

文章编号: 1001-8360(2004)01-0064-04

基于谱相减改进算法的语音增强研究

徐 岩, 杨 静, 王维汉

(兰州交通大学 信息与电气工程学院, 甘肃 兰州 730070)

摘 要: 根据传统的谱相减增强型算法, 提出了抑制噪声的谱减改进算法, 暨根据带噪语音帧频谱功率与噪声帧频谱功率比值, 动态调整 α 、 β 谱减系数, 使之谱减效果在较大提高信噪比的同时, 又能将残留的音乐噪声和语音失真保持在人耳听觉的容忍范围之内, 保证原始语音信号质量达到一定的可懂度和清晰度。借助 MATLAB 语言进行试验仿真, 仿真结果与原算法相比较, 证明语音增强效果十分显著。

关键词: 语音增强; 谱相减算法; 噪声抑制

中图分类号: TN91 **文献标识码:** A

Research of Speech Enhancement Based on the Improved Spectral Subtraction Algorithm

XU Yan, YANG Jing, WANG Wei han

(School of Information and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: According to the traditional spectral subtraction and speech enhancement algorithm, a new improved spectral subtraction algorithm has been put forward to restrain noises. This method gives different spectral estimate values α and β based on the signal spectral power and uses spectral subtraction to realize speech enhancement. At last Emulation by means of the MATLAB language and its analysis result have been given in detail.

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

环境噪声的污染使得许多语音处理系统的性能恶化。基于语音生成模型的低码率参数编码中涉及到 LPC 系数的求解, 该系数对语音的编解码质量有重要的影响。由于噪声的干扰导致 LPC 系数的精确求解十分困难^[1,2], 从而极大地降低了重建语音的质量。目前的语音识别系统虽然已达到较高的识别率, 但其前提是在理想的环境或实验室环境中。在有噪声的环境中, 这些语音识别系统的识别率会普遍地大幅度下降。如一个典型的孤立词识别系统, 用纯净语音训练后, 其识别率可以达到 100%, 但在 90 km/h 速度行驶的列车上其识别率将会降到 30%, 这主要是受到车内外噪声的影响。在常用的增强算法中, 多种增强算法^[3]都有其发展历史, 而基于短时谱估计增强算法中的谱相减法及其变体最为常用, 这是由于它运算量较小, 容易实时实现, 增强效果也比较明显。

1 谱相减改进算法理论分析

1.1 谱相减后对噪声的抑制方法

对通用的谱减法语音增强效果进行分析, 发现在低信噪比下它的增强效果是比较明显的, 而且时域图形中噪声总是在零坐标线上呈均匀分布, 噪声中包含了音乐噪声和少量的清音信号, 可以采用半波整形, 中心限幅的方法^[4,5]对语音信号在时域内进一步处理, 进一步提高信噪比, 消除音乐噪声。

谱相减技术大多数都是在频域进行, 也可用于时域或倒谱域。当语音谱超过估计的噪声谱时, 便产生了间歇短暂的突发声调, 这种具有音乐特性的残余噪声是各帧在随机频率上出现的许多声调的群体结果, 它要比原始语音中的噪声清楚的多。

1.2 谱相减改进算法理论分析

处理宽带噪声最常用谱减法, 它从带噪语音估值中减去噪声频谱估计, 而得到纯净语音的频谱。由于人耳对语音频谱分量的相位不敏感, 这种方法主要用

收稿日期: 2003-05-25; 修回日期: 2003-09-04
基金项目: 铁道部科学技术基金资助项目 (2003X012)
作者简介: 徐 岩 (1963—), 男, 江苏南通人, 副教授, 硕士。
E-mail: xuyan@lzri.edu.cn

