

基于 MATLAB 的谱相减语音增强算法的设计与实现

汤维维

(武汉理工大学, 信息工程学院, 湖北 武汉 430070)

摘要: 谱减法是消除噪声的经典算法, 它是处理宽带噪声的最通用技术, 即从带噪语音估值中减去噪声频谱估值, 而得到纯净语音的频谱。本文基于 MATLAB 对谱减法算法进行设计和实现, 并对传统谱减法和改进形式的谱减法进行比较。仿真表明, 改进形式的谱减法可以有效地降低背景噪声, 提高信噪比。

关键词: 语音增强; 传统谱减法; 改进形式的谱减法; MATLAB

1 引言

人们在语音通信过程中不可避免地会受到来自周围环境、传输媒介引入的噪声、通信设备内部电噪声、乃至其它讲话者的干扰。这些干扰最终将使接收者接收到的语音已非纯净的原始语音信号, 而是受噪声污染的带噪语音信号。

语音增强是解决噪声污染的一种有效方法, 它的一个主要目标是从带噪语音信号中提取尽可能纯净的原始语音。语音增强目前已发展成为语音信号数字处理的一个重要分支。它的应用范围是降低听觉噪声、识别系统的预处理和线性预测编码的预处理。这种技术对语音识别和说话人识别是十分重要的, 可使识别装置在通常环境中的含噪语音下进行工作。然而, 由于干扰通常都是随机的, 从带噪语音中提取完全纯净的语音几乎不可能。在这种情况下, 语音增强的目的主要有两个: 一是改进语音质量, 消除背景噪声, 使听者乐于接受, 不感觉疲劳, 这是一种主观度量; 二是提高语音可懂度, 这是一种客观度量。这两个目的往往不能兼得。

语音增强不但与语音信号数字处理理论有关, 而且涉及到人的听觉感知和语音学。再则, 噪声来源众多, 随应用场合而异, 它们的特性也各不相同, 因此针对不同的噪声, 我们将采取不同的语音增强对策。语音增强中的谱减法是处理宽带噪声的最通用技术, 即从带噪语音估值中减去噪声频谱估值, 而得到纯净语音的频谱。宽带噪声通常可以假定为高斯噪声和白噪

声。它的来源很多, 包括风、呼吸噪声和一般随机噪声源。量化噪声通常作为白噪声来处理, 也可以视为宽带噪声。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。由于人耳对语音频谱分量的相位不敏感, 所以谱减法主要针对短时幅度谱^[1]。

MATLAB 是一种既可交互使用又能解释执行的计算机编程语言, 是解决工程技术问题的计算平台。通过 MATLAB 编程执行谱减法算法进行仿真, 我们可以直观地看到算法思想以及算法的执行效果, 有利于进一步工程的开展。

2 语音增强算法

2.1 传统谱减法

假定语音为平稳信号, 而噪声和语音为加性信号且彼此不相关。此时带噪语音信号可表示为

$$y(n) = s(n) + d(n) \quad 0 \leq n \leq N-1 \quad (1)$$

式中, $s(n)$ 为纯净语音信号, $d(n)$ 为平稳加性高斯噪声。实际上, 为避免分帧时的截断效应, 应对 $y(n)$ 加窗。为了书写方便,

这里用 $y(n)$ 代表加窗后的带噪信号。设 $y(n)$ 的傅里叶系数为 $Y_k = |Y_k| \exp[j\theta_k]$, $s(n)$ 的傅里叶系数为 $S_k = |S_k| \exp[j\alpha_k]$, $d(n)$ 的傅里叶系数为 N_k , 由(1)式, 则有 $Y_k = S_k + N_k$ 。而语音增强的任务就是利用已知的噪声功率谱信息, 从 Y_k 中估计出 S_k 。由于人耳对相位不敏感, 所以只要估计出 $|S_k|$, 然后借用带噪语音相位, 进行反傅里

