

基于双耳线索编码原理的语音增强方法

陈楠, 鲍长春

(北京工业大学信息学部, 北京 100124)

摘要: 借助双耳线索编码原理,通过构建一个语音和噪声的双耳线索先验码书,本文提出一种单通道语音增强方法. 首先,该算法将语音和噪声的双耳线索作为语音和噪声的先验知识,在线下被训练成为先验码书. 之后,在线上通过加权码书映射(Weighted CodeBook Mapping, WCBM)算法估计纯净线索参数,最后,利用双耳线索编码原理增强含噪语音. 此外,本文采用深度神经网络,即堆栈式自编码器(Stacked Auto-Encoders, SAE)代替 WCBM 算法估计纯净线索参数,提出了基于深度神经网络的双耳线索语音增强算法. 进一步提高了增强算法的性能. 客观测试结果表明,本文所提方法优于参考算法.

关键词: 语音增强; 双耳线索编码; 码书驱动; 深度神经网络

中图分类号: TN912.35 **文献标识码:** A **文章编号:** 0372-2112 (2019)01-0227-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2019.01.030

Speech Enhancement Method Based on Binaural Cues Coding Principle

CHEN Nan, BAO Chang-chun

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: In this paper, a single channel speech enhancement method is proposed by constructing a priori binaural cue codebook of speech and noise based on binaural cue coding principle. Firstly, as a priori information, the binaural cues of speech and noise are offline trained to form a priori codebook. Then, the weighted codebook mapping (WCBM) algorithm is used to estimate the clean cue. At last, the noisy speech is enhanced with binaural cue coding (BCC) model. Moreover, an estimation method of the clean cue is proposed for further improving performance based on deep neural network, namely stacked auto-encoders (SAE), instead of WCBM algorithm. Objective test results show that the proposed method is superior to the reference methods.

Key words: speech enhancement; binaural cue coding; codebook driven; deep neural network

1 引言

简单实用的单通道语音增强技术经过四十多年的发展,已经形成了很多经典的算法. 例如谱减法^[1], 维纳滤波法^[2]等等. 在平稳噪声环境中,这些传统的语音增强算法可以获得较好的增强效果. 但是,由于这类传统的语音增强算法没有考虑到语音和噪声的先验信息,其算法性能在非平稳噪声环境中并不理想. 为了解决这一问题, Kleijin 等人提出一种基于码书驱动的语音增强算法^[3,4]. 首先,该算法将语音和噪声的自回归 (Auto-Regressive, AR) 系数作为语音和噪声的先验信息^[5],并在线下将其训练成为先验码书. 之后,通过最

小化真实含噪语音和重构含噪语音间的误差,线上估计语音和噪声的 AR 增益以及其各自的 AR 系数. 最后,根据估计得到的语音和噪声的 AR 模型参数(包括 AR 系数和 AR 增益)构建维纳滤波器来实现语音增强. 该算法在非平稳噪声环境中可以拥有较好的性能,但是仍有一些不足之处. 首先,该算法通过对谱形状建模实现语音增强,忽略了谱细节的作用. 导致该增强算法不能抑制谐波间噪声. 此外,由于 AR 系数与噪声类型相关性较强,致使噪声预分类成为该增强算法必不可少的一步^[6].

双耳线索编码 (Binaural Cue Coding, BCC)^[7,8] 是一种双声道语音编码技术. 该技术在 BCC 编码模块提取

输入声道语音信号的边缘信息. 在 BCC 合成模块通过结合提取的边缘信息, 将输入声道语音信号经过“下混合”得到的混合语音信号恢复成原始输入语音信号. 当 BCC 模型中的双输入信号分别是纯净语音信号和噪声信号时, 其“下混合”后的混合语音信号便是含噪语音. 此时, 通过提取纯净语音和噪声的边缘信息, BCC 模型可以将纯净语音从含噪语音中分离出来, 从而达到语音增强的目的.

本文将 BCC 模型引入到语音增强算法中, 提出了一种基于码书的双耳线索单通道语音增强算法. 该算法将双耳线索引入到单通道语音增强算法中, 利用纯净语音和噪声的双耳线索作为其先验知识进行训练, 并结合 BCC 模型对谱细节建模, 实现语音增强. 本算法中的双耳信号分别为纯净语音和噪声, 并不是真正双耳接收的信号, 因此, 回声的影响未做考虑. 因为作为先验知识的双耳线索与噪声类型的相关性较小, 所以, 所提算法避免了传统码书驱动算法的噪声预分类问题. 同时, 所提算法通过对谱细节建模增强含噪语音, 有效抑制了谐波间噪声. 此外, 为了得到更好的纯净线索估计, 本文采用堆栈式自编码器 (Stacked Auto-Encoders, SAE) 替代加权码书映射 (Weighted CodeBook Mapping, WCBM) 算法^[9]来估计纯净线索参数, 进一步提高了算法性能.

