

◎理论与研发◎

改进相位谱补偿的语音增强方法

吉慧芳, 贾海蓉, 王 雁

太原理工大学 信息与计算机学院, 太原 030024

摘 要:针对传统单通道语音增强方法中用带噪语音相位代替纯净语音相位重建时域信号,使得语音主观感知质量改善受限的情况,提出了一种改进相位谱补偿的语音增强算法。该算法提出了基于每帧语音输入信噪比的 Sigmoid 型相位谱补偿函数,能够根据噪声的变化来灵活地对带噪语音的相位谱进行补偿;结合改进 DD 的先验信噪比估计与语音存在概率算法(SPP)来估计噪声功率谱;在维纳滤波中结合新的语音存在概率噪声功率谱估计与相位谱补偿来提高语音的增强效果。相比传统相位谱补偿(PSC)算法而言,改进算法可以有效抑制音频信号中的各类噪声,同时增强语音信号感知质量,提升语音的可懂度。

关键词:相位谱补偿;功率谱估计;先验信噪比;语音增强

文献标志码:A **中图分类号:**TN912.35 **doi:**10.3778/j.issn.1002-8331.1805-0454

吉慧芳, 贾海蓉, 王雁. 改进相位谱补偿的语音增强方法. 计算机工程与应用, 2019, 55(8):48-52.

Ji Huifang, Jia Hairong, Wang Yan. Speech enhancement method for improving phase spectrum compensation. Computer Engineering and Applications, 2019, 55(8):48-52.

Speech Enhancement Method for Improving Phase Spectrum Compensation

Ji Huifang, Jia Hairong, Wang Yan

College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

Abstract: Aiming to the problem that the clean speech phase is replaced by the noisy speech phase when reconstructing the waveform in the traditional single-channel speech enhancement methods, which leads to the poor subjective perception quality of the speech, a speech enhancement algorithm with improved phase spectrum compensation is proposed. Firstly, the Sigmoid phase-spectral compensation function which is based on the signal to noise ratio of each frame input speech is presented in this paper, it can flexibly compensate the phase spectrum of the noisy according to the change of noise. Next, it estimates noise power spectrum through combining a priori SNR estimation of the improved DD and Speech Presence Probability algorithm(SPP). Finally, Wiener filtering is applied to improve speech enhancement effect by combining the new speech presence probability noise power spectrum estimation and phase spectrum compensation. Compared to the traditional Phase Spectrum Compensation(PSC) algorithm, the improved algorithm can effectively suppress the noise in the audio signal, and then enhance the perceived quality of the speech and improve speech intelligibility.

Key words: phase spectrum compensation; power spectrum estimation; priori signal-to-noise ratio; speech enhancement

1 引言

在正常的语音通信、听力辅助和自动语音识别中,由于语音信号受到不同类型的背景噪声干扰会严重退

化。因此,从退化语音中去除噪声成分一直是研究的主要目的。当前,大多数单通道语音增强方法主要集中在从原始的带噪语音信号中减去噪声的估计量或功率,同

基金项目:国家自然科学基金(No.61371193);山西省自然科学基金(No.201701D121058)。

作者简介:吉慧芳(1993—),女,硕士研究生,研究领域为语音信号处理;贾海蓉(1977—),通讯作者,女,博士研究生,副教授,研究领域为语音信号处理、心理声学模型, E-mail: helenjia722@163.com;王雁(1994—),女,硕士研究生,研究领域为语音信号处理。

收稿日期:2018-05-29 **修回日期:**2018-10-22 **文章编号:**1002-8331(2019)08-0048-05

CNKI网络出版:2018-11-16, <http://kns.cnki.net/kcms/detail/11.2127.TP.20181115.1038.002.html>

