预设的状态空间模型中拟合窄带信号的短时频谱和宽带信号的短时频谱之间的映射关系,并根据预设的误差准则对宽带信号短时频谱高频成分的频谱包络进行估计,结合低频频谱包络和采用适当频谱修补方法扩展后的频谱细节,得到带宽扩展后的多个点声源。
5. 如权利要求4所述的立体声音频的带宽扩展方法,其特征在于,
所述在所述状态空间模型中拟合窄带信号的短时频谱和宽带信号的短时频谱之间的映射关系,并根据预设的误差准则对高频成分的频谱包络进行估计,包括:
利用前一时刻隐藏状态矢量和前一时刻窄带信号的短时频谱,得到所述预设状态空间模型中的隐藏状态矢量;
利用所述预设状态空间模型中的隐藏状态矢量和当前时刻窄带信号的短时频谱,得到宽带信号的短时频谱。
6. 一种立体声音频的带宽扩展装置,其特征在于,包括:分解模块、扩散声扩展模块、直达声分离与扩展模块、重构模块;

所述分解模块,用于将立体声信号分解为直达声和扩散声;

所述扩散声扩展模块,用于按照预设的频带扩展方法对所述扩散声进行带宽扩展;

所述直达声分离与扩展模块,用于将所述直达声分离成多个不同方位的点声源,对多个点声源分别进行带宽扩展,得到带宽扩展后的多个点声源;

所述重构模块,用于将所述带宽扩展后的多个点声源按照预先估计的方位信息进行重新混合,得到带宽扩展后的直达声,根据所述带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号。

7.如权利要求6所述的立体声音频的带宽扩展装置,其特征在于,所述分解模块具体用于:

将所述立体声信号分解为左声道和右声道;

分别将分帧处理后的左声道和右声道进行时频变换,得到立体声信号的左声道短时频谱成分和右声道短时频谱成分;

分别根据所述左声道短时频谱成分和右声道短时频谱成分,得到左右声道信号能量谱之间的和 P_{sum} 、左右声道信号能量谱之间的差 P_{diff} 、左右声道信号能量谱之间的互相关 P_{cc} ;

利用所述 P_{sum} 、 P_{diff} 及 P_{cc} ,通过最小二乘法得到直达声矩阵;

利用所述直达声矩阵从所述立体声信号中分离出直达声;

在所述立体声信号中剔除所述直达声得到扩散声。

8.如权利要求6所述的立体声音频的带宽扩展装置,其特征在于,所述直达声分离与扩展模块具体用于:

计算每一个时频点上直达声的方向信息,对全部时频点的方向信息进行聚类,得到方向信息的聚类中心,所述聚类中心分别对应各个点声源的方向信息;

根据某一时频点上直达声的方向信息和所述方向信息的聚类中心,得到掩蔽矩阵;

利用所述掩蔽矩阵对直达声进行分离,得到多个不同方位的点声源。

9.如权利要求6所述的立体声音频的带宽扩展装置,其特征在于,所述直达声分离与扩展模块具体用于:

将多个点声源分别输入到预设的状态空间模型中拟合窄带信号的短时频谱和宽带信号的短时频谱之间的映射关系,并根据预设的误差准则对宽带信号短时频谱高频成分的频谱包络进行估计,结合低频频谱包络和采用适当频谱修补方法扩展后的频谱细节,得到带宽扩展后的直达声。

10.如权利要求9所述的立体声音频的带宽扩展装置,其特征在于,所述直达声分离与扩展模块具体用于:

利用前一时刻隐藏状态矢量和前一时刻窄带信号的短时频谱,得到所述预设状态空间模型中的隐藏状态矢量;

利用所述预设状态空间模型中的隐藏状态矢量和当前时刻窄带信号的短时频谱,得到宽带信号的短时频谱。

一种立体声音频的带宽扩展方法与装置

技术领域

[0001] 本发明涉及网络技术应用领域,特别涉及一种立体声音频的带宽扩展方法与装置。

背景技术

[0002] 在数字音频信号处理技术中,通常将覆盖人耳可感知的20Hz~20KHz全部频率范围内的音频信号称作全带音频,这类信号主要应用于音乐信号的高保真重现。现阶段的音频即时通信系统无法提供足够的网络传输速率和终端处理能力,不可避免地会限制重建信号的有效带宽,优先量化编码音频信号的低频成分,进而提升音频通信系统的编码效率。

[0003] 传统电话语音通信系统通常传输的是窄带信号,其频率分布在300~3400Hz范围内,采样率为8kHz。相关主观听力测试结果表明,窄带语音中保留了91%的音节可懂度以及99%的语句可理解性。但是相比于真实语音,在实际通话中所传输窄带信号的自然度和主观质量均有明显下降。由于高频成分的缺失,窄带语音无法良好地区分部分的清音或爆破音,并削弱了其描述说话人特性的能力。为了有效地克服窄带音频的不足,宽带音频被广泛应用到了电话语音通信领域中,其有效带宽扩展到50Hz~7kHz,较好地覆盖了表征语音信号重要特性的大部分频谱,实现了接近调幅广播的音质水平。然而受到历史、经济、技术等诸多问题的限制,传统固定和移动通信完全实现从窄带向宽带音频的迈进还需要相当长的一段过渡期。

