DOI:10.19650/j.cnki.cjsi.J2412996

基于谱特征自适应估计的激光相干语音探测信号增强方法

芮小博[1],孔欣玥[1],伍 洲[2,3],张文喜[2,3],曾周末[1]
(1.天津大学精密测试技术及仪器国家重点实验室 天津 300072;2.中国科学院空天信息创新研究院 北京 100094;
3.中国科学院大学 北京 100049)

摘 要:针对激光相干语音探测引入的缓变宽带背景噪声和测振目标造成的信道作用,本文提出了基于分析-重合成框架,针对特定说话人的语音增强方法。该方法首先提取观测语音特征:基音频率、浊音概率、MCEP 系数,其中,MCEP 系数是能够表示谱包络形状的谱包络特征。通过观测语音的谱包络特征和预训练的对应说话人语音谱包络特征 GMM,估计对应的纯净语音谱包络特征,再与观测语音的基音频率和浊音概率一起重新合成语音信号,实现语音增强。噪声和信道参数的估计通过最大化观测语音谱包络特征后验概率的自适应估计实现,然后通过 MMSE 估计得到纯净语音谱包络特征的估计值。合成信号实验和实际信号采集实验检验了本文提出算法在激光相干语音探测场景下的去噪和均衡能力。

关键词:激光相干语音探测;语音增强;混合高斯模型;矢量泰勒级数
中图分类号:TH741 TN911.7 **文献标识码**:A **国家标准学科分类代码**:510.40

Enhancement of speech detected by laser coherent detection method based on spectral feature adaptation

Rui Xiaobo[1], Kong Xinyue[1], Wu Zhou[2,3], Zhang Wenxi[2,3], Zeng Zhoumo[1]
(*1. State Key Laboratory of Precision Measurement Technology and Instrument, Tianjin University, Tianjin 300072, China;*
2. Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: To address the issue of slowly varying broadband background noise and channel effects caused by vibration of the target in laser-coherent speech detection, this paper proposes a speech enhancement method for specific speakers based on an analysis-re-synthesis framework. This method first extracts the features from the observed signal: pitch, voiced speech probability, and MCEP coefficients, where MCEP coefficients represent the spectral envelope features which can capture the shape of the spectral envelope. A GMM trained by speech features of the corresponding speaker is used to help estimate the spectral envelope features of the clean speech from the spectral envelope features of the observed speech, and then the speech signal is resynthesized by combining it with pitch and voiced speech probability estimated from the observed speech to achieve speech enhancement. The estimation of noise and channel parameters is achieved by adaptation, which maximize the posterior probability of the observed speech′s spectral envelope features, and then the estimation of the clean speech spectral envelope features is obtained by MMSE estimation. Both synthesized signal experiments and actual signal acquisition experiments verify the denoising and equalization capabilities of the algorithm in laser coherent speech detection scenarios.

Keywords: laser coherent speech detection; speech enhancement; Gaussian mixture model; vector Taylor series

收稿日期:2024-06-26 Received Date: 2024-06-26

0　引　　言

麦克风通过测量振膜在空气声压作用下的振动以感知语音信号,由于声波能量随着距离增大快速衰减,其探测距离十分有限。激光语音探测方法通过激光光束测量声源附近物体表面微振动以探测语音信号,可以实现超远距离的信号采集并且可深入人员难以到达的区域,在军事安全等领域具有一定的应用价值。

激光相干语音探测指利用激光多普勒测振技术采集语音信号。激光多普勒测振技术基于激光光波的多普勒效应,用相干探测的方法探测多普勒频移,从而解算出测量位置在光束方向的位移/速度/加速度,是实现激光语音探测的一种较为成熟的方案[1]。然而在应激光相干语音探测中,一些因素将会对采集语音信号的质量造成影响:1)声源附近的外界噪声的影响;2)测量系统噪声和光束受环境影响引入噪声;3)非合作测振目标在语音频段的频响往往不是一条平稳的水平直线,这些影响具体表现为引入宽带背景噪声、冲击噪声和信道作用(语音各频段相对能量发生改变)。

因此,激光探测语音信号需要进行处理以提升信号质量,语音增强方法能够一定程度上实现语音和噪声的分离,具有在激光探测语音信号质量提升中进行应用的潜力。单通道语音增强方法包括:谱减法、维纳滤波法、最小均方差(minimum mean square error, MMSE)估计等结合噪声功率谱估计实现的短时谱估计算法[2];在其它变换域进行语音和噪声分离的子空间法[3]和小波去噪法[4];将语音信号或特征建模为随机自回归过程使用卡尔曼滤波[5]或粒子滤波[6]进行估计的方法;引入了语音和噪声的先验模型的基于码本[7]、字典学习[8]、隐马尔可夫模型[9]的语音增强方法;基于语音生成模型,通过带噪语音特征估计纯净语音特征并重新合成语音的分析-重合成语音增强方法[10];基于深度学习的语音增强方法[11]等。通常应用于数字语音通信设备、视听会议、自动语音识别系统等。

目前,激光相干语音探测与语音增强方法结合的相关应用研究较少,吕韬[12]采用最优对数谱(optimally-modified log-spectral amplitude, OMLSA)估计算法与相位补偿相结合进行信号降噪处理,然后使用基于半波整流的语音带宽扩展算法扩展相干侦测语音的带宽,但测振目标造成的信道作用其实并不是简单的窄带宽导致信号高频部分缺失;王亚慧等[13-14]分别提出拉普拉斯分布下 MMSE 谱减语音增强算法和循环生成对抗网络对采集语音进行增强,谱减法只能实现降低噪声的影响,而激光探测信号中复杂多变的噪声干扰使得基于深度学习的方法需要大量不同条件下的采集数据。整体来看,相关研究有待进一步深入,缺乏对侦听等应用场景下特定人语音信号增强等针对性的研究。

针对以上问题与不足,本文提出了基于谱特征自适应估计的激光相干语音探测信号增强方法,引入了纯净语音谱包络特征高斯混合模型,实现噪声和信道的自适应估计,通过信号分析与重合成,一定程度上实现了特定目标人语音激光相干探测信号的信号去噪,以及语音信号在频率响应上的均衡。

1　方　法

本文提出的语音增强算法基于分析-重合成框架,整体流程如图 1 所示。首先提取带噪语音的浊音概率、基音频率、谱包络特征参数;然后根据提前训练好的特定说话人的语音谱包络特征高斯混合模型(Gaussian mixture model, GMM),对噪声和信道参数进行预估计和自适应估计,进而对纯净语音谱包络特征进行估计;最后,用带噪语音的浊音概率、基音频率和估计纯净语音的谱包络特征参数进行语音重合成,得到增强语音。