叶变换后即可得到增强的语音。

$y(n)$ 经FFT后, 有 $Y_k = S_k + N_k$, 由此可得

$$|Y_k|^2 = |S_k|^2 + |N_k|^2 + S_k N_k^* + S_k^* N_k \quad (2)$$

式中, *表示复共轭。因为假定噪声为不相关的, 即 $s(n)$ 与 $d(n)$ 独立, 不会出现信号和噪声的乘积项, 则互谱的统计均值为0。而 N_k 为零均值的高斯分布, 所以有

$$E[|Y_k|^2] = E[|S_k|^2] + E[|N_k|^2] \quad (3)$$

只要从 $|Y_k|^2$ 中减去 $|N_k|^2$ 便可恢复 $|S_k|^2$ 。之所以能这样是基于人耳对语音相位不敏感(相聋)这一特点。因为噪声是局部平稳的, 故可以认为语音前的噪声与发语音期间的噪声功率谱相同, 因而可以利用发语音前(或后)的“寂静帧”来估计噪声。对于一个分析帧内的短时平稳过程, 有

$$|Y_k|^2 = |S_k|^2 + \lambda(k) \quad (4)$$

其中, $\lambda(k)$ 为无语音时 $|N_k|^2$ 的统计平均值, 即 $\lambda(k) = E[|N_k|^2]$ 。由此可得原始语音的估计值

$$\hat{S}_k = \left[|Y_k|^2 - E[|N_k|^2] \right]^{1/2} = \left[|Y_k|^2 - \lambda(k) \right]^{1/2} \quad (5)$$

这就是谱减法的基本原理。谱减法的原理框图如图1所示。

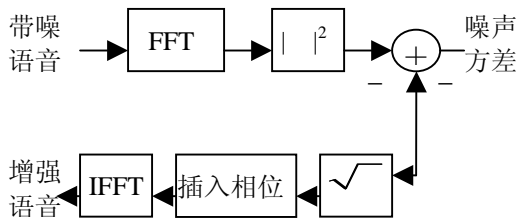


图1 谱减法的原理框图

在实际计算中, 可能会出现 $|Y_k|^2$ 小于 $\lambda(k)$ 的情况, 因为功率谱不能为负, 此时修改式(5)为

$$\hat{S}_k = \max \left\{ \left[|Y_k|^2 - \lambda(k) \right]^{1/2}, \varepsilon \right\} \quad (6)$$

式中 ε 为一个大于0的常数, 它由试验确定。定义第 k 个频谱分量的增益函数

$$G_k = \frac{\hat{S}_k}{|Y_k|}, \text{ 以及后验信噪比}$$

$\gamma_k = |Y_k|^2 / \lambda(k)$, 则式(5)可以改写为

$$G_k = (1 - 1/\gamma_k)^{1/2} \quad (7)$$

由式(7)可以清楚地看出谱相减的物理意义: 它相当于对带噪声语音的每一个频谱分量乘以一个系数 G_k 。当信噪比高时, 含语音的可能性很大, 衰减小。反之, 则认为含有语音的可能性小, 衰减则增大。

2.2 改进的谱减法

式(5)中的 $\lambda(k)$ 是以无声期间的统计平均的噪声方差代替当前分析帧的噪声频谱, 而实际上噪声频谱服从高斯分布

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \quad (8)$$

其中, m 为 x 的均值, σ 为标准偏差。噪声的帧功率谱随机变化范围很宽, 在频域中的最大、最小值之比往往达到几个数量级, 而最大值与均值之比也达6~8倍。因而, 在减去噪声谱后, 会有些较大的功率谱分量的剩余部分, 在频谱上呈现出随机出现的尖峰, 在听觉上形成残留噪声。这种噪声具有一定的节奏性起伏感, 所以称之为“音乐噪声”。

为有效地减少宽带和音乐噪声, 可对谱减法进行改进。噪声的能量往往分布于整个频谱范围, 而语音能量则较集中于某些频率或频段, 尤其在元音的共振峰处。因此可在元音段等幅度较高的时帧去除噪声时, 减去 $n \cdot \lambda(k)$, 则可更好地相对突出语音的功率谱。这种改进也称为被减项权值处理。

