



(12)发明专利申请

(10)申请公布号 CN 109448751 A

(43)申请公布日 2019.03.08

(21)申请号 201811646317.7

(22)申请日 2018.12.29

(71)申请人 中国科学院声学研究所

地址 100190 北京市海淀区北四环西路21号

(72)发明人 李军锋 孙兴伟 夏日升 颜永红

(74)专利代理机构 北京亿腾知识产权代理事务所 11309

代理人 陈霖

(51)Int.Cl.

G10L 21/0232(2013.01)

G10L 25/30(2013.01)

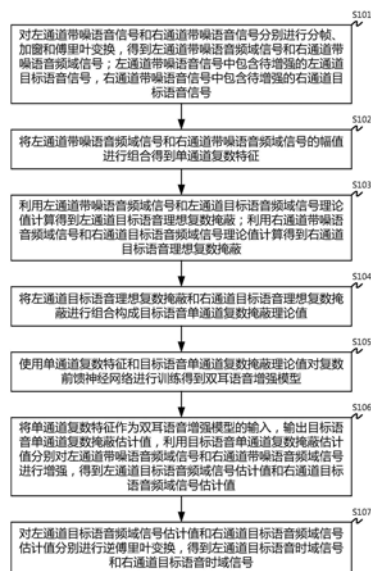
权利要求书2页 说明书4页 附图1页

(54)发明名称

一种基于深度学习的双耳语音增强方法

(57)摘要

本发明公开了一种基于深度学习的双耳语音增强方法,包括:对包含待增强目标语音信号的左/右通道带噪语音信号分别进行处理得到左/右频域信号,并对其幅值进行组合得到单通道复数特征,利用左/右通道的频域信号和对应的目标频域信号理论值分别计算出对应的目标语音理想复数掩蔽,将其组合构成目标语音单通道复数掩蔽理论值,并结合单通道复数特征对复数前馈神经网络进行训练得到双耳语音增强模型,利用模型输出的目标语音单通道复数掩蔽估计值分别处理左/右通道带噪语音信号得到左/右通道频域信号,最后得到对应的目标语音时域信号。本方法可以抑制噪声干扰并保持目标声源的空间信息。充分利用深度神经网络的泛化能力,达到双耳语音的增强。



1. 一种基于深度学习的双耳语音增强方法,其特征在于,包括步骤:

对左通道带噪语音信号和右通道带噪语音信号分别进行分帧、加窗和傅里叶变换,得到左通道带噪语音频域信号和右通道带噪语音频域信号;所述左通道带噪语音信号中包含待增强的左通道目标语音信号,所述右通道带噪语音信号中包含待增强的右通道目标语音信号;

将所述左通道带噪语音频域信号和右通道带噪语音频域信号的幅值进行组合得到单通道复数特征;

利用所述左通道带噪语音频域信号和左通道目标语音频域信号理论值计算得到左通道目标语音理想复数掩蔽;利用所述右通道带噪语音频域信号和右通道目标语音频域信号理论值计算得到右通道目标语音理想复数掩蔽;

将左通道目标语音理想复数掩蔽和右通道目标语音理想复数掩蔽进行组合构成目标语音单通道复数掩蔽理论值;

使用所述单通道复数特征和目标语音单通道复数掩蔽理论值对复数前馈神经网络进行训练得到双耳语音增强模型;

将所述单通道复数特征作为所述双耳语音增强模型的输入,输出目标语音单通道复数掩蔽估计值,利用所述目标语音单通道复数掩蔽估计值分别对左通道带噪语音频域信号和右通道带噪语音频域信号进行增强,得到左通道目标语音频域信号估计值和右通道目标语音频域信号估计值;

对所述左通道目标语音频域信号估计值和右通道目标语音频域信号估计值分别进行逆傅里叶变换,得到左通道目标语音时域信号和右通道目标语音时域信号。

2. 根据权利要求1所述的方法,其特征在于,所述对所述左通道带噪语音信号和右通道带噪语音信号进行分帧、加窗和傅里叶变换的步骤,具体为,

对所述左通道带噪语音信号和右通道带噪语音信号分别进行分帧和加窗处理,取1024个采样点作为一帧信号,若长度不足则先补零到1024点;然后对每一帧信号进行加窗,加窗函数采用汉明窗;最后对每一帧信号进行傅里叶变换。

