

基于近场波束形成的麦克风阵列语音增强方法

王冬霞 殷福亮

(大连理工大学电子与信息工程学院 大连 116024)

摘要: 当麦克风阵列用于封闭环境中非手持式语音拾取时, 必须面对的一个问题是声场为阵列近场的问题。该文在子带自适应波束形成方法的基础上, 引进了一种基于近场波束形成的麦克风阵列语音增强方法。该方法充分利用了近场球面波的波前弯曲率, 有效地衰减了混响和噪声对期望信号的影响。仿真实验结果表明, 在小房间混响条件下, 基于近场波束形成的麦克风阵列语音增强方法取得了较好的噪声抑制效果。

关键词: 麦克风阵列; 语音增强; 近场波束形成

中图分类号: TN912.3

文献标识码: A

文章编号: 1009-5896(2007)01-0067-04

A Nearfield Beamforming Method for Microphone Array Based on Speech Enhancement

Wang Dong-xia Yin Fu-liang

(School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: When using a microphone array for hands-free speech acquisition in enclosure environments, one can be faced with the problem of wave propagation in near-field. In this paper, therefore, a near-field beamforming method for microphone array based speech enhancement is introduced which modified subband adaptive beamforming method. The proposed method takes full advantage of spherical nature of the sound wavefront, in which the distance discrimination reduces the effect of reverberation as well as noise on the desired speech signal. Simulations experimental results demonstrate that the proposed microphone array based speech enhancement method exhibits a better noise reduction performance than other methods.

Key words: Microphone array; Speech enhancement; Near-field beamforming

1 引言

在车载系统、电话会议与多媒体会议等非手持式的智能语音通信系统中, 由于受到环境噪声、混响以及干扰等因素的影响, 麦克风阵列拾取的语音信号质量较差, 而这将直接影响到语音编码及语音识别系统的性能。因此, 需要进行有效的噪声抑制, 以增强封闭环境下的带噪语音信号质量。

大多数的麦克风阵列语音增强方法, 建立在声源位于阵列远场的假设基础上^[1-4]。在这种情况下, 声波的波前弯曲率可以忽略, 阵列的所有入射波均是平面波。该假设简化了分析过程, 在大房间会议室中产生了较好的性能。然而, 当麦克风阵列用于一些特殊环境, 诸如小房间非手持式语音拾取时(麦克风阵列安置在计算机显示器上面), 与前者相比, 声源与麦克风阵列的距离较小, 即声源位于阵列的近场中。此时, 声波的波前弯曲率可以由阵列探测到而不能忽略, 阵列的所有入射波均是球面波。在这种情况下, 如果仍然使用远场波束形成方法来增强带噪语音信号, 那么麦克风阵列语音增强系统的性能急剧下降^[5]。

因此, 近场问题是封闭环境下, 麦克风阵列语音增强方

法必须面对的一个问题。Ryan 等提出了近场环境下, 将麦克风阵列用于非手持式语音拾取的方法, 并取得了较高的阵列输出增益^[6,7]。但是, 该方法是基于统计特性基础上的固定波束形成方法, 因而缺少对环境的适应性。Zheng 等提出了基于奇异值分解基础上的近场自适应波束形成方法, 也取得了较好的阵列输出效果^[8]。但是, 该方法的理论模型基础仍然是广义旁瓣抵消器(GSC)^[1], 所以对非平稳及弱相关噪声的抑制较差。

对于封闭房间里的声源来讲, 近场源的直达波波前比其反射波的波前具有更大的弯曲率。与原始声源相比, 反射波相应的镜像源往往位于距离阵列更远的地方^[6]。如果阵列聚焦于近场声源的话, 那么可以通过距离差来减小位于远场的混响和噪声对纯净语音的影响。因此, 本文则从该角度出发, 充分利用近场球面波的波前弯曲率, 在子带波束形成的基础上^[4], 提出一种基于近场波束形成的麦克风阵列语音增强方法。该方法试图进一步衰减混响和噪声对期望信号的影响, 从而改善近场环境下, 带噪语音的质量。

