

改进谱减法语音增强算法的研究

陈欢, 邱晓晖

(南京邮电大学 通信与信息工程学院, 江苏 南京 210003)

摘要: 语音信号中, 噪声的存在极大地影响了语音质量, 必须采用语音增强技术在一定程度上抑制背景噪声, 最大限度地改善语音通信质量, 提高语音的可懂度。语音增强就是从带噪语音的信号中提取尽可能纯净的原始语音信号。谱减法由于其计算简单, 实时性好, 所以得到了广泛应用。文中介绍了基于谱减法的语音增强算法的基本原理, 并针对传统谱减法残余“音乐噪声”过强的缺点, 分析了两种典型的改进谱减法的原理。实验结果表明改进算法更好地抑制了音乐噪声, 有效提高了语音增强效果, 改善了语音质量。

关键词: 谱减法; 语音增强; 音乐噪声

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2014)04-0069-03

doi: 10.3969/j.issn.1673-629X.2014.04.017

Research on Speech Enhancement of Improved Spectral Subtraction Algorithm

CHEN Huan, QIU Xiao-hui

(College of Telecommunication & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: For the speech signal, voice quality is greatly affected by the noise. Speech enhancement techniques must be used to suppress background noise, improving voice quality and voice clear degrees. The speech enhancement is to get the speech as original as possible from the speech signal with noise. The spectral subtraction speech enhancement is utilized broadly because it is simple and easy for the real-time processing. In this paper, the basic principle of algorithm of the spectral subtraction is presented. Aiming at reducing annoying musical noise in the standard spectral subtraction algorithm, the principle of two kinds of typical improved spectral subtraction is analyzed. Experimental results show that the proposed algorithm reduces the musical noise effectively and improves the speech quality.

Key words: spectral subtraction algorithm; speech enhancement; musical noise

0 引言

噪声的影响不但会使语音通信质量降低, 而且也会使基于特征参数提取的语音处理系统(如语音识别、低速率语音编码等)性能下降。语音增强^[1]就是要提高语音质量, 尽可能地消除带噪语音中的噪声分量, 提升语音的可懂度。语音增强主要存在基于发声模型和基于语音幅度谱估计两种算法, 近年也出现了一些诸如子空间分解、小波变换等新方法。其中, 基于语音幅度谱估计的增强算法一直是研究热点, 主要有谱相减法、最小均方误差(MMSE)估计法以及维纳滤波法等。谱减法^[2-3]是目前最常用的语音增强算法,

特点是容易实时实现, 运算量较小。但也存在着一些问题, 如“音乐噪声”(Musical Noise)等, 严重影响增强效果。近来针对谱减法有很多改进算法, 比如参数可调的谱减法、基于听觉掩蔽效应的谱减法等。文中主要对两种不同的改进算法^[4-5]进行对比研究。

1 谱减法的基本原理

基本谱减法的主要思想^[4]是以加性噪声和短时平稳的语音信号相互独立的假设为前提, 通过带噪语音的功率谱与噪声功率谱相减, 由此获得较为纯净的语音频谱。

收稿日期: 2013-06-03

修回日期: 2013-09-15

网络出版时间: 2014-01-28

基金项目: 江苏省自然科学基金项目(BK2011789); 南京邮电大学自然科学基金(NY211087)

作者简介: 陈欢(1990-), 男, 安徽芜湖人, 硕士研究生, 研究方向为语音信号处理; 邱晓晖, 教授, 研究生导师, 研究方向为现代通信中的智能信号处理、图像处理与模式识别等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140128.1144.027.html>

设 $s(t)$ 为纯净语音信号, $n(t)$ 为噪声信号, $y(t)$ 为含噪语音信号, 这三者关系可由下式表示为:

$$y(t) = s(t) + n(t) \quad (1)$$

用 $S(w)$, $N(w)$, $Y(w)$ 分别表示 $s(t)$, $n(t)$, $y(t)$ 的傅里叶变换, 可以得到:

$$Y(w) = S(w) + N(w) \quad (2)$$

由此可得:

$$|Y(w)|^2 = |S(w)|^2 + |N(w)|^2 + 2\text{Re}[S(w)N^*(w)] \quad (3)$$

由上式可得:

$$E(|Y(w)|^2) = E(|S(w)|^2) + E(|N(w)|^2) + 2E\{\text{Re}[S(w)N^*(w)]\} \quad (4)$$

