

基于联合最大后验概率的语音增强算法^①

李婉玲, 张秋菊

(江南大学 机械工程学院, 无锡 214122)
(江苏省食品先进制造装备技术重点实验室, 无锡 214122)
通讯作者: 李婉玲, E-mail: jxliwanling@163.com

摘要: 针对传统谱减法存在的算法缺陷, 提出一种基于联合最大后验概率的改进谱减法. 传统谱减法通过获取带噪语音与噪声的幅度差值, 并提取带噪语音的相位信息进行语音信号重建. 该方法因为谱相减产生“音乐噪声”, 并因为相位估计不准确, 导致低信噪比下信号增强效果不理想. 为此, 引入多频带谱减法和相位估计, 通过划分频谱, 分别在子频带进行谱减法, 有效降低“音乐噪声”的影响; 同时构建基于最大后验概率的相位估计器, 联合信号幅度函数和相位函数, 通过多次交替迭代得到相位估值. 实验结果表明, 相对于传统谱减法, 在低信噪比下该算法有效提高增强语音的质量感知和可懂度.

关键词: 语音增强; 相位估计; 最大后验概率; 语音可懂度

引用格式: 李婉玲, 张秋菊. 基于联合最大后验概率的语音增强算法. 计算机系统应用, 2018, 27(12): 163-168. <http://www.c-s-a.org.cn/1003-3254/6670.html>

Speech Enhancement Based on Joint Maximum A Posteriori Probability

LI Wan-Ling, ZHANG Qiu-Ju

(College of Mechanical Engineering, Jiangnan University, Wuxi 214122, China)
(Jiangsu Key Laboratory of Advanced Manufacturing Equipment & Technology, Wuxi 214122, China)

Abstract: In order to solve the defect of the traditional spectral subtraction algorithm, an improved spectral subtraction based on the joint maximum a posteriori probability is proposed. The traditional spectral subtraction was used to reconstruct the speech via obtaining difference of the amplitude between the noisy speech and noise and extracting the phase of the noisy speech. “Music noise” was produced by the method, and the effect of signal enhancement under low signal-to-noise ratio was not ideal because of inaccurate phase estimation. For this, the multiband spectral subtraction and phase estimation were introduced, and spectral subtraction was carried out in the subbands which were obtained by spectrum division. And it has worked well on reducing the influence of “music noise”. Meanwhile, the phase estimator based on the maximum a posteriori probability was constructed which was obtained by combining the amplitude function and the phase function of the signal and alternate iteration. The experimental results show that, compared with the traditional spectral subtraction, the proposed algorithm has performed better in terms of the quality perception and intelligibility of the enhanced speech at low signal to noise ratio.

Key words: speech enhancement; phase estimation; maximum posterior probability; speech intelligibility

1 引言

语音增强技术一直是信号处理领域的研究热点,

广泛应用于移动通信、模式识别和助听器等系统中.

语音增强旨在降低加性噪声对语音质量感知和语音可

^① 基金项目: 国家自然科学基金 (51575236)

Foundation item: National Natural Science Foundation of China (51575236)

收稿时间: 2018-05-02; 修改时间: 2018-05-24; 采用时间: 2018-06-05; csa 在线出版时间: 2018-12-03

懂度的影响,常用语音增强算法包括参数法、统计法和非模型法,其中谱减法因其良好的消噪效果且算法简单而得到广泛应用^[1-4].传统谱减法在高信噪比时具有良好的降噪表现,而在低信噪比语音中,因算法固有缺陷,增强语音存在明显的“音乐噪声”和残余噪声.Upadhyay等^[5]提出多频带谱减法,在划分的子频谱上分别进行谱减法,有效降低“音乐噪声”;Wojcicki等^[6]将噪声幅度谱与变化的相位谱相结合,提高语音质量;Mowlaee等^[7]利用基音频率和噪声信噪比信息得到相位估计器;Kulmer等^[8]通过噪声相位谱的相位分解和时间平滑获得谐波相位估计.