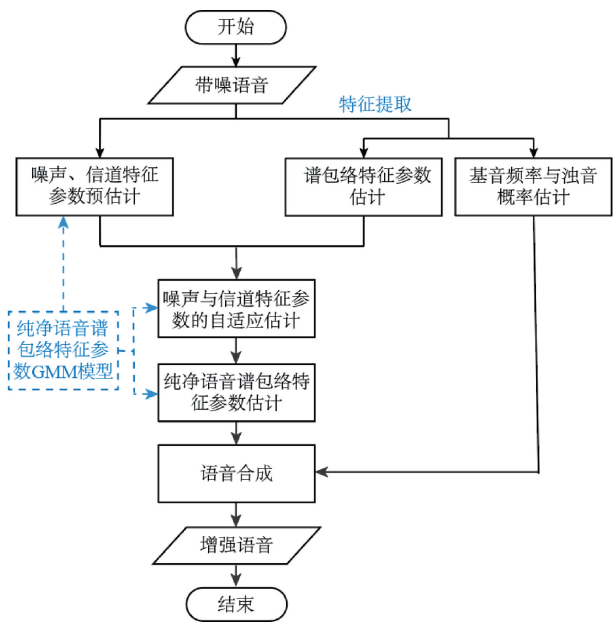


图 1　本文方法整体流程

Fig. 1　Flow chart of the method proposed in this paper

1.1　谐波加噪声语音生成模型

谐波加噪声模型(harmonic plus noise model, HNM)是一类将语音建模成谐波成分与随机成分相叠加的语音生成模型[15]。本文将语音信号表示为基音频率(声调)、浊音(声带振动)概率、谱包络特征参数,然后使用 HNM 算法根据这些参数生成语音信号。理论上,带噪信号的

基音频率和浊音概率与纯净语音一致,因此可以将语音增强问题转换为纯净语音信号谱包络参数的估计。

1)特征提取

通过在强噪声条件下具有鲁棒性的基音频率估计算法 PEFAC(pitch estimation filter with amplitude compression, PEFAC)[16] 估计经短时傅里叶变换(short time Fourier transform, STFT)后每帧信号的基音频率 f_i 和浊音概率 $pv_i, i = 0,1,\cdots,I-1$ 表示帧索引。谱包络特征参数选择 MCEP(mel-cepstral)系数[17],利用该系数可将一帧语音信号的频谱 $S(e^{j\omega})$ 表示为 M 阶 MCEP 系数:

$$S(z) = \exp\sum_{m=0}^{M} c_\alpha[m] z_\alpha^{-m} \tag{1}$$

$$z_\alpha^{-1} = \psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}, \quad |\alpha| < 1 \tag{2}$$

式中:c_α 表示 MCEP 系数,z_α^{-1} 为全通传递函数,α 控制频率的扭曲,如果 α 不为 0,则系统具有梅尔尺度频率特性。一帧语音信号的对数幅度谱可以表示为:

$$\log(|S(e^{j\omega})|) = \sum_{m=0}^{M} c_\alpha[m] real(\psi(e^{j\omega})^m) \tag{3}$$

2)语音合成

语音合成基于 1)特征提取所述的特征,计算谐波成分与随机成分并进行叠加,谐波成分分布于浊音段,表示为:

$$s_h\left[\frac{1}{2}(l_w - 1) + il_h + l\right] = \begin{cases} \sum_{m=1}^{M_i} \hat{A}_{i,m}[l]\cos(\hat{\theta}_{i,m}[l]), & pv_i > 0.5 \\ 0, & pv_i \leqslant 0.5 \end{cases} \tag{4}$$

$pv_i > 0.5$ 时判定这一帧语音信号为有话段,在对应帧的中间部分加入谐波成分。l_w 表示计算帧长,$\hat{A}_{i,m}[l]$,$\theta_{i,m}[l]$ 分别是第 i 帧 m 次谐波的幅度与相位,其中,$l = 0,1,\cdots,l_h - 1$,l_h 表示帧移。$\hat{A}_{i,m}[l]$ 通过三角滤波器估计幅度谱包络在对应谐波频率处的值得到,其它时间点处的幅度经线性插值得到,谐波频率的估计与谐波幅度相同,谐波相位经随机初始化,然后根据谐波频率和采样时间计算得到。

$$s_n[n] = \begin{cases} c\sum_{i=0}^{I-1} w[n - il_h]\mathrm{IDFT}\{\hat{E}_i[k]\mathrm{HPF}\{D_i[k]\}\}, & pv_i > 0.5 \\ c\sum_{i=0}^{I-1} w[n - il_h]\mathrm{IDFT}\{\hat{E}_i[k]D_i[k]\}, & pv_i \leqslant 0.5 \end{cases} \tag{5}$$

式中:$\hat{E}_i[k]$ 为谱包络估计,$D_i[k]$ 为能量归一化的高斯白噪声的频谱,IDFT$\{\cdot\}$ 表示逆离散傅里叶变换操作,HPF$\{\cdot\}$ 表示高通滤波操作,常数 c 和窗 w 用于实现逆短时傅里叶变换。

合成语音信号最终表示为:

$$s[n] = s_h[n] + s_n[n] \tag{6}$$

1.2 语音特征参数(MCEP 系数)增强方法

1)观测信号特征模型

为了从观测语音谱包络特征中估计纯净语音谱包络特征的方法,本文采用了基于矢量泰勒级数(vector Taylor series, VTS)的特征增强方法[18],首先需要建立观测信号模型。

将激光相干语音探测信号简单表示为纯净语音在信道(卷积)和加性噪声的作用下得到观测信号,观测信号经测量系统的采集和离散傅里叶变换转换到频域可得:

$$Y_k = H_k X_k + N_k \tag{7}$$

X_k、H_k、N_k 和 Y_k 分别表示纯净语音信号、信道、噪声、观测信号的 DFT 系数,k 代表频率采样点,由此可得观测信号的功率谱:

$$|Y_k|^2 = |H_k|^2|X_k|^2 + |N_k|^2 + 2\cos\theta_k|H_k||X_k||N_k| \tag{8}$$

其中,θ_k 表示复变量($H_k X_k$) 与 N_k 之间的夹角。

用向量形式表示信号的频谱 $\boldsymbol{Y} = [Y_0, Y_1, \cdots, Y_{K-1}]^{\mathrm{T}}$,同理可得 $\boldsymbol{X}$、$\boldsymbol{N}$、$\boldsymbol{H}$,根据式(3)通过 MCEP 系数估计观测信号的对数功率谱:

$$\log(|\boldsymbol{Y}|^2) \approx \boldsymbol{C}^{-1}\boldsymbol{y} \tag{9}$$

$\boldsymbol{y} = [y_0, y_1, \cdots, y_m]^{\mathrm{T}}$ 表示 MCEP 系数组成的向量,同理可得 $\boldsymbol{x}$、$\boldsymbol{h}$、$\boldsymbol{n}$,$\boldsymbol{C}^{-1}$ 为转化矩阵,其伪逆表示为 $\boldsymbol{C}$,所以观测语音信号的 MCEP 系数可估计为:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{h} + \boldsymbol{C}\log(1 + \exp(\boldsymbol{C}^{-1}(\boldsymbol{n} - \boldsymbol{x} - \boldsymbol{h})) + 2\boldsymbol{\alpha} \cdot \exp(\boldsymbol{C}^{-1}(\boldsymbol{n} - \boldsymbol{x} - \boldsymbol{h})/2)) = \boldsymbol{x} + \boldsymbol{h} + \mathrm{g}(\boldsymbol{x},\boldsymbol{h},\boldsymbol{n}) \tag{10}$$

其中,

$$\mathrm{g}(\boldsymbol{x},\boldsymbol{h},\boldsymbol{n}) = \boldsymbol{C}\log(1 + \exp(\boldsymbol{C}^{-1}(\boldsymbol{n} - \boldsymbol{x} - \boldsymbol{h})) + 2\boldsymbol{\alpha} \cdot \exp(\boldsymbol{C}^{-1}(\boldsymbol{n} - \boldsymbol{x} - \boldsymbol{h})/2)) \tag{11}$$

