

文章编号: 1001-0920(2003)05-0577-04

## 基于简化的 KLT 和小波变换的非平稳宽带噪声语音增强

楼红伟, 胡光锐

(上海交通大学 电子工程系, 上海 200030)

**摘要:** 针对非平稳宽带噪声语音增强存在的困难, 提出一种基于二进小波变换和简化的 Karhunen-Loeve 变换 (KLT) 的语音增强算法。通过基于小波空间和 KLT 矢量空间分解带噪语音信号, 并针对浊音和清音分别进行处理和重构, 在无需进行噪声的白化及信噪比大小的预先计算条件下, 实现了语音增强。在不同的语音增强系统对比实验中, 通过失真测度分析表明了该算法能克服以往方法的固有缺陷, 在非平稳宽带噪声的语音增强中, 对噪声的整形和抑制具有较好的作用。

**关键词:** 语音增强; Karhunen-Loeve 变换; 二进小波变换

**中图分类号:** TP391.42 **文献标识码:** A

## Simplified KLT and wavelet transform based approach for enhancing speech degraded by non-stationary wideband noise

LOU Hongwei, HU Guangrui

(Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract:** Aiming to suppress the non-stationary wideband noise, a novel speech enhancement algorithm based on the dyadic wavelet transform and the simplified Karhunen-Loeve transform is proposed. The noisy speech is decomposed into components by the wavelet space and KLT-based vector space. And the components are processed and reconstructed respectively by distinguishing between voiced speech and unvoiced speech. There are no requirements of noise whitening and SNR precalculating. The experiments and comparison with different speech enhancement systems by means of the distortion measure show that the proposed method overcomes drawbacks existing in the previous methods and performs better shaping and suppressing of the non-stationary wideband noise for speech enhancement.

**Key words:** Speech enhancement; Karhunen-Loeve transform; Dyadic wavelet transform

### 1 引言

语音在通信过程中不可避免地会受到干扰噪声的影响, 使接收者接收到的语音已非纯净的原始语音信号。语音增强是从带噪语音信号中提取尽可能纯净的原始语音。由于噪声来源众多, 随应用场合而异, 它们的特性也各不相同。即使在实验室仿真条件下, 也难以找到一种能适用于各种噪声环境的通用

的语音增强算法, 所以必须针对不同的噪声采取不同的语音增强对策。噪声种类很多, 但研究最多也最为常见的是加性噪声, 诸如周期性噪声、脉冲噪声、宽带噪声以及同声道其他语音的干扰, 其中宽带噪声又分为平稳和非平稳的。由于宽带噪声与语音信号在时域和频域上完全重叠, 因而消除它最为困难。对于平稳的宽带噪声, 可认为它是高斯白噪声, 或许

收稿日期: 2002-05-23; 修回日期: 2002-07-31。

基金项目: 国家自然科学基金资助项目 (69672007)。

作者简介: 楼红伟 (1976—), 男, 浙江永康人, 博士生, 从事语音信号处理与识别、神经智能等研究; 胡光锐 (1938—), 男, 上海人, 博士生导师, 从事语音识别、神经网络和通信系统抗干扰等研究。

可以进行白化处理;而对于非平稳宽带噪声的情况则更为复杂。

小波理论是新近发展起来的时频分析技术,特别适用于非平稳信号的分析<sup>[1]</sup>。运用小波分析进行语音信号消噪处理是小波分析的重要应用之一。近年来,已有许多运用小波变换的降噪方法。尽管基于小波的噪声消除算法对于语音增强具有一定的作用,但它们仍不能满足鲁棒性和准确性的要求,在具有较高的背景噪声中,其性能还有待提高。另外,在小波变换的各成分中不可避免地会产生一些干扰成分,也会影响增强效果。

本文针对复杂的非平稳宽带噪声,提出一种基于简化的 KLT 和二进小波变换(DWT)的语音增强算法。该算法不要求噪声逆矩阵变换,只是通过对带噪的浊音和清音信号的变换系数进行不同处理来达到抑制非平稳宽带噪声的作用。

