

# 耦合 MMSE 和 WEDM 幅度谱估计的语音增强方法

韩 勇 赵宇红

(北方工业大学电子信息工程学院 北京 100144)

**摘 要:**为同时在高 SNR 和低 SNR 条件下获得较好的语音质量,提出了一种耦合 MMSE 和 WEDM 幅度谱估计的语音增强方法。此方法利用 Sigmoid 映射函数将后验 SNR 映射到 $[0,1]$ 之间。然后,根据此映射函数,提出一种自适应的谱恢复增益函数计算方法。在高 SNR 条件下,为避免语音失真,所提方法采用较大的映射函数值,从而使 MMSE 谱估计增益函数为谱恢复增益函数。反之,在低 SNR 条件下,为有效抑制含噪语音中的噪声,所提方法使用较小的映射函数值,从而选择 WEDM 谱估计增益函数为谱恢复增益函数。实验结果表明,所提算法性能在客观性能测试方面要优于参考算法。

**关键词:** 语音增强; 谱恢复; MMSE; WEDM

**中图分类号:** TP391 TN911 **文献标识码:** A **国家标准学科分类代码:** 510

## Coupled MMSE and WEDM spectral amplitude estimations for speech enhancement

Han Yong Zhao Yuhong

(College of Electronic and Information Engineering, North China University of Technology, Beijing 100144, China)

**Abstract:** In order to achieve better speech quality both in high SNR (signal-to-noise ratio) and low SNR conditions, this paper proposes a coupled MMSE (minimum mean square error) and WEDM (weighted euclidean distortion measure) spectral amplitude estimations method for speech enhancement. This method employs sigmoid mapping function to map a posteriori SNR into  $[0, 1]$  range. Thus, according to this mapping function, we develop an adaptive gain function calculation method for spectral restoration. In higher SNR conditions, to prevent speech distortions, the proposed method adopts a larger value of mapping function and makes MMSE spectrum estimation gain function be the gain function for spectral restoration. On the other hand, in lower SNR conditions, to more effectively remove noise from noisy speech, the proposed method uses a smaller value of mapping function and selects the gain function of WEDM spectrum estimation to be the gain function for spectral restoration. Our experiments demonstrate that the proposed method is superior to the reference methods.

**Keywords:** speech enhancement; spectral restoration; MMSE; WEDM

### 1 引 言

语音增强的目的是从含噪语音信号中抑制背景噪声,并尽可能的恢复纯净语音信号。在语音通信、语音识别<sup>[1-2]</sup>和助听器等语音处理系统中,语音增强算法作为预处理器,用于增强语音质量。一般而言,语音增强算法可以分为3类,即滤波算法、谱恢复算法和语音模型技术。首先介绍滤波算法,其目的是设计一个消除噪声的滤波器,从而恢复纯净语音。典型的滤波算法包括时频域的维纳滤波算法<sup>[3-4]</sup>以及参数维纳滤波算法<sup>[3]</sup>。其次,谱恢复算法主要是在频域设计一个抑制噪声的增益函数,进

而从含噪语音谱中获得纯净语音谱。典型的算法包括基于最小均方误差准则的(minimum mean square error, MMSE)谱估计器<sup>[5]</sup>,基于加权欧式距离失真测度(weighted euclidean distortion measure, WEDM)的谱估计器<sup>[3]</sup>,最大后验概率(maximum a posteriori, MAP)谱幅度估计器<sup>[6]</sup>以及广义最大后验概率(generalized maximum a posteriori, GMAP)谱幅度估计器<sup>[7]</sup>。最后,语音模型技术是结合人类语音产生模型和噪声抑制函数从含噪语音信号中消除噪声部分。用于语音增强的典型语音模型,包括谐波模型<sup>[8]</sup>和自回归(auto-regressive, AR)模型<sup>[9-10]</sup>。

收稿日期:2016-07

本文主要研究谱恢复语音增强算法。虽然许多传统的谱恢复算法能够有效的抑制背景噪声,但其仍然存在不足,即不能同时在高信噪比(signal-to-noise ratio, SNR)和低 SNR 下获得较好的增强效果。例如,MMSE 谱估计器在高 SNR 条件下能较好的保留语音成分,但是在低 SNR 条件下残留噪声较多。WEDM 谱估计器是在低 SNR 下能较好的抑制背景噪声,但在高 SNR 下,语音失真较大。因此,本文提出一种耦合 MMSE 和 WEDM (coupled MMSE and WEDM, C\_MMSE\_WEDM)幅度谱估计的语音增强方法来克服上述局限性。此方法的谱恢复增益函数是利用一个单调递增的 sigmoid 函数来自适应的选择谱估计器。在高 SNR 条件下,利用 sigmoid 映射函数使增益函数选择 MMSE 谱估计器,从而较好的保留语音成分。在低 SNR 条件下,利用 sigmoid 映射函数使增益函数选择 WEDM 谱估计器,从而更有效的抑制背景噪声。