本文的具体结构安排如下: 在所选声学模型的基础上, 给出基于近场波束形成的麦克风阵列语音增强方法的整体框图; 接着从理论上对该方法进行推导; 然后给出在仿真环境下, 该方法语音增强的效果及与其他方法的噪声性能比较; 最后给出结论。

2005-06-08 收到, 2005-11-22 改回

国家自然科学基金(603720821, 60172073)和教育部跨世纪优秀人才基金资助课题

2 位于阵列近场的声波传播模型

如图 1 所示, 设 r 为声源与参考麦克风的距离, θ 与 φ 分别表示相应的方位角和仰角。在球坐标系下, 声波 $s(r, t) = s(r, \theta, \varphi, t)$ 的波动方程表示为^[9]

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial s}{\partial r} \right) + \frac{1}{r^2 \sin \phi} \frac{\partial}{\partial \phi} \left(\sin \phi \frac{\partial s}{\partial \phi} \right) + \frac{1}{r^2 \sin^2 \phi} \frac{\partial^2 s}{\partial \theta^2} = \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} \quad (1)$$

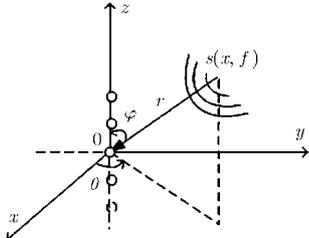


图 1 球面坐标系下的声波与阵列

式中波速 $c = 343 \text{ m/s}$ 。考虑到多数情况下, 近场源的声场表现为各向同性, 即在 θ, φ 方向上没有变化, 式(1)简化为

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial s}{\partial r} \right) = \frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} \quad (2)$$

于是, 得到波动方程的解为

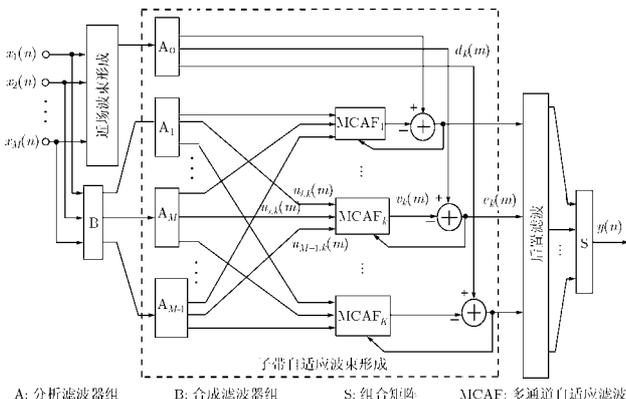
$$s(r, \omega, t) = \frac{A}{r} \exp\{j(\omega t - kr)\} \quad (3)$$

式中 $k = \omega/c$ 为声波波数, 其中 ω 为声波角频率。当 $r \rightarrow \infty$ 时, 球面波 $s(r, \omega, t)$ 的表达式与平面波形式一致。

由式(3)可以看出, 声源距离阵列越远, 信号幅值衰减越大。如果期望声源位于阵列近场(其混响信号的镜像源将位于距离阵列较远的地方), 干扰噪声源处于远场, 那么这一特性可以用来衰减那些相对阵列距离较远混响和噪声, 进一步提高系统对噪声和混响的抑制能力。

3 麦克风阵列语音增强方法的描述

本文以麦克风阵列接收的信号为式(3)所示的球面波为对象, 在子带自适应波束形成的基础上^[4], 采用基于近场波束形成方法, 来解决封闭环境下(期望声源位于近场、噪声源处于远场)带噪语音的增强问题。该方法的系统结构如图 2 所示, 主要有 3 个模块构成, 即近场波束形成、子带自适应波束形成及后置滤波。下面, 分别就这 3 个模块对该方法进行理论阐述。



A: 分析滤波器组 B: 合成滤波器组 S: 组合降噪 MICAF: 多通道自适应滤波

图 2 基于近场波束形成的麦克风阵列语音增强方法框图

3.1 近场波束形成

位于麦克风阵列近场的声源方向矢量表示为

$$\mathbf{a}_s(r_w, \omega) = \frac{r_w}{e^{-jk r_w}} \left[\frac{e^{-jk r_{1w}}}{r_{1w}}, \dots, \frac{e^{-jk r_{iw}}}{r_{iw}}, \dots, \frac{e^{-jk r_{Mw}}}{r_{Mw}} \right]^H \quad (4)$$