## 2 基于 DWT 和 KLT 的语音增强

### 2.1 二进小波变换(DWT)的噪声处理

一个能量有限信号( $x \in L^2$ )的二进小波变换定义为

$$D_j W T_x(n, 2^j) = \frac{1}{2^{j/2}} \int_{-\infty}^{\infty} x(t) \Psi^*(2^{-j}t - n) dt = x(t) \otimes \Psi_2^j(t) \quad (1)$$

式中:  $2^j$  为尺度因子,  $\Psi^*(t)$  为小波函数  $\Psi(t)$  的复共轭。通过取不同的  $j$ , 可以进行母小波  $\Psi(\cdot)$  的伸缩。DWT 是非线性的, 具有时移不变性, 这对语音信号很有用。因为在分析语音信号时, 通常将它模型化为不同时移和具有衰减性的正弦波的线性集合。正是这种性质使其在非平稳信号的分析中占据着重要的位置。原信号可由

$$x(t) = \sum_j \int_z \frac{1}{2^{j/2}} D_j W T_x(n, 2^j) \Psi(2^{-j}t - n) dn \quad (2)$$

进行重构。

一个含噪声的一维信号模型可表示为

$$x(i) = s(i) + n(i), \quad i = 0, \dots, n-1 \quad (3)$$

其中:  $s(i)$  为纯语音信号,  $n(i)$  为噪声。

在本文研究范围内, 噪声通常表现为非平稳的宽带噪声。消噪过程可按如下方法进行:

- 1) 首先选定一个小波系, 并确定一个分解层次  $N$ , 对信号进行小波分解;
- 2) 对第  $1 \sim N$  层的每一层高频系数, 选择一个阈值进行软阈值量化处理;

3) 根据小波分解的第  $N$  层低频系数和经过量化处理后的第  $1 \sim N$  层的高频系数, 进行语音信号的小波重构。

软门限函数可表示为<sup>[2]</sup>

$$\text{Th}(\omega_{i,j}, \gamma_\alpha) = \begin{cases} 0, & |\omega_{i,j}| \leq \gamma_\alpha \\ \text{sgn}(\omega_{i,j}) (|\omega_{i,j}| - \gamma_\alpha), & |\omega_{i,j}| > \gamma_\alpha \end{cases} \quad (4)$$

式中:  $\omega_{i,j} = D_j W T_x(n, 2^j)$  为小波分解系数,  $\gamma_\alpha$  为动态阈值。其中的关键是如何选取阈值和如何进行阈值的量化, 本文采用文献[3]介绍的方法进行  $\gamma_\alpha$  取值。取 Biorthogonal 小波系, 其主要特征体现在具有线性相位性, 在信号的重构中得到了较为广泛的应用。在实际的语音分析中, 语音信号可能包含许多尖峰或突变部分, 信号不是平稳的。传统的傅立叶分析显得无能为力, 因为它是建立在将信号完全转换为频域中进行分析的。由于小波分析能同时在时频中进行, 具有自动变焦区分信号功能, 从而实现信号的消噪。

这里必须注意, 在应用软阈值进行语音增强时, 对非浊音有用信号会产生过分损害。因为非浊音包含许多类似噪声的高频成分, 用小波分解消除这种成分势必会严重损害重建语音的懂度(intelligibility), 因此这里必须针对浊音和清音采取不同的变换系数处理方法。如果判决为浊音信号, 则选取合适的小波系进行  $N$  层分解, 对第  $1 \sim N$  层的每一层高频系数, 选择一个阈值进行软阈值量化处理; 如果判决为非浊音信号, 则同样选取合适的小波系进行  $N$  层分解, 然后对第  $N$  层的高频系数, 选择一个阈值进行软阈值量化处理。

### 2.2 简化的 Karhunen-Loeve 变换噪声处理

上述小波降噪处理措施, 往往不能达到最佳的语音增强效果。有必要在小波非平稳降噪的基础上, 进一步进行噪声处理。Mittal 等<sup>[4]</sup> 所建议的用于有色平稳噪声的语音增强算法只能处理一特定频带下的平稳噪声信号, 而且必须通过长时的平均计算非语音段的噪声功率来求取瞬时信号的信噪比, 在此基础上最终对带噪语音信号帧分类出语音支配帧和噪声支配帧, 分别进行语音增强处理, 导致计算量较大。本文提出一种简化的 KLT 模型, 将小波处理结果直接转到 KLT 算法中。