于短时幅度谱。假定语音为平稳信号, 噪声和语音为加性信号且彼此不相关, 此时带噪声语音信号可表示为

$$y(t) = s(t) + n(t) \tag{1}$$

式中, $s(t)$ 为纯净语音信号; $n(t)$ 为噪声信号。

用 $Y(\omega)$ 、 $S(\omega)$ 和 $N(\omega)$ 表示 $y(t)$ 、 $s(t)$ 和 $n(t)$ 的傅立叶变换, 则有下列关系存在

$$Y(\omega) = S(\omega) + N(\omega) \tag{2}$$

对功率谱则有

$$|Y(\omega)|^2 = |S(\omega)|^2 + |N(\omega)|^2 \tag{3}$$

因为假定噪声为不相关的, 所以不会出现有信号与噪声的乘积项。只要从 $|Y(\omega)|^2$ 中减去 $|N(\omega)|^2$ 便可恢复 $|S(\omega)|^2$ 。因为噪声是局部平稳的, 故可以认为发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的“寂静帧”来估计噪声。由于语音信号是不平稳的, 而且实际上只能使用分段加窗信号, 此式可改写为

$$|Y_w(\omega)|^2 = |S_w(\omega)|^2 + |N_w(\omega)|^2 + S_w(\omega)N_w^*(\omega) + S_w^*(\omega)N_w(\omega) \tag{4}$$

式中, 下标 w 表示加窗信号; $*$ 表示复共轭。

可以根据观测数据估计 $|Y(\omega)|^2$, 其余各项必须近似统计均值。由于 $n(t)$ 和 $s(t)$ 独立, 则互谱的统计均值为 0, 从而原始语音的估值为

$$|S_w(\omega)|^2 = |Y(\omega)|^2 - \langle |N_w(\omega)|^2 \rangle \tag{5}$$

式中, 符号 $\langle \rangle$ 表示估值; $\langle |N_w(\omega)|^2 \rangle$ 为无语音时 $|N(\omega)|^2$ 的统计均值。

为有效减小带宽和噪声, 需对谱减法进行改进。噪声的能量往往分布于整个频率范围, 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帧去除噪声时, 减去 $\beta < |N_w(\omega)|^2$ ($\beta > 1$), 则可更好的相对突出语音的功率谱。这种改进也称为被减项权重值处理。同时, 将原始的功率谱幂次 2 和 1/2 改为 α 和 $1/\alpha$, 这种方法称为功率谱修正处理, 当 $\alpha > 2$ 时, 它具有与被减项权重值处理相同的效果, 从而式 (5) 可修正为

$$|S_w(\omega)|^\alpha = |Y_w(\omega)|^\alpha - \beta \langle |N_w(\omega)|^\alpha \rangle \tag{6}$$

在增强谱减算法中, α 为谱减功率修正系数, 加大这个系数, 将会进一步提高信噪比, 但是也将加大语音信号的失真。 β 为谱减噪声系数, 它的作用是对被减的噪声功率谱进行修正。 β 的加大将会起到减少音乐噪声、更好地相对突出语音频谱的作用, 但是引起的失真也会增大。当 $\alpha = 2$ 、 $\beta = 1$ 时为基本的谱减法。

α 、 β 系数值一旦确定之后, 在谱减处理过程中就一直保持不变, 这并不符合语音增强要求, 因为在噪声段需要减去的噪声功率谱较大, 以提高信噪比, 而在带

噪声语音段减去的噪声功率谱又要相对较小, 以保证较多的清音存在, 提高语音的可懂度。这要求在不同的频率段和不同的语音帧处理过程中动态调整 α 、 β 谱减系数值。在信噪比较低的情况下, 带噪声语音段频谱功率要高于噪声段频谱功率, 因为噪声频谱功率上叠加有语音频谱功率, 可根据带噪声语音帧频谱功率与噪声帧频谱功率比值, 动态调整 α 、 β 谱减系数, 使之谱减效果在较大提高信噪比的同时, 又能将残留的音乐噪声和语音失真保持在人耳听觉的容忍范围之内, 保证原始语音信号质量达到一定的可懂度和清晰度。谱减改进算法原理框图如图 1 所示。