[0004] 作为一种有效的音频增强方法,频带扩展方法可以在不改变窄带信号信源编码和网络传输的前提下,通过分析原始音频信号的时频特性,在接收端从重建的宽带音频中人为地恢复出编码端所截去的高频成分,进而达到增强重建音频听觉质量的目的。对于听力有损人士,频带扩展方法能够进一步改善其音素和语义的分辨能力。近十几年来,许多研究机构与科研人员针对单声道语音信号的频带扩展相继提出了众多解决方案。这些方法通常分别从频谱包络扩展和频谱细节扩展两个方面出发,进而合成信号高频成分,其原理如图1所示。首先根据人耳听觉感知原理对窄带信号进行时频特征提取;接下来,借助边信息或者先验知识所描述高低频特征之间的映射关系来对高频成分的频谱包络和能量进行估计;同时,选择适当的频谱修补方法来扩展频谱细节;最终,结合扩展后的频谱包络和频谱细节,实现宽带音频信号高频成分的有效重建。

[0005] 对于立体声音频,传统频带扩展方法多针对两个声道进行高频成分独立重建,这类方法仅根据单个声道重建信号的主观质量实现对信号带宽的扩展,没有考虑到两个声道中信号能量和相位的相关性,其重建立体声信号严重影响了听者对声源位置和距离的判定。

发明内容

[0006] 鉴于上述问题,本发明提供了一种立体声音频的带宽扩展方法与装置。

[0007] 本发明提供的立体声音频的带宽扩展方法,包括以下步骤:

- [0008] 将立体声信号分解为直达声和扩散声；
- [0009] 按照预设的频带扩展方法对所述扩散声进行带宽扩展；
- [0010] 将所述直达声分离成多个不同方位的点声源，对多个点声源分别进行带宽扩展，得到带宽扩展后的多个点声源；
- [0011] 将所述带宽扩展后的多个点声源按照预先估计的方位信息进行重新混合，得到带宽扩展后的直达声；
- [0012] 根据所述带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号。
- [0013] 本发明还提供了一种立体声音频的带宽扩展装置，包括：分解模块、扩散声扩展模块、直达声分离与扩展模块、重构模块；
- [0014] 所述分解模块，用于将立体声信号分解为直达声和扩散声；
- [0015] 所述扩散声扩展模块，用于按照预设的频带扩展方法对所述扩散声进行带宽扩展；
- [0016] 所述直达声分离与扩展模块，用于将所述直达声分离成多个不同方位的点声源，对多个点声源分别进行带宽扩展，得到带宽扩展后的多个点声源；
- [0017] 所述重构模块，用于将所述带宽扩展后的多个点声源按照预先估计的方位信息进行重新混合，得到带宽扩展后的直达声，根据所述带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号。
- [0018] 本发明有益效果如下：
- [0019] 本发明实施例首先利用声道间的频谱相关性将输入立体声信号分解为直达声和扩散声两种成分，然后扩散声成分直接利用传统频带扩展方法进行扩展；直达声则依据不同声源在时频结构上的稀疏性分离成多个不同方位的点声源，并分别进行带宽扩展，最终扩展后的点声源依照其在原始立体声中方位信息进行重新混合，并结合带宽扩展后的扩散声成分，重建出宽带立体声音频信号。本发明解决了现有技术中仅根据单个声道重建信号的主观质量实现对信号带宽的扩展，没有考虑到两个声道中信号能量和相位的相关性，其重建立体声信号严重影响了听者对声源位置和距离的判定的问题。

附图说明

- [0020] 图1为现有技术中单声道语音信号频带扩展方法的基本流程图；
- [0021] 图2是本发明方法实施例的立体声音频的带宽扩展方法的流程图；
- [0022] 图3是本发明装置实施例的立体声音频的带宽扩展装置的结构示意图；
- [0023] 图4是本发明实例1的立体声音频的带宽扩展方法的原理框图；
- [0024] 图5是本发明实例1中基于深度神经网络的状态空间模型的原理框图。

具体实施方式

[0025] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

[0026] 为了解决现有技术中仅根据单个声道重建信号的主观质量实现对信号带宽的扩展,没有考虑到两个声道中信号能量和相位的相关性,其重建立体声信号严重影响了听者对声源位置和距离的判定的问题,本发明提供了一种立体声音频的带宽扩展方法与装置,以下结合附图以及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不限定本发明。

