

文章编号: 1006-2467(2000) 07-0877-04

基于二阶前向结构和信息最大理论的语音增强算法

胡光锐, 虞 晓

(上海交通大学 电子工程系, 上海 200030)

摘 要: 基于自适应噪声对消(ANC)的思想,提出了一种适合于语音增强任务的二阶 ANC 结构. 利用信息最大原理,推导了基于新结构的相应语音增强算法迭代公式.新结构的主要优点在于其能分离出原始语音信号而没有滤波失真.对真实语音信号进行大量计算机模拟实验,结果表明,新算法在噪声抑制方面具有较好的性能.

关键词: 语音增强;盲源分离;二阶自适应噪声抵消结构;信息最大原理

中图分类号: TN 912.3.12 **文献标识码:** A

Speech Enhancement Based on Second-Order Architecture and Informax Theory

HU Guang-rui, YU Xiao

(Dept. of Electronic Eng., Shanghai Jiaotong Univ., Shanghai 200030, China)

Abstract: Based on the idea of adaptive noise cancellation (ANC), a new second-order feedforward architecture was proposed for speech enhancement. According to the information maximization (Informax) theory, the gradient descend speech enhancement algorithm was proposed. The main advantage of the new architecture is that the clean speech signal can be separated without filter distortion in contrast to the other forward architecture. The computer simulation results demonstrate the good performance of the proposed algorithm by using real speech signals.

Key words: speech enhancement; blind source separation; second-order adaptive noise cancellation architecture; information maximization theory

基于最小均方误差(Minimizing Mean Square Error, MMSE)自适应噪声对消(Adaptive Noise Cancellation, ANC)算法是较为广泛应用的一种语音增强算法^[1].在ANC系统中,当辅助通道输入信号只由相关噪声组成时,这类算法(例如LMS算法)能工作得较好.但当两个通道的输入均是有用信号和噪声的混合时,算法的性能就大大地减弱了,并会在输出信号中加入所谓的“音乐噪声”.而在另一方面,语音增强可以看成是一个盲源分离(Blind Source Separation, BSS)问题^[2~8].目前,大多数

BSS算法用于源信号的即时、线性混合,即混合模式假设为 $X(t) = A \times S(t)$.但在语音增强问题中,由于语音信号传播过程中的时延效应以及空间、器件滤波效应,传感器接收的输入大多是时延和相移源信号的混合.所以语音增强问题应该看成一个带时延和卷积的BSS问题^[6,7].本文基于ANC的思想,提出了一种二阶ANC前向结构用于语音增强,利用BSS中信息最大原理,给出了基于新结构的梯度下降算法.

1 二阶ANC结构

在本文中,语音增强被模式化为从卷积混合的源信号中分离出干净的语音信号以完成语音增强任

收稿日期: 1999-09-21

基金项目: 国家自然科学基金资助项目(69672007)

作者简介: 胡光锐(1938~),男,教授

Electronic Journal Publishing House. All rights reserved. <http://www.cnki.net>

务. 假设其中一个源信号是语音信号, 另一个为干扰噪声信号. 基本的前向结构的盲信号分离结构如图1所示. 两个输入通道可以用来模拟人的两个耳朵. 主通道输入信号可用

$$\left. \begin{aligned} x_1(z) &= A_{11}(z)s_1(z) + A_{12}(z)s_2(z) \\ x_2(z) &= A_{21}(z)s_1(z) + A_{22}(z)s_2(z) \end{aligned} \right\} \quad (1)$$

在 z 变换域中表示. 其中 $A_{ij}(z)$ 是从源 j 传送到传感器 i 的传输通道响应的 z 变换.

