改进的后滤波波束形成器语音增强算法

阎兆立 杜利民
(中国科学院声学研究所 北京 100080)

摘 要 该文提出了一种具有后滤波的波束形成器的语音增强改进算法。该算法主要解决维纳滤波器的理想信号功率谱估计，结合自功率谱减法和互功率谱减法计算出尽可能多的功率谱估计值，以使平均结果更接近于真实值，同时修正了声源移动引起的互功率谱变化。实验结果信噪比提高 5dB 以上，汽车环境中基于隐含马尔可夫模型(HMM)的小词汇量短语识别达到 84%。从信噪比、平均谱距离和语音识别率可以看出该算法有效去除了原始算法中易残留的低频噪声，减少了语音信号失真。

关键词 语音增强，传声器阵列，波束形成器
中图分类号：TN912.3 **文献标识码**：A **文章编号**：1009-5896(2006)12-2269-04

The Modified Post-Filter Beamforming for Speech Enhancement

Yan Zhao-li Du Li-min
(*Institute of Acoustics, Chinese Academy of Sciences, Beijing* 100080, *China*)

Abstract A modified Wiener post-filter beamforming for speech enhancement is presented. The proposed algorithm focuses on the estimation of the ideal signal's power spectrum for the Wiener filter, both auto-correlation and cross-correlation are taken into consideration to obtain as more power spectrum estimations as possible, the average of which produces more accurate result. Meanwhile, the alteration of cross-correlation caused by the moving speaker is revised. The performance of the algorithm has been evaluated from Signal to Noise Ratio (SNR), Average Log Spectrum Distance (ALSD) and speech recognition rate. The SNR is improved by 5 dB. The command recognition rate based on Hidden Markov Model (HMM) in the car environment reaches 84%. The low frequency noise that still remains after the process of the primary method is reduced by the proposed method, and the speech signal distortion is also decreased.

Key words Speech enhancement, Microphone array, Beamforming

1 引言

传声器阵列技术是语音增强中的重要部分，它被广泛地应用于视频会议系统、移动电子系统[1]和助听[2]等领域。由于多通道采样引入了空间信息，它相对于单通道算法具有更大的增强潜力。近年来围绕传声器阵列的语音增强应用研究取得了长足进展，尤其具有小口径和低计算量的阵列被越来越多的人所重视。

具有维纳后滤波的波束形成器由Zelinski提出[3]，他利用空间信息巧妙地解决了维纳滤波器的估计问题。该模型假设不同传声器单元采样的噪声信号之间不相关，不过在实际环境中很难完全成立，导致实际的维纳滤波器输出仍含有明显低频噪声。为了改善存在的问题，此基础上，Meyer引入谱减法直接对低频噪声作消除，然而会有音乐噪声残留[4]；Zhang Xianxian把人耳屏蔽效应与波束形成器算法结合起来，也只是采用二者级联的方式[5]；McCowan运用扩散噪声场的数学模型讨论了解决扩散噪声场中不同通道噪声相关的问题[6]，该算法要求预先得到噪声相干函数，适用范围受到限制。除后滤波算法以外，广义旁瓣消除算法(GSC)也是传声器阵列语音增强的常用方法[7]，Cohen 研究了结合后滤波与GSC消除非平稳噪声的算法[8]，Null-forming自提出后虽然可查文献不多，但也不失为一种好的方法[9]，限于篇幅这里不作讨论了。

本文提出基于传声器阵列的语音增强算法，把互功率谱和自功率谱减法与后滤波有机结合起来，以避免各通道间低频噪声的相关性带来的问题，同时产生更多理想信号功率谱估计值，使得所有估值的平均产生更准确的结果。然而由于说话人的移动引起通道间信号时间延迟发生变化，导致各通道间的互功率谱随之变化，无法预先估计，为谱减法的运用带来问题。这一点将在后面给予分析和修正。

2 具有维纳后滤波的波束形成器

首先回顾一下具有维纳后滤波的波束形成器原理[3]，后面的讨论都是在此基础上发展起来的。图 1 是具有维纳后滤波的波束形成器原理框图，该系统首先补偿了各通道间的信号时间延迟。把含噪语音建模为纯净信号s与加性噪声n之和：

$$x_i(t) = s(t) + n_i(t) \tag{1}$$

其中 $x(t)$ 是含噪语音信号，$s(t)$ 是纯净语音信号，$n(t)$ 是噪

2005-05-10 收到，2005-11-30 改回

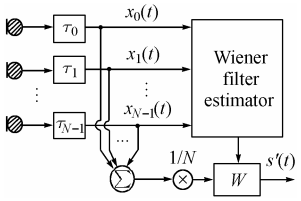

图 1 具有维纳后滤波的波束形成器原理框图

Fig.1 The block of Wiener post-filter beamforming

声，下标 i 是通道标号。则单通道维纳滤波器可表述为

$$W(f)=\frac{\phi_{ss}(f)}{\phi_{xx}(f)}=\frac{\phi_{ss}(f)}{\phi_{ss}(f)+\phi_{nn}(f)} \tag{2}$$

