

一种基于改进谱减法的语音增强方法

王让定

(同济大学 人工智能研究室 上海 200092)
(宁波大学 纵横智能软件研究所 宁波 315211)

柴佩琪

(同济大学 人工智能研究室 上海 200092)

摘要 谱减法是最常用的一种语音增强技术,其特点是计算复杂度低、实时性强、易于实现.谱减法的主要目的是去除语音信号中的噪声干扰,提高语音信号质量.本文在研究基于改进谱减法的基础之上,提出了利用带噪语音的高频区估计噪声谱以及由短时过零率和短时能量组合而成的加权函数去除背景噪声及音乐噪声的语音增强方法.实验表明,这种时频结合的语音增强方法对背景噪声下的语音质量的增强效果明显.

关键词 语音增强,噪声谱,谱减法,加权函数

中图法分类号 TN912.35

1 引言

目前,工作于现实环境中的语音装置(如 Cellular Phones, 语音编码器等)的语音增强问题引起了研究者的广泛关注^[1].语音增强的目的是改善这些系统的性能,去除带噪语音(Noisy Speech)中的噪声,尤其是实际环境中的背景噪声.

多年来,研究者已经提出了许多语音增强算法,从算法的复杂性和实时性角度出发,谱减法(Spectral Subtraction)一直受到了研究者的普遍关注.传统谱减法^[2]的基本思想是短时带噪语音谱减去估计的背景噪声谱,由于此算法的约束是噪声是加性的(即背景噪声叠加在语音信号中),且背景噪声是平稳的,但在实际应用中,噪声很难满足平稳性的假设.因此,从本质上来说,虽然该算法大大去除了背景噪声,但随之而来的是引入了令人厌恶的音乐噪声(Musical Noise),此问题在最常用的单通道系统(语音和噪声在同一通道中传输)中更为突出,也更难处理.

针对谱减后产生的音乐噪声,Djigan 等^[3]的改进语音增强算法利用带噪语音信号的无音段估计背

景噪声谱,该算法当 SNR 较高时,音乐噪声的消除比较明显;Kamath 等^[4]提出了多带(Multi-Band)改进谱减算法,该算法将带噪语音分成若干个频段,在各频段进行谱减法,算法中的实质是估计谱减因子,即在不同的频带减去不同的噪声估计谱,算法结果大大消除了音乐噪声;而 Yamauchi^[5]等的改进谱减算法是在每个被分析的短时带噪语音信号帧的高频段估计噪声谱,与传统谱减算法相比,该算法有效地去除了音乐噪声;当带噪语音信号的 SNR 相当低时($\text{SNR} < 0\text{dB}$),Hartmann 等^[6]提出了利用语音和噪声先验知识的语音增强模型,通过平滑滤波音乐噪声(Filtering of Smoothed Musical Noise)的方法降低谱减后的音乐噪声.另外,基于 AR-HMM 模型^[7]的改进谱减算法有效地估计了带噪语音中的背景噪声,大大消除了谱减后的音乐噪声.上述改进的谱减算法,基本上打破了传统谱减法中关于噪声是平稳性的假设,并且在算法实现中,权衡考虑(Tradeoff)了增强后的语音失真和去噪效果.

在实际环境中,语音增强主要考虑的是算法的实时性、有效性及计算复杂性.本文在研究背景环境中语音增强问题的基础上,结合 Kamath^[4]和 Kotnik 等^[8]的工作,提出了有效消除音乐噪声的改进谱减

收稿日期:2002-03-21;修回日期:2002-11-13

算法,利用带噪语音的高频区估计噪声谱以及短时过零率和短时能量组合而成的加权函数去除音乐噪声.实验表明,所提算法对背景噪声下语音质量的增强效果显著.

