

文章编号:1002-8684(2005)01-0059-03

一种空域、时域与频域相结合的阵列语音增强算法

· 论文 ·

李东红, 曾庆宁

(桂林电子工业学院 通信与信息工程系, 广西 桂林 541004)

【摘 要】提出了一种空域、时域与频域相结合的阵列语音增强算法,对含噪信号进行处理:第一步,先在空域及时域内利用波束形成得到一个初步增强的语音信号;第二步,利用前面得到的增强信号,在时域内采用多通道自适应滤波消除相关噪声;第三步,在频域内利用谱相减消除残余噪声。整个算法的消噪量大,而且阵列的孔径小,能应用于多种噪声环境。

【关键词】传声器阵; 语音增强; 波束形成; 自适应滤波; 谱相减

【中图分类号】TN912.35

【文献标识码】A

An Array Speech Enhancement Algorithm Based on the combination in Spacial, Time and Frequency Domain

LI Dong-hong, ZENG Qing-ning

(Department of Communication and Information Engineering,

Guilin Institute of Electronic Technology, Guilin 541004, China)

【Abstract】A multimicrophone array with the algorithm that combined in spacial domain, time domain and frequency domain is proposed in this paper. The proposed system for noise reduction is processed in three steps: Beamforming is firstly used to get the enhanced speech, then it is applied to the second step, in which an adaptive multichannel filter as the desired component, thus the possible coherent noise components are canceled. The residual additive noise components are reduced in frequency domain by spectral subtraction. The scheme proposed can remove quite a lot of noise with a small size of the array, thus it can be used to many kinds of noise cancellations.

【Key words】microphone array; speech enhancement; beamforming; adaptive filter; spectral subtraction

1 引言

人们已经研究了很多有效去除环境噪声的语音增强算法,其中利用传声器阵消噪成为近年来研究的新热点。但有时由于空间尺寸的限制或为携带方便,往往要求阵列的尺寸不能太大。笔者提出一种基于微型传声器阵的语音增强算法,从而使适用的环境更加广泛。

因为不可能获得混合信号的先验统计特性,因此,实用的消噪系统必须是自适应的,并且这种方法只采用很少数量的传声器就可达到较高的消噪效果^[1]。自适应噪声对消是一种常用的去除传声器信号中相关噪声成分的方法^[2]。传统的对消系统由2个传声器组成,而这里采用的是含多抽头的多通道宽带阵列作为对消系统的参考信号,且主通道信号是用固定波束形成获得的初步的增强语音。

【基金项目】国家自然科学基金(60272038)。

谱相减方法只需计算1路传声器信号,因此以其计算量小、结构简便得到了广泛的应用。它能有效抑制背景噪声,但同时却会产生残余噪声,又被称为音乐噪声^[3]。但谱相减方法只适用于平稳的噪声环境,且该算法对噪声的削弱以损伤语音为代价,因此含噪语音的信噪比越低时,消噪效果越明显,但语音受到的损伤也就越大,增强后的语音信号的可懂度就越低。