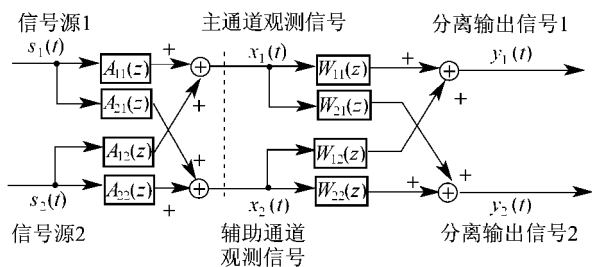


图1 两个输入源的基于前向结构的盲信号分离框图

Fig. 1 The general forward architecture for the BSS algorithm with two sources

理想情况下, 图1的分离源信号输出解可以由式(1)推导出:

$$\left. \begin{aligned} s_1(z) &= [A_{22}(z)x_1(z) - A_{12}(z)x_2(z)]/G(z) \\ s_2(z) &= [-A_{21}(z)x_1(z) + A_{11}(z)x_2(z)]/G(z) \end{aligned} \right\} \quad (2)$$

式中, $G(z) = A_{11}(z)A_{22}(z) - A_{12}(z)A_{21}(z)$. 但由于输出前向滤波网络中的 $W_{11}(z)$ 和 $W_{22}(z)$ 作用, 输出语音信号将被白化, 当算法收敛时, 基于前向结构的盲信号分离算法并不能得到理想解式(2). 如选择 $W_{11}(z)$ 、 $W_{22}(z)$ 归一化为1时, 另外两个滤波器也将成为瞬时白化滤波器^[6,7]. 所以, 基于前向结构的盲信号分离很难恢复原始的输入语音信号. Kari Torkkola^[6]提出了一种反馈网络结构的BSS方法用来处理这个问题, 但是采用反馈结构, 则必须假设输入通道传输滤波器是严格因果的. 在本文中, 假设主通道传感器的输入 $A_{11}(z)s_1(z)$ 为所需要的系统输出. 也即假设 $A_{11}(z)$ 、 $A_{22}(z)$ 为1, 同时在前向结构中设定 $W_{11}(z)$ 、 $W_{22}(z)$ 为1, 由此可以得到前向结构的理想输出为

$$\left. \begin{aligned} y_1(z) &= [1 - A_{12}(z)A_{21}(z)]s_1(z) \\ y_2(z) &= [1 - A_{12}(z)A_{21}(z)]s_2(z) \end{aligned} \right\} \quad (3)$$

由式(3)可见, 输出语音信号将发生失真, 故必须对其进行后处理. 由于未知任何有关传输信道的信息, 这一点较难做到. 为了解决上述问题, 本文提出了一种新的二阶前向结构用于语音信号增强(见图2). 新结构的基本思想来源于自适应噪声对消理

论, 其中第一阶结构用主通道的输入对消辅助通道输入中的语音信号, 理想情况下第一阶输出 $y_2(t)$ 就只是噪声信号的函数, 然后利用第二阶结构对消主通道输入中的噪声信号, 以得到干净的输出语音信号.

基于二阶前向结构的语音增强系统在 z 变换域中可以表示为

$$\left. \begin{aligned} y_1(z) &= x_1(z) - y_2(z)W_2(z) \\ y_2(z) &= x_2(z) - x_1(z)W_1(z) \end{aligned} \right\} \quad (4)$$

由此可得到新结构的最佳语音增强结果:

$$\left. \begin{aligned} y_1(z) &= s_1(z) \\ y_2(z) &= [1 - A_{12}(z)A_{21}(z)]s_2(z) \end{aligned} \right\} \quad (5)$$

很明显新结构的主要优点在于原始语音信号能被分离出来而没有滤波失真, 输出增强语音也不会有解卷积现象存在.

2 信息最大化原理

在语音增强处理中, 一般可以假设输入语音信号和干扰噪声信号是互相独立的. 采用如图2所示的二阶前向ANC结构, 当信号完全分离时, 可以想象输出 $y_1(t)$ 、 $y_2(t)$ 是互相独立的, 并用 $y_1(t)$ 是所需的增强语音输出, 此时输出信号 $y_1(t)$ 、 $y_2(t)$ 的互信息量应该等于零. 根据信息论的基本原理, 除非两个随机变量是联合高斯分布的, 计算两个变量的互信息量需要这两个变量的所有阶数的统计信息. 所以只利用输入信号的二阶统计信息的算法, 如LMS算法在新的二阶结构中难以取得较好的语音增强效果. 最大熵算法(Maximum Entropy, ME)和最小互信息量算法(Minimum Mutual Information, MMI)是两种主要用于盲源分离问题的算法. 两种算法所推导出的权值迭代公式具有相似的表达形式. MMI算法中采用的互信息量是一种最好的代价函数, 它具有对信号的线性或非线性转换不变的特性. 但是