2 BCC 模型

BCC 是一种双声道语音编码技术^[7]. 因为语音编码的目的是用尽可能少的比特数恢复原始输入语音信号, 所以, BCC 通过提取输入声道的边缘信息减少传输的比特数. 其原理框图如图 1 所示.

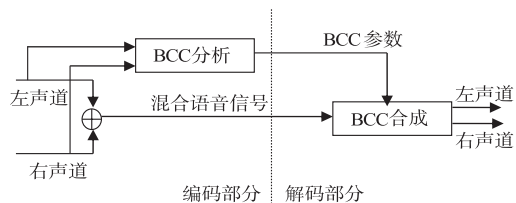


图1 BCC模型原理框图

如图 1 所示, BCC 模型由编码和解码两部分组成. 在编码端, BCC 分析模块提取输入声道的双耳线索参数作为 BCC 参数, 即边缘信息. 在解码端, BCC 合成模块结合 BCC 参数将混合语音信号恢复成为原始的输入语音信号. 其恢复过程如式 (1) 所示.

$$S_{c,k} = F_{c,k} G_{c,k} S_k \quad (1)$$

式 (1) 中, S_k 为混合语音信号频谱, $F_{c,k}$ 为声级差修改值, $G_{c,k}$ 为相位差修改值, $S_{c,k}$ 为恢复的第 c 输入声道的语音信号频谱. 其中, 声级差修改值和相位差修改值是

由提取的 BCC 参数计算求得. 由于相位对于语音可懂度的影响较小, 本文将不对相位差修改值进行介绍.

声级差修改值是由双耳线索参数声道间声级差 (Inter-Channel Level Difference, ICLD) 和声道间互相关 (Inter-Channel Correlation, ICC) 计算得到的^[8], 即

$$F_{c,k} = 10^{(\Delta L_{c,b} + r_{c,k})/20} F_{1,k} \quad (2)$$

其中

$$F_{1,k} = \frac{1}{\sqrt{1 + \sum_{i=1}^{C-1} 10^{(\Delta L_{c,b} + r_{c,k})/10}}} \quad (3)$$

式中, $\Delta L_{c,b}$ 表示第 c 声道、第 b 子带的 ICLD, 其子带分割按照等距带宽的两倍进行分割; $r_{c,k}$ 表示由 ICC 控制的一个随机变量.

3 基于先验码书的双耳线索语音增强算法

将 BCC 模型以及先验码书应用到语音增强中, 本文提出了一种基于先验码书的双耳线索语音增强算法. 图 2 给出了该算法的算法框图.

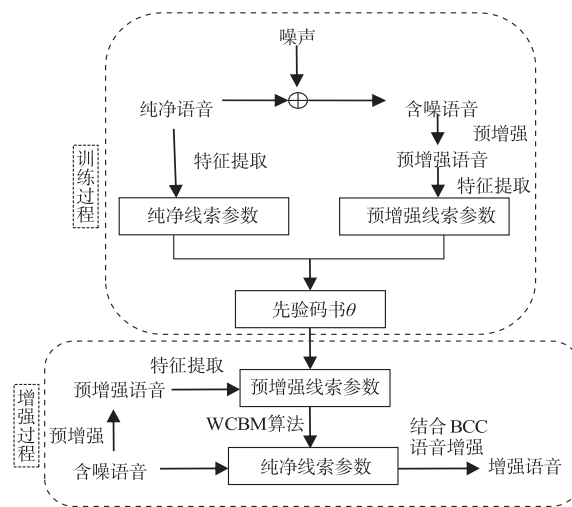


图2 基于双耳线索码书的语音增强算法原理框图

如图 2 所示, 所提算法包括线下训练过程和线上增强过程两部分. 在线下训练部分, 纯净语音被噪声污染成为含噪语音, 之后, 含噪语音通过预处理后得到与纯净语音相对应的预增强语音. 接着, 分别从纯净语音和与其对应的预增强语音中提取纯净线索参数 θ_x 和预增强线索参数 θ_y . 最后, 将提取的线索参数采用 LBG 算法将其训练成为一个联合的先验码书 $\theta(\theta_x, \theta_y)$. 在线上增强部分, 待增强的含噪语音首先进行预处理得到预增强语音, 之后, 从预增强语音中提取线上预增强线索参数 θ_{oy} , 接着, 结合已知的先验码书 $\theta(\theta_x, \theta_y)$ 和线上预增强线索参数 θ_{oy} , 通过 WCBM 算法估计纯净线索参数 $\hat{\theta}_x$. 最后, 结合 BCC 模型, 通过估计的纯净线索参数增强含噪语音. 在所提算法中, 有三个关键的部分: 线索参数定