[0027] 根据本发明的方法实施例,提供了一种立体声音频的带宽扩展方法,图1是本发明方法实施例的立体声音频的带宽扩展方法的流程图,如图1所示,根据本发明方法实施例的立体声音频的带宽扩展方法包括如下处理:

[0028] 步骤201,将立体声信号分解为直达声和扩散声。

[0029] 具体的,步骤201包括以下步骤:

[0030] 将所述立体声信号分解为左声道和右声道;

[0031] 分别将分帧处理后的左声道和右声道进行时频变换,得到立体声信号的左声道短时频谱成分和右声道短时频谱成分;

[0032] 分别根据所述左声道短时频谱成分和右声道短时频谱成分,得到左右声道信号能量谱之间的和 P_{sum} 、左右声道信号能量谱之间的差 P_{diff} 、左右声道信号能量谱之间的互相关 P_{cc} ;

[0033] 利用所述 P_{sum} 、 P_{diff} 、及 P_{cc} 通过最小二乘法得到直达声矩阵;

[0034] 利用所述直达声矩阵从所述立体声信号中分离出直达声;

[0035] 利用所述立体声信号减去所述直达声得到扩散声。

[0036] 更加具体的,所述分别根据所述左声道短时频谱成分和右声道短时频谱成分,得到左右声道信号能量谱之间的和 P_{sum} 、左右声道信号能量谱之间的差 P_{diff} 、左右声道信号能量谱之间的互相关 P_{cc} 、包括:

[0037] 利用所述左声道短时频谱成分 $S_L(t, f)$ 和所述右声道短时频谱成分 $S_R(t, f)$ 根据公式 $P_{sum} = |S_L(t, f)|^2 + |S_R(t, f)|^2$ 计算左右声道信号能量谱之间的和 P_{sum} ;

[0038] 利用所述左声道短时频谱成分 $S_L(t, f)$ 和所述右声道短时频谱成分 $S_R(t, f)$ 根据公式 $P_{diff} = |S_L(t, f)|^2 - |S_R(t, f)|^2$ 计算左右声道信号能量谱之间的差 P_{diff} ;

[0039] 利用所述左声道短时频谱成分 $S_L(t, f)$ 和所述右声道短时频谱成分 $S_R(t, f)$ 根据公式 $P_{cc} = R\{S_L(t, f) S_R^*(t, f)\}$ 计算左右声道信号能量谱之间的互相关 P_{cc} ,其中 $R\{\}$ 为取实部操作。

[0040] 更加具体的,所述利用所述直达声矩阵从立体声信号中分离出直达声,包括:

[0041] 利用所述直达声矩阵 $M_D(t, f)$ 根据公式1从立体声信号 $S(t, f)$ 中分离出直达声 $S'(t, f)$;

[0042] $S'(t, f) = M_D(t, f) [S_L(t, f) S_R(t, f)]^T$ 公式1。

[0043] 步骤202,按照预设的频带扩展方法对所述扩散声进行带宽扩展。

[0044] 具体的,步骤202直接利用传统的频带扩展方法对所述扩散声进行带宽扩展,本发明不作赘述。

[0045] 步骤203,将所述直达声分离成多个不同方位的点声源,对多个点声源分别进行带宽扩展,得到带宽扩展后的多个点声源。

[0046] 具体的,步骤203中将所述直达声分离成多个不同方位的点声源,包括:

[0047] 计算每一个时频点上直达声的方向信息,对全部时频点的方向信息进行聚类,得到方向信息的聚类中心,所述聚类中心分别对应各个点声源的方向信息;

[0048] 根据某一时频点上直达声的方向信息和所述方向信息的聚类中心,得到掩蔽矩阵;

[0049] 利用所述掩蔽矩阵对直达声进行分离,得到多个不同方位的点声源。

[0050] 具体的,所述对多个点声源分别进行带宽扩展,包括:

[0051] 将多个点声源分别输入到预设的状态空间模型中拟合窄带信号的短时频谱和宽带信号的短时频谱之间的映射关系,并根据预设的误差准则对宽带信号短时频谱高频成分的频谱包络进行估计,结合低频频谱包络和采用适当频谱修补方法扩展后的频谱细节,得到带宽扩展后的多个点声源。

[0052] 更加具体的,所述在所述状态空间模型中拟合窄带信号的短时频谱和宽带信号的短时频谱之间的映射关系,并根据预设的误差准则对高频成分的频谱包络进行估计,包括:

[0053] 利用前一刻隐藏状态矢量和前一刻窄带信号的短时频谱,得到状态空间模型中隐藏状态矢量;

[0054] 利用所述状态空间模型中隐藏状态矢量和当前时刻窄带信号的短时频谱,得到宽带信号的短时频谱。

[0055] 步骤204,将所述多个带宽扩展后的点声源按照预设的方位信息进行重新混合,得到带宽扩展后的直达声,根据所述带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号。