$\boldsymbol{\alpha} = [\cos\theta_0, \cos\theta_1, \cdots, \cos\theta_k, \cdots]^{\mathrm{T}}$,本文将 $\boldsymbol{\alpha}(0 \leqslant \alpha_k \leqslant 1)$ 设为定值,$\cdot$ 表示两个向量间的逐元素相乘,使用关于 $\boldsymbol{x}$、$\boldsymbol{h}$、$\boldsymbol{n}$ 的 1 阶 VTS 估计可得:

$$\boldsymbol{y} \approx \boldsymbol{\mu}_x + \boldsymbol{\mu}_h + g(\boldsymbol{\mu}_x,\boldsymbol{\mu}_h,\boldsymbol{\mu}_n) + \boldsymbol{G}(\boldsymbol{x} - \boldsymbol{\mu}_x) + \boldsymbol{G}(\boldsymbol{h} - \boldsymbol{\mu}_h) + (\boldsymbol{I} - \boldsymbol{G})(\boldsymbol{n} - \boldsymbol{\mu}_n) \tag{12}$$

其中,$\boldsymbol{\mu}_x$ $\boldsymbol{\mu}_h$ $\boldsymbol{\mu}_n$ 分别是 $\boldsymbol{x}$、$\boldsymbol{h}$、$\boldsymbol{n}$ 的均值矢量,$\boldsymbol{I}$ 为单位矩阵,

$$\left.\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{\mu}_x,\boldsymbol{\mu}_n,\boldsymbol{\mu}_h} = \left.\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{h}}\right|_{\boldsymbol{\mu}_x,\boldsymbol{\mu}_n,\boldsymbol{\mu}_h} = \boldsymbol{G} \tag{13}$$

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{n}} = \boldsymbol{I} - \boldsymbol{G} \tag{14}$$

$$\boldsymbol{G}=\boldsymbol{I}-\boldsymbol{C}\mathrm{diag}\left(\frac{\exp(\boldsymbol{C}^{-1}(\boldsymbol{\mu}_n-\boldsymbol{\mu}_x-\boldsymbol{\mu}_h))+\boldsymbol{\alpha}\exp(\boldsymbol{C}^{-1}(\boldsymbol{\mu}_n-\boldsymbol{\mu}_x-\boldsymbol{\mu}_h)/2)}{1+\exp(\boldsymbol{C}^{-1}(\boldsymbol{\mu}_n-\boldsymbol{\mu}_x-\boldsymbol{\mu}_h))+2\boldsymbol{\alpha}\exp(\boldsymbol{C}^{-1}(\boldsymbol{\mu}_n-\boldsymbol{\mu}_x-\boldsymbol{\mu}_h)/2)}\right)\boldsymbol{C}^{-1} \tag{15}$$

其中,diag(·)代表构造对角矩阵,其对角分量值分别等于括号中向量的元素。

2)纯净语音特征向量估计

线下训练阶段,用大量纯净语音的 MCEP 系数训练 GMM。GMM 用多个多维高斯分布来拟合多维随机变量的概率分布,利用 GMM 建立纯净语音 MCEP 系数概率分布的统计模型表示为:

$$p(\boldsymbol{x})=\sum_s p(s)\mathrm{N}(\boldsymbol{x}\,|\boldsymbol{\mu}_{x,s},\boldsymbol{\Sigma}_{x,s}) \tag{16}$$

其中,s 代表 GMM 中的聚类,对于聚类 s,$p(s)$ 表示其先验概率,$\mathrm{N}(\boldsymbol{x}\,|\boldsymbol{\mu}_{x,s},\boldsymbol{\Sigma}_{x,s})$ 代表均值为 $\boldsymbol{\mu}_{x,s}$,协方差矩阵为 $\boldsymbol{\Sigma}_{x,s}$ 的高斯函数,训练时令协方差矩阵为对角阵,即 $\boldsymbol{\Sigma}_s=\mathrm{diag}([\sigma_{s,1}^2,\sigma_{s,2}^2,\cdots\sigma_{s,M}^2])$。

纯净语音特征向量估计阶段,根据观测信号的特征向量 $\boldsymbol{y}$ 估计 $\boldsymbol{x}$ 的后验概率分布可得:

$$p(\boldsymbol{x}\,|\boldsymbol{y})=\sum_s p(\boldsymbol{x}\,|\boldsymbol{y},s)p(s\,|\boldsymbol{y}) \tag{17}$$

因此,纯净语音特征的 MMSE 估计量为:

$$\hat{\boldsymbol{x}}=\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y})=\sum_s p(s\,|\boldsymbol{y})\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y},s) \tag{18}$$

通过式(12)将纯净语音特征 GMM 的每个聚类,映射为观测语音 GMM 的对应聚类,所估计的观测语音特征向量 GMM 各聚类的均值向量和协方差矩阵可表示为:

$$\boldsymbol{\mu}_{y,s}\approx\boldsymbol{\mu}_{x,s}+\boldsymbol{\mu}_h+\mathrm{g}(\boldsymbol{\mu}_{x,s},\boldsymbol{\mu}_h,\boldsymbol{\mu}_n) \tag{19}$$

$$\boldsymbol{\Sigma}_{y,s}\approx\boldsymbol{G}_s\boldsymbol{\Sigma}_{x,s}\boldsymbol{G}_s^{\mathrm{T}}+(\boldsymbol{I}-\boldsymbol{G}_s)\boldsymbol{\Sigma}_n(\boldsymbol{I}-\boldsymbol{G}_s)^{\mathrm{T}} \tag{20}$$

通过得到的观测语音特征向量 GMM 来实现式(18)中对于 $p(s\,|\boldsymbol{y})$ 的估计。本文将信道作用视为固定值,因此式(20)中不考虑信道方差。

根据 $\boldsymbol{y}$ 关于 $\boldsymbol{x}$ 的一阶 VTS 估计计算 $\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y},s)$[19],聚类 s 中,将 $\boldsymbol{y}$ 估计为:

$$\boldsymbol{y}\approx\boldsymbol{\mu}_{x,s}+\boldsymbol{\mu}_h+g(\boldsymbol{\mu}_{x,s},\boldsymbol{\mu}_h,\boldsymbol{\mu}_n)+\boldsymbol{G}_s(\boldsymbol{x}-\boldsymbol{\mu}_{x,s})+\boldsymbol{G}_s(\boldsymbol{h}-\boldsymbol{\mu}_h)+(\boldsymbol{I}-\boldsymbol{G}_s)(\boldsymbol{n}-\boldsymbol{\mu}_n) \tag{21}$$

由于 $\boldsymbol{x}$ 与 $\boldsymbol{h}$、$\boldsymbol{n}$ 不相关,聚类 s 中,$\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的联合分布可写作:

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix}_s\sim N\left(\begin{bmatrix}\boldsymbol{\mu}_{x,s}\\ \boldsymbol{\mu}_{y,s}\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}_{x,s} & \boldsymbol{\Sigma}_{xy,s}\\ \boldsymbol{\Sigma}_{yx,s} & \boldsymbol{\Sigma}_{y,s}\end{bmatrix}\right) \tag{22}$$

$$\boldsymbol{\Sigma}_{xy,s}=\boldsymbol{\Sigma}_{x,s}\boldsymbol{G}_s^{\mathrm{T}} \tag{23}$$

由此可得:

$$\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y},s)=\boldsymbol{\mu}_{x,s}+\boldsymbol{\Sigma}_{xy,s}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{y,s})=\boldsymbol{\mu}_{x,s}+\boldsymbol{\Sigma}_{x,s}\boldsymbol{G}_s^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{y,s}) \tag{24}$$