如何估算理想信号的自功率谱 ϕ_{ss} 是研究的重点。这个问题通过附加限制条件得到解决，即假设各传声器单元采集的噪声信号之间不相关，则有

$$\phi_{x_ix_j}(f)=\phi_{ss}(f) \tag{3}$$

其中 $\phi_{x_ix_j}(f)$ 是信号x_i和x_j的互功率谱。

通过计算所有可能阵元组合的互功率谱，更为准确的维纳滤波器估计表述如下

$$\hat{W}(f)=\frac{E\left[R\left\{\sum_{i=0}^{N-2}\sum_{j=i+1}^{N-1}\hat{\phi}_{x_ix_j}\right\}\right]}{E\left[\sum_{i=0}^{N-1}\phi_{x_ix_i}\right]} \tag{4}$$

其中 $R\{\cdot\}$ 表示实部算符，N 是传声器阵列的阵元数目。

然而，在实际声场中各通道噪声并非不相关，在低频部分其相干系数的模值接近于 1，直到高频部分才出现较小的值[10]，也就是说式(4)基于的假设条件不能完全满足，这使得滤波器的输出仍残留部分噪声。

3 改进的具有后滤波的波束形成器

现有维纳后滤波波束形成器的一些不利因素在第 1 节和第 2 节都有介绍，基于以上对维纳后滤波算法的分析，提出一种改进算法，以期弥补前面提到的不足。

每个通道的输入信号建模如下：

$$x_i(t)=s(t-\tau_i)+n_i(t) \tag{5}$$

其中 τ_i 是从说话者到传声器 i 的时间延迟。经过时间延迟补偿后，式(5)的傅里叶变换为

$$\hat{X}_i(f)=S(f)+\hat{N}_i(f) \tag{6}$$

$$\hat{N}_i(f)=N_i(f)e^{j\frac{2\pi}{w}f\tau_i} \tag{7}$$

其中^是时间延迟补偿算符，w 是变换取的帧长。由此得到

$$\hat{\phi}_{x_ix_j}(f)=\phi_{ss}(f)+\hat{\phi}_{n_in_j}(f) \tag{8}$$

$$\hat{\phi}_{n_in_j}(f)=\phi_{n_in_j}(f)e^{j\frac{2\pi}{w}f\tau_{ij}} \tag{9}$$

其中 $\hat{\phi}_{x_ix_j}(f)$ 是信号 $x_i(t+\tau_i)$，$x_j(t+\tau_j)$ 的互功率谱，$\hat{\phi}_{n_in_j}(f)$ 与 $\phi_{n_in_j}(f)$ 分别是延迟补偿前、后噪声 n_i，n_j 的互功率谱，$\tau_{ij}=\tau_i-\tau_j$ 是通道 i 与 j 的信号延迟之差。可以看出，各通道间噪声的互功率谱与 τ_{ij} 有关，因此简单的互功率谱减法不适用于移动说话者的情况。

根据式(8)和式(9)，得到理想语音信号的自功率谱估计器

$$\phi'_{ss}(f)=\hat{\phi}'_{x_ix_j}(f)-\phi'_{n_in_j}(f)e^{j\frac{2\pi}{w}f\tau_{ij}} \tag{10}$$

其中 $(\cdot)'$ 表示估计值，$\phi'_{n_in_j}$ 是时间延迟补偿之前的噪声互功率谱估计值，$e^{j2\pi f\tau_{ij}/w}$ 是补偿系数，它仅包含相位信息，事实上仅式(10)的实部在后续的计算中有效，因此这里的相位补偿是必要的。

根据经典的单通道功率谱减法，我们得到

$$\phi'_{ss}(f)=\phi'_{x_ix_i}(f)-\phi'_{n_in_i}(f) \tag{11}$$

其中 $\phi'_{x_ix_i}(f)$ 与 $\phi'_{n_in_i}(f)$ 分别为含噪信号与噪声的自功率谱估计。把式(10)与式(11)整合到一起，对所有可能的纯净语音信号功率谱估计求均值，得到最终维纳滤波器的估计

$$\hat{W}(f)=\frac{E[R\{A+B\}]}{E\left[R\left\{\sum_{i=1}^{N}\phi'_{x_ix_i}(f)\right\}\right]} \tag{12}$$