同时, 将式(5)中的功率谱计算 $| \cdot |^2$ 及 $| \cdot |^{1/2}$ 改为 $| \cdot |^m$ 和 $| \cdot |^{1/m}$ 计算(这里 m 不一定为整数), 可以增加灵活性。这种方法称为功率谱修正处理。经分析和试验得知, 当 $m > 2$ 时, 它具有与被减项加权处理相同的效果。

综合上面两种处理, 谱减法改进形式的原理框图如图2所示。

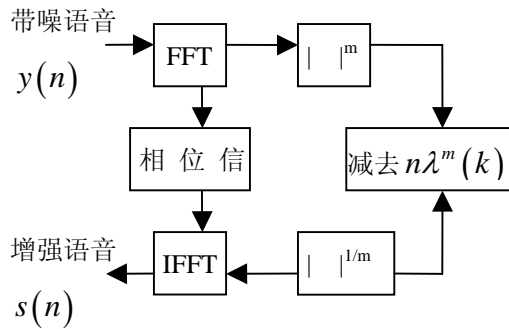


图2 谱减法的改进形式

此时式 (5) 修正为

$$\left| \hat{S}_k \right| = \left(\left| Y_k \right|^m - n \lambda_k^m \right)^{1/m} \quad (9)$$

引入 m, n 两个参数为计算提供了很大的灵活性。当 $m=2, n=1$ 时即变为基本的谱减法。实际的增强试验表明, 适当调节 m, n 可以获得比传统的谱减法更好的增强效果。因此实际的增强过程中, 更多地使用谱减法的改进形式^[2]。

3 基于 MATLAB 的算法实现与仿真

3.1 传统谱减法的实现与仿真

基于MATLAB来实现传统谱减法的仿真, 其程序流程图如图3所以。

读入语音文件用wavread函数实现, 对采样点补足帧的整数倍是方便实现MATLAB的矩阵计算以及最后对帧叠加进行还原处理。对每一帧独立的含噪语音进行语音增强处理, 再将每帧增强语音依次连接起来, 会由于恢复的信号帧间不连续, 使得增强语音含有周期性的“嘟嘟”背景声, 因此采用帧间重叠谱减法进行语音处理。程序中每帧信号256点, 128点重叠, 通过加汉明窗 $w=\text{hamming}(\text{framelength})'$ 实现, 经语音增强处理后, 再进行相反的处理, 恢复连续的语音信号。

帧或语音帧间的“寂静帧”来求功率谱。程序中取语音信号的2, 3, 4帧作为寂静帧进行处理, 最后取3帧的平均值来求噪声的功率谱。其代码为:

```
noise1=s(2*framelength+1:3*framelengt
h);
noise2=s(3*framelength+1:4*framelengt
```

```
h);
noise3=s(4*framelength+1:5*framelengt
h);
noise(1,:)=abs(fft(noise1'))';
noise(2,:)=abs(fft(noise2'))';
noise(3,:)=abs(fft(noise3'))';
ave_noise=mean(noise);
```

对带噪语音信号进行FFT变换可以直接用fft函数实现, 在MATLAB程序中, 这一部分用for循环对每帧循环实现^[3]。

谱相减是算法的关键, 这部分程序代码如下所示, 其中参数 $m=2, n=1$ 。因为涉及到估值, 所以实际中谱相减的差值会为负。因功率谱不能为负, 故可令负值为0或改变其符号, 在本文的实现程序中令负值为0。

```
TT=[];
for i=1:k
kk=(yd(i,:).^m-n*(ave_noise.^m));
pp=find(kk<0);
kk(pp)=0;
kk=kk.^(1/m);
TT=[TT;kk];
End
```