一阶 VTS 估计引入的偏差可能会导致 $\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y},s)$ 与相差过大,因此将 $\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y},s)$ 修正为:

$$\mathrm{E}'(\boldsymbol{x}\,|\boldsymbol{y},s)=\boldsymbol{\mu}_{x,s}+\boldsymbol{\sigma}'_s\odot\tanh\left(\frac{E(\boldsymbol{x}\,|\boldsymbol{y},s)-\boldsymbol{\mu}_{x,s}}{\boldsymbol{\sigma}'_s}\right) \tag{25}$$

$$\boldsymbol{\sigma}'_s=\boldsymbol{\sigma}_{xs}\odot\frac{E(\boldsymbol{x}\,|\boldsymbol{y},s)-\boldsymbol{\mu}_{x,s}}{\|\mathrm{E}(\boldsymbol{x}\,|\boldsymbol{y},s)-\boldsymbol{\mu}_{x,s}\|_2} \tag{26}$$

式(25)和(26)中,⊙表示按位相乘,除号表示按位相除,$\boldsymbol{\sigma}_{xs}$ 为对角阵 $\boldsymbol{\Sigma}_{x,s}$ 中元素开方组成的向量 $\boldsymbol{\sigma}_{xs}=[\sigma_{xs,1},\sigma_{xs,2},\cdots,\sigma_{xs,M}]^{\mathrm{T}}$。

3)噪声与信道特征参数的自适应估计

利用 EM(期望-最大化)算法对信道和噪声特征的均值 $\boldsymbol{\mu}_h$ $\boldsymbol{\mu}_n$ 进行自适应估计[20]。该方法在已知一段观测语音的情况下,通过迭代的方式最大化其后验概率进行参数估计。在迭代估计参数 $\boldsymbol{\mu}_h$、$\boldsymbol{\mu}_n$ 时,将式(19)展开为关于 $\boldsymbol{\mu}_h$、$\boldsymbol{\mu}_n$ 的当前值 $\boldsymbol{\mu}_{h_0}$、$\boldsymbol{\mu}_{n_0}$ 的一阶 VTS。

$$\boldsymbol{\mu}_{y,s}\approx\boldsymbol{\mu}_{x,s}+\boldsymbol{\mu}_{h_0}+g(\boldsymbol{\mu}_{x,s},\boldsymbol{\mu}_{h_0},\boldsymbol{\mu}_{n_0})+\boldsymbol{G}_s(\boldsymbol{\mu}_h-\boldsymbol{\mu}_{h_0})+(\boldsymbol{I}-\boldsymbol{G}_s)(\boldsymbol{\mu}_n-\boldsymbol{\mu}_{n_0}) \tag{27}$$

EM 算法的辅助函数可表示为:

$$Q(\boldsymbol{\lambda}\,|\bar{\boldsymbol{\lambda}})=\sum_t\sum_s p(s_t=s\,|\boldsymbol{y}_t,\bar{\boldsymbol{\lambda}})\log p(\boldsymbol{y}_t\,|s_t=s,\boldsymbol{\lambda}) \tag{28}$$

其中,$\boldsymbol{\lambda}$ 和 $\bar{\boldsymbol{\lambda}}$ 分别为噪声和信道的新旧参数,用 $\gamma_t(s)$ 表示 t 时刻聚类 s 的后验概率:

$$\gamma_t(s)=p(s_t=s\,|\boldsymbol{y}_t,\bar{\boldsymbol{\lambda}}) \tag{29}$$

EM 算法的步骤 M 通过最大化 Q 迭代求解 $\boldsymbol{\mu}_h$ 和 $\boldsymbol{\mu}_n$ 的更新值,加入惩罚项限制更新向量与原始向量之间的距离,避免一阶泰勒估计失效,以 $\boldsymbol{\mu}_h$ 为例,更新值可写为:

$$\boldsymbol{\mu}_h=\underset{\boldsymbol{\mu}_h}{\mathrm{argmin}}((-Q)+\eta_h\|\boldsymbol{\mu}_h-\boldsymbol{\mu}_{h_0}\|_2^2),\ \eta_h>0 \tag{30}$$

辅助函数加上惩罚项之后分别对 $\boldsymbol{\mu}_h$ 和 $\boldsymbol{\mu}_n$ 求导取 0 求最值得:

$$\sum_t\sum_s\gamma_t(s)\boldsymbol{G}_s^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{y,s})+\eta_h(\boldsymbol{\mu}_h-\boldsymbol{\mu}_{h_0})=0 \tag{31}$$

$$\sum_t\sum_s\gamma_t(s)(\boldsymbol{I}-\boldsymbol{G}_s)^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_{y,s})+\eta_n(\boldsymbol{\mu}_n-\boldsymbol{\mu}_{n_0})=0 \tag{32}$$

对两个参数分别进行重估计,可分别得到 $\boldsymbol{\mu}_h$ 和 $\boldsymbol{\mu}_n$ 的更新值:

$$\boldsymbol{\mu}_h=\boldsymbol{\mu}_{h_0}+(\sum_t\sum_s\gamma_t(s)\boldsymbol{G}_s^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}\boldsymbol{G}_s+\eta_h\boldsymbol{I})^{-1}(\sum_t\sum_s\gamma_t(s)\boldsymbol{G}_s^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{x,s}-\boldsymbol{\mu}_{h_0}-g(\boldsymbol{\mu}_{x,s},\boldsymbol{\mu}_{h_0},\boldsymbol{\mu}_{n_0}))) \tag{33}$$

$$\boldsymbol{\mu}_n=\boldsymbol{\mu}_{n_0}+(\sum_t\sum_s\gamma_t(s)(\boldsymbol{I}-\boldsymbol{G}_s)^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{I}-\boldsymbol{G}_s)+\eta_n\boldsymbol{I})^{-1}(\sum_t\sum_s\gamma_t(s)(\boldsymbol{I}-\boldsymbol{G}_s)^{\mathrm{T}}\boldsymbol{\Sigma}_{y,s}^{-1}(\boldsymbol{y}_t-\boldsymbol{\mu}_{x,s}-\boldsymbol{\mu}_{h_0}-g(\boldsymbol{\mu}_{x,s},\boldsymbol{\mu}_{h_0},\boldsymbol{\mu}_{n_0}))) \tag{34}$$

1.3 参数初始值估计

1)噪声参数初始值估计

假设噪声在一个处理时段内保持平稳,在这一个处理时段内估计噪声的特征向量与对角协方差矩阵。假设平稳噪声的短时谱系数符合复高斯分布[21],且短时谱各频带相互独立,可得短时谱系数的概率密度函数为:

$$p_{N_k}(d)=\frac{1}{2\pi\sigma_k^2}\exp\left(-\frac{|d|^2}{2\sigma_k^2}\right) \tag{35}$$

式中:N_k 表示短时谱第 k 个频率采样点处的系数,可以推导得出其幅值 $|N_k|$ 符合瑞利分布:

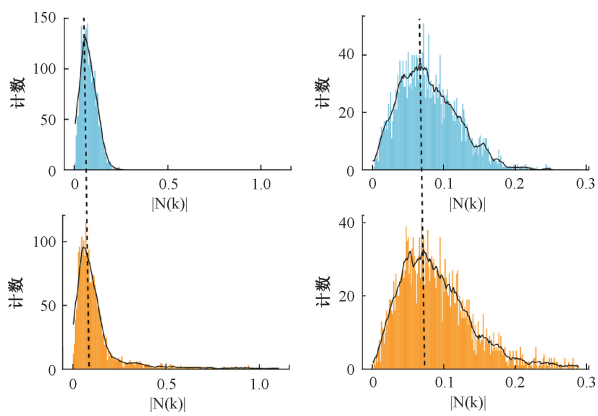
$$p_{|N_k|}(d)=\frac{d}{\sigma_k^2}\exp\left(-\frac{d^2}{2\sigma_k^2}\right) \tag{36}$$