笔者提出了一种综合上述算法的新型的自适应消噪系统,并充分利用空域、时域和频域消噪技术。

2 传声器阵语音增强系统

2.1 结合波束形成、自适应滤波和谱相减算法的语音增强系统

笔者提出的语音增强算法的结构框图如图1所示。其中的传声器阵采用的是一个二维的面阵。语音源离阵列较近,而噪声源离阵列较远,以便区分期望的语

音信号 s 和噪声干扰 n_i 。传声器的间距不能太大,以便于应用自适应算法去除相关噪声。

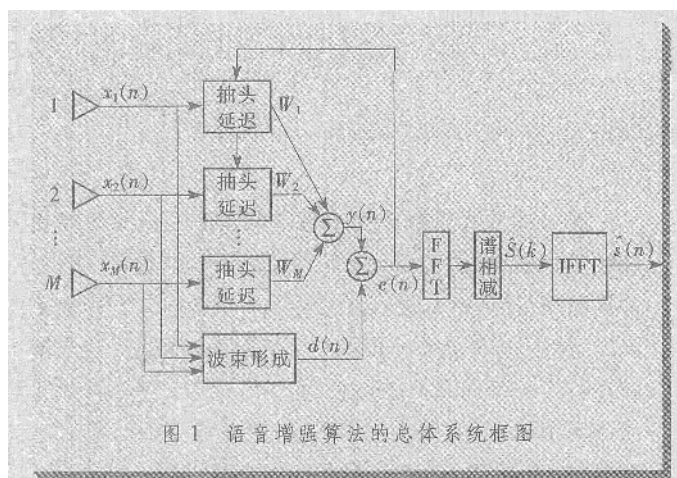
由于采用的微型阵列孔径很小,所以可忽略各个传声器上的语音到达时间的不同。各通道的信号 $x_i(n)$ 就可表示成期望语音与加性噪声的和^[4],即

$$x_i(n)=s(n)+n_i(n), \quad i=1,2,\dots,M \quad (1)$$

其中, $s(n)$ 为期望语音信号, $n_i(n)$ 为第 i 个传声器上的噪声, n 是时域信号的采样序列。然后系统分 2 路对含噪信号进行处理:下面的一路采用固定波束形成得到初步增强的语音 $d(n)$,然后将其作为上面一路自适应滤波的期望信号。通过自适应过程消去相关噪声后,再将得到的增强信号 $e(n)$ 变换频域:先对信号做加窗处理后再做 FFT 以防频谱泄漏,然后就可分帧对信号进行谱相减得到频域增强信号 $\hat{S}(k)$,最后进行 IFFT 后相邻帧重叠相加就可得到最终的增强语音 $\hat{s}(n)$ 。

2.2 波束形成及多通道宽带阵列的自适应滤波

如图 1 所示,采用含多抽头的多通道宽带阵列作为对消系统的参考信号。采用 M 个传声器组成一个多边形面阵,由于采用了波束形成算法,使得这个阵列结构不仅可提供最强的指向性,而且能最大限度地抑制期望方向外的噪声。



波束形成能利用传声器阵进行时域和空域滤波,从而消去主波束方向之外的噪声信号,并保留期望信号^[7]。图1中共有 M 个传声器,即有 M 个输入通道,这里采用时域波束形成方法,即

$$d(n)=\frac{1}{M} \sum_{i=1}^M x_i(n) \quad (2)$$

因此,只要残余噪声的方向不在主波束的方向上,就会相互抵消,达到初步降噪的目的。

然后进入自适应处理过程:首先利用 M 个传声器信号作为参考输入 $x_i(n)$,则自适应结构的输出为

$$y(n)=\sum_{i=1}^{M-1} \sum_{j=1}^N w_{ij}(n) x_{ij}(n) \quad (3)$$

其中, N 代表每个通道抽头的数量。根据 LMS 算法,主通道信号 $d(n)$ 与输出 $y(n)$ 的差值为

$$e(n)=d(n)-\sum_{i=1}^{M-1} \sum_{j=1}^N w_{ij}(n) x_{ij}(n) \quad (4)$$

则每个权值更新的表达式为

$$w_{ij}(n+1)=w_{ij}(n)+\mu x_{ij}(n) e(n) \quad (5)$$

由上述分析可知,这步处理可以对期望方向以外的噪声和干扰实现自适应对消。即当期望方向以外的声场发生变化(包括方向和强度)时,传声器系统都会自适应地补偿这种变化。甚至连微弱的风信号也会被当作干扰加以滤除。

2.3 谱相减

经过前 2 步处理,特别是自适应消噪以后,信噪比有了大幅度的提高,特别是语音的包络已从被淹没状态变得比较清晰,残余噪声也趋于平稳。所以,这时再采用谱相减算法,有声无声检测的误码率将大大降低,噪声帧的判断也更准确,从而可使处理后的增强语音中的音乐噪声降至最低。