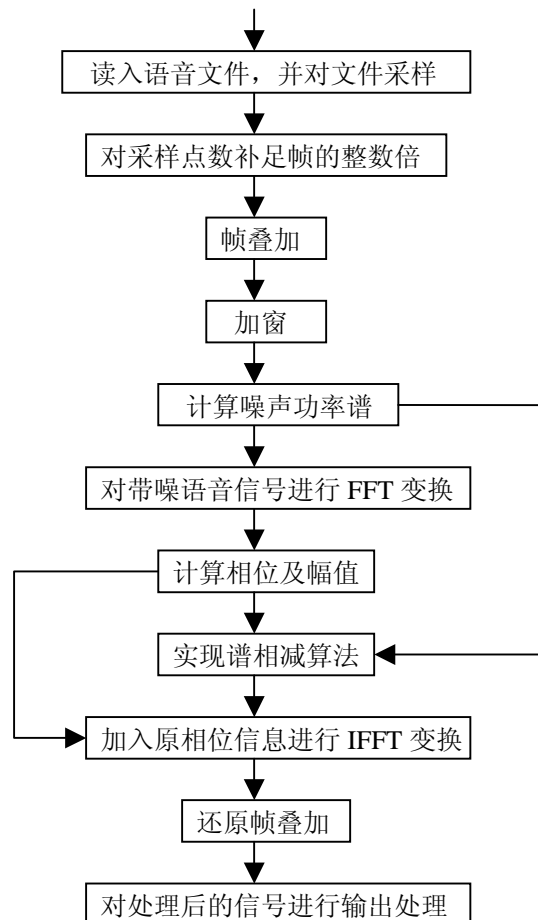


图3 传统谱减法程序流程图

在对FFT变换且谱相减后的信号加上原输入信号的相位信息后,就对信号进行IFFT变换,在MATLAB中用ifft函数实现,即

```
t=TT.*exp(j*angleyy); ss=ifft(t,[],2);
```

算法的最后要对前面的帧重叠进行相反的处理,以保持语音的连续特性。帧重叠进行相反处理的代码如下所示:

```
output=[];  
output=[output,ss(1,1:128)];  
for i=2:k  
    temps=(ss(i,1:128)+ss(i-1,129:256))/2;  
    output=[output,temps];  
end  
output=[output,ss(k,129:256)];
```

语音增强后的信号用wavwrite函数输出到新的wav文件中。

传统谱减法仿真后的结果如图4所示。仿真结果表明,噪声信号得到了一定的抑制,但是还存在一定的具有节奏性起伏感的音乐噪声。

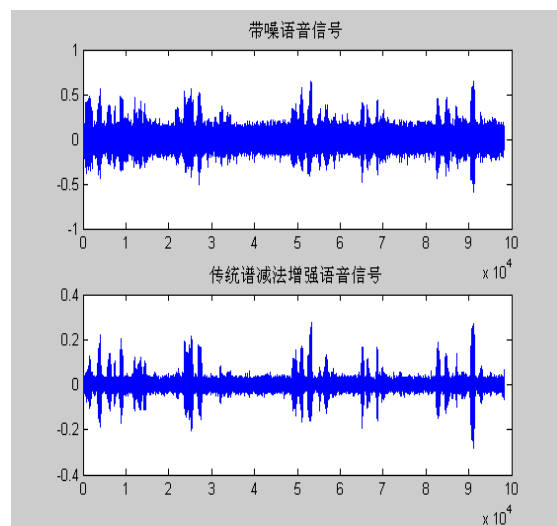


图4 传统谱减法仿真效果

3.2 改进形式的谱减法的实现与仿真

式(8)中的 m, n 两个参数为算法提供了很大的灵活性,适当调节这两个参数,可获得比原始相减法更好的增强效果。试验表明,两个可调参数 m, n 取不同值时,语音增强的效果是不同的: m 不变,随着 n 值的适当增大,宽带噪声及残余的音乐噪声减小,信噪比大大提高,但是语音也逐渐变尖,即有较大的失真; n 不变,随着 m 值的适当增大,宽带噪声及残余的音乐噪声增大,信噪

比的提高减少,但是语音也逐渐自然,即有较小的失真。试听结果表明, $m=0.4, n=0.9$ 时,较好的做到了降噪与提高语音可懂度之间的折衷,噪声得到了适当的抑制,较好地保持了原语音的特性。在MATLAB的实现上,只要在原来传统谱减法的基础上改动参数 m, n 的值即可得到改进后的效果,如图5所示即为改进形式的谱减法仿真效果。