令  $K \times K$  的矩阵  $R_x$  为一矢量信号  $x$  的协方差矩阵。假设  $\hat{x}$  为  $x$  的估计,  $\hat{R}_x$  为  $R_x$  的估计值。对于一个任意给定的矩阵  $A$ ,  $A^T$  为  $A$  的转置矩阵,  $\text{tr}(A)$  为  $A$  的迹(即  $A$  的对角元素之和)。令

$$R_x = U_x \Lambda_x U_x \quad (5)$$

为  $R_x$  的特征值分解,  $\lambda(k)$  为对角阵  $\Lambda_x$  的  $k$  阶对角元素,  $u_{xk}$  为相应的特征矢量。这里还假设  $s, n$  和  $x = s + n$  分别为  $K$  维纯语音、加性噪声和带噪语音噪声矢量。假设语音信号和噪声是相互独立的, 则有  $R_x = R_s + R_n$ 。仿照式(5), 令

$$R_s = U_s \Lambda_s U_s \quad (6)$$

为  $R_s$  的特征值分解。这种基于子空间的分解就是基于 KL T 的方法。假设  $U_s = [U_1 \ U_2]$ , 这里的  $U_1$  为  $R_s$  的  $M$  个正特征值产生的特征向量构成的  $K \times M$  矩阵, 且  $M < K$ 。必须注意  $U_1$  的张成即为语音信号子空间, 而  $U_2$  的张成则正交于  $U_1$ 。这种基于子空间的估计方法源于  $R_s$  能被原始采样数据准确地估计, 所以也叫采样子空间。注意到  $U_1 U_1^T$  和  $U_2 U_2^T$  分别为信号子空间及其正交子空间的正交投影, 则噪声子空间的投影  $U_2 U_2^T x = 0^{[5]}$ , 所以这部分不能提供有关  $s$  的任何信息。为了减轻语音增强对有用信号的失真, 并同时简化非白噪声白化过程, 采用以下一些处理<sup>[4]</sup>。

令语音增强产生的语音失真矢量分别为  $r_s$  和  $r_n$  (残余噪声), 这里  $r_s = (H - I)s$ ,  $I$  为单位对角矩阵。为了使语音失真达到最小, 令  $\epsilon$  为  $r_s$  的方差矩阵的迹, 则有如下一个优化问题

$$\min_H \epsilon \quad (7)$$

这里的  $H$  为基于 KL T 的滤波器。式(7)可转化为

$$E(|u_{sk} r_n|^2) = \alpha_k \sigma_n^2(k), \quad k = 1, 2, \dots, K \quad (8)$$

式中:  $\alpha_k$  为第  $k$  阶约束常数,  $\sigma_n^2(k)$  为  $U_s R_n U_s^T$  的  $k$  阶对角元素。进一步假设

$$H = U Q U_s^T \quad (9)$$

其中  $Q$  为一对角矩阵。注意到如果  $\alpha_k = \lambda(k) / \sigma_n^2(k)$ , 则在式(8)中等号存在的条件下, 能使残余噪声  $r_n$  的谱与纯语音的谱相似。令  $q_{kk}$  为  $Q$  的第  $k$  阶对角元素, 则式(8)中等号成立的条件为

$$q_{kk} = \alpha_k^{1/2} \quad (10)$$

且有

$$\begin{aligned} \epsilon &= \text{tr}(E\{r_s r_s^T\}) = \\ &= \text{tr}(U_s (Q - I) \Lambda_s (Q - I) U_s^T) = \\ &= \sum_{k=1}^K \lambda(k) (1 - q_{kk})^2 \end{aligned} \quad (11)$$