基于传统谱减法的语音增强系统采用不准确的噪声信号估计纯净语音幅谱,并使用带噪语音的相位进行信号重建,导致算法所得的增强语音与实际纯净语音存在较大偏差.一些学者^[6-8]基于相位在频谱上均匀分布且与幅度相互独立的假设,针对该偏差提出某种意义上的最优估计器,这些估计器在低信噪比下取得不错的相位修正效果.Kulmer等^[9]研究发现,语音增强应考虑信号的不确定性;Krawczykbecker等^[10,11]考虑相位分布的不均匀性,提出一种基于冯·米塞斯分布的相位感知信号增强模型.基于此,本文对传统谱减法进行改进,采用多频带谱减法减少“音乐噪声”,并联合相位函数和幅度函数,推导基于最大后验概率(Maximum A Posteriori probability, MAP)的最优估计器,称为联合MAP估计器.

2 基于多频带的改进谱减法

假设噪声信号为平稳加性噪声,则带噪语音信号 $x(n)$ 可用如下公式^[12]表示:

$$x(n) = s(n) + d(n) \quad (1)$$

式中, n 为采样时间标号, $s(n)$ 为纯净语音, $d(n)$ 为噪声信号,假设两者相互独立.对语音信号进行傅里叶变换,则 $X_k = A \exp(j\alpha_k)$, $S_k = B \exp(j\beta_k)$, $D_k = C \exp(j\gamma_k)$ 分别表示带噪信号、纯净信号和噪声的傅里叶形式,三者之间的关系^[13]可表示为:

$$S_k = X_k - D_k = A \exp(j\alpha_k) - C \exp(j\gamma_k) \quad (2)$$

将其扩展到功率谱域,则纯净语音的信号谱可表示为:

$$|S_k|^2 = |X_k|^2 - |D_k|^2 - 2\text{Re}\{S_k D_k^*\} \quad (3)$$

假设噪声信号均值为零,则式(3)的乘积项为零.

为消除谱相减残留的功率谱分量(音乐噪声),引入多频带谱减法^[5],划分语音信号频谱,并对各子频带的语音频谱独立进行谱减法,改进谱减法对纯净语音幅度的估值为:

$$|\hat{S}_i(\omega_k)| = \left(|X_i(\omega_k)|^2 - \alpha_i \delta_i |\hat{D}_i(\omega_k)|^2 \right)^{1/2}, \quad b_i \leq \omega_k \leq e_i \quad (4)$$

式中, i 为子频带序号, α 为自适应过减系数,与信噪比相关, δ 为子频带噪声抑制因子, b_i 、 e_i 为子频带的始末频率,噪声为语音间隙计算所得的数学期望,是一个估值.实验表明,该算法可以有效抑制“音乐噪声”.

谱减法基于人耳相干特性^[14]进行噪声相位估计,采用带噪语音的相位近似描述噪声相位,导致纯净语音的相位估值与真实值之间存在较大误差.带噪语音与纯净语音的相位偏差用 φ_{dev} 表示:

$$\varphi_{dev} = \alpha - \beta \quad (5)$$

从图1可以看到,不准确的相位估计对信号幅频估值产生较大影响,信号的幅度与相位不完全独立,因此,在进行相位预估时,应充分考虑信号幅度的影响.

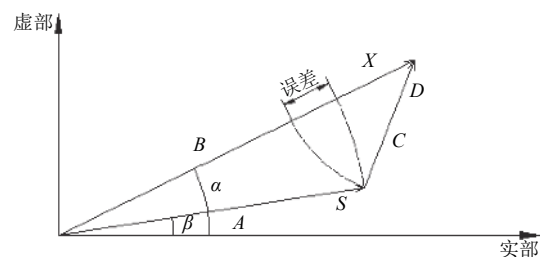


图1 信号的相位偏差图

3 基于联合MAP的改进谱减法

谱减法的本质是一种似然估计,采用带噪语音的相位近似噪声相位和纯净语音相位,造成相位偏差,影响低信噪比下语音增强效果.本文基于此,采用最大后验概率法,提出联合相位函数和幅度函数的MAP估计器对相位进行估计^[15],并与多频带谱减法结合,取得良好的增强效果.