义、应用 WCBM 估计纯净线索参数和 BCC 模型增强语音。下面,对这三个要点一一进行介绍。

3.1 线索参数的定义

在 BCC 模型中,双耳线索参数通常用于重构原始输入语音信号。本文的基本思想是当将纯净语音和噪声作为 BCC 模型的输入信号时,纯净语音和噪声的双耳线索可以作为线索参数用来增强含噪语音。但是,在线上增强过程中,纯净语音和噪声的双耳线索不能直接得到,只能通过预增强语音和噪声的双耳线索间接获得。因此,本文定义的线索参数包含两部分:纯净线索参数和预增强线索参数。两种线索参数是一一对应的。本文利用该对应关系通过预增强线索参数估计纯净线索参数。下面,对纯净线索参数和预增强线索参数进行定义。

纯净线索参数包含两个参数,即纯净语音和噪声声级差(Speech and Noise Level Difference, SNLD)和纯净语音和噪声互相关(Speech and Noise Correlation, SNC)。其计算公式分别为

$$SNLD(t, k) = 10 \log_{10} \frac{X_t(k) X_t^*(k)}{W_t(k) W_t^*(k)} \quad (4)$$

$$SNC = \frac{\left| \sum_{k=A_b}^{A_{b+1}-1} X_t(k) W_t^*(k) \right|}{\sqrt{\left(\sum_{k=A_b}^{A_{b+1}-1} X_t(k) X_t^*(k) \right) \left(\sum_{k=A_b}^{A_{b+1}-1} W_t(k) W_t^*(k) \right)}} \quad (5)$$

式中,符号“*”表示共轭, t 表示帧索引, k 表示频率索引, $X_t(k)$ 为纯净语音信号频谱, $W_t(k)$ 为估计的噪声频谱, A_b 为第 b 子带所对应的频点下界限。

预增强线索包含四个参数:预增强语音和噪声声级差(Pre-enhanced speech and Noise Level Difference, PNLD),预增强语音和噪声互相关(Pre-enhanced speech and Noise Correlation, PNC),后验信噪比(Posterior SNR, PSNR)以及语音存在概率(Speech Presence Probability, SPP)。其中,PNLD 和 PNC 表示预增强语音和噪声的双耳线索参数,其计算方法同式(4)和式(5),只需将纯净语音频谱更改为预增强语音的频谱即可。PSNR 和 SPP 这两个参数是在预处理过程中得到的,本文将它们作为预增强线索参数用来增加预增强线索参数和纯净线索参数的匹配度。其中,PSNR 的计算公式为

$$PSNR(t, k) = 10 \log_{10} \left(\frac{Y(t, k)^2}{\lambda_w(t, k)} \right) \quad (6)$$

式中, $Y(t, k)$ 表示含噪语音信号的 DFT 谱, $\lambda_w(t, k)$ 表示估计的噪声信号的方差。

SPP 的计算过程^[10]为

$$\hat{S}_{pp}(t, k) = \alpha_p \hat{S}_{pp}(t-1, k) + (1 - \alpha_p) \hat{S}_{pp}(t, k) \quad (7)$$

其中

$$S_{pp}(t, k) = \begin{cases} 1, & \text{if } \frac{S_f(t, k)}{S_{\min}(t, k)} \geq \eta \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$S_f(t, k) = \alpha_s S_f(t-1, k) + (1 - \alpha_s) \sum_{i=-2}^2 w(i) |Y_t(k-i)|^2 \quad (9)$$

式中, η 表示语音是否存在的决策阈值, $S_{\min}(t, k)$ 表示搜索窗内 $S_f(t, k)$ 的最小值, α_s 和 α_p 为平滑因子, $w(i)$ 是长度为 5 的归一化汉明窗。

3.2 基于 WCBM 的纯净线索估计

在估计纯净线索过程中,首先,在线下将提取的纯净线索参数和预增强线索参数训练成联合码书,之后,线上提取预增强线索参数,已知线上预增强线索参数和先验码书,本文通过 WCBM 算法^[9]估计纯净线索参数。其框图如图 3 所示。