2 算法描述

2.1 谱相减法

设背景噪声信号 $d_m(n)$ 是短时平稳的,语音信号 $s_m(n)$ 也是短时平稳的,且 $d_m(n)$ 是与 $s_m(n)$ 不相关的加性噪声,则在此背景噪声下的带噪语音信号 $y_m(n)$ 可表示为

$$y_m(n) = s_m(n) + d_m(n) \quad m = 1, 2, \dots; \\ n = 0, 1, \dots, N-1. \quad (1)$$

式中 N 表示语音分析帧长.若以 $Y_m(\omega)$ 、 $S_m(\omega)$ 、 $D_m(\omega)$ 分别代表 $y_m(n)$ 、 $s_m(n)$ 及 $d_m(n)$ 的能量密度谱(Magnitude Spectrum),则

$$Y_m(\omega) = S_m(\omega) + D_m(\omega), \quad (2)$$

其中 $Y_m(\omega)$ 、 $S_m(\omega)$ 及 $D_m(\omega)$ 分别对应 $y_m(n)$ 、 $s_m(n)$ 及 $d_m(n)$ 的 FFT 变换,于是有

$$|Y_m(\omega)|^2 = |S_m(\omega)|^2 + |D_m(\omega)|^2 + S_m(\omega) \cdot D_m^*(\omega) + S_m^*(\omega) \cdot D_m(\omega), \quad (3)$$

式中 $S_m^*(\omega)$ 、 $D_m^*(\omega)$ 分别是 $S_m(\omega)$ 及 $D_m(\omega)$ 的复共轭, $|S_m(\omega)|^2$ 、 $|D_m(\omega)|^2$ 分别代表语音信号和噪声信号的短时能量谱.基于短时谱幅度估计的语音增强技术的目的就是设法得到 $|S_m(\omega)|$ 的估计 $|\hat{S}_m(\omega)|$,并由此得到 $s_m(n)$ 的估计 $\hat{s}_m(n)$,即增强后的语音.

由含噪语音信号经加窗、FFT 变换等运算可直接得到 $|Y_m(\omega)|^2$.但 $|D_m(\omega)|^2$ 、 $S_m(\omega) \cdot D_m^*(\omega)$ 、 $S_m^*(\omega) \cdot D_m(\omega)$ 无法精确得到,因而分别以三者各自的平均能量 $E[|D_m(\omega)|^2]$ 、 $E[S_m(\omega) \cdot D_m^*(\omega)]$ 及 $E[S_m^*(\omega) \cdot D_m(\omega)]$ 来近似代替.由于 $s_m(n)$ 与 $d_m(n)$ 独立,所以 $S_m(\omega)$ 与 $D_m(\omega)$ 独立,因此 $E[S_m(\omega) \cdot D_m^*(\omega)]$ 及 $E[S_m^*(\omega) \cdot D_m(\omega)]$ 均为 0.这样, $|S_m(\omega)|^2$ 的估计 $|\hat{S}_m(\omega)|^2$ 由下式获得

$$|\hat{S}_m(\omega)|^2 = |Y_m(\omega)|^2 - E[|D_m(\omega)|^2]. \quad (4)$$

此式是传统的谱减法,在(4)的基础上,引入两个参数 α 、 γ 得下式:

$$|\hat{S}_m(\omega)|^\alpha = |Y_m(\omega)|^\alpha - \gamma |\lambda_m(k)|^\alpha, \quad (5)$$

其中 $\lambda_m(k) = E[|D_m(\omega)|^2]$. (5) 式是基于传统谱减法的改进形式,当 $\alpha = 2$, $\gamma = 1$ 时,这个算法是传统的谱相减法.适当调整 α 和 γ 两个参数,可以得到较好的增强效果.实验表明 α 和 γ 的取值范围为 $2 \leq \alpha$

≤ 4 , $1 \leq \gamma \leq 7$ ^[9],具体取何值,由实验测定.