## 2 谱恢复技术

### 2.1 谱恢复技术原理

假设时域纯净语音  $x(n)$  被不相关的加性噪声  $d(n)$  所污染。这样,含噪语音  $y(n)$  可以表示为:

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

式中: $n$  为离散时间序号,对其两边分别进行快速傅里叶变换(fast fourier transform, FFT),可得其频域表达式为:

$$Y(k) = X(k) + D(k) \quad (2)$$

式中: $k$  为频点索引, $Y(k)$ 、 $X(k)$  和  $D(k)$  分别为含噪语音、纯净语音和噪声信号的 FFT 系数。在极坐标下,式(2)可以表示为:

$$R_Y(k)e^{j\alpha_Y(k)} = R_X(k)e^{j\alpha_X(k)} + R_D(k)e^{j\alpha_D(k)} \quad (3)$$

式中: $R_Y(k)$ 、 $R_X(k)$  和  $R_D(k)$  分别表示含噪语音、纯净语音和噪声信号的幅度谱, $\alpha_Y(k)$ 、 $\alpha_X(k)$  和  $\alpha_D(k)$  分别为其所对应的相位谱。

如图 1 所示为谱恢复语音增强原理,其可以分为噪声估计和增益估计两个模块。噪声估计模块是从含噪语音幅度谱  $R_Y(k)$  中计算噪声的功率谱,进而获得先验 SNR 和后验 SNR。增益估计模块用于计算基于先验和后验 SNR 的增益函数  $G(k)$ ,含噪语音幅度谱通过此增益函数滤波,就可以得到增强语音的幅度谱。

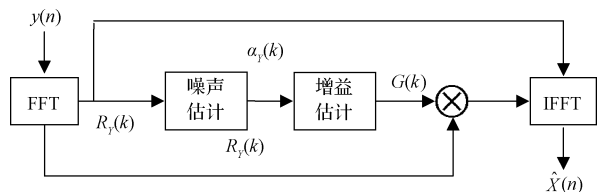


图 1 谱恢复语音增强原理

首先,介绍纯净语音的相位谱估计。文献[5]通过求解如下约束最小化问题来计算纯净语音相位谱  $\hat{\alpha}_X(k)$  的

MMSE 估计:

$$\min_{\exp(j\hat{\alpha}_X(k))} E[|\exp(j\alpha_X(k)) - \exp(j\hat{\alpha}_X(k))|^2] \quad (4a)$$

其中满足:

$$|\exp(j\hat{\alpha}_X(k))| = 1 \quad (4b)$$

利用拉格朗日乘数法,可以获得如下最优解:

$$\exp(j\hat{\alpha}_X(k)) = \exp(j\alpha_Y(k)) \quad (5)$$

这样,增强语音谱就可以表述为:

$$\hat{X}(k) = \hat{R}_X(k)e^{j\alpha_Y(k)} \quad (6)$$

对得到的增强语音  $\hat{X}(k)$  进行 IFFT 变换,就可以获得时域增强语音。

接下来,介绍两种典型的谱估计器增益函数,即 MMSE 谱估计器和 WEDM 谱估计器。

### 2.2 MMSE 和 WEDM 谱估计器

在推导 MMSE 谱估计和 WEDM 谱估计时,研究人员考虑了两个假设:

1) 语音和噪声傅里叶的实部和虚部都服从高斯分布,并且其均值为 0;

2) 傅里叶系数统计无关。

谱恢复语音增强中使用的先验 SNR 定义为  $\xi(k) = \lambda_x(k)/\lambda_d(k)$ ,后验 SNR 定义为  $\gamma(k) = Y^2(k)/\lambda_d(k)$

其中  $\lambda_x(k) = E[|X(k)|^2]$ ,  $\lambda_d(k) = E[|D(k)|^2]$ 。文献[5]分别给出了  $\xi(k)$  和  $\gamma(k)$  的计算表达式,即:

$$\hat{\xi}(l) = \begin{cases} \Delta + (1 - \Delta)\max[\gamma(l) - 1, 0] & l = 1 \\ \Delta \frac{\hat{R}_X^2(l-1)}{\hat{R}_D^2(l-1)} + (1 - \Delta)\max[\gamma(l) - 1, 0] & \text{其他} \end{cases} \quad (7)$$