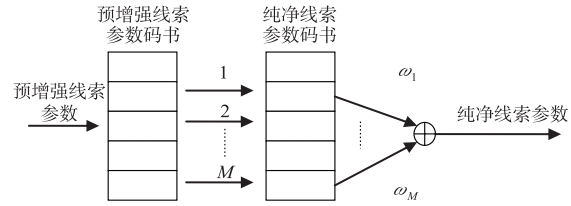


图3 基于WCBM的纯净线索估计原理框图

如图 3 所示,首先,已知线上预增强线索参数 θ_{oy} 和先验码书 $\theta(\theta_x, \theta_y)$, 计算 θ_{oy} 和先验码书中的所有码矢量中存储的预增强线索参数 θ_y 的欧式失真距离(Euclidean Distance, ED)。即

$$d_i = \text{sqrt} \left(\sum_{j=1}^4 (\theta_{oy}(j) - \theta_y(j))^2 \right) \quad \forall i \in [1, N] \quad (10)$$

式中, N 表示训练码书的大小,本文中 $N = 256$; d_i 表示第 i 个码矢量的 ED。

之后,根据 ED 最小原则,从先验码书中选取 M 个失真距离最小的码矢量 $(\theta_1, \theta_2, \dots, \theta_M)$, 其对应的 ED 为 (d_1, d_2, \dots, d_M) 。

接着,根据选取码矢量的 ED, 计算选取的 M 个码矢量的隶属度

$$\rho_i = \left\{ \sum_{j=1}^M \left(\frac{d_i}{d_j} \right)^2 \right\}^{-1} \quad \forall i \in [1, M] \quad (11)$$

式中, M 表示选取的码矢量的个数,本文中 $M = 20$; 已知隶属度,可计算选取的 M 个码矢量的权值

$$\omega_i = \frac{\rho_i}{\sum_{j=1}^M \rho_j} \quad \forall i \in [1, M] \quad (12)$$

最后,利用纯净线索参数和预增强线索参数的一一对应关系,根据预增强线索参数计算的权值对选取的 M 个码矢量中存储的纯净线索参数加权,获取纯净

线索参数估计 $\hat{\theta}_x = [SNLD, SNC]$. 其公式为

$$\hat{\theta}_x = \sum_{i=1}^M \omega_i \theta_{x,i} \quad (13)$$

式中, $\theta_{x,i}$ 为第 i 个码矢量中存储的纯净线索参数.

3.3 结合 BCC 模型实现语音增强

得到纯净线索参数的估计后, 根据第二章给出的式(1)~(3), 可以得到实现语音增强的增益函数

$$G_i(k) = \frac{10^{SNLD(t,k)/20}}{\sqrt{1 + 10^{SNLD(t,k)/10}}} \quad (14)$$

其中

$$SNLD'(t,k) = SNLD(t,k) + [1 - SNC(t,k)] \times \tau(t,k) \quad (15)$$

式中, $\tau(t,k)$ 是均值为 0、方差为 1 的随机函数.

基于增益函数, 可以通过式(16)得到增强语音的频谱

$$\hat{X}_i(k) = G_i(k) Y_i(k) \quad (16)$$

其中, $Y_i(k)$ 为含噪语音的频谱, $\hat{X}_i(k)$ 表示增强语音的频谱. 最后, $\hat{X}_i(k)$ 通过 IDFT 变换后可以得到增强语音的时域信号.

4 基于深度神经网络的双耳线索语音增强

本文在上一章提出一种基于先验码书的双耳线索语音增强方法. 该方法通过 WCBM 算法估计纯净线索参数. 但是, 当码矢量选择有误时, 该算法估计纯净线索参数的准确率会下降. 因此, 本章提出一种基于深度神经网络的双耳线索语音增强算法. 该算法采用深度神经网络替代 WCBM 算法估计纯净线索参数, 通过深度神经网络的特征学习能力训练出预增强线索参数和纯净线索参数间的匹配函数, 从而直接利用预增强线索参数估计纯净线索参数, 提高估计的准确率. 图 4 给出了所提算法的原理框图.

如图 4 所示, 本章所提算法分为线下训练和线上增强两部分. 在线下训练部分, 将上一章介绍的纯净线索参数和与其对应的预增强线索参数分别作为深度神经网络的输出和输入进行训练, 得到含有预增强线索参数和纯净线索参数匹配函数的深度神经网络: SAE^[11]. 在线上增强部分, 将提取的预增强线索参数直接作为训练得到的 SAE 的输入, 其网络输出即为估计的纯净线索参数. 最后, 根据上一章 BCC 模型中的增益函数增强含噪语音. 这里采用 SAE 作为深度神经网络进行训练. 图 5 给出了 SAE 的模型框图.