式中 H 表示共轭转置, M 为麦克风数目。假设在二维投影平面内, $x_w = (r_w, \theta_w)$ 为声源位置, $x_i = (r_i, \theta_i)$ 为麦克风位置, r_{iw} 为声源与第 i 只麦克风之间的距离, 即

$$r_{iw} = \sqrt{r_w^2 + r_i^2 - 2r_w r_i \cos(\theta_i - \theta_w)} \quad (5)$$

这是信号的空间表示形式。考虑到语音信号是宽带信号, 则阵列信号的延迟矢量为

$$\mathbf{a}_t(r_w, \omega) = \mathbf{a}_t = [1, \dots, e^{-jk(L-1)cT}]^H \quad (6)$$

这是信号的时间表示形式。于是, 式(4)的方向矢量可以进一步表示成空时二维信号矢量形式, 即

$$\begin{aligned} \mathbf{a}(r_w, \omega) &= \mathbf{a}_s(r_w, \omega) \otimes \mathbf{a}_t(r_w, \omega) \\ &= \frac{r_w}{e^{-jk r_w}} \left[\frac{e^{-jk r_{1w}}}{r_{1w}}, \dots, \frac{e^{-jk r_{Mw}}}{r_{Mw}}, \frac{e^{-jk(r_{1w} + cT)}}{r_{1w}}, \dots, \right. \\ &\quad \left. \frac{e^{-jk(r_{Mw} + (L-1)cT)}}{r_{Mw}} \right]^H \end{aligned} \quad (7)$$

式中 \otimes 为 Kronecker 积。

麦克风阵列的近场波束形成输出信号为

$$\begin{aligned} d(n) &= \mathbf{C}^H \mathbf{a}(r_w, \omega) = \frac{1}{M} \sum_{i=1}^M \sum_{l=1}^{L-1} \mathbf{a}_{il}(r_w, \omega) \mathbf{C}_{il}^* \\ &= \frac{1}{M} \sum_{i=1}^M \sum_{l=1}^{L-1} \mathbf{a}_{il}(r_w, \omega) \frac{r_{is}}{r_s} \exp\{-jk(r_s - r_{is} - (l-1)cT \\ &\quad + r_i \cos(\theta_i - \theta_s))\} \end{aligned} \quad (8)$$

式中符号 $*$ 表示复数共轭, 近场补偿函数矢量 \mathbf{C} 为

$$\begin{aligned} \mathbf{C} &= [\mathbf{C}_{il}]_{ML \times 1} \\ &= \left[\frac{r_{is}}{M r_s} \exp\{jk(r_s - (l-1)r_{is} + r_i \cos(\theta_i - \theta_s))\} \right]_{ML \times 1} \end{aligned} \quad (9)$$

由式(9)可以看出, 通过对球形波的波前补偿, 各路期望语音信号进行相关求和($r_w = r_s$), 因此, 期望信号得到增强。包括与直达波语音信号具有相同波达方向的混响和干扰噪声在内的其它信号($r_w \neq r_s$), 由于幅值和(或)距离的差别而得到初步衰减。

3.2 分析合成滤波器

如图 2 所示的子带自适应波束形成模块中, 分析滤波器组 A 将全带信号均分成 K 个子带, 抽取因子为 D 。经过子带自适应波束形成处理, 合成滤波器组 S 在系统的输出端将各子带信号合成为全带表示。

本文选取内插因子和抽取因子相等的均匀滤波器组, 即 $I = D$ 。并且, 为了减小子带信号间的混叠效应, 选取 $D < K$ 的过采样方式。鉴于语音为实信号, 本文选取分析和合成滤波器分别是原型低通滤波的余弦调制形成, 即

$$a_k(n) = 2a(n) \cos \left[(2k+1) \frac{\pi}{2K} \left(n - \frac{L_f - 1}{2} \right) + (-1)^k \frac{\pi}{4} \right] \quad (10)$$

$$b_k(n) = 2b(n) \cos \left[(2k+1) \frac{\pi}{2K} \left(n - \frac{L_f - 1}{2} \right) - (-1)^k \frac{\pi}{4} \right] \quad (11)$$