3.1 联合MAP估计

基于相位分布不均匀性,本文提出联合MAP估计器,通过交替迭代相位估计函数和幅度估计函数,得到纯净语音的最优相位估计器.

假设带噪信号为已知信号,可得纯净语音的幅度和相位估计为:

$$\{\hat{B}, \hat{\beta}\} = \operatorname{argmax}_{B, \beta} \frac{p(X|B, \beta)p(B, \beta)}{p(X)} \quad (6)$$

假设条件概率密度函数服从以纯净语音为中心的高斯分布, 则信号的条件概率密度为:

$$p(X|B, \beta) = \frac{1}{\pi\sigma_d^2} \exp\left(-\frac{|X - Be^{j\beta}|^2}{\sigma_d^2}\right) \quad (7)$$

假设信号谱的幅度服从 Γ 分布, 其先验密度函数为:

$$P(B) = \frac{\mu^{\nu+1}}{\Gamma(\nu+1)\sigma_s^{\nu+1}} \exp\left(-\frac{\mu B}{\sigma_s}\right) \quad (8)$$

式中, ν 为形状参数, μ/σ_s 为尺度参数。

考虑到语音信号的不确定性, 本文假设信号相位服从冯·米塞斯分布, 为不均匀分布, 则相位先验密度函数为:

$$p(\beta) = \frac{\exp(\kappa\cos(\beta - \beta_\mu))}{2\pi I_0(\kappa)} \quad (9)$$

式中, β_μ 为循环均值, κ 为集中密度, $I_0(\cdot)$ 为零阶修正贝塞尔函数, 联立方程组 (6)~(9), 可得纯净语音幅度和相位的联合 MAP 估计为:

$$\{\hat{B}^{MAP}, \hat{\beta}^{MAP}\} = \operatorname{argmax}_{B, \beta} L_1(B, \beta) \quad (10)$$

其中:

$$L_1(B, \beta) = B^\nu \exp\left(-\frac{|X - Be^{j\beta}|^2}{\sigma_d^2} - \frac{\mu B}{\sigma_s} + \kappa\cos(\beta - \beta_\mu)\right) \quad (11)$$

为简化偏导计算, 忽略与自变量无关的函数 $p(X)$, 并对 $L_1(B, \beta)$ 取对数似然函数, 可得:

$$L_2(B, \beta) = \nu \log(B) - \frac{|X - Be^{j\beta}|^2}{\sigma_d^2} - \frac{\mu B}{\sigma_s} + \kappa\cos(\beta - \beta_\mu) \quad (12)$$

为获取最大后验相位估计, 取对数似然函数对相位的一阶偏导为零, 可得:

$$\frac{\partial L_2(B, \beta)}{\partial \beta} = \frac{2AB}{\sigma_d^2} \sin(\alpha - \beta) - \kappa \sin(\beta - \beta_\mu) = 0 \quad (13)$$

由式 (13) 和余弦定理可得, 基于联合 MAP 估计器的 β 估计为:

$$\hat{\beta}^{MAP} = g(B) = \tan^{-1}\left(\frac{2AB\sin\alpha + \kappa\sigma_d^2\sin\beta_\mu}{2AB\cos\alpha + \kappa\sigma_d^2\cos\beta_\mu}\right) \quad (14)$$

式中, 相位估计与信噪比 A/σ_d^2 、先验统计参数 β_μ 和集

中密度 κ 相关。

取对数似然函数对幅度的一阶偏导为零:

$$\frac{\partial L_2(B, \beta)}{\partial B} = \frac{\nu}{B} - \frac{2B - 2A\cos(\alpha - \beta)}{\sigma_d^2} - \frac{\mu}{\sigma_s} = 0 \quad (15)$$

即:

$$B^2 - \left(A\cos(\alpha - \beta) - \frac{\mu\sigma_d^2}{2\sigma_s}\right)B - \frac{\nu\sigma_d^2}{2} = 0 \quad (16)$$