3. 根据权利要求1所述的方法,其特征在于,所述单通道复数特征为:

$$X_C = |X_L| + j|X_R|$$

其中, j 为复数虚部单位, $|X_L|$ 为左通道带噪语音频域信号的幅值, $|X_R|$ 为右通道带噪语音频域信号的幅值。

4. 根据权利要求1所述的方法,其特征在于,

所述左通道目标语音理想复数掩蔽为:

$$M_L = \frac{S_L}{X_L} = \frac{S_{Lr} + jS_{Li}}{X_{Lr} + jX_{Li}} = \frac{X_{Lr}S_{Lr} + X_{Li}S_{Li}}{X_{Lr}^2 + X_{Li}^2} + j \frac{X_{Lr}S_{Li} - X_{Li}S_{Lr}}{X_{Lr}^2 + X_{Li}^2}$$

其中, j 为复数虚部单位, X_L 为复数,为左通道带噪语音频域信号, S_L 为复数,为左通道目标语音频域信号理论值, r 和 i 表示取复数的实部和虚部;

所述右通道目标语音理想复数掩蔽为:

$$M_R = \frac{S_R}{X_R} = \frac{S_{Rr} + jS_{Ri}}{X_{Rr} + jX_{Ri}} = \frac{X_{Rr}S_{Rr} + X_{Ri}S_{Ri}}{X_{Rr}^2 + X_{Ri}^2} + j \frac{X_{Rr}S_{Ri} - X_{Ri}S_{Rr}}{X_{Rr}^2 + X_{Ri}^2}$$

其中, j 为复数虚部单位, X_R 为复数, 为右通道带噪语音频域信号, S_R 为复数, 为右通道目标语音频域信号理论值, r 和 i 表示取复数的实部和虚部。

5. 根据权利要求1或4任一权利要求所述的方法, 其特征在于, 所述目标语音单通道复数掩蔽理论值为:

$$M_C = M_L + jM_R$$

其中, j 为复数虚部单位, M_L 为左通道目标语音理想复数掩蔽, M_R 为右通道目标语音理想复数掩蔽。

6. 根据权利要求1所述的方法, 其特征在于, 所述使用所述单通道复数特征和目标语音单通道复数掩蔽理论值对复数前馈神经网络进行训练得到双耳语音增强模型的步骤, 具体为,

所述复数前馈神经网络为一个4层的全连接神经网络, 网络中每层有1024个隐层复数节点。每个神经元的激活函数使用的是线性修正单元, 并且分别作用在复数节点的实部和虚部上, 其表达式为 $f(x) = \max(0, x)$ 。

将所述单通道复数特征进行前后帧扩展, 得到单通道复数扩展特征, 并作为所述复数前馈神经网络的输入, 输出目标语音单通道复数掩蔽估计值, 将目标语音单通道复数掩蔽理论值作为所述复数前馈神经网络的训练目标, 通过迭代不断使目标语音单通道复数掩蔽估计值与目标语音单通道复数掩蔽理论值的均方误差减小。

7. 根据权利要求1所述的方法, 其特征在于, 所述单通道复数掩蔽估计值为:

$$M_C' = M_L' + jM_R'$$

其中, j 为复数虚部单位, M_L' 为左通道目标语音理想复数掩蔽的估计值, M_R' 为右通道目标语音理想复数掩蔽的估计值。

8. 根据权利要求1或7所述的方法, 其特征在于,

所述左通道目标语音频域信号估计值:

$$X'_L = M'_L * X_L$$

其中, M'_L 为左通道目标语音理想复数掩蔽的估计值, X_L 为左通道带噪语音频域信号;

所述右通道目标语音频域信号估计值:

$$X'_R = M'_R * X_R$$

其中, M'_R 为右通道目标语音理想复数掩蔽的估计值, X_R 为右通道带噪语音频域信号。

一种基于深度学习的双耳语音增强方法

技术领域

[0001] 本发明涉及语音增强技术领域,尤其涉及一种基于深度学习的双耳语音增强方法。

背景技术

[0002] 目前,语音增强技术主要是去除语音信号中背景噪声以及指向性噪声干扰,提高语音的质量和可懂度,从而在语音识别和人耳理解方面取得更好的表现。在以单通道语音为输出的增强技术中,通过利用单通道输入时频域中语音和噪声的不同特性可以抑制背景噪声,利用多通道输入中目标语音和干扰信号的空间信息,可以更好的去除指向性噪声。而在双耳听觉中,人耳可以利用双通道语音中目标与干扰信号的空间信息差异提高对语音的理解度,并可以利用目标声源的空间信息进行定位。在大多数传统的以双通道为输出的语音增强中,大多只考虑了去除干扰,对于目标语音的空间信息保持没有特殊处理,并且对于非平稳的噪声的抑制效果欠佳。