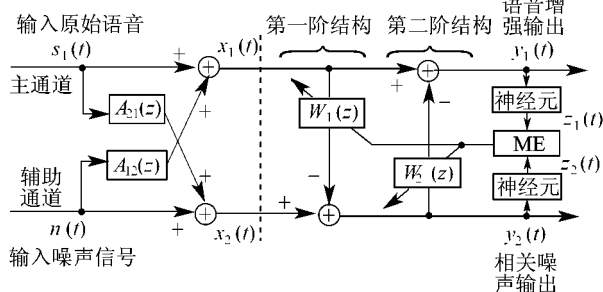


图2 基于二阶前向结构的盲信号处理语音增强算法框图

Fig. 2 The second-order feedforward architecture

MMI 算法需要大量的计算用于近似信号的高阶统计量或累积量, 并且对于语音这种复杂信号的统计估计也不可能是十分准确的. 另一方面, 尽管在除了特定情况下, ME 算法的有效性还没有得到严格的数学证明, 但在实际应用中, 特别是用在诸如语音信号等的输入中, ME 算法获得了很好的性能^[3, 6, 7]. 已有理论证明, 盲源分离问题中, 当输入独立源信号都为零均值时, ME 算法可以得到局域最佳解.

利用 A. J. Bell 的信息最大原理 (Information Maximization, Informax)^[3], 如图 2 所示, 将输出信号 $y_i(t)$ 通过一个模拟神经元, 经过非线性变换 $g(y_i(t))$ 成为 $z_i(t)$. 假设在时间 t , 两个通道的神经元输出 $z_i(t)$ 可以看成两个随机变量, 那么非线性神经元输出 $z_i(t)$ 之间的联合熵为

$$H(Z, W) = - \int f(Z, W) \log f(Z, W) dZ \quad (6)$$

根据 ME 的基本原理, 如果输入源信号是超高斯的 (Super-Gaussian), 最小化输出的互信息量就相当于最大化神经元输出的联合熵. 在实际应用中, 许多信号如语音信号都是超高斯的. 假设在时间 t , 传感器观测值 $x_1(t)$ 、 $x_2(t)$ 的联合条件概率分布为

$$f_x(x_1(t)x_2(t) | x_1(t-1)x_1(t-2) \dots x_1(t-M)x_2(t-1)x_2(t-2) \dots x_2(t-M)) \quad (7)$$

式中 M 为自适应 FIR 滤波器 $W_1(z)$ 、 $W_2(z)$ 的阶数. 由此可以得到神经元输出 $z_1(t)$ 、 $z_2(t)$ 的联合概率分布为

$$f_z(z_1(t), z_2(t)) = f_x(x_1(t)x_2(t)) / |J| \quad (8)$$

式中, $|J| = \left| \frac{\partial z_1}{\partial x_1} \frac{\partial z_2}{\partial x_2} - \frac{\partial z_1}{\partial x_2} \frac{\partial z_2}{\partial x_1} \right|$ 是雅可比矩阵行列式值的绝对值. 因此式(6)可表示为

$$H(z_1, z_2; W) = E[\ln |J|] - E[\ln f_x(x_1, x_2)] \quad (9)$$

所以, 为了使输出信号互相独立, 即最小化输出信号 $Y_1(t)$ 和 $Y_2(t)$ 之间的互信息量等价于最大化神经元输出 $Z_i(t)$ 的联合熵, 也就是使期望值 $E[\ln |J|]$ 最大.

3 梯度下降算法

在时域条件下重写二阶前向结构系统输出的表达式为

$$\left. \begin{aligned} y_1(t) &= x_1(t) - \sum_{k=0}^M w_2(k) y_2(t-k) = \\ & x_1(t) - w_2(0) y_2(t) - \sum_{k=1}^M w_2(k) y_2(t-k) \\ y_2(t) &= x_2(t) - \sum_{k=0}^M w_1(k) x_1(t-k) \end{aligned} \right\} \quad (10)$$