其中

$$A=\sum_{i=0}^{N-2}\sum_{j=i+1}^{N-1}\left(\hat{\phi}'_{x_ix_j}(f)-\hat{\phi}'_{n_in_j}(f)e^{j\frac{2\pi}{w}f\tau_{ij}}\right) \tag{13}$$

$$B=\sum_{i=0}^{N-1}\left[\phi'_{x_ix_i}(f)-\phi'_{n_in_i}(f)\right] \tag{14}$$

其中 $R\{\cdot\}$ 表示复数的实部。自功率谱必须满足非负实数的条件，因此还要对功率谱估计作进一步的半波整形。

式(12)用自功率谱和互功率谱减法，直接得到理想信号功率谱估计，避免了 Zelinski 后滤波中由于低频噪声相关性导致理想信号功率谱估计不准的问题。基于统计意义上的信号功率谱计算在实现时多由单帧数据得到，存在偶然误差问题，因此把所有传声器单元组合的自功率谱减法和互功率谱减法得到的理想信号功率谱估计进行平均，将得到更接近真实值的结果。算法中参与平均的功率谱估计值的个数 M 是

$$M=N+C_N^2 \tag{15}$$

例如在下面的实验中，N 取 4，则 M 等于 10，也就是利用 4 个传声器单元的阵列得到 10 个纯净信号的功率谱估计。

事实上，在实际环境中各个通道的噪声仅在低频部分表现出相关性，据此该维纳滤波器估计可按图 2 实现：滤波器分为两部分估计，高频部分仍按式(4)处理，而低频部分则由式(12)实现。这里我们参考了Meyer的做法[4]。后面的实验中，频率分界点取为 1kHz。这样做既保证了算法的消噪性能，又降低了运算量，当然与基本维纳后滤波算法相比，运算量是增加的。

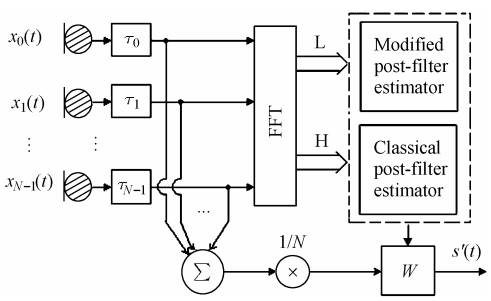

图 2 改进维纳后滤波估计器的实现框图

Fig.2 The block of the modified Wiener post-filter estimator

4　实验结果

为了评估文中提出的改进算法性能，从以下 3 个方面对其与一些现有算法进行了比较。

4.1　信噪比 SNR

该实验在办公室环境下进行的，噪声源来自周围环境中的空调、计算机风扇和扬声器等；阵列系统包含 4 个传声器单元，两两间距分别为 8.2cm，12.3cm，8.2cm；说话者在阵列系统前方，扬声器放置在侧旁位置。首先用扬声器播放平稳背景噪声，图 3 是实验结果的波形图，从图 3(a)到图 3(d)信噪比依次为 7.6dB，15dB，21dB 和 32dB。图 4 则是非平稳背景噪声(轻音乐)的实验结果，从图 4(a) 到图 4(d) 的信噪比依次为 6.5dB，10dB，22dB 和 27dB。所有输出都经过一个 200Hz 高通滤波器的处理。不难看出本文提出的算法对信噪比有了很大提高，尤其对于平稳背景噪声的情况。

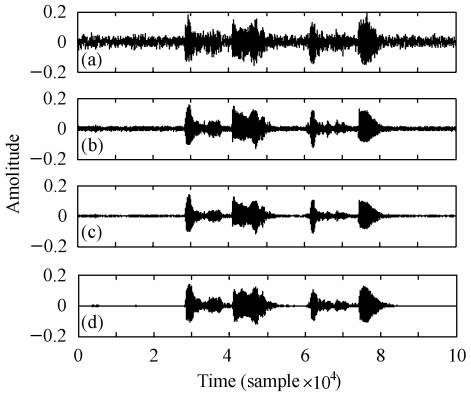

图 3　(a)原始含噪语音(平稳背景噪声)
(b)延迟相加波束形成器
(c)具有维纳后滤波的波束形成器
(d)本文提出改进算法

Fig.3　(a) The primary speech with stationary noise.
(b) Delay & Sum beamforming
(c) Wiener post-filter beamforming
(d)The proposed modified method