由式(11)可知

$$q_{kk} = \min(1, \alpha_k^{1/2}) \quad (12)$$

是式(10)成立的必要条件。

根据以上分析, 可得到基于 KL T 算法求取的滤

波矩阵  $H$ , 使得无需预白化便可增强语音信号

$$\hat{s} = H x \quad (13)$$

算法如下:

1) 取一帧带噪语音信号  $x$ , 并取得无语音下的噪声  $n$ , 分别求取它们的方差矩阵  $R_x$  和  $R_n$  (具体细节详见文献[4]);

2) 估计语音信号的方差  $\hat{R}_s = \hat{R}_x - R_n$ , 进而求取其特征值和特征矢量, 组成  $U_s = [U_1 \ U_2]$ , 其中  $U_1$  为  $R_s$  的  $M$  个正特征值所产生的特征向量构成的  $K \times M$  矩阵, 即  $U_1 = \{u_{sk} \mid \lambda(k) > 0\}$ ;

3) 令

$$\Lambda = \text{diag}(\alpha_k^{1/2}) \quad (14)$$

$$\alpha_k = \exp\left[-\frac{v \sigma_{n_T}^2(k)}{\lambda(k)}\right], \quad k = 1, 2, \dots, M \quad (15)$$

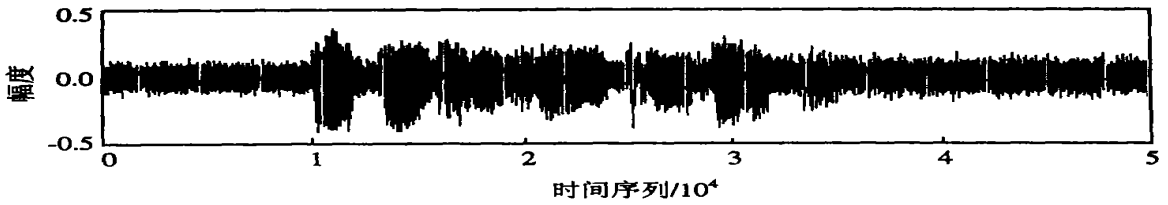
其中:  $v$  为先验常数,  $\sigma_{n_T}^2(k)$  为  $R_{n_T}$  的第  $k$  阶对角元素,  $n_T = U_s n$ ,  $R_{n_T} = U_s R_n U_s^T$ 。令  $Q = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}$ , 则可证明  $H = U_1 \Lambda U_1^T$ 。由此可知,  $U_2$  的求解可以省略, 从而简化了 KL T 的计算过程。

### 3 性能评估

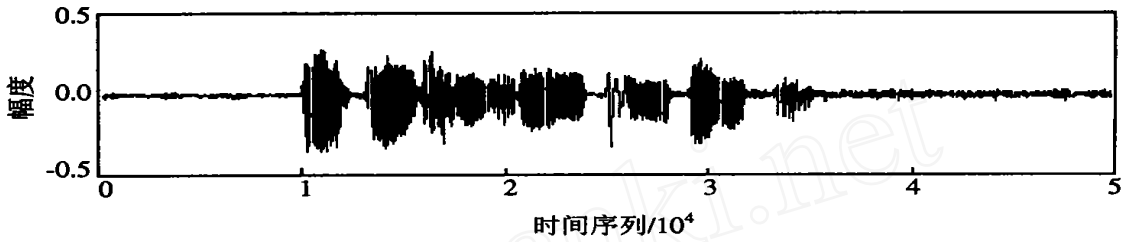
本文算法的目的是为了增强在非平稳宽带噪声下受损的语音信号。首先主观地考察一下该语音增强系统的性能, 为此做如下实验: 在一段女声发音的录音(把电视声调大一点)的基础上, 加上一阶自回归噪声和不同时段有不同信噪比的高斯噪声, 平均信噪比为 4 dB, 实验结果见图 1。从图中可以看出, 尽管噪声环境较为复杂, 其中包含有各种信噪比和不同统计特性的背景噪声, 但经本文方法处理后, 在这种恶劣的噪声环境下仍然具有显著的语音增强效果。回放增强后的语音信号, 感觉不到典型的音乐冗余噪声的存在。