则可以推导出雅可比矩阵行列式

$$J = \frac{\partial z_1}{\partial y_1} \frac{\partial y_1}{\partial x_1} \frac{\partial z_2}{\partial y_2} \frac{\partial y_2}{\partial x_2} - \frac{\partial z_1}{\partial y_1} \frac{\partial y_1}{\partial x_2} \frac{\partial z_2}{\partial y_2} \frac{\partial y_2}{\partial x_1} = \frac{\partial z_1}{\partial y_1} \frac{\partial z_2}{\partial y_2} \quad (11)$$

如果选择神经元的非线性函数 $g(y_i(t))$ 为 sigmoid 函数, 则可得

$$\frac{\partial z}{\partial y} = \frac{e^{-y}}{(1 + e^{-y})^2} > 0$$

由此可得 $|J| = J$. 值得注意的是, 在一般前向网络结构中^[4-6], 雅可比多项式中包含了滤波器权值的函数 D , 因此需要判定雅可比多项式的符号以决定其绝对值. 而在本文提出的二阶前向结构中, 不存在这样的两难问题.

用雅可比多项式的瞬时值代替其统计期望值, 最大化 $\ln J$, 则在二阶结构中权值的迭代式为

$$\left. \begin{aligned} \Delta w_1(k) &= [(1 - 2z_1(t))w_2(0) - \\ & (1 - 2z_2(t))]x_1(t-k) \\ \Delta w_2(k) &= [-(1 - 2z_1(t))]y_2(t-k) \end{aligned} \right\} \quad (12)$$

式中 $k = 0, 1, \dots, M$. 与基于前向网络的权值迭代公式相比, 新算法的学习规则中不包含权值矩阵的逆 W^{-1} , 因此新算法计算简单而且效果较好.

4 仿真实验结果

在实验模拟中, 采用真实的语音信号经过计算机混合, 由模拟产生的相关噪声信号构成系统的两个主辅通道输入信号, 其两个通道的混合输入如下式:

$$\left. \begin{aligned} x_1(t) &= s_1(t) - 0.6s_2(t-2) \\ x_2(t) &= s_2(t) + 0.8s_1(t-3) \end{aligned} \right\} \quad (13)$$

式中: $s_1(t)$ 为一段真实的干净语音信号; $s_2(t)$ 为模拟噪声信号. 图 3 表示了原始语音信号和噪声信号以及两个通道的输入混合信号. 利用本文中提出的二阶前向 ANC 结构和相应的梯度下降算法, 系统的输出结果如图 4. 在实验中, 两个权值滤波器均选用阶数为 36 的 FIR 滤波器, 步长选取为 0.000 3. 经仿真得到当主通道输入信噪比 $SNR_{in} = -3.301$ dB 时, 采用本节提出的语音增强算法, 当权值收敛时, 输出信噪比 $SNR_{out} = 12.565$ dB. 所以本文提出的二阶前向结构语音增强系统表现出较好的算法性能. 表 1 给出了当主通道输入源信号保持不变的情况下, 不同输入信噪比时的输出信噪比较. 图 5 给出了权值滤波器 $W_2(z)$ 中第二个权值 $w_2(2)$ 的收敛曲线. 由图可见, 本文提出的语音增强算法中权值的收敛速度是相当快的, 新算法可以较好地完成任务. 在模拟中发现步长的选取对于权值收敛的速度和收敛后权值的波动有较大的关系. 故实际应

用中,一般采用指数下降因子迭代并有最小下限步长值.对于不同混合输入进行的大量计算机模拟实验结果表明,本文提出的二阶前向结构及语音增强新算法均能很好地分离出较干净的原始独立语音输入信号,完成语音增强任务.