SAE 是由逐层贪婪法^[12]训练得到的. 如图 5 所示, 该网络是由若干个自编码器 (Auto-Encoder, AE) 堆叠而成的一种深度神经网络. SAE 的训练可以分为三个模块: 预处理、网络堆叠和网络微调. 在预处理模块, 首先采用无监督的方式训练第一个 AE; 然后前一个 AE 的

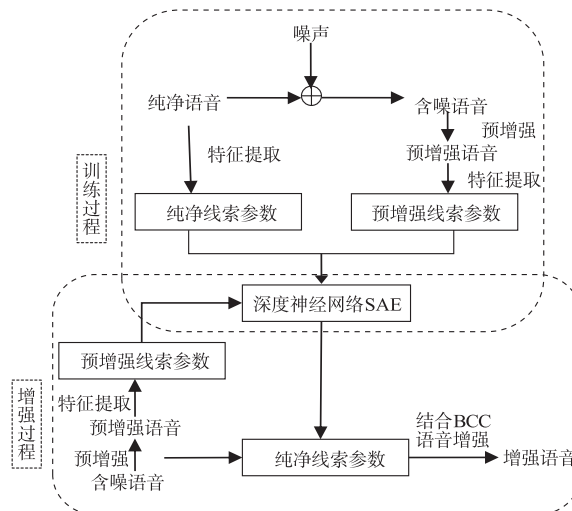


图4 基于深度神经网络的双耳线索语音增强原理框图

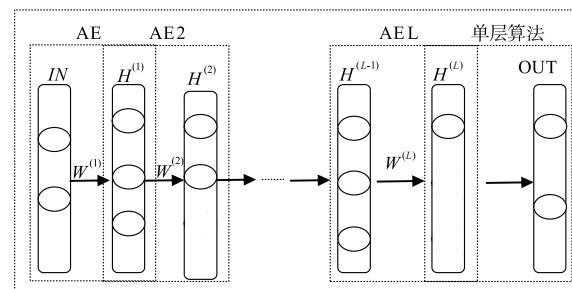


图5 SAE模型框图

输出作为下一个 AE 的输入继续训练下一个 AE, 以此类推, 直到训练出 L 个 AE; 最后, 将训练好的 L 个 AE 的权值作为网络的初始值, 完成网络的预处理模块. 在网络堆叠模块, L 个训练得到的 AE 按照图 5 的方式堆叠到一起, 并在最后一层添加一个单层算法, 得到深度神经网络——SAE. 最后, 在网络微调模块, 根据训练数据中的期望输出, 采用反向传播法对网络进行微调.

当隐含层节点数过大时, AE 的训练会出现参数过拟合的问题. 为解决这一问题, 稀疏限制被引入到 AE 中, 得到了稀疏自编码器, 其重建误差的表达式为

$$R_e(x, y) = \frac{1}{R} \sum_{i=1}^R \sum_{j=1}^P \|x^i(j) - y^i(x(j))\|^2 + \lambda \sum_{j=1}^Q KL(\rho \| \hat{\rho}_j) \quad (17)$$

式(17)中, R 表示小批量数据的大小, P 表示输入向量的大小, Q 表示隐含层节点数, λ 表示稀疏性惩罚因子的权重, $KL(\rho \| \hat{\rho}_j)$ 表示 KL 散度, 其计算公式为

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (18)$$

式中, ρ 表示目标的稀疏度, 它是一个接近于 0 的正数; $\hat{\rho}_j$ 表示该隐含层在 j 节点的所有样本数据输出值的平均值, 其计算表达式为

$$\hat{\rho}_j = \frac{1}{R} \sum_{i=1}^R y_j(x^i) \quad (19)$$

已知重建误差,本文可以根据梯度下降法训练网络权值.完成稀疏自编码器的训练.最后,利用训练得到的稀疏自编码器按照图 5 堆叠成 SAE.

在本章所提算法中,SAE 被引入到基于双耳线索的语音增强算法.其中,预增强线索参数作为网络的输入,纯净线索参数作为网络的输出,通过训练集数据的训练,得到一个能够直接由预增强线索参数估计纯净线索参数的深度神经网络.最后,训练得到的 SAE 用于线上估计纯净线索参数.并且采用上一章所提的增益函数增强含噪语音,实现语音增强.