在式(4)、(5)中,当带噪语音信号的能量小于估计的噪声信号的能量时,谱减法的结果为负值.为了保证相减幅度值为正,我们采用了如下的算法处理.

$$|\hat{S}_m(\omega)|^\alpha = \begin{cases} |Y_m(\omega)|^\alpha - \gamma |D_m(\omega)|^\alpha, & |Y_m(\omega)|^\alpha \geq |D_m(\omega)|^\alpha \\ \beta * |Y_m(\omega)|^\alpha, & |Y_m(\omega)|^\alpha \leq |D_m(\omega)|^\alpha \end{cases} \quad (6)$$

式中当 $|Y_m(\omega)|^\alpha \leq |D_m(\omega)|^\alpha$ 时,因子 β 由实验确定.但须注意的是因子太小,以致接近零时,又引入了相当高的音乐噪声,太大会产生很大的语音失真^[9].

2.2 消除音乐噪声算法

(6) 式的改进谱减法有效地消除了加性背景噪声,虽然可通过调整因子 β 消除部分音乐噪声,但增强后的信号 $\hat{s}_m(n)$ 中仍然存在很大的音乐噪声,特别在无声段及 SNR 较低的语音信号段,音乐噪声的影响更为突出.为了有效地消除音乐噪声,本文在利用(6) 式进行谱减法前,首先估计背景噪声谱,采用的方法是利用带噪语音的高频段估计法,然后利用带噪语音信号 $y(n)$ 的短时过零率和短时能量组成的加权函数,从时域消除音乐噪声.

2.2.1 噪声谱估计

假设带噪语音的采样频率在 20KHz 以上,从文献[5]可知,带噪语音信号的噪声谱主要集中在高频区(大于 8KHz),在进行谱减法前,估计噪声谱,算法为

$$\mu = \text{average}[|Y_m(f)|] \quad 8\text{KHz} < f < F_s/2 \quad (7)$$

$$|D_m(\omega)| = \mu, \quad (8)$$

其中 $\omega = 2\pi f/F_s$. 利用(7)、(8) 式估计的噪声谱基本上反应了实时语音段噪声的非平稳性,谱减后估计的“干净”语音的质量有了很大提高.为了进一步提高语音质量,提出了利用时域特征消除音乐噪声的算法.

2.2.2 加权函数法

根据短时过零率和短时能量鉴别语音信号端点的特点,定义一个加权函数^[8]:

$$F_s(m) = \log(E_s(m)/Z_s(m)), \quad (9)$$

其中 $Z_s(m)$ 为过零率、 $E_s(m)$ 为短时能量, $F_s(m)$ 为加权函数.利用此函数,在实验分析的基础上,我们提出了消除音乐噪声的算法.

设最终得到的去噪后的语音信号为 $z(n)$,则

$$z(n) = \begin{cases} k_1 \hat{s}_m(n) & F_s(m) < \lambda_1; \\ k_2 \hat{s}_m(n) & \lambda_1 \leq F_s(m) < \lambda_2; \\ k_3 \hat{s}_m(n) & \lambda_2 \leq F_s(m) < \lambda_3; \\ k_4 \hat{s}_m(n) & \lambda_3 \leq F_s(m) \leq \lambda_4; \\ k_5 \hat{s}_m(n) & F_s(m) > \lambda_4. \end{cases} \quad (10)$$

其中 k_1, k_2, k_3, k_4, k_5 是加权系数, 经验取值分别是: 0.2, 0.65, 0.9, 1.3, 0.75. $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 是实验设定值, 初始值随意设置, 其后由函数 $F_s(m)$ 自动调

整, 其调整算法为:

$$\begin{aligned} \lambda_1 &= 0.12 * \max(F_s(m)) + 0.88 \min(F_s(m)); \\ \lambda_2 &= 0.4 * \max(F_s(m)) + 0.6 * \min(F_s(m)); \\ \lambda_3 &= 0.6 * \max(F_s(m)) + 0.4 * \min(F_s(m)); \\ \lambda_4 &= 0.88 * \max(F_s(m)) + 0.12 * \min(F_s(m)). \end{aligned} \quad (11)$$