发明内容

[0003] 本发明的目的在于解决现有技术存在的缺陷。

[0004] 为达到上述目的,本发明公开了一种基于深度学习的双耳语音增强方法,包括步骤:

[0005] 对左通道带噪语音信号和右通道带噪语音信号分别进行分帧、加窗和傅里叶变换,得到左通道带噪语音频域信号和右通道带噪语音频域信号;左通道带噪语音信号中包含待增强的左通道目标语音信号,右通道带噪语音信号中包含待增强的右通道目标语音信号;

[0006] 将左通道带噪语音频域信号和右通道带噪语音频域信号的幅值进行组合得到单通道复数特征;

[0007] 利用左通道带噪语音频域信号和左通道目标语音频域信号理论值计算得到左通道目标语音理想复数掩蔽;利用右通道带噪语音频域信号和右通道目标语音频域信号理论值计算得到右通道目标语音理想复数掩蔽;

[0008] 将左通道目标语音理想复数掩蔽和右通道目标语音理想复数掩蔽进行组合构成目标语音单通道复数掩蔽理论值;

[0009] 使用单通道复数特征和目标语音单通道复数掩蔽理论值对复数前馈神经网络进行训练得到双耳语音增强模型;

[0010] 将单通道复数特征作为双耳语音增强模型的输入,输出目标语音单通道复数掩蔽估计值,利用目标语音单通道复数掩蔽估计值分别对左通道带噪语音频域信号和右通道带噪语音频域信号进行增强,得到左通道目标语音频域信号估计值和右通道目标语音频域信号估计值;

[0011] 对左通道目标语音频域信号估计值和右通道目标语音频域信号估计值分别进行

逆傅里叶变换,得到左通道目标语音时域信号和右通道目标语音时域信号。

[0012] 优选地,对左通道带噪语音信号和右通道带噪语音信号进行分帧、加窗和傅里叶变换的步骤,具体为,

[0013] 对左通道带噪语音信号和右通道带噪语音信号分别进行分帧和加窗处理,取1024个采样点作为一帧信号,若长度不足则先补零到1024点;然后对每一帧信号进行加窗,加窗函数采用汉明窗;最后对每一帧信号进行傅里叶变换。

[0014] 优选地,单通道复数特征 $X_C = |X_L| + j|X_R|$,其中, j 为复数虚部单位, $|X_L|$ 为左通道带噪语音频域信号的幅值, $|X_R|$ 为右通道带噪语音频域信号的幅值。

[0015] 优选地,左通道目标语音理想复数掩蔽为:

$$[0016] \quad M_L = \frac{S_L}{X_L} = \frac{S_{Lr} + jS_{Li}}{X_{Lr} + jX_{Li}} = \frac{X_{Lr}S_{Lr} + X_{Li}S_{Li}}{X_{Lr}^2 + X_{Li}^2} + j \frac{X_{Lr}S_{Li} - X_{Li}S_{Lr}}{X_{Lr}^2 + X_{Li}^2}$$

[0017] 其中, j 为复数虚部单位, X_L 为复数,为左通道带噪语音频域信号, S_L 为复数,为左通道目标语音频域信号理论值, r 和 i 表示取复数的实部和虚部;

[0018] 优选地,右通道目标语音理想复数掩蔽为:

$$[0019] \quad M_R = \frac{S_R}{X_R} = \frac{S_{Rr} + jS_{Ri}}{X_{Rr} + jX_{Ri}} = \frac{X_{Rr}S_{Rr} + X_{Ri}S_{Ri}}{X_{Rr}^2 + X_{Ri}^2} + j \frac{X_{Rr}S_{Ri} - X_{Ri}S_{Rr}}{X_{Rr}^2 + X_{Ri}^2}$$

[0020] 其中, j 为复数虚部单位, X_R 为复数,为右通道带噪语音频域信号, S_R 为复数,为右通道目标语音频域信号理论值, r 和 i 表示取复数的实部和虚部。