其中 $a(n)$ 和 $b(n)$ 分别是分析与合成滤波器组的原型低通线性滤波器, 滤波器长度 L_f 为 $2K$ 的整数倍。在满足完全重构条件基础上, 确保原型滤波器的阻带能量最小化, 从而进一步减小频带间的混叠效应^[10]。

为表示子带处理带来信号采样率的变化, 本文假定 n 为原始的时域单位, m 为子带波束形成过程中的时域单位。

3.3 子带自适应波束形成

假定第 k 个子带自适应波束形成的输入信号数据矢量为 $\mathbf{u}_k(m) = [u_{1,k}(m), \dots, u_{M-1,k}(m)]^T$, 固定波束形成器输出的子带信号为 $d_k(m)$, 空间滤波器矢量为 $\mathbf{w}_k^{(p)}(m) = [w_{1,k}^{(p)}(m), \dots, w_{M-1,k}^{(p)}(m)]^T$, 则子带多通道自适应波束形成器(MCAF)的输出信号为

$$v_k(m) = \sum_{p=0}^{L_w-1} [\mathbf{w}_k^{(p)}(m)]^T \mathbf{u}_k(m-p) \quad (12)$$

式中 T 表示转置, L_w 为子带滤波器的长度。子带自适应波束形成的误差信号为

$$e_k(m) = d_k(m) - v_k(m) \quad (13)$$

基于 NLMS 自适应子带滤波器的权值更新方程为

$$\mathbf{w}_k^{(p)}(m+1) = \mathbf{w}_k^{(p)}(m) + \frac{\alpha}{P_k[\mathbf{u}_k(m)]} \mathbf{u}_k^*(m) e_k(m) \quad (14)$$

式中 α 为步长。

为了跟踪语音的时变特性, 并考虑到功率谱估计偏差对系统的影响, 输入数据矢量的功率估计 $P_k[\mathbf{u}_k(m)]$ 采用递推方式得到, 即

$$P_k[\mathbf{u}_k(m)] = \beta P_k[\mathbf{u}_k(m-1)] + (1-\beta) \sum_{l=1}^{M-1} \|\mathbf{u}_{l,k}(m)\|^2 \quad (15)$$

式中 $\beta \in [0, 1]$ 为遗忘因子。

3.4 后置滤波

考虑到子带自适应波束形成去除的是强相关的噪声, 其输出端仍存在部分残留噪声, 因此本文引进多通道后置滤波器(如图 2 所示), 去除那些非平稳和弱相关的噪声。

第 k 个子带的滤波器频率响应的表达式为

$$H_k(e^{j\omega}) = \frac{2}{(M-1)(M-2)} \sum_{i=0}^{M-3} \sum_{j=i+1}^{M-2} P_{u_i u_j}(e^{j\omega}) / P_{e_k}(e^{j\omega}) \quad (16)$$

式中 $P_{e_k}(e^{j\omega})$ 表示自功率谱的估计值。为了跟踪语音的时变特性, 并考虑到功率谱估计偏差对系统的影响, 输入数据矢量的功率估计 $P_{e_k}(e^{j\omega})$ 仍然采用递推方式得到, 即

$$P_{e_k}(l+1, e^{j\omega}) = \eta P_{e_k}(l, e^{j\omega}) + |e_k(l, e^{j\omega})|^2 \quad (17)$$

式中 η 为遗忘因子, l 表示帧数。

麦克风阵列语音增强系统的输出端信号, 即合成滤波器

输出信号为

$$y(n) = \begin{cases} \sum_{k=1}^K [F^{-1}(H_k(e^{j\omega})) * e_k(m)] * b_k(m), & n = mD \\ 0, & \text{其它} \end{cases} \quad (18)$$