由上式计算可得幅度 B 的估计值为:

$$\hat{B}^{MAP} = f(\beta) = \left(A\cos(\alpha - \beta) - \frac{\mu\sigma_d^2}{2\sigma_s}\right) + \sqrt{\left(A\cos(\alpha - \beta) - \frac{\mu\sigma_d^2}{2\sigma_s}\right)^2 + 2\nu\sigma_d^2} \quad (17)$$

设 m 为迭代次数, 则联合 MAP 估计器中:

$$\begin{cases} \hat{\beta}_{m+1} = f(\hat{\beta}_m, \hat{B}_m) \\ \hat{B}_{m+1} = g(\hat{\beta}_m, \hat{B}_m) \end{cases} \quad (18)$$

设相邻迭代间复值偏差的功率谱为 E :

$$E_{m+1} = \sum \left| \hat{B}_{m+1} \exp(j\hat{\beta}_{m+1}) - \hat{B}_m \exp(j\hat{\beta}_m) \right|^2 \quad (19)$$

随着迭代次数的增加, 偏差功率谱 E 逐渐衰减, 设 ΔE 的阈值为 -20 dB 时, 迭代终止, 此时迭代次数 $m=2$, 如图 2 所示。从图中可以看出能量逐步衰减, 算法终将达到平衡状态。根据李雅普诺夫第二法, 该系统渐近稳定。

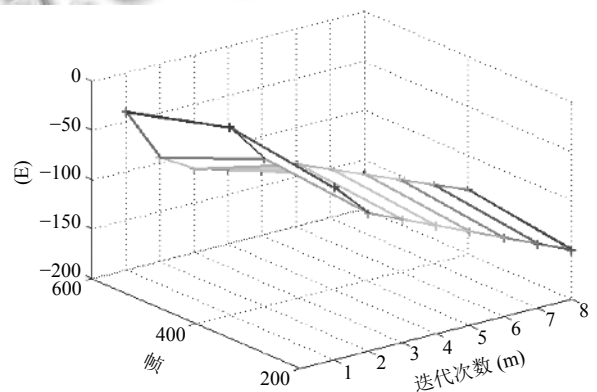


图 2 联合 MAP 交替迭代情况

3.2 基于联合 MAP 的多频带谱减法

本文算法步骤如下:

- (1) 进行预处理并根据傅里叶变换, 转换带噪语音

到频域, 计算其功率谱;

(2) 划分频带并进行联合 MAP 估计, 估计各个子频带的相位谱;

(3) 计算子频带噪声和带噪语音的功率谱;

(4) 按式 (4) 分别计算各个子频带增强语音的幅度谱;

(5) 重建信号并进行反傅里叶变换.

4 实验与分析

本文实验数据为 Cool Edit 软件在安静环境下录制的 50 句语音, 选用 noiseX-92 数据库中的白噪声 (white)、粉红噪声 (pink) 和人群嘈杂声 (babble) 作为实验噪声数据. 采样频率均为 8 kHz. 选择纯净语音中的一句分别加入三种噪声构成信噪比为 0 dB 的带噪语音, 语句内容为“江南大学是一座美丽的校园”, 共 12 个字. 采用 MATLAB 进行改进谱减法和本文算法的对比实验, 实验结果如图 3~图 5 所示. 实验结果表明, 多频带谱减法有效抑制“音乐噪声”, 在低信噪比情况下, 本文算法所得波形图与纯净语音接近, 语音增强效果明显.