[0056] 具体的,所述预先估计的方位信息根据所述方向信息的聚类中心估计得到,所述估计的方法为本领域的常规技术手段,本发明对此不作赘述。

[0057] 具体的,利用公式2根据所述带宽扩展后的直达声结合带宽扩展后的扩散声重建出宽带立体声音频信号;

[0058]

$$\hat{S}(t, f) = \hat{S}'(t, f) + \hat{S}''(t, f) \quad \text{公式 2;}$$

[0059] 在公式2中, $\hat{S}(t, f)$ 表示带宽扩展后立体声信号的短时频谱; $\hat{S}'(t, f)$ 表示带宽扩展后直达声的短时频谱; $\hat{S}''(t, f)$ 表示带宽扩展后扩散声的短时频谱。

[0060] 与本发明的方法实施例相对应,提供了一种立体声音频的带宽扩展装置,图3是本发明装置实施例的立体声音频的带宽扩展装置的结构示意图,如图3所示,根据本发明装置实施例的立体声音频的带宽扩展装置包括:分解模块30、扩散声扩展模块32、直达声分离与扩展模块34、重构模块36,以下对本发明实施例的各个模块进行详细的说明。

[0061] 具体地,所述分解模块30,用于将立体声信号分解为直达声和扩散声;

[0062] 所述扩散声扩展模块32,用于按照预设的频带扩展方法对所述扩散声进行带宽扩展;

[0063] 所述直达声分离与扩展模块34,用于将所述直达声分离成多个不同方位的点声源,对多个点声源分别进行带宽扩展,得到带宽扩展后的多个点声源;

[0064] 所述重构模块36,用于将所述带宽扩展后的多个点声源按照预先估计的方位信息进行重新混合,得到带宽扩展后的直达声,用于根据所述带宽扩展后的直达声结合带宽扩

展后的扩散声重建出宽带立体声音频信号。

[0065] 所述分解模块30具体用于：

[0066] 将所述立体声信号分解为左声道和右声道；

[0067] 分别将分帧处理后的左声道和右声道进行时频变换，得到立体声信号的左声道短时频谱成分和右声道短时频谱成分；

[0068] 分别根据所述左声道短时频谱成分和右声道短时频谱成分，得到左右声道信号能量谱之间的和 P_{sum} 、左右声道信号能量谱之间的差 P_{diff} 、左右声道信号能量谱之间的互相关 P_{cc} ；

[0069] 利用所述 P_{sum} 、 P_{diff} 、及 P_{cc} 通过最小二乘法得到直达声矩阵；

[0070] 利用所述直达声矩阵从所述立体声信号中分离出直达声；

[0071] 利用所述立体声信号减去所述直达声得到扩散声。

[0072] 所述直达声分离与扩展模块34具体用于：

[0073] 计算每一个时频点上直达声的方向信息，对全部时频点的方向信息进行聚类，得到方向信息的聚类中心，所述聚类中心分别对应各个点声源的方向信息；

[0074] 根据某一时频点上直达声的方向信息和所述方向信息的聚类中心，得到掩蔽矩阵；

[0075] 利用所述掩蔽矩阵对直达声进行分离，得到多个不同方位的点声源。

[0076] 所述直达声分离与扩展模块34具体用于：

[0077] 将多个点声源分别输入到预设的状态空间模型中拟合窄带信号的短时频谱和宽带信号的短时频谱之间的映射关系，并根据预设的误差准则对宽带信号短时频谱高频成分的频谱包络进行估计，结合低频频谱包络和采用适当频谱修补方法扩展后的频谱细节，最终得到带宽扩展后的直达声。

[0078] 为了更加详细的说明本发明的技术方案，给出实例1，图4是本发明实例1的立体声音频的带宽扩展方法的原理框图，如图4所示，一种立体声音频的带宽扩展方法包括以下步骤：

[0079] 1. 直达声/扩散声分离

[0080] 本文所提出的立体声扩展系统采用离散傅里叶变换或者正交镜像滤波器组将分帧后的左右声道音频信号各自转换到频域，并根据人耳听觉感知原理划分为多个均匀子带或临界频带。那么，输入立体声信号的短时频谱 $S(t, f)$ 可以表示为 $S(t, f) = [S_L(t, f) S_R(t, f)]^T$

[0081] 其中， t 和 f 分别表示信号的时间帧和子带序号； $S_L(t, f)$ 和 $S_R(t, f)$ 则分别表示立体声信号的左右声道短时频谱成分。

[0082] 为了有效地分离直达声和扩散声，系统还需要分别计算左右声道信号能量谱之间的和 P_{sum} 和差 P_{diff} 以及两个声道的互相关 P_{cc} 。