因此 $\log|N_k|$ 符合对数瑞利分布[22],其均值与方差分别表示为 $\mu_k=\sigma_k+\ln 2/2-\gamma/2$、$\nu_k=\pi^2/24$,$\gamma$ 表示欧拉常数。因此噪声特征的均值初始值和方差可估计为:

$$\boldsymbol{\mu}_n=2\boldsymbol{C}[\mu_0,\mu_1,\cdots,\mu_k,\cdots]^{\mathrm{T}} \tag{37}$$

$$\boldsymbol{\Sigma}_n=\frac{\pi^2}{6}\boldsymbol{C}\boldsymbol{I}\boldsymbol{C}^{\mathrm{T}} \tag{38}$$

因为语音的短时幅度谱具有稀疏性,对于一段带噪语音的短时幅度谱的每个频带来说,所有时间点的取值大部分都是该频带的噪声幅值,因此每个频带处带噪语音分布与噪声分布相近。如图 2 所示是噪声为白噪声,SNR 为 5 dB 时,采样频率为 16 kHz 的带噪语音信号和对应噪声信号分帧加窗后经 640 点的离散傅里叶变换得到一系列频谱中第 16 个和第 6 个频率采样点处的幅值的分布直方图,图中曲线表示均值滤波后得到的分布曲线。根据式(36),噪声幅度谱概率分布的最大值在 σ_k 处取得,因此 σ_k 的估计通过直方图统计带噪语音第 k 个频率采样点短时幅度谱分布,经均值滤波后取分布曲线最大值处的横坐标实现。



(a) k=16 时噪声和带噪语音幅度分布 (b) k= 60时噪声和带噪语音幅度分布

(a) Amplitude distribution of noise and noisy speech When k=16 (b) Amplitude distribution of noise and noisy speech When k=60

图 2 噪声与对应带噪语音幅度谱分布

Fig. 2 Amplitude spectrum distribution of noise and corresponding noisy speech

2)信道参数估计

根据式(7)可得:

$$|H_k|^2|X_k|^2=|Y_k|^2+|N_k|^2-2\cos\beta_k|Y_k||N_k| \tag{40}$$

式中:β_k 表示复变量 Y_k 与 N_k 之间的夹角,为了方便起见,用定值 $\beta(0<\beta<1)$ 代替,所以将 $|H_k|^2$ 估计为:

$$\mathrm{E}(|H_k|^2)\approx\frac{\mathrm{E}(|Y_k|^2)+\mathrm{E}(|N_k|^2)-2\cos\beta\mathrm{E}(|Y_k|)\mathrm{E}(|N_k|)}{\mathrm{E}(|X_k|^2)} \tag{41}$$

由于 $\boldsymbol{X}$ 的概率密度分布经 GMM 建模,根据式(9)、式(16)和对数高斯分布的期望值代数表达式可得:

$$\mathrm{E}(|\boldsymbol{X}|^2)=\sum_s p(s)(\boldsymbol{C}^{-1}\boldsymbol{\mu}_{x,s}+0.5\boldsymbol{C}^{-1}\boldsymbol{\Sigma}_{x,s}(\boldsymbol{C}^{-1})^{\mathrm{T}}) \tag{42}$$

根据式(35)可得 $\mathrm{E}(|N_k|^2)=2\sigma_k^2$,$\mathrm{E}(|Y_k|^2)$ 的值由带噪语音功率谱的均值估计,最终得到信道特征参数的初始估计值为:

$$\boldsymbol{\mu}_h=\boldsymbol{C}\log(\mathrm{E}(|\boldsymbol{H}|^2)) \tag{43}$$

2 实验方案

2.1 算法的实现与评估

本文训练纯净语音 GMM 的训练数据集来自 THCHS-30 数据库[23],该数据库语音采样频率为 16 kHz。用特定说话人的时长 30 min 的语音进行训练,对这段语音进行分帧加窗并计算 25 阶 MCEP 系数,选取帧长为 512,帧移为 64,窗函数为汉明窗,计算 MCEP 系数时选取 $\alpha=0.5$。同时,用 PEFAC 算法计算语音帧的浊音概率,将语音分为浊音部分和清音/无语音段部分,用 Kmeans 算法分别分为 160 个和 40 个簇,然后以这些簇的中心为 GMM 中每个高斯成分的均值的初始值,借助 Scikit-learn 库训练协方差矩阵为对角阵,高斯分布的聚类数量为 200 的 GMM。

选用语音信号评价中的客观指标:对数似然比(log-likelihood ratio, LLR)和语音质量感知评估(perceptual evaluation of speech quality, PESQ)[24]对算法效果进行评估。LLR 是一种语音失真的客观评价方法,通过评估纯净语音信号和处理后语音信号的全极点模型之间的差异实现,数值越低代表语音质量越好。PESQ 在 2001 年成为国际电信联盟推出的 P. 862 标准,PESQ 得分与主观评测相关度较高,能较好得预测语音的主观视听感受,分值区间为-0.5~4.5,得分越高说明语音质量越好。

2.2 合成信号实验

合成信号实验用向纯净语音添加信道作用和特定信噪比的噪声形成的信号对算法进行测试。信号合成方法如图 3 所示。

以不在训练集中但属于同一说话人的语音作为测试

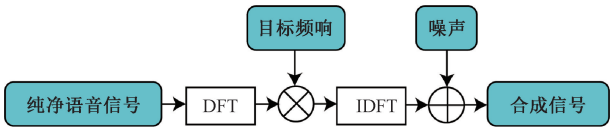
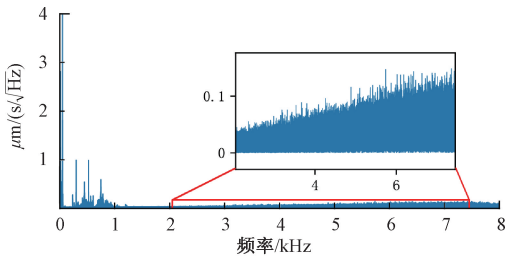


图 3　信号合成方法

Fig. 3　The way to synthesize signal

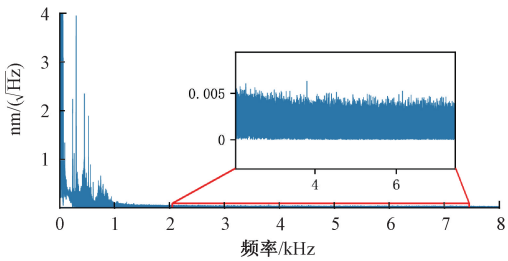
集的纯净语音信号,信道作用的添加方法如下:首先用白噪声激励,激光测振仪采集的方式测量目标物上某一测点的频响,将一段测试集纯净语音信号进行傅里叶变换后,乘以该点的幅频响应,再进行逆傅里叶变换。

用激光多普勒测振仪在测振目标上进行无语音信号的空采时,在外界环境与测振目标稳定的条件下,观察到采集到的信号存在散粒噪声——不同强度的白噪声(测量位移信号时)或幅度谱随频率线性上升的噪声(测量速度信号时)[25],此外,低频段还明显存在可能来源于目标物振动与仪器的其它噪声。静止纸盒上一点的空采速度信号和位移信号如图 4 所示。为了实验方便选取白噪声和白噪声差分噪声(幅度谱随频率线性上升的噪声)用于合成信号。