[0021] 优选地,目标语音单通道复数掩蔽理论值 $M_C = M_L + jM_R$,其中, j 为复数虚部单位, M_L 为左通道目标语音理想复数掩蔽, M_R 为右通道目标语音理想复数掩蔽。

[0022] 优选地,使用单通道复数特征和目标语音单通道复数掩蔽理论值对复数前馈神经网络进行训练得到双耳语音增强模型的步骤,具体为,

[0023] 复数前馈神经网络为一个4层的全连接神经网络,网络中每层有1024个隐层复数节点。每个神经元的激活函数使用的是线性修正单元,并且分别作用在复数节点的实部和虚部上,其表达式为 $f(x) = \max(0, x)$ 。

[0024] 将单通道复数特征进行前后帧扩展,得到单通道复数扩展特征,并作为复数前馈神经网络的输入,输出目标语音单通道复数掩蔽估计值,将目标语音单通道复数掩蔽理论值作为复数前馈神经网络的训练目标,通过迭代不断使目标语音单通道复数掩蔽估计值与目标语音单通道复数掩蔽理论值的均方误差减小。

[0025] 优选地,单通道复数掩蔽估计值 $M_C' = M_L' + jM_R'$,其中, j 为复数虚部单位, M_L' 为左通道目标语音理想复数掩蔽的估计值, M_R' 为右通道目标语音理想复数掩蔽的估计值。

[0026] 优选地,左通道目标语音频域信号估计值 $X'_L = M'_L * X_L$,其中, M'_L 为左通道目标语音理想复数掩蔽的估计值, X_L 为左通道带噪语音频域信号;

[0027] 优选地,右通道目标语音频域信号估计值 $X'_R = M'_R * X_R$,其中, M'_R 为右通道目标语音理想复数掩蔽的估计值, X_R 为右通道带噪语音频域信号。

[0028] 本发明的优点在于:通过利用左通道和右通道的理想复数掩蔽构造单通道复数掩蔽,并通过复数前馈神经网络对其进行估计,达到对左通道和右通道联合处理的目的,进而在抑制噪声干扰的同时保持目标声源的空间信息。通过在训练数据中包含足够多的噪声类

型和方位,可以充分利用深度神经网络的泛化能力,提高模型的鲁棒性,达到双耳语音增强的目的。

附图说明

[0029] 为了更清楚说明本发明实施例的技术方案,下面将对实施例描述中所需使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0030] 图1为一种基于深度学习的双耳语音增强方法的流程图。

具体实施方式

[0031] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0032] 图1为一种基于深度学习的双耳语音增强方法的流程图。如图1所示,包括:

[0033] 步骤S101:对左通道带噪语音信号和右通道带噪语音信号分别进行分帧、加窗和傅里叶变换,得到左通道带噪语音频域信号和右通道带噪语音频域信号。

[0034] 左通道带噪语音信号中包含待增强的左通道目标语音信号,右通道带噪语音信号中包含待增强的右通道目标语音信号。

[0035] 在一个具体实施例中,对左通道带噪语音信号和右通道带噪语音信号分别进行分帧和加窗处理,取1024个采样点作为一帧信号,若长度不足则先补零到1024点;然后对每一帧信号进行加窗,加窗函数采用汉明窗;最后对每一帧信号进行傅里叶变换,得到左通道带噪语音频域信号和右通道带噪语音频域信号。

[0036] 步骤S102:将左通道带噪语音频域信号和右通道带噪语音频域信号的幅值进行组合得到单通道复数特征。

[0037] 具体地,单通道复数特征 $X_C = |X_L| + j|X_R|$,其中, j 为复数虚部单位, $|X_L|$ 为左通道带噪语音频域信号的幅值, $|X_R|$ 为右通道带噪语音频域信号的幅值。

[0038] 步骤S103:利用左通道带噪语音频域信号和左通道目标语音频域信号理论值计算得到左通道目标语音理想复数掩蔽;利用右通道带噪语音频域信号和右通道目标语音频域信号理论值计算得到右通道目标语音理想复数掩蔽。