试验表明,改进的谱减法消除了周期性干扰在电话语音输出端引起的啸叫噪声,除清晰度略有下降外,语音的可懂度和自然度未受损失。该方法对白噪声干扰消噪效果更好,但当环境恶劣,信噪比极低时,信噪比虽然提高不少,但语音的可懂度和自然度却受到一定的损失,实际听起来像机器人的声音。因此,在实时实现时需注意运算精度、帧间重叠的实现和噪声功率谱估计等问题。

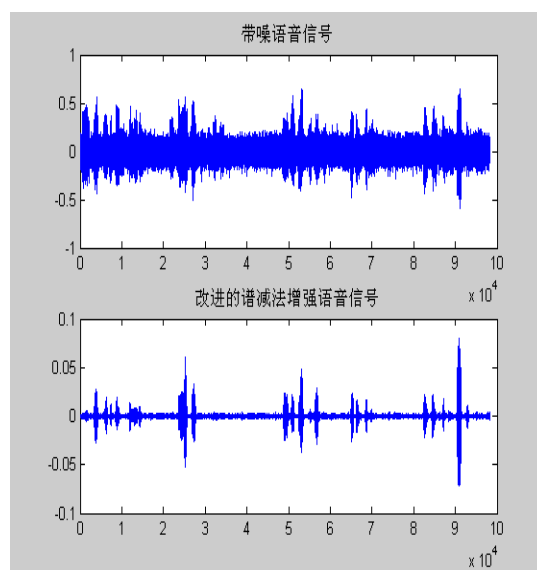


图5 改进的谱减法仿真效果

4 结论

谱减法是一种常用的单通道语音增强经典算法,传统谱减法和它的改进形式可以归纳为一个通用的谱减法参数公式,参数 m, n 的取值和增强效果有着密切的关系,当 m, n 取值适当,则可获得较为理想的增强效果。谱减法有效地降低了背景噪声,提高了信噪比,但是还存在如何更好地改善降噪量与语音可懂度之间的矛盾。采用单端输入的语音增强处理的研究还在发展阶段,相信更多更成熟的研究成果将不断涌现。

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,1995,385-411.
- [2] Hyoun-Gook Kim, Markus Schwab, Nicolas Moreau. Speech Enhancement of Noisy Speech Using Log-Spectral Amplitude Estimator and Harmonic Tunneling[EB/OL].
<http://www.nue.tu-berlin.de/publications/papers/IWANEC2003.pdf>
- [3] J.H.L. Hansen, B.Pellom. An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms[EB/OL].
<http://cslr.colorado.edu/rspl/PUBLICATIONS/PDFs/CP-icslp98-SpchEnhansProto-SL980917-Wcover.PDF>
- [4] Dominc K.C.Ho, Tong Wang. Speech Enhancement Concept and Methodology [EB/OL].http://meru.cecs.missouri.edu/mm_seminar/ho.pdf
- [5] Kotta Manohar. Single Channel Enhancement of Noisy Speech[EB/OL].
http://www.ee.iitb.ac.in/~esgroup/es_mtech02_sem/es02_sem_rep_kotta.pdf
- [6] Gautam Moharir, Pushkar Patwardhan, Preeti Rao. Spectral enhancement Preprocessing for the HNM Coding of Noisy Speech[EB/OL].
<http://www.ee.iitb.ac.in/uma/~daplab/papers/SL021575.pdf>

Design and Realization of Speech Enhancement Using Spectral Subtraction Algorithm Based on MATLAB

Tang Weiwei

Wuhan University of Technology 430070

Abstract

The spectral subtraction is a classical algorithm to dispel the noise. It is the common technology of handling broadband noise, viz. subtracting noise spectral estimation from the speech estimation including noise, to receive the frequency spectrum of clean speech. This paper designs and realizes the algorithms of spectral subtraction based on MATLAB, and compare the improved form of spectral subtraction with the traditional spectral subtraction. It shows that the improved form of spectral subtraction and can efficiently decrease background noise and increase the SNR.

Keywords: speech enhancement; traditional spectral subtraction; improved form of spectral subtraction; MATLAB

作者简介: 汤维维(1982-), 女, 广西桂林市荔浦县人, 武汉理工大学信息工程学院硕士研究生。