因为 $s(t)$ 与 $n(t)$ 独立, 故 $S(w)$ 和 $N(w)$ 也独立, 又 $N(w)$ 服从零均值的高斯分布, 所以 $E\{\text{Re}[S(w)N^*(w)]\} = 0$, 因而有

$$E(|Y(w)|^2) = E(|S(w)|^2) + E(|N(w)|^2) \quad (5)$$

对一帧内的短时平稳过程, 有:

$$|Y(w)|^2 = |S(w)|^2 + |N(w)|^2 \quad (6)$$

由于在发声前和发声期间可以认为平稳噪声的功率谱几乎不产生变化, 从而噪声的功率谱 $|N(w)|^2$ 可以通过发声前的所谓“寂静段”来估计, 故而有

$$|S(w)|^2 = |Y(w)|^2 - |N(w)|^2 \quad (7)$$

则原始语音的估计值为

$$|S(w)| = [|Y(w)|^2 - |N(w)|^2]^{1/2} \quad (8)$$

上式是分帧计算的, 假如式中出现负值的情况, 可将其替换为 0 值。为了避免分帧时产生的截断效应, 应对 y 做加窗处理^[6]。

利用人耳对相位变化不敏感的特点, 可以将估计之后的语音信号相位用原带噪语音信号的相位直接取代, 进而得到增强后的语音信号^[7]。基本谱减法的原理框图如图 1 所示。

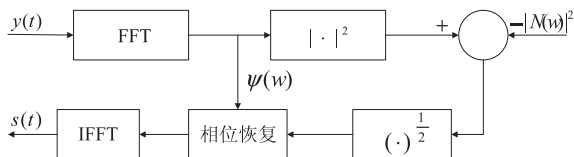


图 1 基本谱减法的原理框图

2 改进的谱减算法

在很多实际的应用场合, 基本谱减算法被改进使用, 下面分析两种典型的改进算法。

2.1 近似估计非零项法

式(4)中相关项 $E\{\text{Re}[S(w)N^*(w)]\}$ 等于 0 是假设加性噪声为零均值的高斯分布才成立。考虑到实际应用中噪声一般不会是零均值的高斯分布, 噪声或许是非高斯分布的, 或许是非零均值的。因而, 就不可

以忽略此非零项, 特别是在低信噪比条件下更不可忽略。近似估计此非零项的方法如下^[4]:

首先:

$$\text{Re}[S(w)N^*(w)] = |S(w)| |N(w)| \cos(\theta_s - \theta_N) \quad (9)$$

其中 $S(w) = |S(w)| \exp(j\theta_s)$, $N(w) = |N(w)| \exp(j\theta_N)$, 故有:

$$E\{\text{Re}[S(w)N^*(w)]\} = E\{|S(w)| |N(w)| \cos(\theta_s - \theta_N)\} \quad (10)$$

假设 $|S(w)|$, $|N(w)|$ 和 $\cos(\theta_s - \theta_N)$ 是相互独立的, 故:

$$E\{\text{Re}[S(w)N^*(w)]\} = E(|S(w)|) E(|N(w)|) E[\cos(\theta_s - \theta_N)] \quad (11)$$

由于不等式 $E(X^2) \geq [E(X)]^2$ 成立, 故有: $E(|S(w)|) = \alpha [E(|S(w)|^2)]^{1/2}$ 和 $E(|N(w)|) = \beta [E(|N(w)|^2)]^{1/2}$ 。

其中 $0 \leq \alpha, \beta \leq 1$, 故式(11)变为:

$$E\{\text{Re}[S(w)N^*(w)]\} = \alpha\beta E[\cos(\theta_s - \theta_N)] [E(|S(w)|^2) E(|N(w)|^2)]^{1/2} = \gamma [E(|S(w)|^2) E(|N(w)|^2)]^{1/2} \quad (12)$$

其中设 $\gamma = \alpha\beta E[\cos(\theta_s - \theta_N)]$,

故 $0 \leq \gamma \leq 1$ 。所以式(4)变为:

$$E(|Y(w)|^2) = E(|S(w)|^2) + E(|N(w)|^2) + 2\gamma [E(|S(w)|^2) E(|N(w)|^2)]^{1/2} \quad (13)$$