首先,假设经过前面的算法后的增强语音由清洁语音信号 $s(n)$ 和非相关加性噪声 $n(n)$ 组成,即

$$e(n)=s(n)+n(n) \quad (6)$$

将输入信号按帧处理,对每帧含噪信号先加窗,前后帧之间重叠 50%,然后进行 FFT 变换,则频域的含噪语音信号的对应表达式为

$$E(k)=S(k)+N(k) \quad (7)$$

又由于语音和噪声不相关,其功率谱的表达式为^[5]

$$|E(k)|^2 \approx |S(k)|^2 + |N(k)|^2 \quad (8)$$

由于无法直接获得纯净噪声的能量谱 $|N(k)|^2$,一般采用将无声阶段的数帧相邻噪声信号进行能量谱加权平均得到噪声能量谱估计 $|\hat{N}(k)|^2$,则纯净语音信号的功率谱为

$$|\hat{S}(k)|^2 = |E(k)|^2 - |\hat{N}(k)|^2 \quad (9)$$

由于噪声能量谱估计与含噪语音中的噪声存在差异,所以式(9)可能出现负值,为避免能量谱出现负值,可将这些负值设为零,这一处理过程称为半波整流(half-wave rectification)。通过半波整流,纯净语音能量谱估计可表示为

$$|\hat{S}(k)|^2 = \begin{cases} |\hat{S}(k)|^2 & |\hat{S}(k)|^2 > 0 \\ 0 & \text{其它} \end{cases} \quad (10)$$

上述谱相减方法的最大缺点是它通常会带来“音乐噪声”，因此 Berouti 等人在通用谱相减的基础上又提出了一种降低音乐噪声的改进的谱相减算法^[6]。该算法中的增强信号能量谱 $|\hat{S}(k)|^2$ 的表达式为

$$|\hat{S}(k)|^2 = |E(k)|^2 - \alpha |\hat{N}(k)|^2 \quad (11)$$

其中, $\alpha = \alpha_0 - \frac{3}{20} \text{SNR}$ ($-5 \text{ dB} \leq \text{SNR} \leq 20 \text{ dB}$) 是过减因子, 它是时变的, 用于控制噪声的削减程度, α_0 是信噪比等于 0 dB 时所期望的过减因子 α 值, 一般取 $\alpha_0 = 4$ 。

为避免式(11)出现负值, 仍采用半波整流, 则

$$|\hat{S}(k)|^2 = \begin{cases} |\hat{S}(k)|^2 & |\hat{S}(k)|^2 > \beta |\hat{N}(k)|^2 \\ \beta |\hat{N}(k)|^2 & \text{其它} \end{cases} \quad (12)$$

其中, β 为频谱基 (spectral floor), 引入过减因子和频谱基的谱相减算法是通过减少增强信号的谱偏移来降低音乐噪声。频谱基 β 是为防止增强信号的谱分量低于 $\beta |\hat{N}(k)|^2$ 而在谱峰之间引入宽带噪声谱来掩蔽邻近的音乐噪声分量。这种改进方法在一定程度上减小了音乐噪声, 但以语音的部分失真为代价。然而, 由于其听觉效果更好, 笔者就采用了这种改进的谱相减方法。

由于人耳对相位不敏感, 可采用输入信号中的相位 $\theta(k)$ 作为重建后增强语音信号的相位, 即

$$\hat{S}(k) = e^{j\angle\theta(k)} \sqrt{|\hat{S}(k)|^2} \quad (13)$$

最后采用 IFFT 对增强的语音信号进行逐帧变换, 并采用重叠相加, 从而得到最终的时域增强语音 $\hat{s}(n)$ 。

3 实验及仿真结果

采用以上描述的语音增强系统, 在室外 (反射较弱) 进行了一系列不同信噪比的实验。其中的噪声环境是一段在公园里录制的环境噪声, 且噪声源距传声器阵较远。语音源是一段女性说话人的语音, 说话人站在距传声器阵较近的地方, 且正对传声器阵。采用了 $M=4$ 个普通传声器, 组成一个孔径为 5 cm 的四边形面阵。采样频率是 8 kHz, 采用长度为 32 ms 的 Hanning 窗加在每帧信号上, 相邻的帧采用 50% 重叠相加, 因此每帧数据的长度是 256 个采样点。