(a) 空采速度信号幅度谱

(a) Amplitude spectrum of detected velocity signal when no speech exists



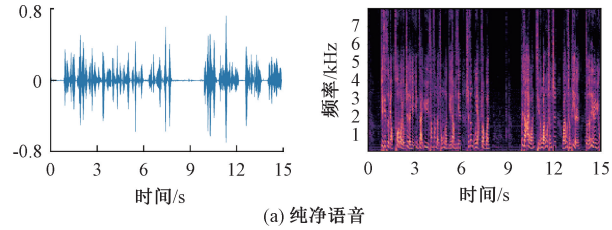
(b) 空采位移信号幅度谱

(b) Amplitude spectrum of detected displacement signal when no speech exists

图 4　纸盒上一点空采信号测量

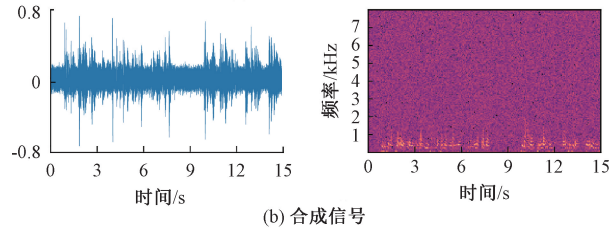
Fig. 4　Measurement on a paper box when no speech exists

分别随机选取纸盒、水瓶、窗帘上一个测点的频响添加到随机截取的长度为 15 s 同一说话人未出现在训练集中的语音信号上,再添加白噪声、白噪声差分噪声进行实验,并使得信噪比分别为 0、5、10 dB,图 5 所示为添加水瓶频响和白噪声的信噪比为 5 dB 的合成信号与对应纯净语音的对比。用 2.1 节所述指标进行评价,选用吕韬使用的相位补偿 OMLSA 算法[12]作为对比算法。



(a) 纯净语音

(a) Clean speech signal



(b) 合成信号

(b) Synthetic signal

图 5　合成信号与纯净语音对比

Fig. 5　Synthetic signal compared with clean signal

2.3　真实实验

采用 Polytec LDV(PSV-500)激光多普勒测振仪实测扬声器播放的不在训练集中但属于同一说话人的语音信号引起的不同目标物表面微振动,以此对算法性能进行验证,如图 6 所示。通过调节测振仪焦距改变回光强度和调节扬声器音量,可以模拟不同的测振距离和激励强度得到信噪比不同的观测信号。可以用 PESQ 指标对算法处理效果进行评估,此外将算法处理后语音与原语音进行互相关运算,可以实现信号波形对齐,进行波形图、语谱图和 LLR 指标的对比。

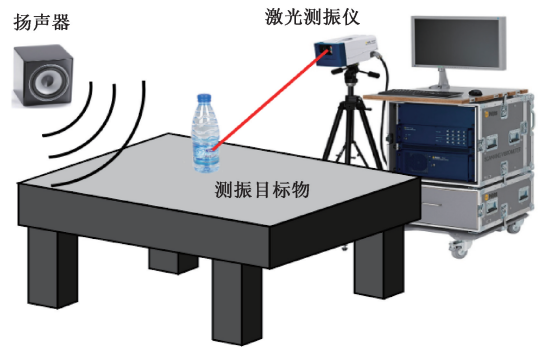


图 6　激光相干语音信号采集

Fig. 5　Laser coherent speech signal acquisition

3　结果分析

3.1　特征提取方法的验证

用激光多普勒测振仪实际采集的语音信号验证 PEFAC 算法对于浊音概率和基音频率的估计。如图 7 所示为从窗帘上探测到的一段语音信号的浊音概率与基音频率的估计结果,最上方为浊音概率 pv 的估计;中间

为观测信号波形图,浊音概率大于 0.5 和小于等于 0.5 的部分分别用深色和浅色标注,最下方为观测信号 0~1 000 Hz 频段的语谱图,浊音概率大于 0.5 的部分所估计的基音频率用实线在图中标出。总得来说,PEFAC 算法在 50~1 000 Hz 的低频噪声能量不过于高的情况下可以较好地实现激光相干探测语音的浊音概率与基音频率的估计。

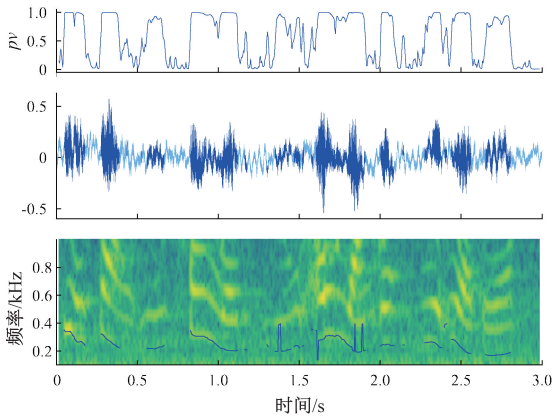


图 7 浊音概率与基音频率的估计

Fig. 7 Estimation of pitch and voiced probability

观测信号分帧加窗后每帧信号计算 25 阶 MCEP 系数,形成 26 维特征向量,用特征向量恢复该帧信号的对数幅度谱包络与原语音对数幅度谱的对比如图 8 所示,其中,频率扭曲系数 α 使得 MCEP 系数对于对数幅度谱低频部分的形状刻画更为精细。

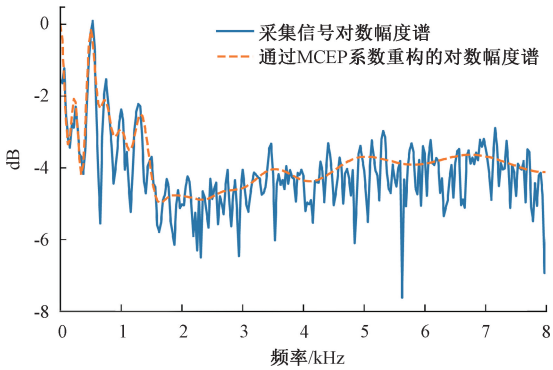


图 8 MCEP 系数恢复观测信号短时谱包络

Fig. 8 Sort-time spectrum envelope of observed signal recovered by the MCEP coefficients

3.2 合成信号实验结果分析

合成信号实验由于信道和噪声已知,可以将对信道和噪声的初始估计和自适应估计迭代后的结果与实际值进行对比。图 9 所示为添加纸盒上一点处频响和白噪声差分噪声,SNR 为 5 dB 时,信道和噪声的短时对数功率谱均值、初始估计 $\boldsymbol{\mu}_h$ 和 $\boldsymbol{\mu}_n$ 恢复的对数功率谱、噪声和信道参数经自适应估计分别经迭代 2 次后 $\boldsymbol{\mu}_h$ 和 $\boldsymbol{\mu}_n$ 恢复的对数功率谱的对比结果。可以看出信道和噪声参数的自适应估计可以使信道和噪声的估计在一定程度上更接近于真实值。由于语音能量集中在 30~3 000 Hz,语音能量过小的频段会存在 1.3 中第 2)部分信道估计方法得到的估计值偏差过大,经迭代修正也无法很好地估计的情况。