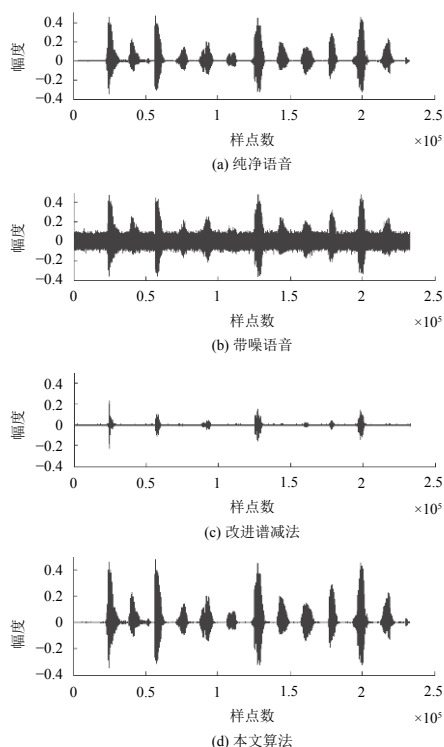


图 3 white

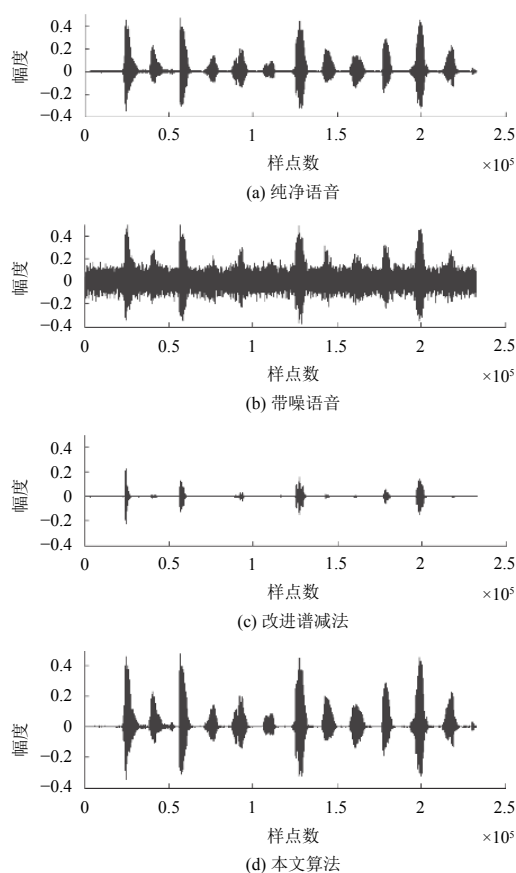


图 4 pink

将纯净语音与不同噪声进行混合, 带噪信号的信噪比分别为 0、5、15 dB, 取各语句实验结果的平均值作为评价数据, 引入语音质量感知评价 (PESQ) 对增强信号的语音质量进行评价, 如图 6 所示. PESQ 在 1.0~4.5 范围内产生得分, 分值越高说明语音的质量越好. 同时采用 STOI 对增强信号的语音可懂度进行评价, 验证降噪后语音信号的可懂度. 在 (0,1) 范围内, STOI 算法所得数值越大, 增强信号的可懂度越高, 如图 7 所示.

从图 6 可以看到, 在低信噪比情况下, 本文算法的 PESQ 得分明显高于传统谱减法和改进谱减法, 有效降低相位偏差导致的残留噪声, 增强信号的语音质量较高. 图 7 显示的 STOI 测评结果表明本文算法提高了带噪语音的可懂度, 在低信噪比情况下有较优的表现. 实验结果显示, 本文算法在取得良好降噪效果的同时, 保证了语音信息量, 增强了语音可懂度.