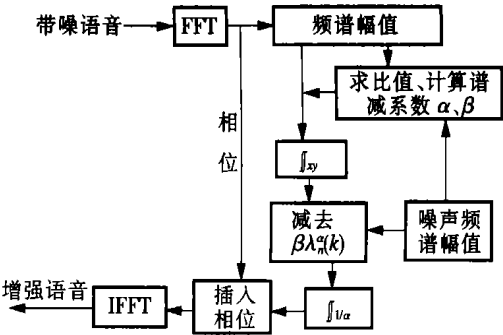


图 1 谱相减改进算法原理示意图

1.3 算法中 α 、 β 值的计算

式 (7) 的推导是通过对大量数据的分析实现的, 首先对语音频谱放大幅值, 同时减小被减噪声谱以保证少量清音不被滤除; 其次对纯噪声的频谱缩小幅值, 同时放大被减噪声谱, 以提高信噪比。 α 、 β 值不能通过设置统一的门限值来推导, 通过实验发现, 有些清音能量稍大, 有些非常小, 叠加到白噪声上, 功率谱很难清楚的分辨, 因此推导只能采用比值法。

$$\begin{cases} \alpha = \alpha_0 \times (1 + p/p_{\max}) \\ \beta = \beta_0 \times q / (1 + p/p_{\max}) \end{cases} \tag{7}$$

式中, α_0 、 β_0 为设定增强型谱减法系数, p 为信号帧频谱功率, p_{\max} 为信号帧频谱功率最大值, q 为修正系数, 经反复比较, 取 q 值 1.3 为最佳。

实验中测得一个合适的帧噪声频谱功率值是非常困难的, 即使在实验室条件下, 帧与帧之间的差异都非常大。选取帧频谱最大功率值为基数与每帧频谱功率作比较, 这个比值就为 α_0 系数值的增加值, 帧频谱功率大时, α_0 值变大, 帧频谱功率小时, α_0 值变小。 β_0 的改正值取为这个比值的倒数, 当帧频谱功率大时, 减去的噪声频谱功率反而小, 帧频谱功率小时, 减去的噪声频谱功率反而大, 系数值 q 是为了突出 β 值的作用。

在实现中可逐帧比较, 取出每一帧频谱的值与已有的最小频谱值作比较, 取出相应点的最小值作为最小频谱值, 依次类推, 这个最小频谱值就是噪声功率谱

值。同理,可计算出每帧频谱幅值累加和与已有的最大值作比较,如果新值比最大值还大,则最大值被新值替换,这里的最大值就是上面公式中的 p_{\max} ,一直循环比较下去。

2 谱相减改进算法数字仿真

2.1 数字仿真的实现

在实验中采用 8 000 Hz, 16 bit, 单声道的 PCM 格式的语音录制条件,纯净语音录入是在安静的夜晚环境下,读“我们的国家”,时间大约为 4 s,使用 MATLAB 函数^[6,7,8] wavread() 调用,语音段的开始有停顿以留出寂静帧,寂静帧是用来计算信噪比的。

噪声是用两种方式产生,一种是录制的,包括了计算机自己的“嗡嗡”声,这不是严格意义上的高斯白噪声,但仿真结果表明在提高语音信号的信噪比的同时,它具有较好的可懂度;另一种为 MATLAB 函数 randn() 产生随机数,这是真正意义上的高斯白噪声,信号产生

语句为 $yy = \text{randn}(1:\text{length}(y))$ 。
仿真中取输入信号为纯信号与 k 倍的噪声之和,即 $y = y + k * yy$,改变 k 值就可以得到不同的输入信噪比,经过实验,选取 0.530、0.298、0.168 为输入信噪比-5 dB、0 dB、5 dB 的 k 值。主观试听效果采用函数 sound() 来播放。

2.2 谱相减改进算法流程图

谱相减改进算法流程图如图 2 所示。在众多窗函数中选用了哈明窗,因为它可以有效的降低吉布斯效应的震荡程度。窗口长度为抽样频率(8 000) * 系数,256 点为帧长,实际窗口长度为 256/过采样恒量,实际帧长为 64 点。窗口如果不滑动,文献表明^[9]结果会出现“嘟,嘟”的间断音,滑动速度取为实际窗口的 1/4 较合适,即 16 点。信号 Y 经 FFT() 函数由时域转为频域后,频谱功率计算式为 $Y = Y * \text{CONJ}(Y)$,噪声频谱值取噪声帧中对应每一点的最小值作为被减噪声频谱值。