为了进一步考察该语音增强系统的性能, 还需进行客观评定。可懂度是语音增强效果的一个客观测度, 其中一个重要的衡量方式是失真测度。下面介绍一种谱失真测度方法<sup>[4]</sup>。假设有一纯语音信号  $s$ , 受噪声污染后的信号为  $x$ , 所加噪声的平均信噪比设为 snr, 经过语音增强得到的信号为  $\hat{s}$ , 并假设该  $s$  长为  $N$ 。为了得到  $s$  和  $\hat{s}$  之间的失真测度, 事先采取了两个步骤: 首先分别对两个  $N$  维矢量进行归一化处理, 使之具有单位能量(0 dB), 然后加上 -30 dB 的白噪声  $n_0$ , 以防止出现奇异值  $\log(0)$ 。该谱失真度为

$$DM(\text{snr}, s, \hat{s}) =$$



(a) 带噪语音



(b) 增强后语音

图 1 非平稳宽带噪声语音增强

$$P \sum_{p=1}^{255} \sum_{k=0} \left| \log |\hat{S}_p(k)| - \log |\tilde{S}_p(k)| \right| \quad (16)$$

$$\begin{cases} \hat{s} = s / \hat{s} + n_0 \\ \tilde{s} = s / \tilde{s} + n_0 \end{cases} \quad (17)$$

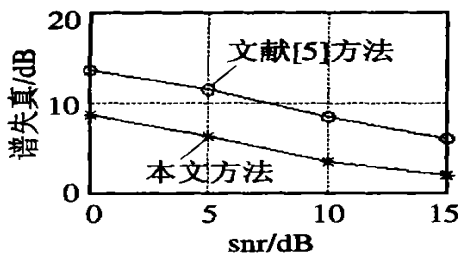
这里的  $\hat{S}_p(k)$  和  $\tilde{S}_p(k)$  分别为  $\hat{s}$  和  $\tilde{s}$  在  $p$  帧下离散傅立叶变换的第  $k$  个频率分量。式(16)表征了在一定信噪比的噪声下所增强的语音与纯语音信号的失真测度。为了评估该增强系统的性能,对本文提出的

语音增强系统 Ephraim 等<sup>[5]</sup>提出的基于信号子空间的语音增强系统进行比较。所测试的语音由 10 个句子组成,分别由 3 个男声和 3 个女声录音所得,再添加不同信噪比(本文取 0, 5, 10, 15)的噪声信号后进行语音增强对比实验,对比结果见图 2。该实验采用了两种噪声信号,第 1 种含有一阶自回归宽带噪声,第 2 种侧重于中高频段的宽带噪声。图中的每一点都是对 6 个人说的 10 个句子共 60 个失真测度 DM (snr,  $\hat{s}$ ,  $\tilde{s}$ ) 求取平均得到的。从图 2 可以看出,在两种噪声的情况下,所采用的语音增强算法都得到了显著提高。从图中还可以看出,该语音增强系统对中高频段也具有较好的增强效果。由于语音原有信号中清音的噪声频谱与外加的干扰噪声相重叠,一般的语音增强系统很容易对原有的有用信号过增强,导致语音信号的严重失真,可懂性较低。由于本文方法在浊音与清音之间事先进行了判别,并分别予以不同的小波系数进行加权处理,克服了这种以往语音增强系统固有的缺陷,实现了对非平稳宽带噪声的语音增强。

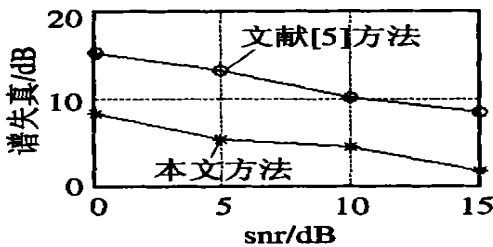
### 4 结 论

本文提出一种结合小波变换和简化 KLT 变换的语音增强算法。通过比较实验,验证了该算法的可靠性。该算法能克服以往一些方法的固有缺陷,在非平稳宽带噪声的语音增强中,对噪声的整形和抑制具有较好的作用。

(下转第 589 页)