和

$$\hat{\gamma}(l) = \frac{R_Y^2(l)}{\hat{R}_D^2(l)} \quad (8)$$

为了简化,在式(7)和式(8)中省略了频率索引, $l$  为帧索引,  $\hat{R}_X^2(l-1)$  和  $\hat{R}_D^2(l-1)$  分别为前一帧估计的语音功率谱和噪声功率谱。

在 MMSE 和 WEDM 谱估计的语音增强算法中,其目标是通过最小化一个代价函数  $C(R_X(k), \hat{R}_X(k))$  的期望得到语音幅度谱  $R_X(k)$  的估计  $\hat{R}_X(k)$ , 即:

$$\hat{R}_X(k) = \arg \min E\{C(R_X(k), \hat{R}_X(k))\} = G(k) \times R_Y(k) \quad (9)$$

式中: $E\{\cdot\}$  为统计期望,  $G(k)$  增益函数。下面分别给出这两种谱估计器的代价函数和增益函数。

#### 2.2.1 MMSE 谱估计器

在文献[5]中,基于 MMSE 谱估计器的代价函数定义为:

$$C(R_X(k), \hat{R}_X(k)) = (R_X(k) - \hat{R}_X(k))^2 \quad (10)$$

把式(10)代入式(9),并对式(9)中的  $\hat{R}_X(k)$  求偏导

数,令其偏导数为0,可以获得:

$$\hat{R}_X(k) = R_Y(k) \frac{\sqrt{\pi v(k)}}{2\gamma(k)} \exp\left(-\frac{v(k)}{2}\right) \times \left[ (1+v(k)) I_0\left(\frac{v(k)}{2}\right) + v(k) I_1\left(\frac{v(k)}{2}\right) \right] \quad (11)$$

式中:  $I_0(\cdot)$  和  $I_1(\cdot)$  分别为第零阶修正贝塞尔函数和第一阶修正贝塞尔函数,  $v(k)$  定义如下:

$$v(k) = \frac{\xi(k)}{1+\xi(k)} \gamma(k) \quad (12)$$

由式(11)可以获得 MMSE 谱估计的增益函数  $G_{MMSE}(k)$  为:

$$G_{MMSE}(k) = \frac{\sqrt{\pi v(k)}}{2\gamma(k)} \exp\left(-\frac{v(k)}{2}\right) \times \left[ (1+v(k)) I_0\left(\frac{v(k)}{2}\right) + v(k) I_1\left(\frac{v(k)}{2}\right) \right] \quad (13)$$

### 2.2.2 WEDM 谱估计器

在文献[3]中,基于 WEDM 谱估计器的代价函数定义为:

$$C(R_X(k), \hat{R}_X(k)) = \frac{(R_X(k) - \hat{R}_X(k))^2}{R_X(k)} \quad (14)$$

把式(14)代入式(9),并对式(9)中的  $\hat{R}_X(k)$  求偏导数,令其偏导数为0,可以获得:

$$\hat{R}_X(k) = \frac{\sqrt{v(k)}}{\sqrt{\pi\gamma(k)}} \frac{\exp\left(\frac{v(k)}{2}\right)}{I_0\left(\frac{v(k)}{2}\right)} R_Y(k) \quad (15)$$

由式(15)可以获得 WEDM 谱估计的增益函数  $G_{WEDM}(k)$  为:

$$G_{WEDM}(k) = \frac{\sqrt{v(k)}}{\sqrt{\pi\gamma(k)}} \frac{\exp\left(\frac{v(k)}{2}\right)}{I_0\left(\frac{v(k)}{2}\right)} \quad (16)$$

## 3 耦合 MMSE 和 WEDM 幅度谱估计的语音增强

由于 MMSE 谱估计器算法在高 SNR 的条件下能较好的保留语音成分,而 WEDM 算法在低 SNR 的条件下能更好的抑制背景噪声,因此,本文提出了一种基于 MMSE 和 WEDM 耦合的语音增强算法,下面对其进行具体介绍。