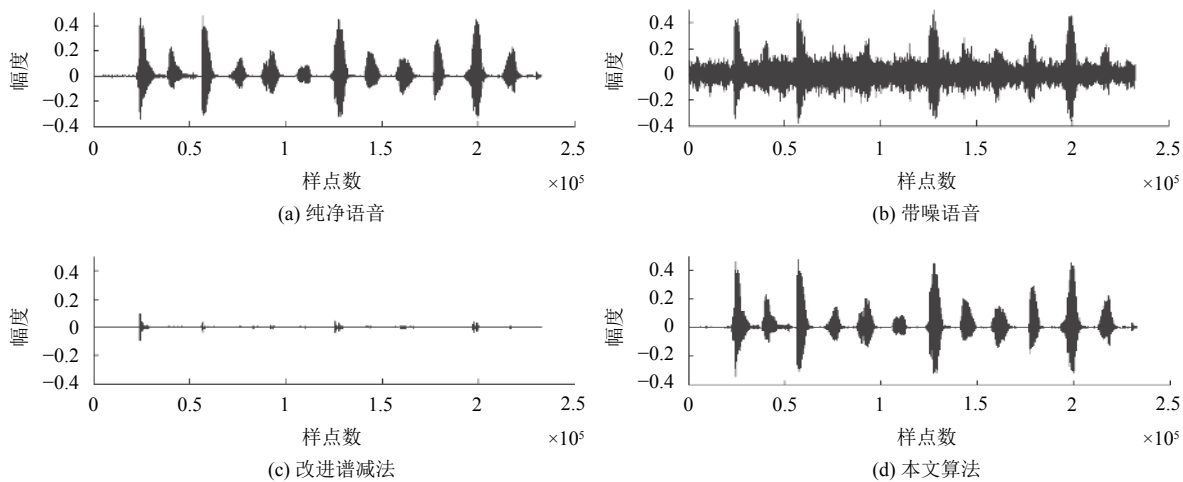


图5 babble

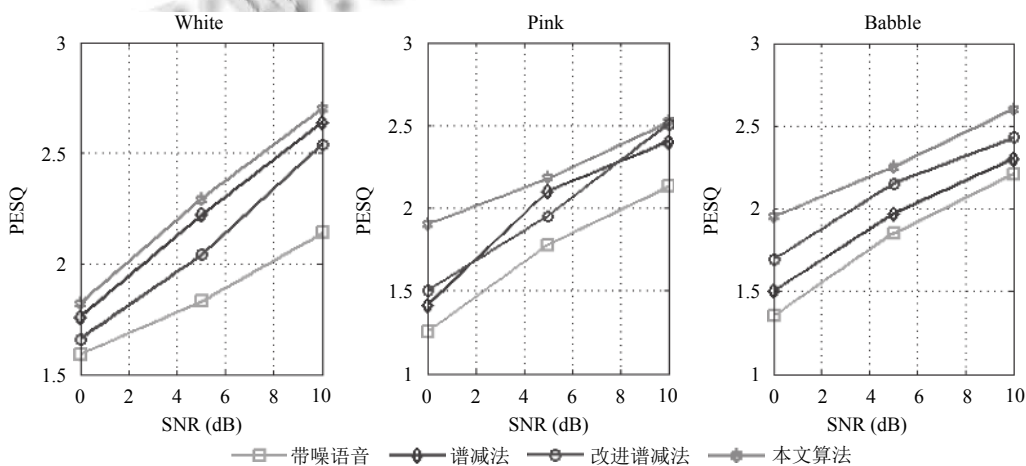


图6 不同背景噪声下各算法的 PESQ 得分

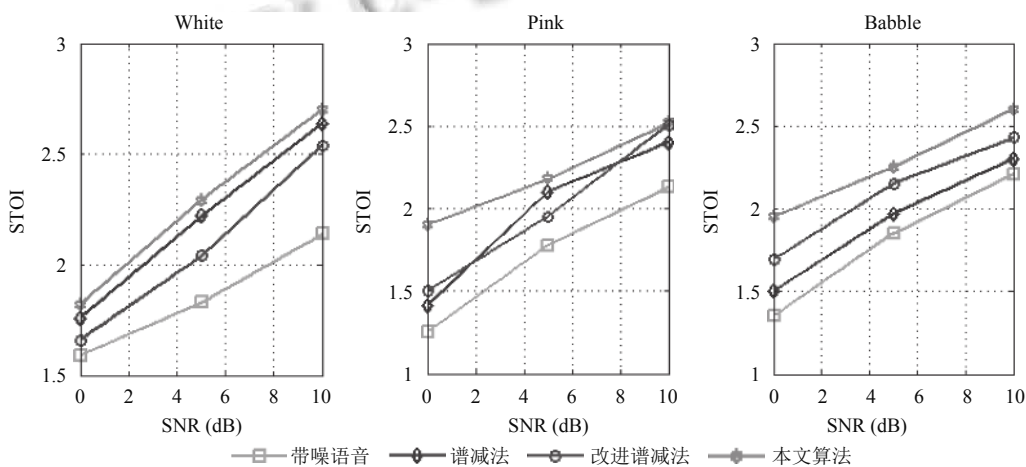


图7 不同背景噪声下各算法的 STOI 值

5 结论

针对传统谱减法因为谱相减和相位估计偏差带来的“音乐噪声”和残留噪声问题,本文提出一种基于联合最大后验概率的谱减语音增强算法.采用冯·米塞斯分布体现信号相位分布的不均匀性,并假设条件概率密度服从高斯分布,信号幅度服从 Γ 分布,根据最大后验概率得到幅度和相位的估值,并进行交替迭代,得到联合MAP的相位估计.将联合MAP估计与多频带谱减法结合,重建增强语音.实验结果表明,相比于传统谱减法和多频带谱减法,在低信噪比情况下,本文算法具有更好的语音质量感知和更高的语音可懂度,有效提高语音增强效果.

参考文献

- 1 韦高梧,冯祖勇.基于去噪技术的DSP语音识别系统设计.传感器与微系统,2017,36(1):108–111.
- 2 Hendriks RC, Gerkmann T, Jensen J. DFT-domain based single-microphone noise reduction for speech enhancement: A survey of the state of the art. Synthesis Lectures on Speech and Audio Processing, 2013, 9(1): 1–80.
- 3 Kleijn WB, Crespo JB, Hendriks RC, *et al.* Optimizing speech intelligibility in a noisy environment: A unified view. IEEE Signal Processing Magazine, 2015, 32(2): 43–54. [doi: 10.1109/MSP.2014.2365594]
- 4 Loizou PC. Speech enhancement: Theory and practice. Boca Raton, FL, USA: CRC Press, 2013.
- 5 Upadhyay N, Karmakar A. An improved multi-band spectral subtraction algorithm for enhancing speech in various noise environments. Procedia Engineering, 2013, 64: 312–321. [doi: 10.1016/j.proeng.2013.09.103]
- 6 Wojcicki K, Milacic M, Stark A, *et al.* Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement. IEEE Signal Processing Letters, 2008, 15: 461–464. [doi: 10.1109/LSP.2008.923579]
- 7 Mowlaee P, Kulmer J. Harmonic phase estimation in single-channel speech enhancement using phase decomposition and SNR information. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(9): 1521–1532. [doi: 10.1109/TASLP.2015.2439038]