5 性能测试

本章通过实验测试结果和分析对所提语音增强算法进行性能分析.在实验中,训练数据和测试数据的采样率均为 8kHz.训练数据来自 NNT 数据库,长度为 30 分钟.噪声训练集选自 NOISE 92 数据库,包含 12 种噪声.之后利用纯净语音训练集和噪声训练集得到含噪语音训练集,其信噪比从 0dB 到 20dB,步长为 2dB.此外,测试语音也选自 NTT 数据库,但在训练集之外.测试语音包含 9 段语音,每段 8s.测试噪声类型包括白噪声(white),说话人噪声(babble),工厂噪声(factory)以及街道噪声(street)四种噪声类型.测试数据的输入信噪比分别为 0dB,5dB 和 10dB.在第三章所提算法中,在线下通过 LBG 算法训练提取线索参数,得到 8bit 的先验码书,码书中的矢量维数为 6.在第四章所提算法中,在线下通过学习训练提取线索参数,得到含有预增强线索参数和纯净线索参数匹配函数的先验深度神经网络 SAE.其中,SAE 含有 3 个隐含层,其隐含层节点数分别为 30、30、30.训练迭代次数为 1000 次.

为了验证本文方法的性能,本文采用客观质量测试方法对增强算法进行性能测试,并给出语谱图比较示例.客观质量测试指标有分段信噪比(Segment SNR, SSNR)^[13]、PESQ(Perceptual Evaluation of Speech Quality)^[14]和对数谱失真(Log-Spectral Distortion, LSD)^[15].这里我们用经典的 MMSE 估计算法^[16](称为 Ref. A)和 AR 谱码书驱动算法^[3](Ref. B)做为参考方法.第三章所提算法称之为 Pro. 1,第四章所提算法称之为 Pro. 2.下面给出客观测试结果.

表 1~表 3 给出了所提算法的客观测试结果,从结果可以看出所提算法的性能优于参考算法.由于 Pro. 2 算法利用 SAE 可以得到更加准确的纯净线索参数估计,所以,Pro. 2 算法的表现较之 Pro. 1 更好.为了更加直观的观察算法性能,本文针对四种噪声类型给出了

柱状图比较结果,如图 6、图 7 和图 8 所示.

从图 6 到图 8 中可以看出,由本文方法得到的增强语音的客观质量明显高于参考算法的.为易于观察,本文给出了语谱图对比,如图 9 所示.从语谱图可以看出,与参考算法相比较,所提算法不仅可以抑制更多的背景噪声,而且能够较好地消除谐波间的噪声.因此,从语谱图结果可知,所提算法优于参考算法.

表 1 SSNR 提高量结果

噪声类型	方法	信噪比		
		0dB	5dB	10dB
White	Ref. A	11.01	9.74	8.22
	Ref. B	10.37	9.69	8.94
	Pro. 1	11.92	10.95	9.81
	Pro. 2	15.10	13.98	12.68
Babble	Ref. A	8.74	7.75	6.60
	Ref. B	5.68	6.27	4.43
	Pro. 1	7.49	7.28	6.94
	Pro. 2	8.47	8.45	8.09
Street	Ref. A	10.60	9.56	8.38
	Ref. B	12.86	11.5	10.19
	Pro. 1	11.59	11.09	10.31
	Pro. 2	13.08	12.73	11.80
Factory	Ref. A	10.17	8.99	7.84
	Ref. B	8.54	7.81	6.96
	Pro. 1	11.46	10.81	9.95
	Pro. 2	13.63	13.03	11.94

表 2 PESQ 结果

噪声类型	方法	信噪比		
		0dB	5dB	10dB
White	Noisy	1.40	1.59	1.97
	Ref. A	1.62	1.99	2.33
	Ref. B	1.87	2.25	2.56
	Pro. 1	2.01	2.35	2.66
	Pro. 2	2.20	2.52	2.79
Babble	Noisy	1.80	2.12	2.49
	Ref. A	2.04	2.45	2.75
	Ref. B	1.82	2.24	2.55
	Pro. 1	2.02	2.42	2.78
	Pro. 2	2.01	2.44	2.80
Street	Noisy	2.30	2.64	2.94
	Ref. A	2.63	2.88	3.09
	Ref. B	2.57	2.83	3.06
	Pro. 1	2.79	3.04	3.25
	Pro. 2	2.84	3.09	3.31
Factory	Noisy	1.85	2.27	2.61
	Ref. A	2.32	2.62	2.86
	Ref. B	2.24	2.56	2.83
	Pro. 1	2.52	2.81	3.05
	Pro. 2	2.59	2.88	3.11