对一个分析帧内的短时平稳过程, 有:

$$|Y(w)|^2 = |S(w)|^2 + |N(w)|^2 + 2\lambda [|S(w)|^2 |N(w)|^2]^{1/2} \quad (14)$$

由上式得到:

$$|S(w)|^2 = |Y(w)|^2 - (1 - 2\lambda^2) |N(w)|^2 - 2\lambda [|Y(w)|^2 |N(w)|^2 - (1 - \lambda^2) |N(w)|^4]^{1/2} \quad (15)$$

先通过发声前的“寂静段”来估计噪声 $n(t)$ 的功率谱 $|N(w)|^2$, 再由原带噪语音信号 $y(t)$ 可求得功率谱 $|Y(w)|^2$, 参数 λ 可取 $[0, 1]$ 之间的某个值, 根据式(15)可以得出 $|S(w)|^2$, 进一步可求得原始语音信号的估计值, 最后就得出降噪后的语音时域信号^[8]。参数 λ 以实验结果为依据来确定, 为了得到最好的去噪效果也可以由噪声的大小来动态地确定参数 λ 的取值。

2.2 引入参数法

另一种谱减法的改进形式为:

$$|S(w)| = [|Y(w)|^\alpha - \beta |N(w)|^\alpha]^{1/\alpha} \quad (16)$$

这种改进方式相比于基本的谱减方法,差别是引入了 α 和 β 这两个参数^[9]。 α 为谱减功率修正系数,增大这个系数在提高信噪比的同时也会使语音信号的失真加大。取不同的 α 值进行测试,可以发现原始信号信噪比较低的时候 α 对去噪效果有一定的影响,而在原始信号信噪比较高的时候 α 对去噪效果影响不大,不过一般情况下 α 可取2使运算简便。 β 为谱减噪声系数。由于在谱减法计算中,当前分析帧各频率点的噪声频谱分量是用无声期间的统计平均的噪声方差来替代。又噪声频谱服从高斯分布,其幅度会有很宽的随机变化范围。所以相减时,当该帧某个频率点噪声分量比较大的时候,会残留很大一部分分量,在频谱上表现出随机出现的尖峰,而在听觉上则产生有节奏性起伏的类似音乐的残留噪声。为了解决这个问题,在语音能量较高的地方,可令 $\beta>1$,通过人为的多减去一些噪声分量,来更好地突出语音的频谱,当然语音失真的可能性也会由此变得更大^[5]。因此,必须要依据实验结果来确定参数值的选取,当然也可以由语音当时的大小来动态地确定其值,调节参数^[10]也能够获得类似的效果。

3 实验结果及分析

通过分别计算增强前后语音的信噪比来验证上文中近似估计非零项改进谱减算法的有效性。实验中语音信号和噪声信号均按照8 kHz采样,16 bit量化。所选取的语音材料为在安静的实验室条件下录制的一段纯净语音,噪声选择了NOISEX-92数据库中F16战斗机噪声。按比例将语音和噪声信号线性相加生成5种不同的信噪比:-15 dB,-10 dB,-5 dB,0 dB,5 dB。含噪语音采用汉明窗进行分帧,每帧512个采样点,帧间叠加128个采样点。在这5种不同信噪比的条件下,对含噪语音分别采用基本谱减法和近似估计非零项改进谱减算法进行语音增强的实验,参数 λ 分别取为0.4和0.7来计算,实验结果如表1所示。

表1 语音增强实验结果比较

输入信噪比/dB	5	0	-5	-10	-15
基本谱减法信噪比	8.77	4.46	3.15	0.14	-0.25
改进谱减法信噪比/ $\lambda=0.4$	9.73	5.64	4.66	2.06	2.27
改进谱减法信噪比/ $\lambda=0.7$	9.79	5.69	4.78	2.12	2.38

由表1可以得出:用基本谱减法增强以后,信噪比得到提高的同时背景噪声也被较大地削弱了;用近似估计非零项改进谱减算法的增强效果相比基本谱减法更优,噪声被更好地抑制,信噪比获得了进一步的提高,特别在低信噪比时增强效果更明显;还可以看出参数 λ 取0.7时的增强效果要好于 λ 取0.4时的增强效果,这说明参数 λ 的取值对改善增强效果而言是非常