[0039] 具体地,左通道目标语音理想复数掩蔽为:

$$[0040] \quad M_L = \frac{S_L}{X_L} = \frac{S_{Lr} + jS_{Li}}{X_{Lr} + jX_{Li}} = \frac{X_{Lr}S_{Lr} + X_{Li}S_{Li}}{X_{Lr}^2 + X_{Li}^2} + j \frac{X_{Lr}S_{Li} - X_{Li}S_{Lr}}{X_{Lr}^2 + X_{Li}^2}$$

[0041] 其中, j 为复数虚部单位, X_L 为复数,为左通道带噪语音频域信号, S_L 为复数,为左通道目标语音频域信号理论值, r 和 i 表示取复数的实部和虚部。

[0042] 右通道目标语音理想复数掩蔽为:

$$[0043] \quad M_R = \frac{S_R}{X_R} = \frac{S_{Rr} + jS_{Ri}}{X_{Rr} + jX_{Ri}} = \frac{X_{Rr}S_{Rr} + X_{Ri}S_{Ri}}{X_{Rr}^2 + X_{Ri}^2} + j \frac{X_{Rr}S_{Ri} - X_{Ri}S_{Rr}}{X_{Rr}^2 + X_{Ri}^2}$$

[0044] 其中, j 为复数虚部单位, X_R 为复数, 为右通道带噪语音频域信号, S_R 为复数, 为右通道目标语音频域信号理论值, r 和 i 表示取复数的实部和虚部。

[0045] 步骤S104: 将左通道目标语音理想复数掩蔽和右通道目标语音理想复数掩蔽进行组合构成目标语音单通道复数掩蔽理论值。

[0046] 具体地, 目标语音单通道复数掩蔽理论值 $M_C = M_L + jM_R$, 其中, j 为复数虚部单位, M_L 为左通道目标语音理想复数掩蔽, M_R 为右通道目标语音理想复数掩蔽。

[0047] 步骤S105: 使用单通道复数特征和目标语音单通道复数掩蔽理论值对复数前馈神经网络进行训练得到双耳语音增强模型。

[0048] 在一个具体实施例中, 复数前馈神经网络为一个4层的全连接神经网络, 网络中每层有1024个隐层复数节点。每个神经元的激活函数使用的是线性修正单元, 并且分别作用在复数节点的实部和虚部上, 其表达式为 $f(x) = \max(0, x)$ 。

[0049] 将单通道复数特征进行前后帧扩展, 得到单通道复数扩展特征, 并作为复数前馈神经网络的输入, 输出目标语音单通道复数掩蔽估计值, 将目标语音单通道复数掩蔽理论值作为复数前馈神经网络的训练目标, 通过迭代不断使目标语音单通道复数掩蔽估计值与目标语音单通道复数掩蔽理论值的均方误差减小。

[0050] 步骤S106: 将单通道复数特征作为双耳语音增强模型的输入, 输出目标语音单通道复数掩蔽估计值, 利用目标语音单通道复数掩蔽估计值分别对左通道带噪语音频域信号和右通道带噪语音频域信号进行增强, 得到左通道目标语音频域信号估计值和右通道目标语音频域信号估计值。

[0051] 具体地, 单通道复数掩蔽估计值 $M_C' = M_L' + jM_R'$, 其中, j 为复数虚部单位, M_L' 为左通道目标语音理想复数掩蔽的估计值, M_R' 为右通道目标语音理想复数掩蔽的估计值。

[0052] 左通道目标语音频域信号估计值 $X'_L = M'_L * X_L$, 其中, M'_L 为左通道目标语音理想复数掩蔽的估计值, X_L 为左通道带噪语音频域信号。

[0053] 右通道目标语音频域信号估计值 $X'_R = M'_R * X_R$, 其中, M'_R 为右通道目标语音理想复数掩蔽的估计值, X_R 为右通道带噪语音频域信号。

[0054] 步骤S107: 对左通道目标语音频域信号估计值和右通道目标语音频域信号估计值分别进行逆傅里叶变换, 得到左通道目标语音时域信号和右通道目标语音时域信号。

[0055] 本发明提供了一种基于深度学习的双耳语音增强方法, 通过利用左通道和右通道的理想复数掩蔽构造单通道复数掩蔽, 并通过复数前馈神经网络对其进行估计, 达到对左通道和右通道联合处理的目的, 进而在抑制噪声干扰的同时保持目标声源的空间信息。通过在训练数据中包含足够多的噪声类型和方位, 可以充分利用深度神经网络的泛化能力, 提高模型的鲁棒性, 达到双耳语音增强的目的。