### 3.1 算法原理

如图2所示为耦合 MMSE 和 WEDM 幅度谱估计的语音增强原理。首先对输入的含噪语音进行 FFT 变换,得到其幅度谱和相位谱。其次,对含噪语音进行噪声估计获得噪声的功率谱,进而得到先验 SNR 和后验 SNR。接下来,通过 MMSE 谱估计器和 WEDM 谱估计器分别获得增益函数  $G_{MMSE}(k)$  和  $G_{WEDM}(k)$ 。为进一步抑制低 SNR 条件下的背景噪声,把  $G_{WEDM}(k)$  通过增益函数  $\tilde{G}(k)$  得到  $G'_{WEDM}(k)$ 。然后再把获得的两个增益函数分别与 sigmoid 函数  $\alpha(k)$  和  $(1-\alpha(k))$  相结合,得到耦合增益函数  $G(k)$ ,含噪语音经此增益函数就可以得到增强的幅度谱。

最后利用增强的幅度谱和含噪语音相位谱,经 IFFT 变换得到时域的增强语音。

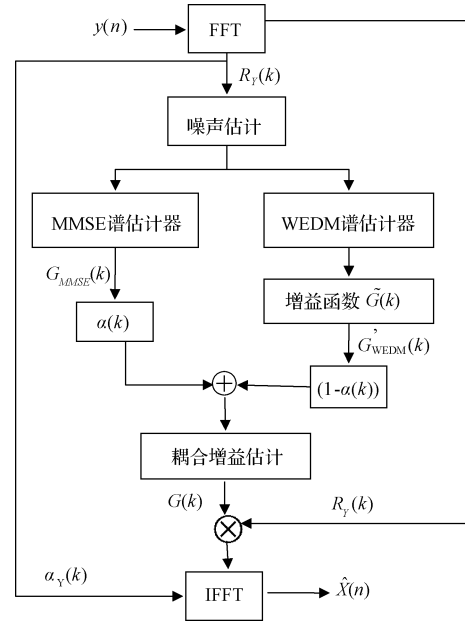


图2 耦合 MMSE 和 WEDM 幅度谱估计的语音增强原理

### 3.2 MMSE 和 WEDM 耦合算法

首先,为了进一步抑制 WEDM 算法在低 SNR 条件下的噪声,做如下处理:

$$G'_{WEDM}(k) = G_{WEDM}(k) \times \tilde{G}(k) \quad (17)$$

式中:  $\tilde{G}(k)$  定义如下:

$$\tilde{G}(k) = \begin{cases} \beta_1, & \text{if } \bar{\xi} < 0.25 \\ \beta_2, & \text{otherwise} \end{cases} \quad (18)$$

式中:  $\bar{\xi} = \sum_{k=0}^{K-1} \hat{\xi}(k)/K$  是当前帧先验 SNR 频域均值。

接下来,利用式(13)和式(17),耦合增益函数可以表述为:

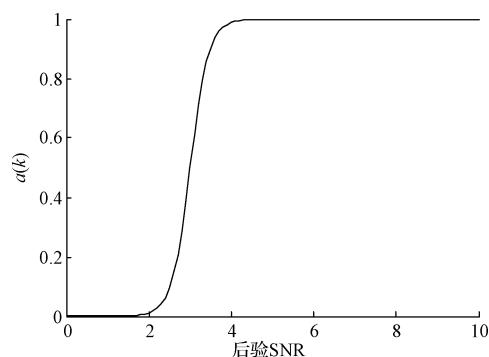
$$G(k) = \alpha(k) \times G_{MMSE}(k) + (1-\alpha(k)) \times G'_{WEDM}(k) \quad (19)$$

式中:  $\alpha(k)$  为 sigmoid 的函数,其是后验 SNR  $\gamma(k)$  的单调递增函数,定义为:

$$\alpha(k) = \frac{1}{1 + e^{b(\gamma(k)+c)}} \quad (20)$$

这里使用 sigmoid 函数主要目的是自适应的调节耦合增益函数  $G(k)$ ,使其在不同的 SNR 条件下选择最佳的增益滤波器。

如图3所示,实线为 sigmoid 函数  $\alpha(k)$  随后验 SNR  $\gamma(k)$  的变化情况。从图中可以看出,当后验 SNR 较低时,  $\alpha(k)$  取值较小并趋近于0,此时式(19)中的增益函数  $G(k)$  近似于  $G'_{WEDM}(k)$ ,这样能较好的抑制低 SNR 下的背景噪声。而当后验 SNR 较高时,  $\alpha(k)$  取值较大并趋近于1,此时式(19)中的增益函数  $G(k)$  近似于  $G_{MMSE}(k)$ ,这样能在高 SNR 下能较好的保留语音成分。