重要的,要依据实验的结果来确定。

为了验证上文中引入参数改进谱减算法的有效性,实验采用同一段纯净语音,叠加上给定的高斯白噪声,实验结果如图2所示(其中 α 取2, β 取1.5,横轴表示采样点数目/个,纵轴表示归一化幅度)。

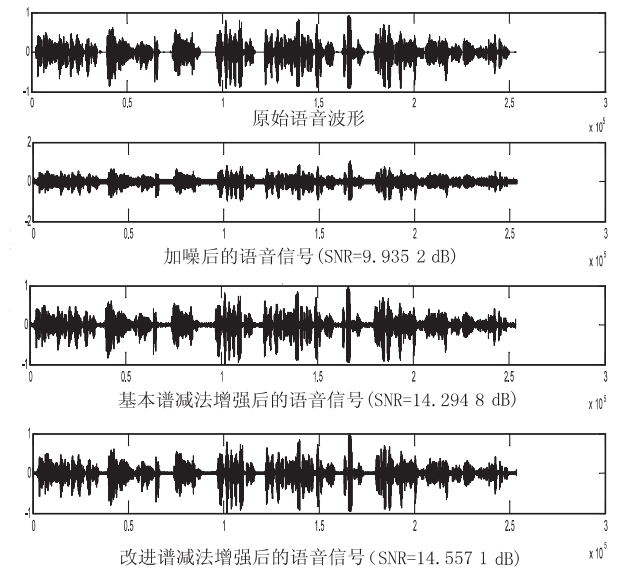


图2 增强实验结果图

4 结束语

基本谱减法对含噪语音进行增强后,虽然能明显地消除背景噪声,并提高语音信号的信噪比,但又产生了令人讨厌的有节奏的音乐噪声^[11]。针对基本谱减算法中关于噪声是零均值的高斯分布的假设,提出了近似估计非零项改进谱减算法,使增强效果得到了有效的提高,并且更好地抑制了音乐噪声,改善了语音质量。实验表明引入参数改进谱减算法中 α 对效果的影响不是很大,而且当 α 取2时,其物理意义比较明显,所以一般在研究中将 α 固定取值为2^[12]。

参考文献:

[1] 杨行峻,迟惠生. 语音信号数字处理[M]. 北京:电子工业出版社,1995.

[2] Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise[C]//Proc of IEEE international conference on acoustics, speech, and signal processing. [s. l.]:[s. n.],1979:208-211.

[3] Boll S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE transaction on acoustic, speech and signal processing,1979,27(2):113-120.

[4] 钱国青,赵鹤鸣. 基于改进谱减算法的语音增强新方法[J]. 计算机工程与应用,2005,41(35):42-43.

[5] 陶智,葛良. 基于减谱法的语音增强和噪声消除的研究[J]. 苏州大学学报(自然科学),2002,18(3):58-61.

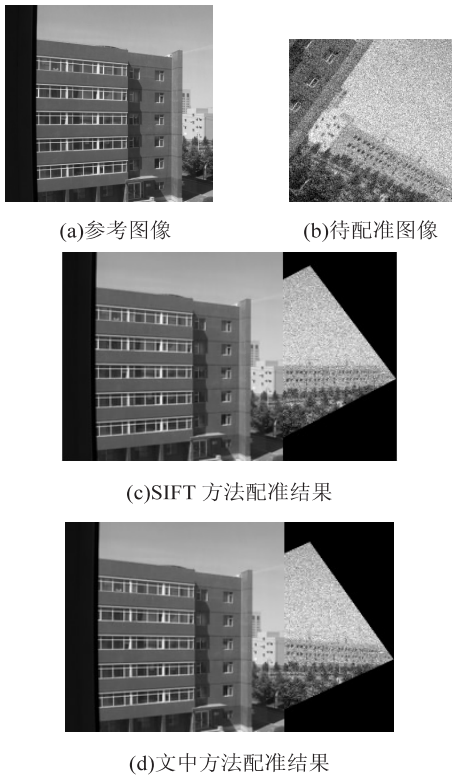


图 2 实物图像的对比实验结果

化伪 Zernike 矩相结合的图像配准方法。该方法利用尺度相互作用的墨西哥帽小波提取图像中的特征点,而且采用了在墨西哥帽小波中加入尺度因子的方法来解决不同尺度图像特征点定位不一致的问题。之后利用归一化伪 Zernike 矩计算特征点周围的圆形邻域的特征描述矢量,该矢量具有严格的旋转不变性以及尺度不变性。特征点匹配过程利用双向匹配策略,提高了匹配的可靠性,在估计几何变换参数时用迭代加权最小二乘法减小误匹配点的影响,并给出了图像配准的评价准则,精度分析可以看出对应点匹配的均方根误差在一个像素范围之内。

参考文献:

[1] 张红颖,张加万,孙济洲.改进 Demons 算法的非刚性医学

图像配准[J]. 光学精密工程,2007,15(1):145-150.