[0083] $P_{sum} = |S_L(t, f)|^2 + |S_R(t, f)|^2$

[0084] $P_{diff} = |S_L(t, f)|^2 - |S_R(t, f)|^2$

[0085] $P_{cc} = R\{S_L(t, f) S_R^*(t, f)\}$

[0086] 其中， $R\{\}$ 为取实部操作。为了改善分离算法的稳定性，分别对计算得到的 P_{sum} 、 P_{diff} 和 P_{cc} 进行时间平滑。

[0087] 立体声左右声道中的直达声成分之间高度相关,并可表示为由某一方向传播来的点声源信号。据此,本文所提系统利用一个直达声矩阵从原始立体声双声道信号 $S(t, f)$ 中直接分离出直达声成分 $S'(t, f)$,如下式所示,

$$[0088] \quad S'(t, f) = [S_L'(t, f) \ S_R'(t, f)]^T = M_D(t, f) [S_L(t, f) \ S_R(t, f)]^T = M_D(t, f) S(t, f)$$

[0089] 其中, $S_L'(t, f)$ 和 $S_R'(t, f)$ 分别表示直达声的左右声道短时频谱成分, $M_D(t, f)$ 为直达声矩阵。根据文献【M Vinton, D McGrath, C Robinson, P Brown, next generation surround decoding and upmixing for consumer and professional applications. AES 57th International conference, USA, 2015】所述,直达声矩阵 $M_D(t, f)$ 可以利用最小二乘方法获得,从而使得估计得到的直达声成分和真实成分之间的期望平方误差最小,即

$$[0090] \quad M_D(t, f) = \arg \min_{M_D(t, f)} \left\{ E \left[S'(t, f) - M_D(t, f) S(t, f) \right]^2 \right\}$$

[0091] 则直达声矩阵 $M_D(t, f)$ 可以由下式计算得到,

$$[0092] \quad M_D(t, f) = \begin{bmatrix} \frac{\left(\sqrt{|P_{cc}|^2 + |P_{diff}|^2} + P_{diff} \right)}{2P_{sum}} & \frac{P_{cc}}{2P_{sum}} \\ \frac{P_{cc}}{2P_{sum}} & \frac{\left(\sqrt{|P_{cc}|^2 + |P_{diff}|^2} - P_{diff} \right)}{2P_{sum}} \end{bmatrix}$$

[0093] 而扩散声成分 $S''(t, f)$ 则可以表示为原始立体声信号和直达声成分之差,

$$[0094] \quad S''(t, f) = S(t, f) - S'(t, f)$$

[0095] 2. 直达声成分的声源分离

[0096] 根据 $S'(t, f) = [S_L'(t, f) \ S_R'(t, f)]^T$ 利用公式3得到某一时频点上直达声 $S'(t, f)$ 的方向信息 $\theta(t, f)$,所述某一时频点上直达声 $S'(t, f)$ 的方向信息 $\theta(t, f)$ 与点声源的方向信息 θ_i 相同:

$$[0097] \quad \theta(t, f) = \arctan \left(\frac{|S_L'(t, f)|}{|S_R'(t, f)|} \right) = \theta_i, \quad i \in [1, 2, \dots, N] \text{ 公式 3;}$$

[0098] 对全部时频点的方向信息 $\theta(t, f)$ 进行聚类,得到方向信息的聚类中心 $C_i, i = 1, 2, \dots, N$;这些聚类中心分别对应各个点声源 $S_1(t, f), S_2(t, f), S_3(t, f) \dots S_N(t, f)$ 的方向信息 $\theta_1, \theta_2, \theta_3 \dots \theta_N$;

[0099] 根据某一时频点上直达声 $S'(t, f)$ 的方向信息 $\theta(t, f)$ 和聚类中心 C_i 得到掩蔽矩阵 $m_i(t, f)$;

$$[0100] \quad m_i(t, f) = \begin{cases} \begin{bmatrix} \cos \theta_i & \sin \theta_i \end{bmatrix} & |\theta(t, f) - C_i| \leq |\theta(t, f) - C_j|, j \neq i \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \text{other} \end{cases}$$

[0101] 利用所述掩蔽矩阵 $m_i(t, f)$ 根据公式4对直达声 $S'(t, f)$ 进行分离,得到直达点声源 $\hat{S}_i(t, f)$;

$$[0102] \quad \hat{S}_i(t, f) = m_i(t, f) S'(t, f) \text{ 公式 4.}$$

[0103] 3. 带宽扩展

[0104] 根据上文所述方法,分别从立体声信号中分离出扩散声成分 $S''(t, f)$ 和直达声成分 $S'(t, f)$,并利用时频稀疏性进一步将直达声成分 $S'(t, f)$ 分离成多个点声源 $\hat{S}_i(t, f)$, $i = 1, 2, \dots, N$ 。接下来可以根据单声道频带扩展方法分别对扩散声 $S''(t, f)$ 和直达点声源 $\hat{S}_i(t, f)$, $i = 1, 2, \dots, N$ 进行独立的带宽扩展。