式中 F^{-1} 表示傅里叶反变换, $*$ 表示线性卷积, D 是内插因子。

4 仿真实验

4.1 仿真环境

实验选取的仿真房间结构为小房间(5m×4m×3m), 且各面墙的反射系数相同。按照 Eyring 公式, 混响时间 T_{60} 可以表示为墙壁反射系数的函数^[6], 即

$$T_{60} = \frac{0.163V}{-S \log(1-\gamma)} \quad (19)$$

式中 V , S 和 γ 分别表示房间的体积、总面积与墙壁的反射系数。线性麦克风阵列由 5 个间距为 10cm 的麦克风组成。目标语音声源的入射方向为 90° , 与麦克风阵列的距离符合近场的假设条件, 即满足

$$r_s < 2R^2/\lambda \quad (20)$$

式中 R 表示阵列孔径, λ 为工作波长。干扰噪声位于阵列远场, 信号的入射方向为 30° 。

第 i 只麦克风接收的信号为

$$x_i(n) = h_{1,i}(n) * s(n) + h_{2,i}(n) * n_I(n) \quad (21)$$

式中 $s(n)$, $n_I(n)$, $h_{1,i}(n)$ 和 $h_{2,i}(n)$ 分别为目标语音信号、干扰噪声、以及目标语音信号、干扰噪声与第 i 只麦克风之间的房间冲激响应(采用 Image 模型^[12]得到)。实验所采用的语音文本“第一课认识新同学”由男生朗读, 信号采样率为 8kHz, 每个样值用 16bit 二进制进行编码。在语音文本中加入高斯白噪声^[13]以组成不同输入信噪比输入信号。

在整个麦克风阵列语音增强方法实现的过程中, 窗函数选用长度为 256 的汉宁窗, 帧移为 128。原型滤波器的长度为 $L = 128$, 子带数为 $K = 16$, 抽取(内插)因子为 $D = 8$ 。

4.2 实验结果

图 3 与图 4 分别表示了麦克风阵列的近场和远场波束形成图。当期望声源位于阵列的近场时($r = r_1$), 由图中可以看到, 采用近场波束形成方法, 麦克风阵列取得了较大的增益, 同时衰减了那些距离较远的其它声源($r = 2r_1$ 和 $r = 15r_1$)。同样条件下, 如果采用远场波束形成方法, 则取得了较差的结果。从而, 进一步证明了前面理论分析所得出的结论。

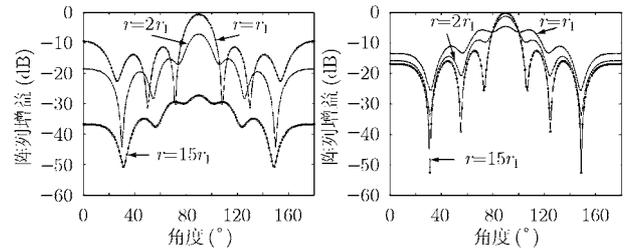


图 3 近场波束形成图

图 4 远场波束形成图

图5表示混响时间 $T_{60} = 300\text{ms}$, 信噪比为 0 dB 情况下各种信号的波形。将图5(b)与图5(a)相比较不难看出, 混响和噪声严重畸变了原来纯净的语音信号。与 Griffiths-Jim 提出的广义旁瓣抵消器 GJGSC 方法^[1]、作者提出的 CSGSCPF 方法^[5]相比, 本文提出的近场波束形成方法取得较好的去噪效果。非正式的听音测试已经证明了相应的结论。由于分析滤波器组对信号的预滤波作用, 该方法在有色噪声情况下, 也得到类似结论。

图6表明, 在近场环境下, 随着输入信噪比的增加, 几种方法的噪声抑制性能均逐渐下降。与 Griffiths-Jim 提出的广义旁瓣抵消器 GJGSC 方法, CSGSCPF 方法^[5]相比, 本文的近场波束形成方法取得较好的噪声抑制能力。

5 结束语

本文在子带自适应波束形成方法的基础上, 提出了一种基于近场波束形成的麦克风阵列语音增强方法。该方法充分利用了近场球面波的波前弯曲率, 在提高收敛速度的同时, 有效地衰减了混响和噪声对期望信号的影响。无论从实验得出的噪声抑制情况还是从非正式的听音测试都已证明, 小房间混响条件下, 基于近场波束形成的麦克风阵列语音增强方法明显改善了带噪语音信号的质量。但是, 该方法仅限于声