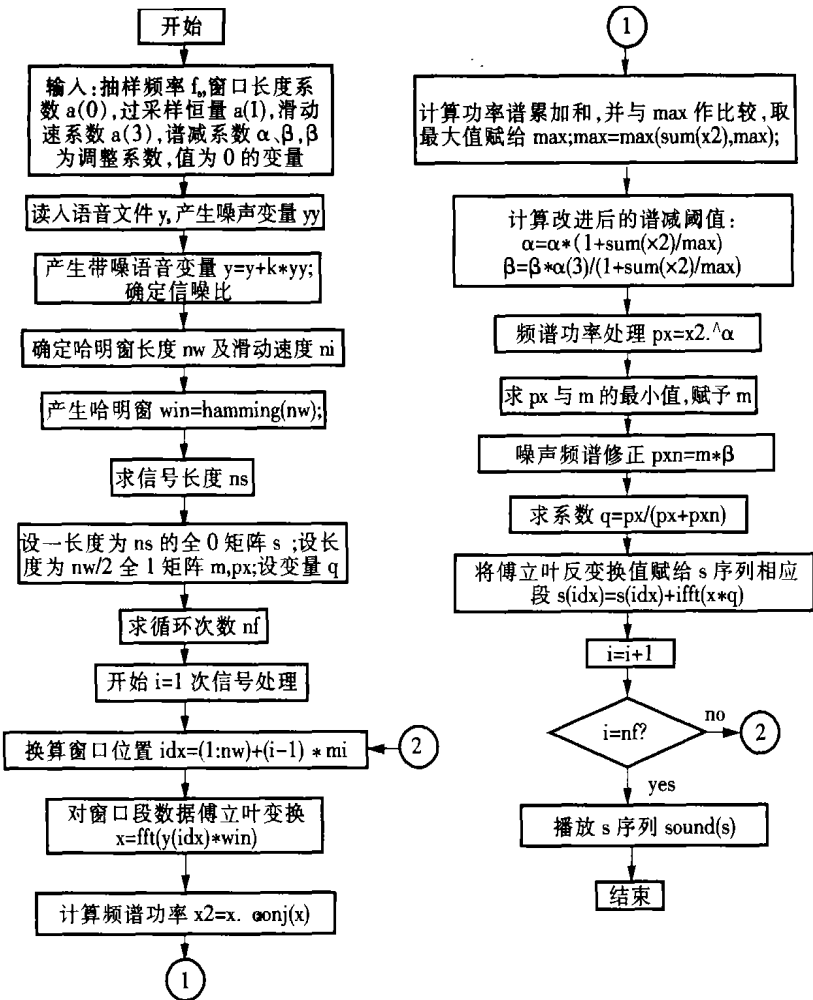


图 2 谱相减改进算法流程图

2.3 仿真结果及其分析

在仿真中选用-5 dB 和+5 dB 的信噪比语音信

号进行实验,算法中 α 、 β 值经式(7)计算都为 3,噪声抑制的被减幅度值统一取为 0.01。时域原始及结果

信号图形分别如图 3、图 4、图 5 所示, 实验比较结果量化值见表 1。

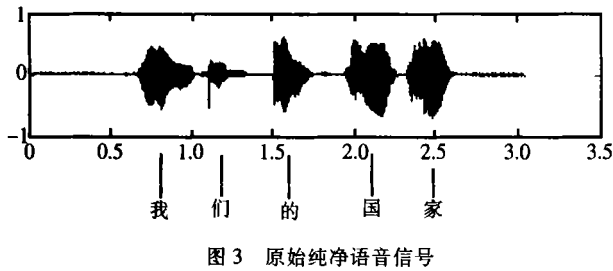
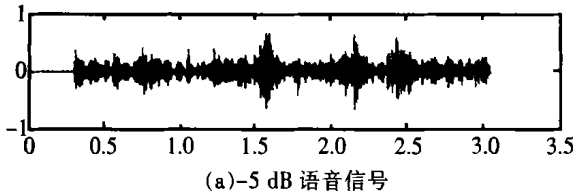
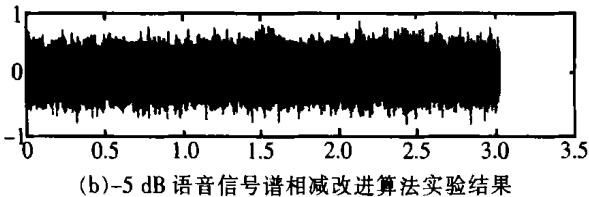