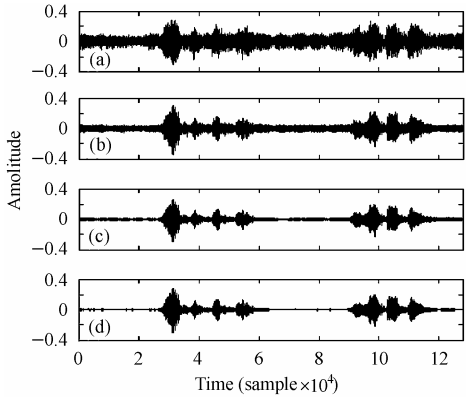

图 4　(a)原始含噪语音(音乐背景噪声)
(b)，(c)，(d)的说明同图 3

Fig.4　(a) The primary speech with music background noise. The notes of (b), (c), (d) are the same as Fig.3

4.2　平均 log 域谱距离(Average Log-Spectrum Distance, ALSD)

ALSD 是两个信号之间的距离测度，它从频域描述了两个信号的相似程度。该实验中理想信号与实际滤波器输出结果之间的平均谱距离，从另一方面对波束形成器的性能进行了验证。实验以模拟方式进行，模拟房间尺寸为 7m×4m×2.75m，混响时间 0.2s；各个传声器单元的坐标分别为(3.0, 2.0, 1.5), (3.08, 2.0, 1.5), (3.16, 2.0, 1.5) 和 (3.24, 2.0, 1.5)；噪声源坐标为(5.0, 3.0, 1.5)。以上坐标单位都是米。从说话者及噪声源到传声器单元的声学传递函数可由 Allen 的映像算法得到[11]，处理后的语音和噪声信号加在一起得到传声器单元的模拟输入。

平均谱距离 ALSD 定义如下：

$$D(\omega)=\frac{1}{P}\sum_{i=0}^{P-1}\left|\ln(|X_i(\omega)|)-\ln(|\hat{X}_i(\omega)|)\right| \tag{16}$$

其中 $X_i(\omega)$ 和 $\hat{X}_i(\omega)$ 是两个需要测度的信号频谱，P 是参与计算的帧数。图 5 是实验结果，从图 5 可看出，改进算法性能较好，同时图 5(c)显示低频噪声被保留，这印证了维纳后滤波算法的不足，同时也是改进算法中界定高、低频分界点的依据。由于数据经过一次上采样，所以图 5 中 8kHz 到 16kHz 部分 ALSD 保持较低的值。

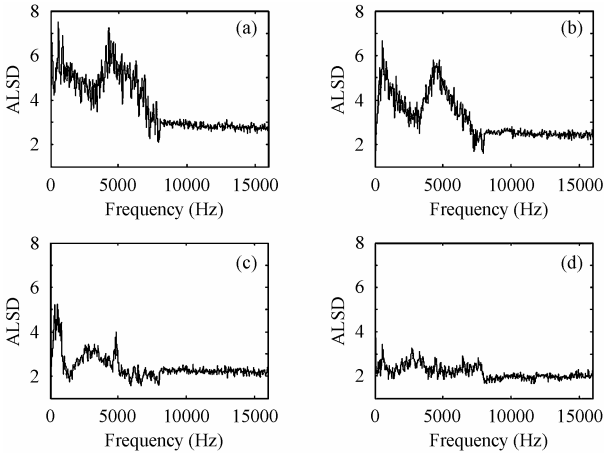

图 5　纯净语音信号与(a)原始信号 (b)延迟相加(DS)波束形成器输出 (c)具有维纳后滤波的波束形成器输出 (d)本文提出的改进算法输出之间的平均谱距离

Fig.5　ALSD between the pure speech and (a)the primary noisyspeech (b)the output of Delay & Sum beamforming (c)the output of Wiener post-filter beamforming (d)the output of the proposed method

4.3　语音识别(Automatic Speech Recognition, ASR)

实验配置如下：录音环境是以 60km/h 速度行驶的 Passat 汽车；阵列系统同实验 4.1 节，被安放在副驾驶正前的风挡玻璃上方；采样率是 32kHz；采样精度 16bit；录得 250 短句包括 16 条女声控制命令，如打开窗户、打开空调、左转灯、右转灯等；信噪比为 0dB 左右。

4.3.1 *话者相关语音识别实验*　用 Sensory 公司的语音识别芯片 RSC-164 做识别实验，识别前先对所有数据做高频提

升。16 句训练数据取自增强后的较为纯净的语音，所有语音都通过式(17)作高频提升：

$$H(z)=1-\alpha z^{-1}, \qquad \alpha \le 1 \tag{17}$$

实验数据由人工统计完成，得到控制命令识别正确的句子个数除以测试的句子总数得到识别率百分比。表 1 是原始含噪语音、高通滤波结果和本文提出改进算法输出的语音识别率比较。经过增强算法后，识别正确率(Corr. Rate)得到很大提高。