图3  $\alpha(k)$ 随后验 SNR 变化示意

最后,利用式(19)中的耦合增益函数对含噪语音进行增强处理,可得增强语音谱为:

$$\hat{R}_X(k) = R_Y(k)G(k) \quad (21)$$

把得到的增强语音谱  $\hat{R}_X(k)$  代入式(6),再经 IFFT 变化就可得时域增强语音。

## 4 实验

为验证所提 C\_MMSE\_WEDM 语音增强方法的性能,选择文献[5]所提的 MMSE 谱估计器 (MMSE) 和文献[3]所提 WEDM 谱估计器 (WEDM) 作为参考算法,将这些方法与本章所提的 C\_MMSE\_WEDM 算法进行性能比较。此外,所有参考算法和 C\_MMSE\_WEDM 方法都使用式(7)和式(8)来估计先验 SNR 和后验 SNR,并且都利用最小值控制递归平均 (minima controlled recursive averaging, MCRA) 算法估计噪声功率谱<sup>[11]</sup>。

在测试实验中,噪声来源于 Noise\_x92 数据库,主要包括 babble、white、street 和 office 4 种噪声。10 段测试纯净语音序列来自 NTT 标准语音数据库中的中文子库。所有测试语音序列和噪声源的采样率都为 8 kHz,帧长为  $N=256$  个样本,并对这些样本加正弦窗,且相邻帧之间进行 50% 叠接。测试语音的信噪比分别为 0 dB, 5 dB 和 10 dB。对于各语音增强算法性能的衡量,采用分段信噪比 (segmental signal-to-noise ratio, SSNR)<sup>[12]</sup> 和对数谱失真 (log-spectral distortion, LSD)<sup>[13]</sup> 作为客观质量评价方法,并且还对各增强算法得到的增强语音语谱进行了比较分析。

接下来首先介绍实验参数设置,然后介绍所提方法和参考方法的测试结果并对其进行比较分析。

### 4.1 实验参数设置

本文所提 C\_MMSE\_WEDM 方法的基本参数设置如表 1 所示。

表1 基本参数设置表

$\Delta$	$\beta_1$	$\beta_2$	$b$	$c$
0.98	0.3	1	-4.5	-3

### 4.2 实验结果及分析

#### 4.2.1 语谱图比较

对于各增强语音语谱图的比较,如图 4 所示给出了含噪语音语谱图和不同增强算法得到的增强语音语谱对比示例。从图 4(a)、(b) 和 (c) 中可以看出, WEDM 算法和 MMSE 算法都能有效的抑制背景噪声,但在低 SNR 下 WEDM 算法相比于 MMSE 算法噪声衰减量更大,而从 (b) 和 (d) 中看出所提 C\_MMSE\_WEDM 算法的噪声衰减量相比于 WEDM 算法更大。

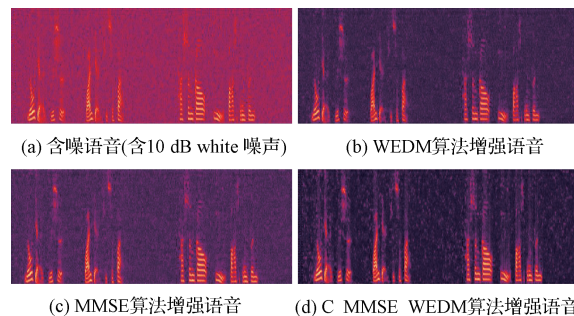


图4 不同增强算法语谱图对比示例

#### 4.2.2 客观测试

如表 2 所示为增强算法相比于含噪语音的 SSNR 提高量测试结果,对于 4 种噪声任一种输入信噪比,所提 C\_MMSE\_WEDM 算法的 SSNR 提高量大于参考算法。如表 3 所示为 LSD 性能测试结果,在各种信噪比条件下,3 种增强算法的 LSD 值相比于含噪语音信号都有较大程度的降低,其中,与参考算法相比,所提 C\_MMSE\_WEDM 算法的 LSD 值更低。

表2 SSNR 提高量性能测试结果

噪声类型	方法	0 dB	5 dB	10 dB
white	MMSE	12.61	11.51	10.11
	WEDM	13.57	11.68	9.33
	C_MMSE_WEDM	15.87	14.51	12.66
babble	MMSE	9.86	9.00	8.00
	WEDM	11.30	9.56	7.58
	C_MMSE_WEDM	11.64	10.75	9.57
street	MMSE	13.29	12.35	11.34
	WEDM	13.87	11.82	9.33
	C_MMSE_WEDM	16.59	15.32	13.72
office	MMSE	11.64	10.95	10.05
	WEDM	12.72	10.92	8.73
	C_MMSE_WEDM	14.23	13.36	12.12