时保持带噪语音相位信息的完整性。这是因为早期的研究表明相位谱在信噪比较高的情况下感知效果不明显。最近研究发现相位谱中也包含了很多与语音可懂度相关的信息,这些信息对于语音增强起到一定的作用^[1-2]。Mowlae和Saeidi^[3]提出了一种闭环迭代相感知的语音增强方法,利用几何约束(Mowlae et al., 2012)和相感知幅度估计器(Gerkmann和Krawczyk, 2013)来估计相位,在降噪的同时提高了语音信号质量;Stephen So等人^[4]结合卡尔曼滤波和PSC算法得到语音与噪声对数谱和语音相位谱的时间/帧间的非线性关系,提高了不同信噪比下的语音增强效果。Dang X等^[5]通过高分辨率的STFT估计出相位谱的近似值,提出了一种语音增强的相位谱估计方法,实验结果表明将相位谱补偿与传统语音增强方法相结合,能够有效且显著地提升语音质量和清晰度。Mowlae P等^[6]提出了一种结合非均匀相位分布假设和最大后验幅度的相位估计器(ijMAP),实验结果表明该方法能够有效地从带噪语音中恢复纯净语音的幅度和相位;Md Tauhidul Islam等^[7]应用信噪比相关方法确定相位谱上的补偿量,修正带噪语音的相位谱^[8],与谱减法结合后在不同信噪比下的性能优于传统语音增强方法。

在此基础上,提出了一种改进PSC算法。针对传统的相位谱补偿算法中补偿因子固定,无法对含噪语音相位谱灵活补偿的问题,提出了一种基于每帧语音输入信噪比的Sigmoid型相位谱补偿函数,并通过维纳滤波结合新的语音存在概率噪声功率谱估计算法来增强语音。

2 传统相位谱补偿语音增强算法(PSC)

假设 $x(t)$ 表示纯净语音, $v(t)$ 表示平稳加性高斯噪声,且 $x(t)$ 与 $v(t)$ 相互独立,则含噪语音 $y(t)$ 的时域表达式为:

$$y(t) = x(t) + v(t) \quad (1)$$

对式(1)做短时傅里叶变换,其频域表示式为:

$$Y(n, k) = \sum_{m=-\infty}^{\infty} y(m) \omega(n-m) \exp(-j2\pi km/N) \quad (2)$$

其中, k 为频率索引, n 为帧索引, N 为DFT帧大小, $\omega(n)$ 为窗函数。含噪语音 $Y(n, k)$ 的极坐标形式为:

$$Y(n, k) = |Y(n, k)| \exp(j\angle Y(n, k)) \quad (3)$$

其中, $|Y(n, k)|$ 为带噪语音信号第 n 帧第 k 个频率点的幅度谱, $\angle Y(n, k)$ 为带噪语音信号第 n 帧第 k 个频率点的相位谱。传统PSC算法中,相位谱补偿函数的表示式是:

$$\Lambda(n, k) = \lambda \psi(k) |\hat{D}(n, k)| \quad (4)$$

其中, λ 是补偿因子,文献[9]中选取 $\lambda = 3.14$ 为其最优值; $\psi(k)$ 为判决函数,其表达式为:

$$\psi(k) = \begin{cases} 1, & 0 < k/N < 0.5 \\ -1, & 0.5 < k/N < 1 \\ 0, & \text{其他} \end{cases} \quad (5)$$

式(4)中 $|\hat{D}(n, k)|$ 是估算出的第 n 帧第 k 个频率点的噪声幅度值。补偿后的频谱表达式为:

$$Y_{\Lambda}(n, k) = Y(n, k) + \Lambda(n, k) \quad (6)$$

其中, $Y(n, k)$ 为带噪语音信号第 n 帧第 k 个频率点的频谱, $\Lambda(n, k)$ 为相位谱补偿函数。对补偿后的频谱取相位得到相位谱:

$$\angle Y_{\Lambda}(n, k) = \arg[Y_{\Lambda}(n, k)]$$

其中, $\arg(\cdot)$ 表示取复数幅角函数。将补偿后的相位谱与短时傅里叶变换的幅度谱相结合,得到语音增强后的频谱表达式为:

$$S_{\Lambda}(n, k) = |Y(n, k)| \exp(j\angle Y_{\Lambda}(n, k)) \quad (7)$$

3 改进相位谱补偿的语音增强算法

文中主要围绕新的Sigmoid型相位谱补偿函数、新的改进DD先验信噪比估计、新的基于语音存在概率(SPP)的噪声估计算法进行算法描述,具体算法框架如图1所示。

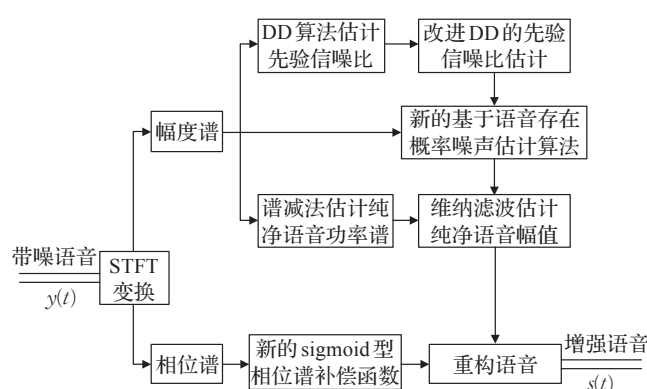


图1 改进算法的结构图

3.1 新的相位谱补偿函数

传统相位谱语音增强算法增强语音主要依赖于补偿函数 $\Lambda(n, k)$,而补偿因子 λ 是一个固定经验值,不能随噪声变化而灵活变化,因此在实际中PSC算法无法获得最优值。因此,提出一种新的相位谱补偿函数,对 $\Lambda(n, k)$ 中的补偿因子 λ 进行改进,将其设置为一个随含噪语音信噪比变化而相应变化的Sigmoid型函数,该函数的表达式为:

$$\lambda_{\text{new}} = c \left[\frac{1}{1 + e^{-\frac{(|Y(n, k)| - |D(n, k)|)^2}{|D(n, k)|^2}}} \right] \quad (8)$$

其中, c 为固定经验值,取值3.5; $|D(n, k)|$ 为噪声信号第 n 帧第 k 个频率点处的幅度谱。将 λ_{new} 代入式(4)中,计算出新的相位谱补偿函数表示式是:

$$\Lambda_{\text{new}}(n, k) = c \left[\frac{1}{1 + e^{-\frac{(|Y(n, k)| - |D(n, k)|)^2}{|D(n, k)|^2}}} \right] \psi(k) |\hat{D}(n, k)| \quad (9)$$

带噪语音的能量基本集中于低频和中值频点处,特别是当信号为浊音帧时,因此在这些区域信噪比非常高。检测到这些区域,为了快速跟踪语音的变化,权值因子可以适当设置为较小的值。由式(9)可知,由于Sigmoid函数随自变量单调递增的性质,在语音存在的区域信噪比很高补偿因子会相对较小,因此可以跟踪突然的信噪比变化,对含噪语音的频谱进行补偿;反之类似。对结合式(9)与式(6)得到的频谱取相位可得新的相位谱为:

$$\angle Y_{\text{new}}(n, k) = \arg[Y_{\text{new}}(n, k)] = \arg[Y(n, k) + \Lambda_{\text{new}}(n, k)] \quad (10)$$

3.2 新的改进DD先验信噪比估计结合基于语音存在概率的噪声估计算法

为实现纯净语音信号的恢复,需要从带噪语音信号中估计出谱幅值和相位信息^[10]。上述相位谱补偿算法中,式(7)用含噪语音幅度谱直接代替纯净语音幅度谱与补偿后的相位谱叠加^[11],且式(4)中用带噪语音的幅度谱替代噪声的幅度谱^[12],并没有估计噪声幅度谱,这样会使语音增强效果不佳。因此,就需要从带噪语音信号中估计出噪声功率谱从而得到纯净语音信号的幅度谱。本文提出了一种新的改进决策导向(Decision-Directed, DD)结合基于语音存在概率的噪声估计算法^[13],估计噪声功率谱用一个后验子带信噪比计算,基于该子带的后验信噪比的一个Sigmoid型权值用于增强听觉域中的语音频谱。具体步骤如下:

(1)根据贝叶斯公式计算出语音存在下的后验概率 $P(H_1|Y)$ 。用 H_1 代表语音存在,用 H_0 代表语音缺失,进而得到 $P(H_1|Y)$:

$$P(H_1|Y) = \frac{P(H_1)P(Y|H_1)}{P(H_1)P(Y|H_1) + P(H_0)P(Y|H_0)} \quad (11)$$

式(11)中假设语音存在和语音缺失的概率相等,即 $P(H_1) = P(H_0) = 0.5$ 。由于STFT系数服从复高斯分布,概率 $P(Y|H_1)$ 和 $P(Y|H_0)$ 可近似表示为:

$$P(Y|H_m) = \frac{1}{\pi |\hat{D}(n, k)|^2 (1 + \xi_{H_m})} \exp\left(-\frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2 (1 + \xi_{H_m})}\right) \quad (12)$$

其中, $m=0, 1$; ξ 为先验信噪比,定义为:

$$\xi = \frac{P_s(n, k)}{P_n(n, k)}$$

其中, $P_s(n, k)$ 和 $P_n(n, k)$ 分别为语音的DFT幅度和噪声方差, n 表示帧索引, k 表示频率索引。

当语音缺失时,噪声功率根据当前含噪语音第 n 帧第 k 个频率点功率 $|Y(n, k)|^2$ 和之前噪声第 $n-1$ 帧第 k 个频率点功率的估计值 $|\hat{D}(n-1, k)|^2$ 更新,而当语音存在时则使用之前第 $n-1$ 帧第 k 个频率点的估计。文献[14]中先验信噪比是一个经验常数 $10 \lg \xi_{H_1} = 15 \text{ dB}$

而不是随噪声变化灵活变化的值,不能跟踪突然的噪声起伏。故引入新的改进决策导向(DD)的先验信噪比估计算法,即

$$\hat{\xi}_{H_1} = \mu G \frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2} + (1 - \mu) \left[\frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2} - 1 \right] \quad (13)$$

其中 μ 为一个基于后验信噪比的Sigmoid型权值,其表达式为:

$$\mu = \frac{1}{1 + \exp\left[-b \left(\frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2} - 1 \right)\right]} \quad (14)$$

G 为增益函数,由先验信噪比计算得到^[15],其计算式为:

$$G = \frac{\xi_{H_1}}{1 + \xi_{H_1}} \quad (15)$$

其中 ξ_{H_1} 为DD算法估计的先验信噪比。DD算法是由Ephraim和Malah提出的基于前一帧语音先验信噪比和当前帧语音后验信噪比估计的加权求和的决策导向(Decision-Directed)方法。其表达式为:

$$\xi_{H_1} = \alpha \frac{|Y(n-1, k)|^2}{|\hat{D}(n, k)|^2} + (1 - \alpha) \max\left[\frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2} - 1, 0\right] \quad (16)$$

其中 α 为时频相关平滑因子,取 $\alpha = 0.5$ 为其经验值。结合式(11)~式(13),得到一种新的语音存在下的后验概率,即

$$P(H_1|Y) = \frac{\frac{1}{1 + \hat{\xi}_{H_1}} \exp\left[-\frac{1}{1 + \hat{\xi}_{H_1}} \frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2}\right]}{1 + \frac{1}{1 + \hat{\xi}_{H_1}} \exp\left[-\frac{1}{1 + \hat{\xi}_{H_1}} \frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2}\right]} = \left[1 + (1 + \hat{\xi}_{H_1}) \exp\left[-\frac{|Y(n, k)|^2}{|\hat{D}(n, k)|^2} \frac{\hat{\xi}_{H_1}}{1 + \hat{\xi}_{H_1}}\right] \right]^{-1} \quad (17)$$

(2)对噪声功率谱进行初步估计:

$$|\tilde{N}|^2 = P(H_0|Y) |Y(n, k)|^2 + P(H_1|Y) |\hat{D}(n, k)|^2 \quad (18)$$

当 $PH_{1_mean} > 0.9$ 时 $P(H_1|Y) = PH_{1_mean}$ 其中:

$$PH_{1_mean} = (1 - I) \times PH_{1_mean} + I \times P(H_1|Y)$$

I 为语音存在决策,其表达式为:

$$I = \begin{cases} 1, & P(H_1|Y) > \text{mean}(P(H_1|Y)) \\ 0, & P(H_1|Y) < \text{mean}(P(H_1|Y)) \end{cases} \quad (19)$$