[0105] 本文采用状态空间模型来直接拟合窄宽带频谱参数之间的映射关系,并在实际扩展中根据一定的误差准则对高频成分的频谱包络进行估计,

$$[0106] \quad S_Y(t, f) = F[S_X(t, f)]$$

[0107] 式中, $S_X(t, f)$ 和 $S_Y(t, f)$ 分别表示窄带和宽带信号的短时频谱, $F[\]$ 表示映射(或估计)函数。

[0108] 根据状态空间模型,映射函数 $F[\]$ 可以由状态演变函数 $F_{state}[\]$ 和观察函数 $F_{obs}[\]$ 两个过程来描述,如下式所示,

$$[0109] \quad S_{hidden}(t, f) = F_{state}[S_{hidden}(t-1, f), S_X(t-1, f), N_1(t, f)]$$

$$[0110] \quad S_Y(t, f) = F_{obs}[S_{hidden}(t, f), S_X(t, f), N_2(t, f)]$$

[0111] 其中, $S_{hidden}(t, f)$ 为模型中隐藏状态矢量, $N_1(t, f)$ 和 $N_2(t, f)$ 分别描述状态演变函数 F_{state} 和观测函数 F_{obs} 的误差。上述模型中,当前时刻的隐藏状态矢量 $S_{hidden}(t, f)$ 由前一时刻隐藏状态矢量 $S_{hidden}(t-1, f)$ 和前一时刻窄带信号的短时频谱 $S_X(t-1, f)$ 所决定,而当前时刻宽带信号短时频谱 $S_Y(t, f)$ 则进一步由当前时刻隐藏状态矢量 $S_{hidden}(t, f)$ 和当前时刻窄带信号的短时频谱 $S_X(t, f)$ 决定。利用状态空间模型中蕴含的隐藏状态递归结构能够更加精确地拟合窄宽带频谱参数之间的复杂映射关系,该模型可以采用广义卡尔曼滤波方法实现,也可以采用两个相互独立的深度神经网络来实现。基于深度神经网络的状态空间模型基本原理如图5所示。此处,状态演变函数 F_{state} 和观测函数 F_{obs} 可以采用堆栈自编码器、多层感知器、延时递归网络、长短时记忆网络等各种前向和递归深度神经网络实现。

[0112] 4. 立体声信号合成

[0113] 采用单声道频带扩展方法可以分别对扩散声 $S''(t, f)$ 和直达点声源 $\hat{S}_i(t, f)$, $i = 1, 2, \dots, N$ 进行扩展,从而得到相应的宽带频谱 $S_Y(t, f)$ 。接下来,可以利用各个点声源方向信息 θ_i 来重现宽带直达声 $\hat{S}'(t, f)$,

$$[0114] \quad \hat{S}'(t, f) = \begin{bmatrix} \cos \theta_1 & \cos \theta_2 & \cos \theta_3 & \dots & \cos \theta_N \\ \sin \theta_1 & \sin \theta_2 & \sin \theta_3 & \dots & \sin \theta_N \end{bmatrix} \begin{bmatrix} \hat{S}_1(t, f) \\ \hat{S}_2(t, f) \\ \hat{S}_3(t, f) \\ \vdots \\ \hat{S}_N(t, f) \end{bmatrix}$$

[0115] 其中, $\hat{S}_1(t, f)$, $\hat{S}_2(t, f)$, $\hat{S}_3(t, f)$, \dots , $\hat{S}_N(t, f)$ 为扩展后的点声源宽带频谱。 $\hat{S}'(t, f)$ 为扩展后宽带直达声的短时频谱。最终,结合扩展后的宽带扩散声 $\hat{S}''(t, f)$,可以实现宽带立体声信号 $\hat{S}(t, f)$ 的重现,

$$[0116] \quad \hat{S}(t, f) = \hat{S}'(t, f) + \hat{S}''(t, f)$$

[0117] 本发明实施例首先利用声道间的频谱相关性将输入立体声信号分解为直达声和扩散声两种成分,然后扩散声成分直接利用传统频带扩展方法进行扩展;直达声则依据不

同声源在时频结构上的稀疏性分离成多个不同方位的点声源,并分别进行带宽扩展,最终扩展后的点声源依照其在原始立体声中方位信息进行重新混合,并结合带宽扩展后的扩散声成分,重建出宽带立体声音频信号。本发明解决了现有技术中仅根据单个声道重建信号的主观质量实现对信号带宽的扩展,没有考虑到两个声道中信号能量和相位的相关性,其重建立体声信号严重影响了听者对声源位置和距离的判定的问题。

[0118] 以上所述仅为本发明的实施例而已,并不用于限制本发明,对于本领域的技术人员来说,本发明可以有各种更改和变化。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,均应包含在本发明的权利要求范围之内。