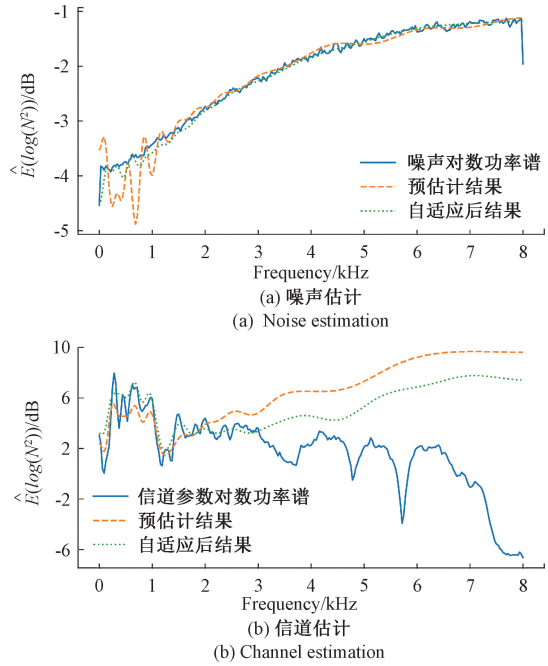


图 9 信道与噪声估计

Fig. 9 Channel and noise estimation

信道和噪声参数经自适应估计后,用 1.2 节中第 2)部分中所述 MMSE 估计方法估计纯净语音 MCEP 系数。图 10 展示了一帧带噪语音、其对应纯净语音和本文算法根据带噪语音估计的 MCEP 系数恢复的对数幅度谱包络,可见本文算法通过 MCEP 系数的估计能得到与纯净语音相近的对数幅度谱包络估计。

每种添加了不同信道和噪声的信号取 10 组用本文

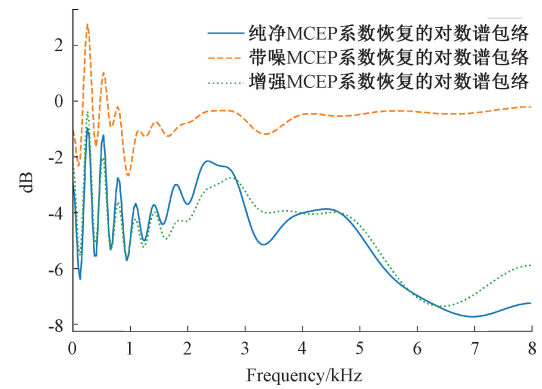


图 10 MCEP 系数估计

Fig. 10 Estimation of MCEP coefficients

算法进行处理,并用 2.1 所述指标进行评价,将评价值取平均,结果如表 1 所示,每种合成信号每种指标的 3 行数据从上到下分别为原始合成信号、相位补偿 OMLSA 算法处理后、本文算法处理后的结果,其中较好的结果经加粗标注。本文提出算法处理信号的 LLR 指标在所有情况下较好,这是由于这个算法的机制使得处理后的信号的短时对数幅度谱包络与原语音信号较为接近。在本文的测试范围内,所提出算法处理信号的 PESQ 指标显示,在有些低信噪比的情况下,相较相位补偿 OMLSA 算法本文提出算法有时能更好地提高语音质量;然而,在高信噪比情况下本文提出算法则效果欠佳,分析原因在于语音合成算法比较粗糙,且算法各模块存在太多近似环节。因此,在噪声不至于影响基音频率检测的情况下,本文算法更适用于较低信噪比时的语音信号质量增强。

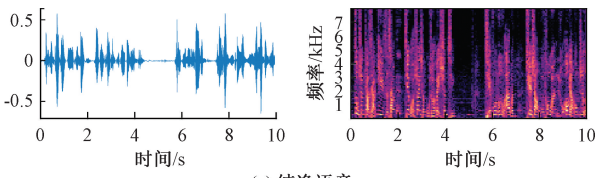
表 1　合成信号实验客观评价指标对比

Table 1　Objective measures of the proposed method against OMLSA with phase compensation

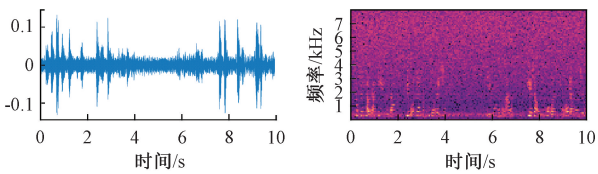
评价指标	输入 SNR/dB	信道+噪声					
		纸盒+白	纸盒+差分	水瓶+白	水瓶+差分	窗帘+白	窗帘+差分
PESQ	0	1.18	1.53	1.03	1.25	1.32	1.66
		1.39	**1.98**	1.47	1.72	1.60	**2.05**
		1.53	1.82	**1.48**	**1.80**	1.77	2.03
	5	1.43	1.72	1.32	1.56	1.64	1.91
		1.93	**2.29**	1.73	2.17	**2.10**	**2.52**
		1.75	1.95	1.83	1.90	1.99	2.16
	10	1.75	1.93	1.56	1.87	1.90	2.17
		2.37	**2.67**	**2.14**	**2.63**	**2.49**	**2.91**
		1.95	2.07	1.96	2.05	2.09	2.29
LLR	0	1.94	1.96	1.87	1.87	1.89	1.88
		1.92	1.93	1.88	1.89	1.90	1.90
		1.23	**1.03**	**1.10**	**1.03**	**1.09**	**1.01**
	5	1.90	1.94	1.84	1.86	1.86	1.84
		1.88	1.92	1.86	1.88	1.87	1.86
		1.07	**1.01**	**1.10**	**1.07**	**1.02**	**0.98**
	10	1.88	1.89	1.81	1.85	1.82	1.83
		1.85	1.86	1.82	1.87	1.83	1.85
		1.01	**1.06**	**1.04**	**1.01**	**1.01**	**0.97**

3.3　真实实验结果

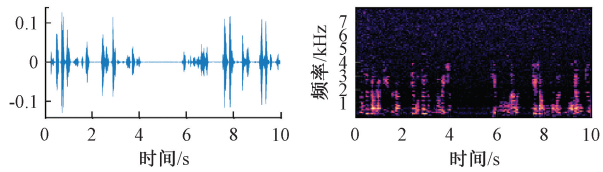
图 11 是测振目标物为纸盒时的一组实际采集信号的波形图和语谱图,将纯净语音信号、观测信号、相位补偿 OLMSA 算法增强后的信号和本文算法增强后的信号



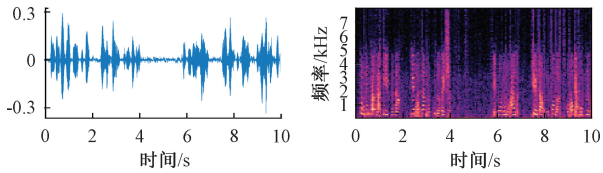
(a) 纯净语音
(a) Clean speech signal



(b) 激光相干探测语音
(b) Speech signal acquired by LDV



(c) 相位补偿OMLSA增强信号
(c) Signal enhanced by OMLSA with phase compensation



(d) 本文算法增强信号
(d) Signal enhanced by algorithm proposed in this paper

图 11　语音采集实验结果

Fig. 6　Results of speech signal acquisition experiments

对比,可以看出本文算法的去噪和均衡效果。该组观测信号、相位补偿 OLMSA 算法增强后的信号和本文算法增强后的信号 PESQ 分别 1.39、1.63 和 1.79,LLR 分别为 1.88、1.96 和 1.15。

4 结 论

为了提高激光相干语音探测的信号质量,本文创新性地提出了一种基于分析-重合成框架和噪声与信道参数自适应估计的语音增强算法。通过对观测语音进行特征提取,结合预训练的目标人物语音谱包络特征参数 GMM 和观测语音中包含的信道与噪声信息,估计观测信号谱包络特征所对应的纯净语音谱包络特征,进而用纯净语音谱包络特征的估计值进行语音重合成。合成信号实验与实际实验证明了本文提出的算法可以在一定程度上实现激光相干探测语音信号的去噪和均衡。然而该算法由于语音合成模块比较粗糙,且算法各模块存在太多近似环节,导致其几乎不可能得到质量很高的语音信号,且实验表明本文算法与传统语音增强算法一样,并不能实现语音可懂度的提高,作者将把改善这些问题作为之后的工作目标。

参考文献

[1] LYU T, GUO J, ZHANG H Y, et al. Acquirement and enhancement of remote speech signals[J]. Optoelectronics Letters, 2017, 13(4):275-278.