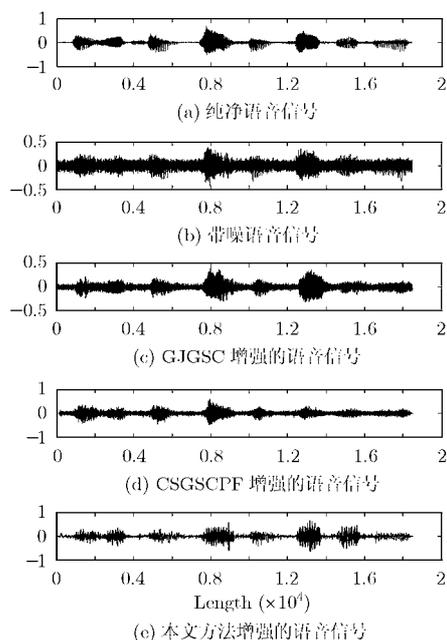


图5 白噪声情况下各信号波形

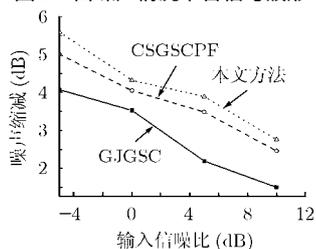


图6 几种方法的噪声抑制性能比较

源和麦克风阵列之间位置相对固定的情况。因此, 如何改善该方法的韧性将是未来要开展的工作。

参考文献

- [1] Griffiths L J and Jim C W. An alternative approach to linearly constrained adaptive beamforming [J]. *IEEE Trans. on Antennas and Propagation*, 1981, AP-30(1): 27-34.
- [2] Marro C, Mahieux Y, and Simmer K U. Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering [J]. *IEEE Trans. on Speech and Audio Processing*, 1998, (3): 240-259.
- [3] Gannot S and Cohen I. Speech enhancement based on the general transfer function GSC and postfiltering [J]. *IEEE Trans. on Speech and Audio Processing*, 2004, 12(6): 561-571.
- [4] Wang Dongxia, and Yin Fuliang. A subband adaptive learning algorithm for microphone array based speech enhancement [A]. *Lecture Notes in Computer Science* [C], Springer-Verlag GmbH, 2005, 3497: 592-597.
- [5] Ryan G J. Criterion for the minimum source distance at which plane-wave beamforming can be applied [J]. *Journal of the Acoustic Society of America*, 1998, 104(1): 595-598.
- [6] Ryan J G and Goubran R A. Array optimization applied in the near field of a microphone array [J]. *IEEE Trans. on Speech and Audio Processing*, 2000, 8(2): 173-176.
- [7] Ryan J G and Goubran R A. Application of near field optimum microphone arrays to hands free mobile telephony [J]. *IEEE Trans. on Vehicular Technology*, 2003, 52(2): 390-400.
- [8] Zheng Y R and Goubran R A. Robust near-field adaptive beamforming with distance discrimination [J]. *IEEE Trans. on Speech and Audio Processing*, 2004, 12(5): 478-488.
- [9] 杨善元, 杨燕译著. 波[M]. 北京: 地震出版社, 1988: 4-15.
- [10] Klierer J and Mertins A. Design of paraunitary oversampled cosine-modulated filter banks [A]. *IEEE International Conference on Acoustics, Speech, and Signal Processing* [C], Munich, Germany, 1997, 3: 2073-2076.
- [11] Everest A F. *The Master Hand of Acoustics*, 2nd ed [M]. New York: McGraw-Hill, 1989: 131-155.
- [12] Allen J B and Berkley D A. Image method for efficiently simulating small room acoustics [J]. *Journal of the Acoustic Society of America*, 1979, 65(4): 943-950.
- [13] 殷福亮, 宋爱民. 数字信号处理C语言程序集[M]. 沈阳: 辽宁科学出版社, 1997: 30-34.

王冬霞: 女, 1976年生, 博士生, 讲师, 研究方向为语音信号处理和阵列信号处理的理论与应用。

殷福亮: 男, 1962年生, 博士生导师, 教授, 主要研究方向为数字信号处理、语音处理、图像处理及现代通信技术。