图 3 原始纯净语音信号

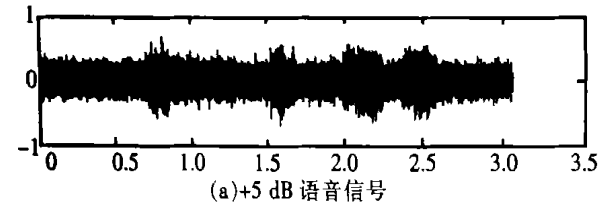


(a)-5 dB 语音信号

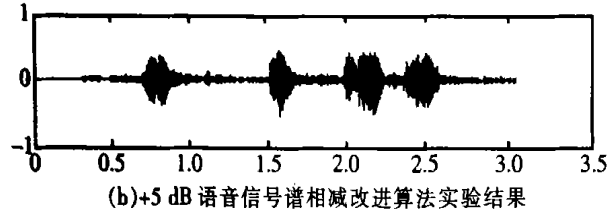


(b)-5 dB 语音信号谱相减改进算法实验结果

图 4 -5 dB 信号实验结果比较图



(a)+5 dB 语音信号



(b)+5 dB 语音信号谱相减改进算法实验结果

图 5 +5 dB 信号实验结果比较图

表 1 谱相减算法处理后的输出信噪比较

输入信噪比/dB	- 5	提高值	+ 5	提高值
谱相减普通算法	8.39	13.39	19.16	14.16
谱相减增强算法	10.74	15.74	21.7	16.7
谱相减改进算法	12.75	17.75	25.69	20.69

3 结论

经过以上仿真实验可得出以下结论:

(1) 谱减算法有一定的去噪功能, 且在低信噪比下效果明显, 信噪比提高一般都在十几分贝以上, 但经主观试听留有明显的类似流水声的音乐噪声。

(2) 谱相减增强算法在调节 α 、 β 系数后可以取得比传统谱减法高的信噪比, 经主观试听噪声有一定减弱。

(3) 噪声影响主观试听的懂度和信噪比, 半波整形, 中心限幅的解决方案对噪声的去除是有决定性作用的, 主观试听可以感受到背景噪声变成了细小的白噪声声音, 语音清晰度明显增强。但被减幅值过大, 噪声去除的越多, 信噪比越高, 同时一部分清音成分被减去, 反而降低了语音的懂度, 因此被减幅值只能适当选取。

(4) 谱相减改进算法在输入信噪比较低的情况下, 输出信噪比始终比谱相减增强算法高, 但比较明显的是噪声减小, 同时语音懂度得到了一定的提高, 但在输入信噪比较高的情况下, 效果更好。

在实验中证实, 这种方法可以明显抑制噪声, 噪声由流动的水声变为细小的嗡嗡声。同时, 实验也表明, 被减幅值不能取得过大, 否则噪声虽被滤除, 信噪比提高, 但大量的清音也被滤除, 这会影响语音的懂度。

参考文献:

[1] 徐耀华, 王刚, 郭英, 门向生. 一种基于噪声生成模型的语音消噪算法[J]. 空军工程大学学报, 2001, 2(1): 45—48.

[2] 易克初, 田斌, 付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000.

[3] 胡广书. 数字信号处理——理论、算法与实现[M]. 北京: 清华大学出版社, 2000.

[4] 蔡汉添, 袁波涛. 一种基于听觉掩蔽模型的语音增强算法[J]. 通信学报, 2002, 8: 93—98 .

[5] 刘晓晖, 周炜, 郑明, 吴道梯. 语音增强系统中宽带噪声的滤除[J]. 西安交通大学学报, 1998, 12: 11—16.

[6] 何强, 何英. MATLAB 扩展编程[M]. 北京: 清华大学出版社, 2002.

[7] 王立宁, 乐光新, 詹菲. Matlab 与通信仿真[M]. 北京: 人民邮电出版社, 2000.

[8] 胡昌华, 张军波, 夏军, 张伟. 基于 MATLAB 的系统分析与设计[M]. 西安: 西安电子科技大学出版社, 2001.

[9] 李宏伟, 段艳丽, 郭英. 基于帧间重叠谱减法的语音增强算法及实现[J]. 空军工程大学学报, 2001, 5: 48—54.

(责任编辑 姚家兴)