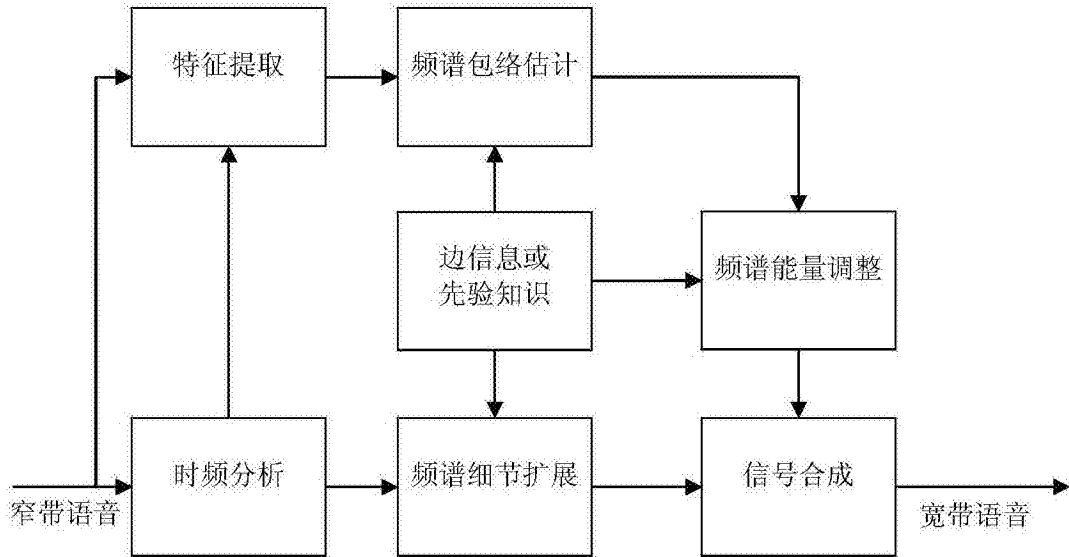


图1

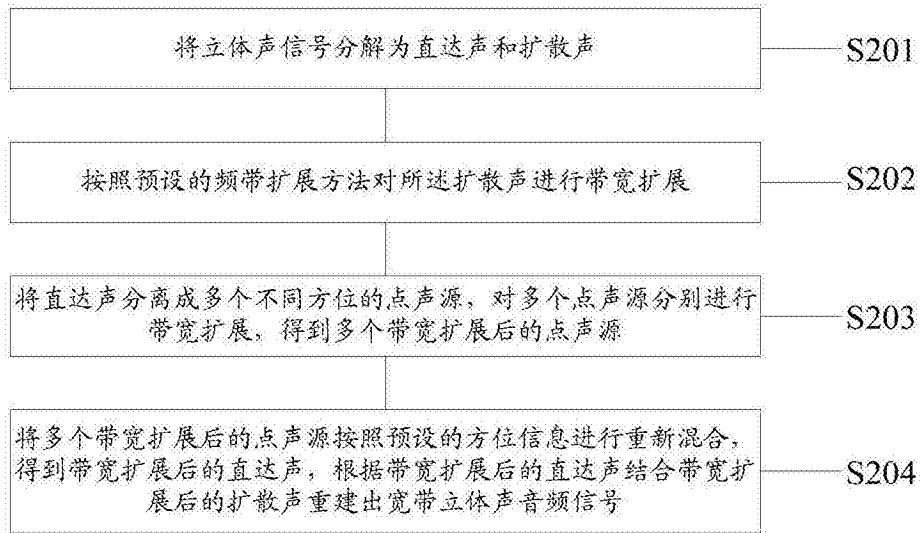


图2

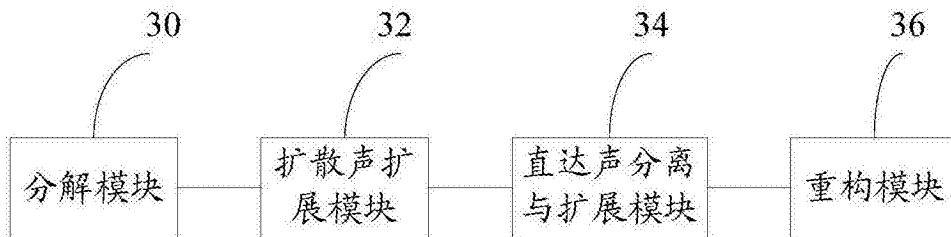