表 3 LSD 结果

噪声类型	方法	信噪比		
		0dB	5dB	10dB
White	Noisy	18.33	16.20	14.19
	Ref. A	12.33	10.86	9.53
	Ref. B	12.09	10.63	9.23
	Pro. 1	11.91	10.41	9.04
	Pro. 2	10.42	9.08	7.83
Babble	Noisy	14.34	12.45	10.71
	Ref. A	10.52	9.02	7.62
	Ref. B	11.60	10.05	8.61
	Pro. 1	10.83	9.17	7.66
	Pro. 2	10.35	8.62	7.18
Street	Noisy	12.78	11.01	9.37
	Ref. A	8.88	7.54	6.29
	Ref. B	9.15	7.81	6.69
	Pro. 1	8.57	7.15	5.90
	Pro. 2	7.82	6.35	5.19
Factory	Noisy	13.94	12.07	10.34
	Ref. A	9.76	8.29	6.98
	Ref. B	10.31	8.87	7.51
	Pro. 1	9.14	7.71	6.43
	Pro. 2	8.11	6.64	5.49

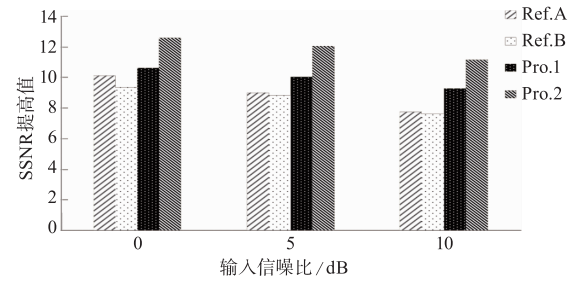


图6 四种噪声下的平均SSNR提高量

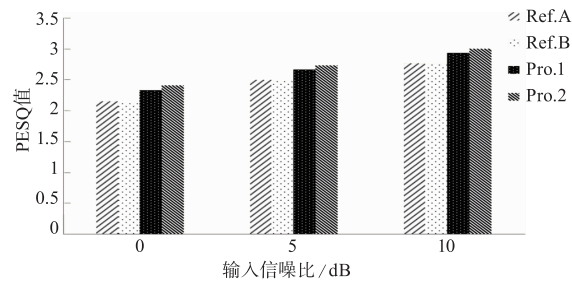


图7 四种噪声下的平均PESQ

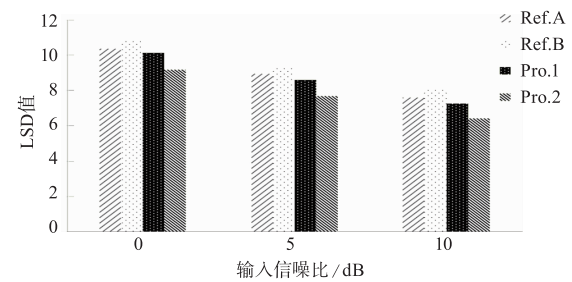


图8 四种噪声下LSD

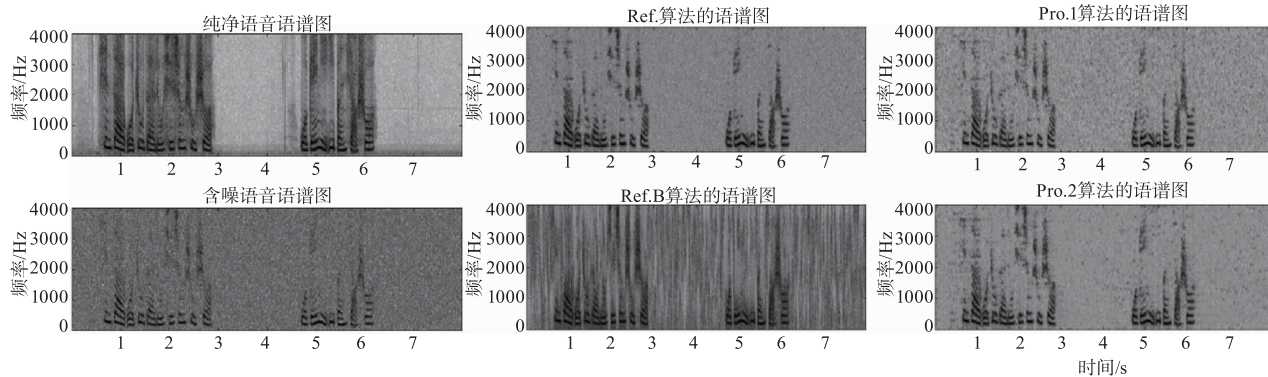


图9 10dB白噪声下的语谱图比较

6 结论

本文提出一种基于先验码书的双耳线索单通道语音增强算法. 该算法采用语音和噪声的双耳线索作为其先验知识, 避免了噪声预分类的问题, 同时, 该算法通过对谱细节建模来实现语音增强, 能够有效地抑制谐波间噪声. 此外, 本文提出了基于神经网络的双耳线索语音增强算法. 该算法采用神经网络算法代替 WCBM 算法估计纯净线索参数, 提高了估计的准

确率. 最后, 通过给出的性能测试, 可以发现所提算法的性能优于参考算法.