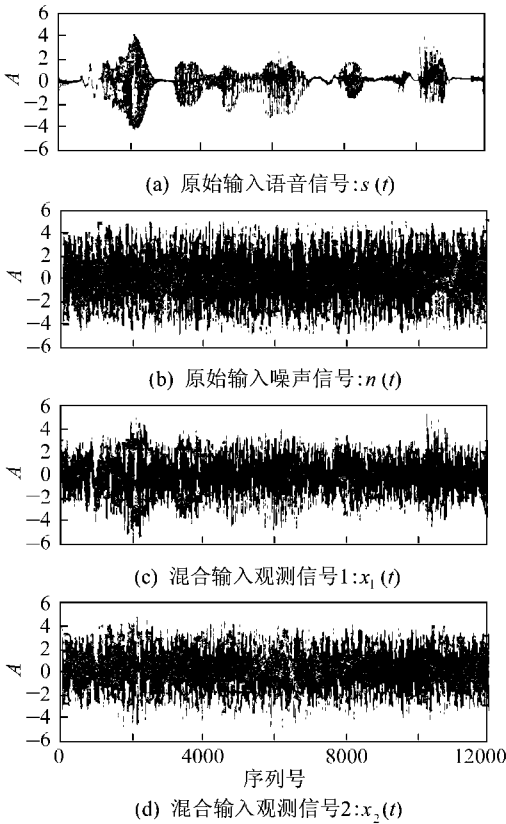


图3 原始输入源信号和混合后观测输入信号的幅度波形
Fig.3 The waveforms of the original sources and the input mixed signals

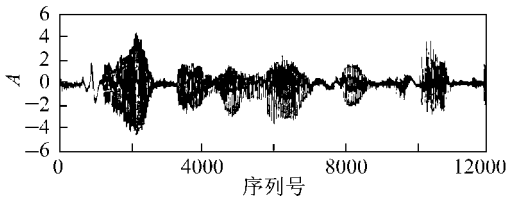


图4 采用语音增强算法收敛后的语音输出波形
Fig.4 The waveform of the output speech by using the proposed algorithm

表1 语音增强新算法的输入与输出信噪比

Tab.1 The input SNRs and output SNRs by using the proposed algorithm

SNR _{in} /dB	-16.342	-9.321	-3.301	2.720	7.157
SNR _{out} /dB	6.201	9.162	12.565	15.306	18.347

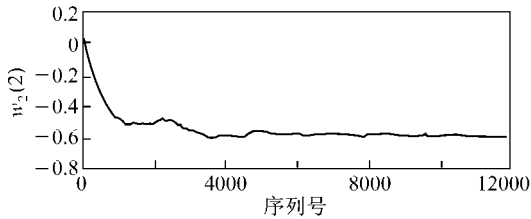


图5 权值 $w_2(2)$ 的收敛曲线

Fig.5 The convergence contour of the weight $w_2(2)$

5 结 语

基于二阶前向结构和 ME 基本原理,本文提出了一种新的语音增强算法,该算法十分简洁有效.利用真实的语音信号输入,经大量的计算机模拟实验结果表明,本文提出的语音增强新算法具有较好的算法性能.新结构的主要优点在于能分离出干净的原始语音输入信号而不受系统滤波效应的影响.在本文中假设输入混合是线性的,故新算法采用 FIR 滤波结构,进一步的研究将着重于采用非线性滤波器来改进实际环境情况下的语音增强效果.

参考文献:

- [1] Widrow B. Adaptive signal processing [M]. New York: Prentice-Hall, Englewood Cliffs, 1985.
- [2] Amari S, Cichocki A, Yang H H. A new learning algorithm for blind signal separation [J/OL]. In: Touretzky D, Mozer M, Hasselmo M, eds. Advances in Neural Information Processing System, Cambridge, MA: MIT Press, 1996. 757 ~ 763. <http://www.bip.riken.go.jp/irl/hhy/acyNIPS95.ps.Z>.
- [3] Bell A J, Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution [J]. Neural Computation, 1995, 7 (6): 1129 ~ 1159.
- [4] Comon P. Independent component analysis, a new concept? [J]. Signal Processing, 1994, 36 (3): 287 ~ 314.
- [5] Jutten C, Herault J. Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture [J]. Signal Processing, 1991, 24 (1): 1 ~ 10.
- [6] Torkkola K. Blind separation of delayed sources based on information maximization [A]. In Proc ICASSP [C]. Atlanta, GA, May 7-10, 1996. <ftp://ftp.cnl.salk.edu/pub/tony/OtherPapers/karil.ps>.
- [7] Lee T-W, Orlmeister R. Blind source separation of real-world signals [A]. Proc of ICNN Houston [C]. USA, 1997. <ftp://ftp.cnl.salk.edu/pub/tony/icnn97.ps.Z>.
- [8] Cardoso J F, Laheld B. Equivariant adaptive source separation [J]. IEEE Trans on Signal Processing, 1996, 45 (2): 434 ~ 444.