表3 LSD性能测试结果

噪声类型	方法	0 dB	5 dB	10 dB
white	Noisy	17.53	15.48	13.38
	MMSE	10.98	9.59	8.20
	WEDM	10.42	9.18	7.92
	C_MMSE_WEDM	9.22	8.04	7.09
babble	Noisy	13.50	11.63	9.91
	MMSE	9.41	7.87	6.52
	WEDM	9.11	7.72	6.48
	C_MMSE_WEDM	8.83	7.20	5.98
Street	Noisy	11.95	10.19	8.56
	MMSE	7.29	5.99	4.90
	WEDM	7.11	5.94	5.04
	C_MMSE_WEDM	6.14	5.07	4.49
Office	Noisy	12.34	10.56	9.00
	MMSE	8.10	6.67	5.43
	WEDM	7.80	6.50	5.45
	C_MMSE_WEDM	7.28	5.96	4.99

由于本文所提 C\_MMSE\_WEDM 算法在低 SNR 下相比于 MMSE 算法能更好的抑制噪声。同时,在高 SNR 下相比于 WEDM 算法保留了更多的语音成分,所以就两项客观测试而言,所提算法都有一个较大的提高。

## 5 结 论

本文提出一种 C\_MMSE\_WEDM 幅度谱估计的语音增强方法,该算法利用一个 sigmoid 映射函数  $\alpha(k)$  来自适应的调节谱恢复增益函数。文中利用语谱图和客观性能测试来验证所提算法性能。从语谱图比较结果可以证明,所提算法相比于 MMSE 算法和 WEDM 算法噪声衰减量更大。从客观性能测试结果可以证明,所提算法的 SSNR 和 LSD 性能测试结果要优于参考算法。

## 参 考 文 献

- [1] 杨金宵,沈天飞,滕秋霞. 基于声门激励的语音语速、音量调整方法[J]. 电子测量技术, 2016, 39(2): 72-75.
- [2] 何侃,田亚清,李强,等. 基 LD3320 的语音识别智能垃圾桶设计[J]. 国外电子测量技术, 2015, 34(6): 85-88.
- [3] PHILIPPOSC L. Speech enhancement: theory and practice [M]. CRC, 2007.

- [4] SCALARTP, FILHO J V. Speech enhancement based on a priori signal to noise estimation[C]. In Proceedings of ICASSP, 1996: 629-632.
- [5] EPHRAIMY, MALAH D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1984, 32(6): 1109-1121.
- [6] LOTTERT, VARY P. Speech enhancement by maximum a posteriori spectral amplitude estimation using a super-Gaussian speech model[J]. Eurasip Journal on Applied Signal Processing, 2005(7): 1110-1126.
- [7] SU Y C, TSAO Y, WU J E, et al. Speech enhancement using generalized maximum posteriori spectral amplitude estimators [C]. In Proceedings of ICASSP, 2013:7467-7471.
- [8] MCAULAY RJ, QUATIERI T F. Speech analysis/synthesis based on a sinusoidal representation [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1986, 34(4): 744 - 754.
- [9] SRINIVASAN S, SAMUELSSON J, KLEJIN W B. Codebook-based Bayesian speech enhancement for nonstationary environments [J]. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(2): 441-451.
- [10] NAIDUD H R, SRINIVASAN S. Robust Bayesian estimation for context-based speech enhancement [J]. Eurasip Journal on Audio, Speech and Music Processing, 2014.
- [11] COHENI. Noise estimation by minima controlled recursive averaging for robust speech enhancement [J]. IEEE Signal Process. Letters, 2002, 9(1): 12-15.
- [12] QUACKENBUSHS R, BARNWELL T P, CLEMENTS M A. Objective Measures of Speech Quality[M]. Englewood Cliffs, NJ: Prentice Hall, 1988.
- [13] ABRAMSONA, COHEN I. Simultaneous detection and estimation approach for speech enhancement [J]. IEEE Transactions on Audio, Speech and Language Processing, 2007, 15(8): 2348-2359.

## 作 者 简 介

韩勇(通讯作者),1990 年出生,硕士研究生,主要研究方向为语音增强等。

E-mail:hy905190@sina.com

赵宇红,1962 年出生,工学博士,副教授,主要研究方向为多媒体通信技术与现代电路理论。

E-mail:zhylh@ncut.edu.cn