参考文献

- [1] Boll S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1979, 27(2): 113-120.
- [2] Lim J S, Oppenheim A V. Enhancement and bandwidth compression of noisy speech[J]. Proceedings of the IEEE,

- 1979, 67(12):1586-1604.
- [3] Srinivasan S, Samuelsson J, Klejin W B. Codebook driven short-term prediction parameter estimation for speech enhancement[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 2006, 14(1):163-175.
- [4] Srinivasan S, Samuelsson J, Klejin W B. Codebook-based bayesian speech enhancement for nonstationary environments[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 2007, 15(2):441-451.
- [5] 何玉文, 鲍长春, 夏丙寅. 基于 AR-HMM 在线能量调整的语音增强方法[J]. *电子学报*, 2014, 42(10):1991-1997.
HE Yu-wen, BAO Chang-chun, XIA Bing-yin. Online energy adjustment using AR-HMM for speech enhancement [J]. *Acta Electronica Sinica*, 2014, 42(10):1991-1997. (in Chinese)
- [6] 梁岩, 鲍长春, 夏丙寅, 何玉文, 李娜. 基于高斯混合模型的压缩域语音增强方法[J]. *电子学报*, 2012, 40(10):2031-2038.
LIANG Yan, BAO Chang-chun, HE Yu-wen, LI Na. Compressed domain speech enhancement based on gaussian mixture model[J]. *Acta Electronica Sinica*, 2012, 40(10):2031-2038. (in Chinese)
- [7] Faller C, Baumgarte F. Binaural cue coding-part1: Psychoacoustic fundamentals and design principles [J]. *IEEE Transactions on Audio, Speech, Language Processing*, 2003, 11(6):509-519.
- [8] Faller C, Baumgarte F. Binaural cue coding-part2: Schemes and applicarions[J]. *IEEE Transactions on Audio, Speech, Language Processing*, 2003, 11(6):520-531.
- [9] 张勇, 胡瑞敏. 基于高斯混合模型的语音频带扩展所发的研究[J]. *声学学报*, 2009, 34(5):471-480.
ZHANG Yong, HU Rui-min. Speech wideband extension based on gaussian mixture model [J]. *Chinese Journal of Acoustics*, 2009, 34(5):471-480. (in Chinese)
- [10] 孟宪波, 鲍长春. 基于最小控制 GARCH 模型的噪声估计方法[J]. *电子学报*, 2016, 44(3):747-752.
MENG Xian-bo, BAO Chang-chun. Noise estimate algorithm based on minima controlled GARCH model [J]. *Acta Electronica Sinica*, 2016, 44(3):747-752. (in Chinese)
- [11] Araki S, Araki T. Exploring multi-channel features for denoising-autoencoder-based speech enhancement [A]. *Proceedings of the 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [C]. Brisbane, Australia: IEEE Press, 2015. 116-120.
- [12] Hinton G E, Osindero S. A fast learning algorithm for deep belief nets [J]. *Neural Computation*, 2006, 18(7):1527-1554.
- [13] Quackenbush S R, Barnwell T P, Clements M A. *Objective Measures of Speech Quality* [M]. Englewood Cliffs, NJ: Prentice Hall, 1988.
- [14] ITU-T, Recommendation P. 862. *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Coders* [S]. 2001.
- [15] Abramson A, Cohen I. Simultaneous detection and estimation approach for speech enhancement [J]. *IEEE Transactions on Audio, Speech and Language Processing*, 2007, 15(8):2348-2359.
- [16] Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator [J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1985, 23(2):443-445.

作者简介



陈楠 女, 1992 年生于北京密云. 北京工业大学硕士研究生. 主要研究方向为语音增强.
E-mail: chennan12@emails.bjut.edu.cn



鲍长春 男, 1965 年生于内蒙古赤峰. 现为北京工业大学教授、博士生导师. 主要研究方向为语音与音频信号处理.
E-mail: chchbao@bjut.edu.cn