[0056] 以上的具体实施方式, 对本发明的目的、技术方案和有益效果进行了进一步详细说明, 所应理解的是, 以上仅为本发明的具体实施方式而已, 并不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, 所做的任何修改、等同替换、改进等, 均应包含在本发明的保护范围之内。

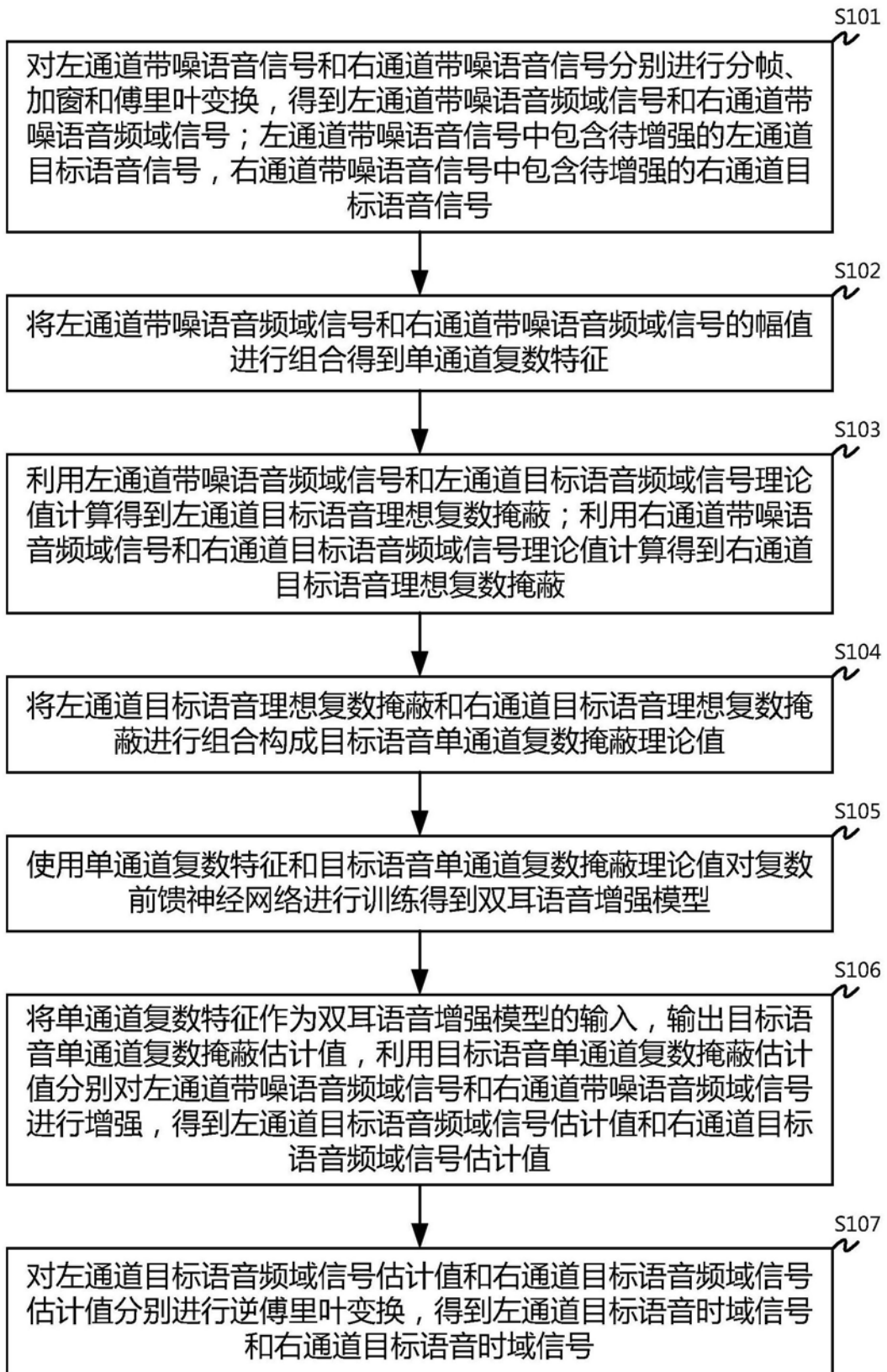


图1