(3)最后,更新噪声功率谱估计:

$$|\hat{D}(n, k)|^2 = \beta |\hat{D}(n-1, k)|^2 + (1 - \beta) |N(n, k)|^2 \quad (20)$$

其中, β 是平滑系数,选取0.9为其经验常数。通过以上步骤计算出每一帧噪声的功率谱估计值 $|\hat{D}(n, k)|^2$ 。

为了评估该基于语音存在概率噪声估计方法的性能,将其和传统方法进行比较。实验中,语音分帧后的第1帧语音的后验概率初始值为 $10 \lg \xi_{H_1} = 15 \text{ dB}$,之后

结合式(17)和式(19)对噪声功率谱进行初步估计,再由式(20)对噪声功率谱进行计算并更新。选自NOIZEUS语音库的纯净语音,Noise-92语音库的F16噪声用于测试对比,以SP01为例,含噪声语音输入信噪比设置为5 dB,用噪声功率谱估计值曲线图来对噪声估计方法的性能进行评价如图2所示。

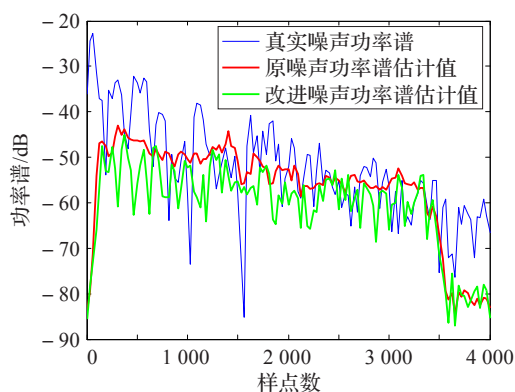


图2 真实噪声功率谱与噪声功率谱估计值曲线图

由图可知,不同于传统的语音存在概率噪声估计方法,新的算法在每个频率点根据语音输入信噪比计算语音存在的先验概率而不是使用一个固定值。在噪声急剧变化时仍能够对其进行实时跟踪,相比传统的SPP估计方法而言,整体谱包络与原噪声功率谱更为接近。

3.3 改进相位谱补偿的语音增强算法

在维纳滤波中应用新的 Sigmoid 型相位谱补偿算法和新的基于语音存在概率(SPP)的噪声估计算法得到改进相位谱补偿的语音增强算法。本文结合文献[16]中估计出的纯净语音幅度谱 $|\hat{S}_k(n,k)|$ 的方法与式(10)中新提出的补偿相位谱,得到在频域上增强后的语音信号为:

$$S(n,k) = |\hat{S}_k(n,k)| \exp(j\angle Y_{\text{new}}(n,k)) \quad (21)$$

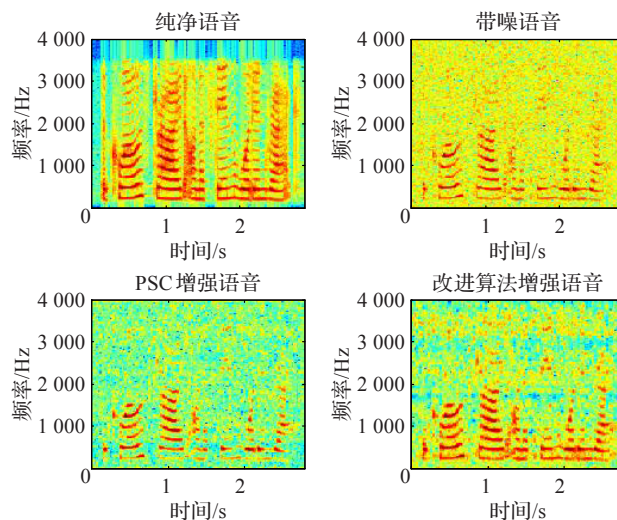
对式(21)作傅里叶逆变换,得到最终增强后的时域信号为:

$$s(t) = T_{\text{IFFT}}(S(n,k)) \quad (22)$$

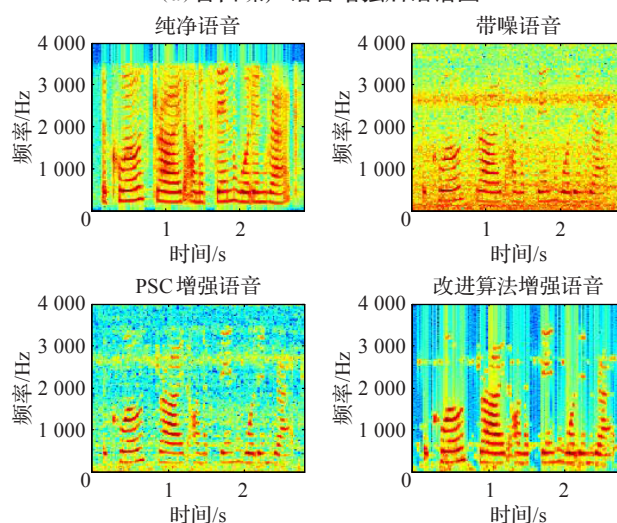
4 实验仿真

实验中从 Noise-92 语音库选择的三种不同的背景噪声用于测试,其中包括:白噪声、F16 噪声和 Pink 噪声。带噪声语音信号的输入信噪比设置为 0 dB、5 dB、10 dB。以 SP15 为例,采样率取 8 000 Hz,选用汉宁窗为其窗函数,帧长取 256,帧移取 64。使用信噪比和感知语音质量评估测度来对增强语音进行评价,对比结果如表 1 和表 2 所示。

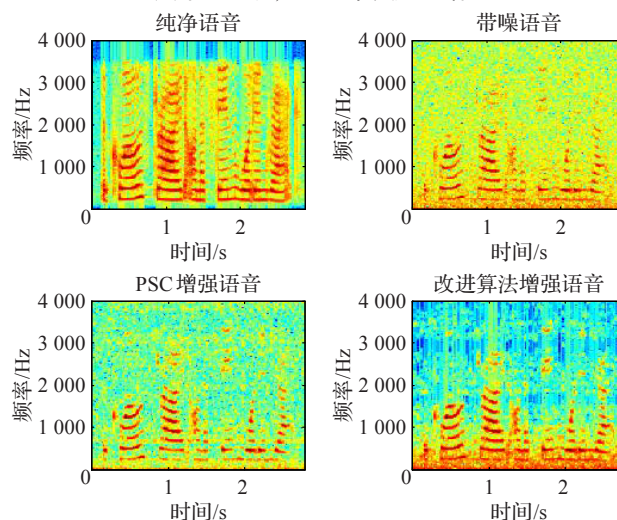
为了更好地从语谱图的角度比较改进算法与传统相位谱补偿算法,下面以 SP15 为例,将白噪声、F16 噪声和 Pink 噪声作为背景噪声,在信噪比为 5 dB 的条件下,语谱图的对比结果如图 3 所示。



(a) 含白噪声语音增强后语谱图



(b) 含 F16 噪声语音增强后语谱图



(c) 含 Pink 噪声语音增强后语谱图

图3 不同噪声背景下语谱图效果

5 结束语

针对在语音增强中传统相位谱补偿方法使语音主观感知质量改善受限问题,提出了一种新的随噪声急剧变化而灵活变化的 Sigmoid 型相位谱补偿因子来改进传

表1 两种语音增强算法的信噪比结果

噪声类型	输入信噪比/dB	PSC	改进算法
White 噪声	0	7.198 9	8.757 2
	5	9.839 8	12.267 6
	10	11.485 5	14.407 0
Pink 噪声	0	6.956 9	8.532 6
	5	9.666 7	12.159 1
	10	11.398 4	15.855 6
F16 噪声	0	7.238 2	8.783 9
	5	9.711 3	12.328 3
	10	11.243 2	16.366 9