(a) 加一阶自回归宽带噪声



(b) 加中高频段宽带噪声

图 2 对比结果

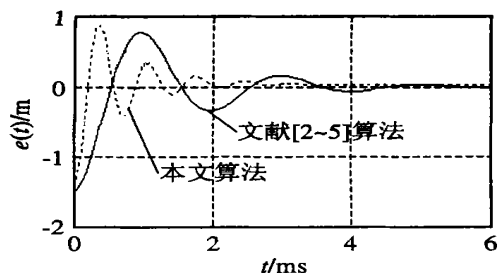


图 3 两种方法的响应曲线对比

$Z_m$  所进行的模型简化操作部分, 忽略了与“主导子系统”关联不深的信息, 而这些信息对于改善控制效果可能是至关重要的。

## 5 结 论

针对文献[1~5]介绍的 T-S 型模糊控制器设计方法难以求解的问题, 本文给出了一种模型简化算法, 并将该算法应用于一类非线性系统镇定问题的求解。研究表明, 所提出的模型简化算法减少了求解过程的约束条件, 增加了问题的可解性, 使得更容易得到问题的可行解。仿真结果验证了这种方法的有效性和正确性。

然而, 如何处理简化模型与原模型之间的误差, 以及简化模型与原模型之间的关系仍是需要进一步研究的问题。

## 参考文献(References):

[1] Tanaka K, Sano M. A robust stabilization problem of

fuzzy controller systems and its applications to backing up control of a trucktrailer[J]. *IEEE Trans on Fuzzy Systems*, 1994, 2(1): 119-134

[2] Tanaka K, Ikeda T, Wang H O. Fuzzy regulators and fuzzy observers: Relaxed stability conditions and LM I based design[J]. *IEEE Trans on Fuzzy Systems*, 1998, 6(2): 250-265

[3] Tanaka K, Ikeda T, Wang H O. Robust stabilization of a class uncertain nonlinear system via fuzzy control[J]. *IEEE Trans on Fuzzy Systems*, 1996, 4(1): 1-13

[4] Wang H O, Tanaka K, Griffin M. An approach to fuzzy control of nonlinear systems: Stability and design issues [J]. *IEEE Trans on Fuzzy Systems*, 1996, 4(1): 14-23

[5] Lee H J, Park J B, Chen Guanrong. Robust fuzzy control of nonlinear systems with parametric uncertainties [J]. *IEEE Trans on Fuzzy Systems*, 2001, 9(2): 369-379

[6] 诸静. 模糊控制原理与应用[M]. 北京: 机械工业出版社, 1995: 18-90

[7] Lam H K, Leung F H F, Tam P K S. Stable and robust fuzzy control for uncertain nonlinear system [J]. *IEEE Trans on Systems, Man, and Cybernetics*, 2000, 30(4): 825-840

[8] Ying H. An analytical study on structure, stability and design of general nonlinear Takagi-Sugeno fuzzy control system [J]. *Automatica*, 1998, 34(2): 157-167.

[9] Ying H. The Takagi-Sugeno fuzzy controllers using the simplified linear control rules are nonlinear variable gain controllers[J]. *Automatica*, 1998, 34(12): 1617-1623

(上接第 580 页)

## 参考文献(References):

[1] 杨福生. 小波变换的工程分析与应用 [M]. 北京: 科学出版社, 2000

[2] Qin Linmei, Hu Guangrui. A new speech enhancement method [A]. *Proc of 2001 Int Symposium on Intelligence Multimedia, Video and Speech Processing* [C]. Hong Kong, 2001. 92-94

[3] David L, Donoho. Denoising by soft-thresholding [J]. *IEEE Trans on Information Theory*, 1995, 41(3): 613-

627.

[4] Udar Mittal, Nam Phamdo. Signal/noise KLT based approach for enhancing speech degraded by colored noise [J]. *IEEE Trans on Speech and Audio Processing*, 2000, 8(3): 159-167.

[5] Yariv Ephraim, Harry L, Van Trees. A signal subspace Approach for speech enhancement [J]. *IEEE Trans on Speech and Audio Processing*, 1995, 3(4): 251-266