这里只给出一组信噪比较低的仿真结果, 因为它更能说明此系统的优越性。最初的 4 路含噪语音如图

2 所示, 语音几乎完全淹没在噪声中, 只能听见有人说话, 但很难听清其内容。

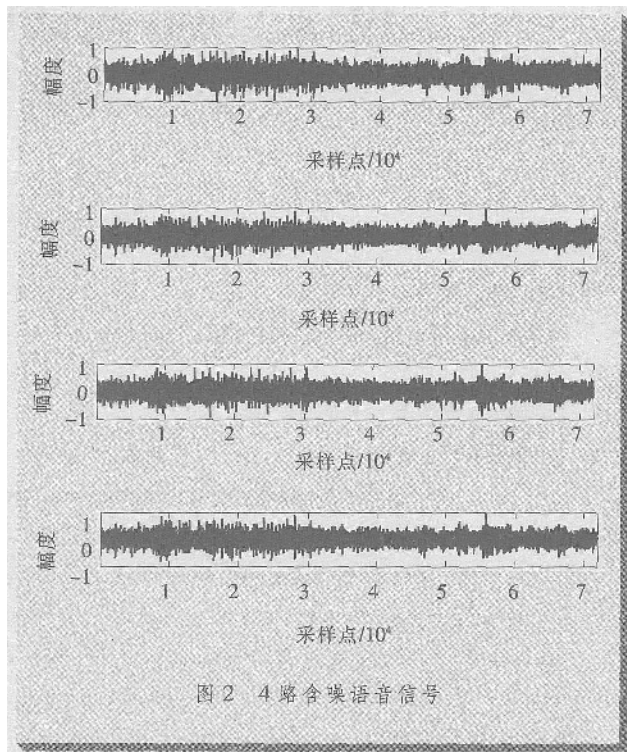


图2 4路含噪语音信号

图 3 是处理后的增强语音。从图中可以看出最后获得的增强语音的质量有显著的提高, 尽管残余的背景噪声还没有完全被消除, 但试听时, 语音的内容已经能够听清楚了。经过整个系统的处理后, 噪声的能量共下降了 25 dB, 其中包括波束形成和自适应滤波得到的 18 dB 和谱相减得到的 7 dB。

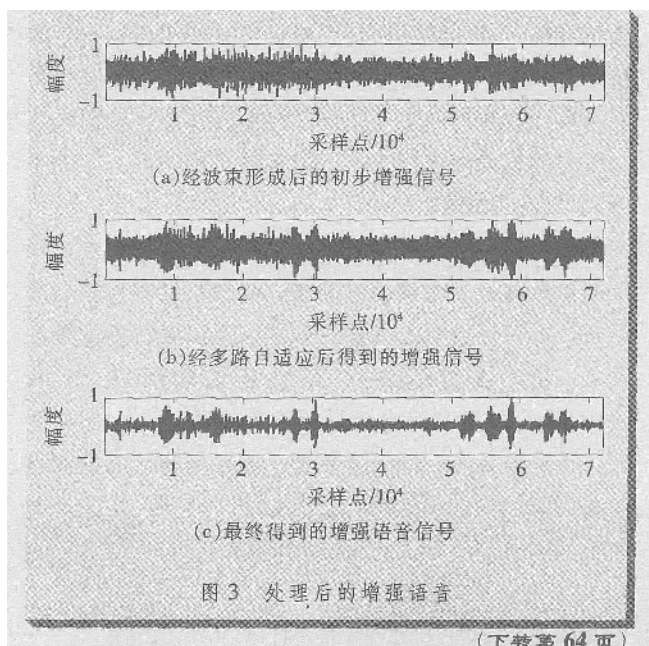


图3 处理后的增强语音

(下转第 64 页)