图3

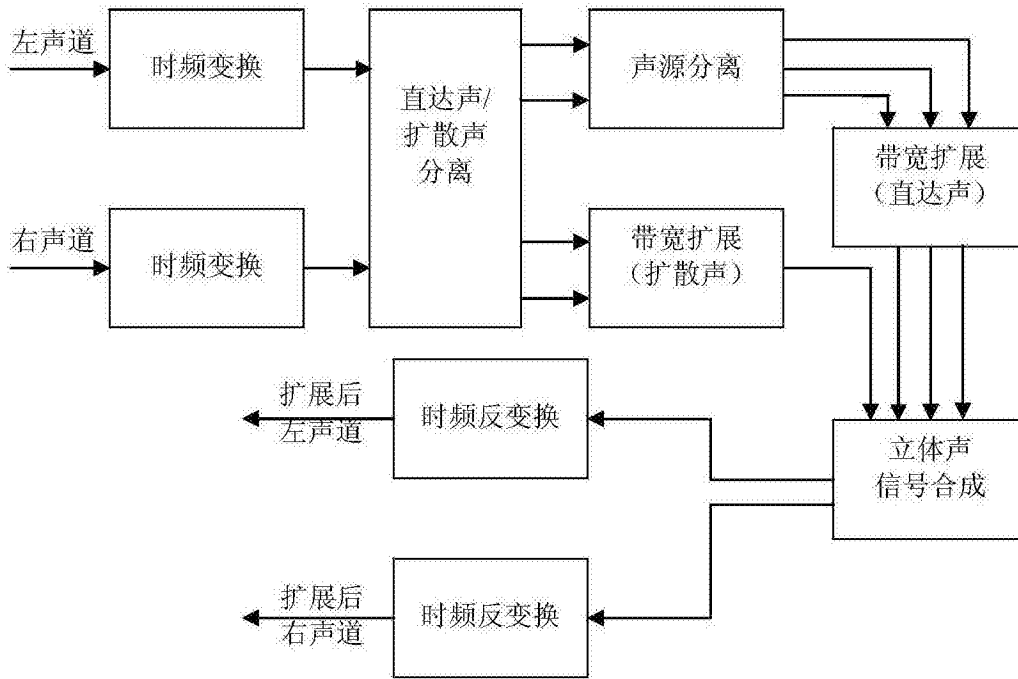


图4

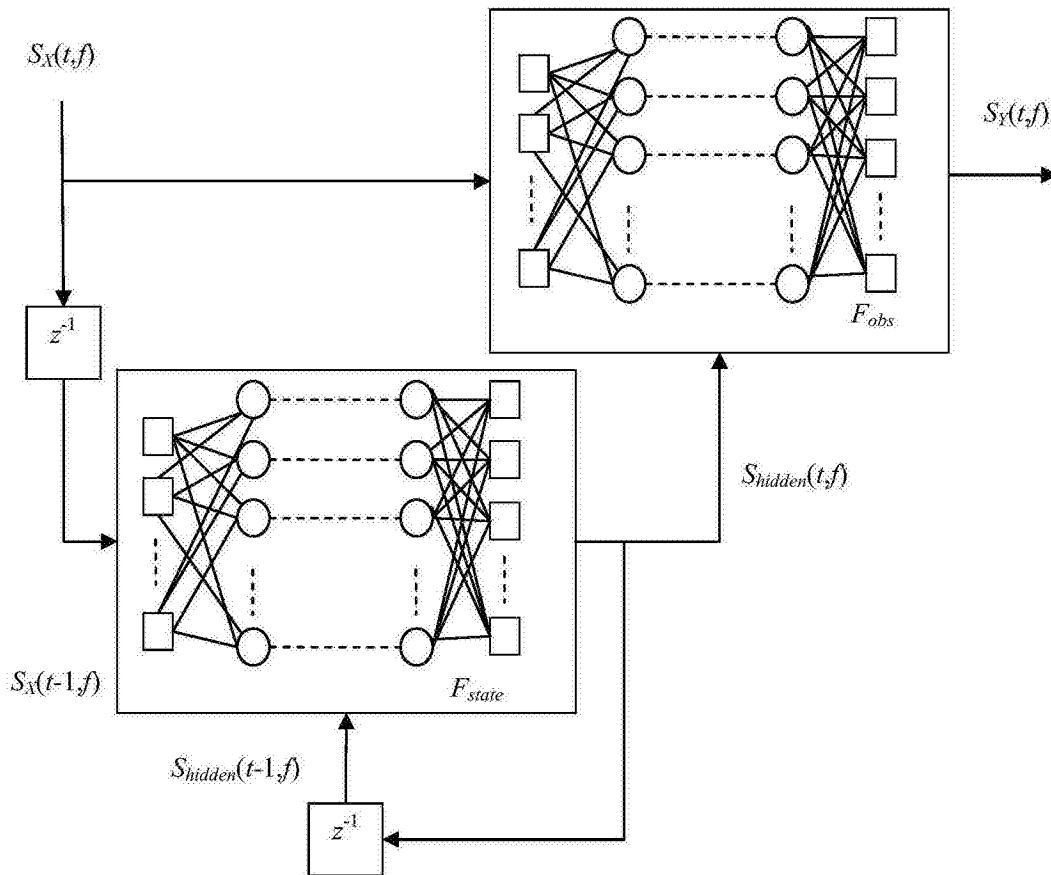


图5