- [2] 覃凤清,何小海,陈为龙,等.一种图像配准的超分辨率重建[J]. 光学精密工程,2009,17(2):409-416.
- [3] 聂宏宾,侯晴宇,赵明,等.基于似然函数 EM 迭代的红外与可见光图像配准[J]. 光学精密工程,2011,19(3):657-663.
- [4] Zitova B, Flusser J. Image registration methods: A survey[J]. Image and vision computing, 2003, 21(11):977-1000.
- [5] David G L. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [6] Bay H, Tuvellars T, Van G L. SURF: Speeded up robust features[C]//Proceedings of the European conference on computer vision. [s. l.]: [s. n.], 2006:404-417.
- [7] Manjunath B S, Shekhar C, Chellappa R. A new approach to image feature detection with applications[J]. Pattern recognition, 1996, 29(4):627-640.
- [8] Dobbins A, Zucker S W, Cynader M S. Endstopping and curvature[J]. Vision research, 1989, 29(10):1371-1387.
- [9] Antoine J P, Vanderghyest P, Murenzi R. Two-dimensional directional wavelets in image processing[J]. International journal of imaging systems and technology, 1996, 7(3):152-165.
- [10] Bhattacharjee S, Kutter M. Compression tolerant image authentication[C]//Proc of IEEE international conference on image processing. Chicago, USA: [s. n.], 1998:435-439.
- [11] 朱世元,王凤英.一种基于特征点和伪 Zernike 矩的数字图像水印算法[J]. 山东理工大学学报(自然科学版), 2007, 21(5):88-92.
- [12] 丁南南,刘艳滢,朱明.尺度相互作用墨西哥帽小波提取图像特征点[J]. 液晶与显示, 2012, 27(1):125-129.
- [13] The C H, Chin R T. On image analysis by the methods of moments[J]. IEEE transaction on pattern analysis and machine intelligence, 1988, 10(4):496-513.
- [14] 杨迪,孙劲光,何巍,等.基于伪 Zernike 矩归一化的人脸特征提取方法[J]. 计算机工程与应用, 2010, 46(35):163-165.

(上接第 71 页)

- [6] 王欣,罗代升,王正勇.基于改进谱减算法的语音增强研究[J]. 成都信息工程学院学报, 2007, 22(2):201-204.
- [7] 职振华,马建芬.改进的谱减法在语音增强中的应用[J]. 电声技术, 2008, 32(2):46-48.
- [8] 洪晓芬.基于谱减法的改进语音增强方法[J]. 计算机工程与设计, 2007, 28(22):5453-5454.
- [9] 林琴,张道信,吴小培.一种基于改进谱减法的语音去噪新方法[J]. 计算机技术与发展, 2007, 17(7):63-66.
- [10] 丁伟,吴小培.基于改进谱减方法的语音增强研究[J]. 计算机技术与发展, 2008, 18(9):98-100.
- [11] Liu Huanhuan, Yu Xiaoqing, Wan Wanggen, et al. An improved spectral subtraction method[C]//Proc of 2012 international conference on audio, language and image processing. [s. l.]: [s. n.], 2012:790-793.
- [12] 张仁志,崔慧娟.谱相减法语音增强技术中“音乐噪声”的抑制[J]. 电声技术, 2005(5):35-38.

改进谱减法语音增强算法的研究

作者：[陈欢](#)，[邱晓晖](#)，[CHEN Huan](#)，[QIU Xiao-hui](#)

作者单位：[南京邮电大学 通信与信息工程学院](#)，[江苏 南京](#)，[210003](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

ISTIC

年，卷(期)：2014(4)

本文链接：http://d.wanfangdata.com.cn/Periodical_wjfz201404017.aspx