[2] ZHENG CH SH, ZHANG H Y, LIU W ZH, et al. Sixty years of frequency-domain monaural speech enhancement: From traditional to deep learning methods[J]. Trends in Hearing, 2023, 27:23312165231209913.

[3] GHORPADE K, KHAPARDE A. Single-channel speech enhancement using single dimension change accelerated particle swarm optimization for subspace partitioning[J]. Circuits, Systems, and Signal Processing, 2023, 42(7): 4343-4361.

[4] 李如玮,鲍长春,窦慧晶. 基于双正交小波包分解的自适应阈值语音增强[J]. 仪器仪表学报,2008(10):2135-2140.
LI R W, BAO CH CH, DOU H J. Speech enhancement using adaptive threshold based on bi-orthogonal wavelet packet decomposition[J]. Chinese Journal of Scientific Instrument, 2008(10):2135-2140.

[5] XUE W, MOORE A H, BROOKES M, et al. Speech enhancement based on modulation-domain parametric multichannel Kalman filtering[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 29: 393-405.

[6] 朱志宇. 基于高斯粒子滤波器和 TVAR 模型的语音增强技术[J]. 仪器仪表学报,2008(9):1903-1907.
ZHU ZH Y. Speech enhancement technique based on gaussian particle filter and time-varying autoregressive model[J]. Chinese Journal of Scientific Instrument, 2008(9):1903-1907.

[7] SONG Y J, MADHU N. Investigations on the optimal estimation of speech envelopes for the two-stage speech enhancement[J]. Sensors, 2023, 23(14):23146438.

[8] KWON K, SHIN J W, KIM N S. NMF-based speech enhancement using bases update[J]. IEEE Signal Processing Letters, 2014, 22(4):450-454.

[9] XIANG Y, SHI L M, HOJVANG J L, et al. A novel NMF-HMM speech enhancement algorithm based on poisson mixture model[C]. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. 2021:721-725.

[10] LIU B, TAO J H, WEN ZH Q, et al. Speech enhancement based on analysis-synthesis framework with improved parameter domain enhancement[J]. Journal of Signal Processing Systems, 2016, 82(2):141-150.

[11] 袁文浩,屈庆洋,梁春燕,等. 基于感知条件网络的可控语音增强模型[J]. 仪器仪表学报,2023,44(5):53-60.
YUAN W H, QU Q Y, LIANG CH Y, et al. Controllable speech enhancement model based on perceptual conditional network[J]. Chinese Journal of Scientific Instrument, 2023,44(5):53-60.

[12] 吕韬. 远距离激光相干语音信号侦测技术研究[D]. 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2019.
LYU T. Research on the remote laser coherent speech signal detection technology[D]. Changchun: University of Chinese Academy of Sciences(Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences), 2019.

[13] 王永彪,张文喜,王亚慧,等. 拉普拉斯分布下的 MMSE 谱减语音增强算法[J]. 计算机应用,2020,40(3):878-882.
WANG Y B, ZHANG W X, WANG Y H, et al. Speech enhancement algorithm based on MMSE spectral subtraction with Laplacian distribution[J]. Journal of Computer Applications, 2020,40(3):878-882.

[14] 王亚慧. 远距离激光相干语音探测系统解调及增强技术研究[D]. 北京邮电大学, 2022.
WANG Y H. Research on demodulation and enhancement technology of long-distance laser coherent speech detection system[D]. Beijing University of Posts and Telecommunications, 2022.

[15] ERRO D, SAINZ I, NAVAS E, et al. Harmonics plus noise model based vocoder for statistical parametric speech synthesis[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 8(2):184-194.

[16] GONZALES S, BROOKES M. PEFAC-a pitch estimation algorithm robust to high levels of noise[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014, 22(2):518-530.

[17] YOSHIMURA T, TAKAKI S, NAKAMURA K, et al. Embedding a differentiable mel-cepstral synthesis filter to a neural speech synthesis system [C]. 2023 IEEE International Conference on Acoustics, Speech and Signal Processing, 2023:11816-11820.

[18] LOWEIMI E, BARKER J, HAIN T. Channel compensation in the generalised vector Taylor series approach to robust ASR[C]. Interspeech. 2017:2466-2470.

[19] LI J Y, SELTZER M L, GONG Y F. Improvements to VTS feature enhancement[C]. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing, 2012:4677-4680.

[20] LI J Y, DENG L, YU D, et al. A unified framework of HMM adaptation with joint compensation of additive and convolutive distortions [J]. Computer Speech & Language, 2009, 23(3):389-405.

[21] NIELSEN J K, CHRISTENSEN M G, BOLDT J B. An analysis of traditional noise power spectral density estimators based on the Gaussian stochastic volatility model[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31:2299-2313.

[22] LEE H, LEE M H, YOUN S, et al. Speckle reduction via deep content-aware image prior for precise breast tumor segmentation in an ultrasound image[J]. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. 2022, 69(9):2638-2650.

[23] WANG D, ZHANG X W. THCHS-30: A free Chinese speech corpus[J]. ArXiv preprint arXiv: 1512.01882, 2015.

[24] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ) a new method for speech quality assessment of telephone networks and codecs [C]. 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, 2001, 2:749-752.

[25] JIANG L A, ALBOTA M A, HAUPT R W, et al. Laser vibrometry from a moving ground vehicle[J]. Applied Optics, 2011, 50(15):2263-2273.

作者简介

芮小博,2021 年于天津大学获得博士学位,现为天津大学副研究员、硕士生导师,研究方向为声学检测技术。

E-mail:ruixiaobo@tju.edu.cn

Rui Xiaobo received his Ph. D. degree from Tianjin University in 2021. He is currently an Associate Researcher and master supervisor with Tianjin University. His research interest is acoustic detection technology.

孔欣玥,2022 年于天津大学获得学士学位,现为天津大学在读研究生,研究方向为语音信号处理。

E-mail:kxy7372517@163.com

Kong Xinyue received her B. Sc. degree from Tianjin University in 2022. She is currently a M. Sc. candidate at Tianjin University. Her research interest is speech signal processing.

伍洲(通信作者),2019 年于中国科学院大学获得博士学位,现为中国科学院空天信息创新研究院研究员、博士生导师,研究方向为光学精密测量技术。

E-mail:wz@aircas.ac.cn

Wu Zhou (Corresponding author) received her Ph. D. degree from Chinese Academy of Sciences in 2019. She is currently a researcher and Ph. D. supervisor at the Aerospace Information Research Institute, Chinese Academy of Sciences. Her research focus is on optical precision measurement technology.