表2 两种语音增强算法的 PESQ 结果

噪声类型	输入信噪比/dB	纯净 SP15 语音	含噪语音	PSC	改进方法
White 噪声	0	4.500 00	1.295 8	1.716 3	1.920 5
	5	4.500 00	1.608 0	2.063 4	2.407 3
	10	4.500 00	1.964 9	2.457 0	2.658 3
Pink 噪声	0	4.500 00	1.394 1	1.826 5	1.947 3
	5	4.500 00	1.744 5	2.247 1	2.374 3
	10	4.500 00	2.116 9	2.669 0	2.752 8
F16 噪声	0	4.500 00	1.420 4	1.926 0	2.002 4
	5	4.500 00	1.783 3	2.372 2	2.459 3
	10	4.500 00	2.166 9	2.786 3	2.888 1

统PSC算法。同时,提出了一种新的改进先验信噪比的语音存在概率噪声估计方法,将其应用在维纳滤波中得到纯净语音幅度谱并与新的相位谱结合。大量实验表明,该算法在不同信噪比条件下,相比原算法而言,语音质量显著改善,同时语音可懂度也有了明显提升。

参考文献:

- [1] Paliwal K, Wójcicki K, Shannon B. The importance of phase in speech enhancement[J]. Speech Communication, 2011, 53(4): 465-494.
- [2] Mowlaee P, Saeidi R, Stylianou Y. Advances in phase-aware signal processing in speech communication[J]. Speech Communication, 2016, 81: 1-29.
- [3] Mowlaee P, Saeidi R. Iterative closed-loop phase-aware single-channel speech enhancement[J]. IEEE Signal Processing Letters, 2013, 20(12): 1235-1239.
- [4] So S, Wójcicki K K, Lyons J G, et al. Kalman filter with phase spectrum compensation algorithm for speech enhancement[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), Taipei, Taiwan, China, 2009: 4405-4408.
- [5] Dang X, Khan M I A, Nakai T. Noise reduction of speech signal based on phase spectrum estimation[C]//International Conference on Informatics, Electronics & Vision, 2013: 1-4.
- [6] Mowlaee P, Stahl J, Kulmer J. Iterative joint MAP single-channel speech enhancement given non-uniform phase prior ☆[J]. Speech Communication, 2017, 86: 85-96.
- [7] Islam M T, Shahnaz C, Zhu W P, et al. Speech enhancement in adverse environments based on non-stationary noise-driven spectral subtraction and SNR-dependent phase compensation[J]. Electrical Engineering and Systems Science, 2018.
- [8] Maly A, Mowlaee P. On the importance of harmonic phase modification for improved speech signal reconstruction[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2016: 584-588.
- [9] Stark A P, Wójcicki K, Lyons J, et al. Noise driven short time phase spectrum compensation procedure for speech enhancement[J]. Proceedings Interspeech, 2008: 549-552.
- [10] 王栋, 贾海蓉. 改进相位谱补偿的语音增强算法[J]. 西安电子科技大学学报(自然科学版), 2017, 44(3): 83-88.
- [11] Krawczyk M, Gerkmann T. STFT phase reconstruction in voiced speech for an improved single-channel speech enhancement[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2014, 22(12): 1931-1940.
- [12] 王虎, 李晶, 赵恒森, 等. 稀疏低秩模型及相位谱补偿的语音增强算法[J]. 计算机工程与应用, 2018, 54(5): 150-155.
- [13] 沈锁金, 刘伟, 高颖. 基于语音存在概率的先验信噪比估计算法的研究[J]. 中国集成电路, 2016, 25(12): 26-30.
- [14] Gerkmann T, Krawczyk M, Rehr R. Phase estimation in speech enhancement—unimportant, important, or impossible?[C]//Electrical & Electronics Engineers in Israel, 2012: 1-5.
- [15] Chinaev A, Haeb-Umbach R. A generalized log-spectral amplitude estimator for single-channel speech enhancement[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, 2017: 4980-4984.
- [16] 容强, 肖汉. 基于MMSE维纳滤波语音增强方法研究与Matlab实现[J]. 计算机应用与软件, 2015, 32(1): 153-156.