- 8 Kulmer J, Mowlaee P. Phase estimation in single channel speech enhancement using phase decomposition. IEEE Signal Processing Letters, 2015, 22(5): 598–602. [doi: 10.1109/LSP.2014.2365040]
- 9 Mowlaee P, Kulmer J. Phase estimation in single-channel speech enhancement: Limits-potential. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(8): 1283–1294. [doi: 10.1109/TASLP.2015.2430820]
- 10 Krawczyk-Becker M, Gerkmann T. An evaluation of the perceptual quality of phase-aware single-channel speech enhancement. The Journal of the Acoustical Society of America, 2016, 140(4): EL364–EL369. [doi: 10.1121/1.4965288]
- 11 Krawczyk-Becker M, Gerkmann T. On MMSE-based estimation of amplitude and complex speech spectral coefficients under phase-uncertainty. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(12): 2251–2262. [doi: 10.1109/TASLP.2016.2602549]
- 12 Kulmer J, Mowlaee P. Harmonic phase estimation in single-channel speech enhancement using von mises distribution and prior SNR. Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. Brisbane, QLD, Australia. 2015. 5063–5067.
- 13 杜志然,周萍,景新幸,等.基于谱熵的耳语音增强研究.传感器与微系统,2012,31(6):69–72. [doi: 10.3969/j.issn.1000-9787.2012.06.021]
- 14 吴进.语音信号处理实用教程.北京:人民邮电出版社,2015.287–298.
- 15 Mowlaee P, Stahl J, Kulmer J. Iterative joint MAP single-channel speech enhancement given non-uniform phase prior. Speech Communication, 2017, 86: 85–96. [doi: 10.1016/j.specom.2016.11.008]