表 1 校准结果

校准当地灵敏度		$\Delta S_p + \Delta S_m$		校准标准灵敏度		产品灵敏度		误差	
$/(mV \cdot Pa^{-1})$	/dB	$/(mV \cdot Pa^{-1})$	/dB	$/(mV \cdot Pa^{-1})$	/dB	$/(mV \cdot Pa^{-1})$	/dB	$/(mV \cdot Pa^{-1})$	/dB
42.54	-27.42	1.05	0.41	43.59	-27.01	44.33	-27.07	-0.74	0.06
42.82	-27.37			43.87	-26.96	44.56	-27.02	-0.69	0.06
39.24	-28.13			40.29	-27.72	40.93	-27.76	-0.64	0.04
46.64	-26.66			47.69	-26.25	48.55	-26.28	-0.86	0.03

mm电容传声器,若采用较专业的精密校准法——耦合腔互易校准法,那么在 63 Hz~10 kHz 频率范围的误差小于 0.07 dB^[3],参考文献[4]中的多功能声校准器的校准不确定度是 0.08 dB。可见该校准方法虽然简单,但几乎具有专门校准设备的精度。

4 结论

笔者提出的用灵敏度准确的标准传声器作参考传声器,对 $\Phi 12.7$ mm 测试电容传声器灵敏度进行比对校准方法简单易行,可用于需要经常校准的测试传声

器的工作场所。该方法得出的当地灵敏度可直接使用。另外,如果已知传声器的标准灵敏度,那么通过测量测试环境的气压、气温和相对湿度,再根据式(4)可计算得到当地灵敏度。

参考文献

- [1] 徐占朝,张盛林,赵国志等. 传声器的技术特性与种类. 中国医学教育技术, 2002, 16(1): 58—60.

- [2] 张宝峰,许智钦,王宇华等. 测试电容传声器声压灵敏度校准方法研究. 天津大学学报, 2001, 34(4): 521—524.
- [3] 陈剑林,张美娥. 12.7 mm 电容传声器耦合腔互易法校准. 现代计量测试, 1999, (5): 13—19.
- [4] 陈剑林,杨增丰,张美娥等. 多功能声校准器. 现代计量测试, 1997, (3): 29—34.

作者简介

董明荣,硕士,主要研究方向为飞行器噪声声场技术研究。

[收稿日期] 2004-07-18

(上接第 61 页) 这里必须指出,由于改进的谱相减算法使得最后的增强语音会有一点失真,但因人耳对相位不敏感,所以对其听觉效果并没有什么影响。

4 结论

提出了一种新的适用于室外环境下的自适应语音增强系统。理论分析和实验结果都表明这种新的消噪算法可以有效地去除环境噪声,且阵列的体积小,因此可适用于多种噪声环境。又因为整个处理过程结构简单,计算量小,因此该算法就更容易应用在 DSP 系统中进行实时信号处理,具有更大的实用价值。

参考文献

- [1] Akihiko Sugiyama(NEC Corporation). An Adaptive Microphone Array with Good Sound Quality Using Auxiliary Fixed Beamformers and Its DSP Implementation Osamu Hoshuyama, 1999, 2: 949—952.
- [2] Liu Tienan, Wang Liguang, Xu Baochang, Xie Aihua, Zhang Hang. Adaptive Noise Canceller and Its Applications for Systems with Time-variant Correlative Noise. Intelligent Control and Automation. Proceedings of the

4th World Congress. on, 2002(2): 1 412—1 415.

- [3] Boll S F. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Trans.ASSP, 1979, ASSP-27(2): 113—120.
- [4] Griffiths L, Jim C. An Alternative Approach to Linearly Constrained Adaptive Beamforming. Antennas and Propagation. IEEE Transactions on [legacy, pre - 1988], 1982, 30 (1): 27—34.
- [5] Marro C, Mahieux Y, Simmer KU. Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfiltering. IEEE Trans. Speech and Audio Processing, 1998, (6): 240—259.
- [6] Berouti M, Schwartz R, Makhoul J. Enhancement of Speech Corrupted by Acoustic Noise. Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1979-04. 208—211.
- [7] Van Veen B D, Buckley K M. Beamforming: A Versatile Approach to Spatial Filtering. IEEE ASSP Magazine, 1988, (4): 4—24.

[收稿日期] 2004-08-30