4.3.2 基于 HMM 的语音识别实验 识别系统采用 triphone HMM 模型，每个 triphone 模型有 3 个状态，每个状态用 6 混合高斯分布描述。HMM 训练数据取自 863 数据库，男女声各 79 人，每人 650 句子。表 2 是基于 HMM 的小词汇量短语识别结果(共 18 个词组成控制命令)。由于行驶中的汽车环境可近似看作一扩散噪声场，McCowan 的算法也放到结果比较中。改进算法在基于 HMM 的小词汇量短语识别中，无论是音节正确率、准确率(Acc.)还是句子正确率都取得较好结果。

表 1 话者相关语音识别率比较

Tab.1 The speaker independent speech recognition rates

	原始含噪语音	高通滤波 (200Hz)	改进算法
Corr. Rate	25%	30.8%	97.6%

表 2 基于 HMM 的语音识别率比较

Tab.2 The speech recognition rates based on HMM

	原始含噪语音	延迟-相加波束形成器	维纳后滤波波束形成器	散射噪声场维纳后滤波	改进算法
音节 Corr.	52.0%	65.1%	64.8%	68.9%	87.2%
Acc.	47.7%	60.0%	62.3%	66.2%	86.5%
句子 Corr.	35.3%	49.1%	52.2%	59.8%	84.4%

5 结束语

本文提出了一种改进的具有维纳后滤波的波束形成器，并通过实验分别从信噪比(SNR)，平均谱距离(ALSD)和语音识别(ASR)等 3 个方面对该算法性能进行评估。尽管谱减法能减去低频噪声，但避免直接用谱减法，而是维纳滤波方式，既去除了低频噪声又避免了音乐噪声；且通过对多个理想信号功率谱估值的平均，得到较为准确的维纳滤波器估计，从而使语音信号得到更好的还原，信噪比被提高，处理后语音与纯净语音之间的谱距离得到改善，语音识别率也较其他一些算法有一定提高。

参 考 文 献

[1] Åhgren P. Teleconferencing, system identification and array processing. [Master Thesis], Uppsala University, 2001.

[2] Spriet A. Stochastic gradient-based implementation of spatially preprocessed speech istortion weighted multichannel Wiener filtering for noise reduction in hearing aids. *IEEE Trans. on Signal Processing*, 2005, 53(3): 911−925.

[3] Zelinski R. A microphone array with adaptive post–filtering for noise reduction in reverberant rooms. IEEE Int. Conf. Acoustic Speech Signal Process, New York, 1988: 2578−2581.

[4] Meyer J, Simmer K U. Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction. IEEE Int. Conf. Acoustic Speech Signal Process, Munich, Germany, 1997: 1167−1170.

[5] Zhang Xianxian, Hanson J H L, Arehart Kathryn. Speech enhancement based on a cobined multi-channel array with constrained iterative and auditory masked processing. IEEE Int. Conf. Acoustic Speech Signal Process, Montreal, Canada, 2004: 229−230.

[6] McCowan I A, Bourland H. Microphone array post-filter for diffuse noise field. IEEE Int. Conf. Acoustic Speech Signal Process, Orlando, USA, 2002: 905−908.

[7] Griffiths L J, Jim C W. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagat*, 1982, 30(1): 27−34.

[8] Gannot S, Cohen I. Speech enhancement based on the general transfer function GSC and postfiltering. *IEEE Trans. on Speech and Audio Processing*, 2004, 12(6): 561−571.

[9] Luo Fa-Long, Yang Jun, Pavlovic C, Nehorai A. Adaptive null-forming scheme in digital hearing aids. *IEEE Trans. on Signal Processing*, 2002, 50(7): 1583−1590.

[10] Fischer S, Kammeyer K D, Simmer K U. Adaptive microphone arrays for speech enhancement in coherent and incoherent noise fields. Invited talk at Third Joint Meeting of the Acoustical Society of America & the Acoustical Society of Japan, Hawaii, 1996. http://www.ant.uni-bremen.de/pub/speech/

[11] Allen J B, Berkley D A. Image method for efficiently simulating small room acoustics. *J. Acoust. Soc. Am.*, 1979, 65(4): 943−950.

阎兆立： 男，1974 年生，博士，助理研究员，目前主要研究多通道声信号处理和采集.

杜利民： 男，研究员，博士生导师，从事信号与信息处理技术研究.