本文的改进语音增强算法的实现总体框图如图 1 所示。

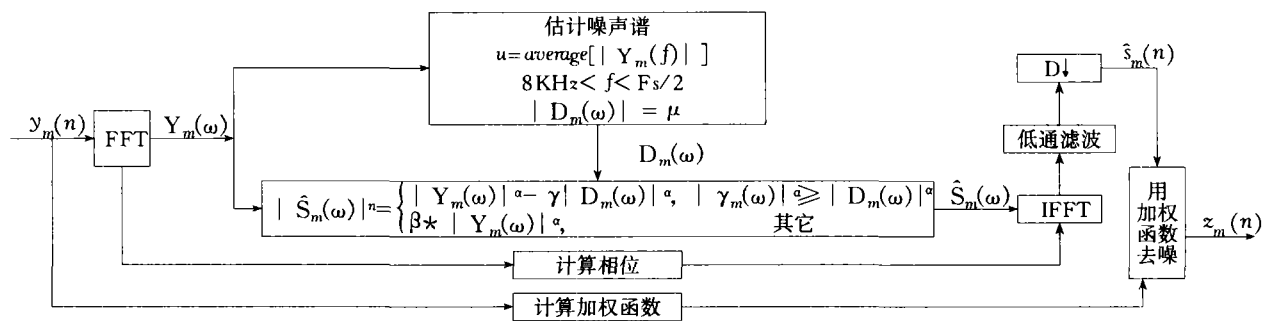


图 1 语音增强算法实现总框图

依图可知, 谱减完成后, 经 $IFFT$ 变化至时域, 然后低通滤波, 下采样, 使其采样频率降至正常语音信号的范围。

$$SNR_{improved} = SNR_{output} - SNR_{input}. \quad (13)$$

图 2 给出了平稳噪声环境下的实验结果, 图 3 是非平稳环境的实验结果。

3 实 验

为了评价上述语音增强算法的效果, 我们分别进行了在平稳和非平稳噪声环境下求增强前后语音的 SNR 的差值实验以及听觉实验。

本实验条件是: 采样频率为 22.05KHz, 加 Hamming 窗, 窗长为 512, 帧间重叠 50%, FFT 变换长度为 1024, 下采样 $D = 2$, 重构语音的采样频率为 11.025KHz, 带噪语音的 SNR 分别为 -5dB , 0dB 和 5dB . 实验分别是传统谱减法 (不含加权函数法), 即 SS (Spectral Subtraction) 法, 算法中的各参数取值为 $\alpha = 2$, $\gamma = 1$, $\beta = 0$; 改进增减法 (不含加权函数法), 即 ISS (Improved Spectral Subtraction) 法, 算法中的各参数取值为 $\alpha = 2$, $\gamma = 4$, $\beta = 0.075$; 以及含加权函数的改进谱减法, 本文称为扩展谱减法, 即 ESS (Extended Spectral Subtraction) 法, 算法中的各参数同 ISS 法. 三类实验中的噪声估计都采用本文 2.2.1 的方法。

评价指标 SNR 的定义如下:

$$SNR = 10 \log_{10} \sum_{m=1}^N \frac{s_m(k)^2}{d_m(k)^2}, \quad (12)$$

式中 N 为所分析的语音帧数, 语音增强前后的 SNR 差值定义为

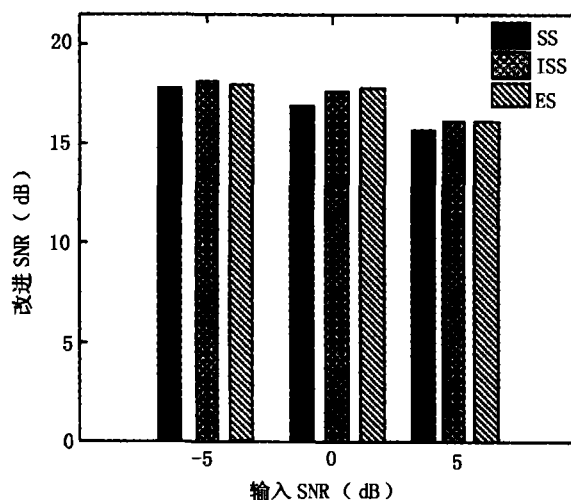


图 2 平稳噪声下 SNR 的改进

从图 2 可知, 在平稳噪声环境下, SS 、 ISS 、 ESS 对 SNR 的贡献差别不是很明显, 主要原因是增强算法中噪声谱的估计较为准确, 而从图 3 中可知, ISS 、 ESS 算法可获得较高的 SNR . 从图示实验结果可知, 与 ISS 法相比, ESS 算法对 SNR 的改进略有降低, 实验结果进一步说明了 ISS 和 ESS 算法的差别主要在改善语音的听觉质量上, 即消除音乐噪声上。

图 3、图 4 的实验结果说明了语音增强算法对

SNR 的贡献,但未能说明增强后语音质量的改善.为了进一步验证所提算法的有效性,我们进行了听觉实验.算法的实验条件同上,参加听觉实验的对象是 5 男 3 女,每个参加者分别对 SS、ISS 和 ESS 算法的增强效果按语音质量好(5 分)、良好(4 分)、较差(2.5 分)和差(1 分)打分,然后求其平均得分.实验结果如表 1 所示.

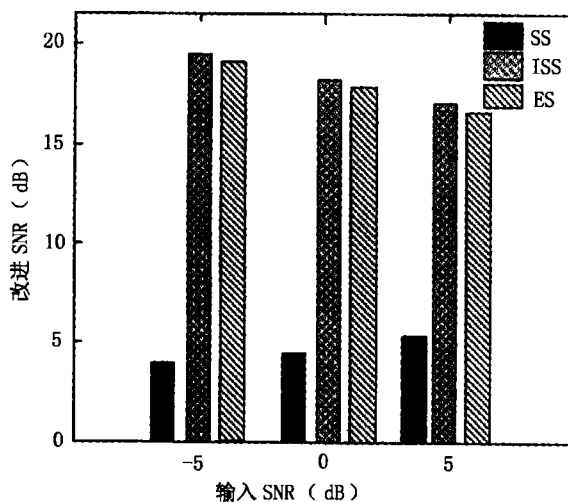


图 3 非平稳噪声下 SNR 的改进

表 1 语音质量评测结果

	输入语音(非平稳噪声)SNR(dB)		
	-5	0	5
SS	2.02	2.75	3.48
ISS	3.12	4.23	4.45
ESS	3.74	4.43	4.63

从听觉实验结果分析,当 SNR 较低时,效果不是很理想,但 ESS 法基本达到可听的效果;当 SNR 较高时 ESS 比 ISS 算法略有提高,说明所提算法对语音质量有一定的贡献.

4 讨论与小结

通过实验,我们认为:本文提出的语音增强算法对噪声背景下的语音增强有一定的效果.尤其对非平稳噪声污染的语音,其效果更为明显.算法的实时性强,计算复杂性低,便于硬件实现,适当调整算法中引入的系数因子,可获得较好的语音增强效果.

在本文研究的基础上,我们结合多带(Multi-Band)方法进一步研究在单通道系统中,如何有效

增强语音的方法,同时结合 DSP^[10]对本文算法的实现,研究所提算法中有关系数因子的自适应调整策略.文中算法在 MATLAB6.1 下实现.

参 考 文 献

- [1] Poruba J. The Study of Spectral Subtraction. <http://www.electronicletter.com/papers/2001/0021/paper.asp>
- [2] Boll S F. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Trans on Acoustics, Speech & Signal Processing*, 1979, 27(2): 113-120
- [3] Djigan V I, Sovka P, Cmejla R. Modified Spectral Subtraction Based Speech Enhancement. <http://amber.feld.cvut.cz/user/cmejla/papers/iwaenc.pdf>
- [4] Kamath J S D, Loizou P C. A Multi-Band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise. <http://www.utdallas.edu/~loizou/speech/mbss2002.pdf>
- [5] Yamauchi J, Shimamura T. Noise Estimation Using High Frequency Regions for Spectral Subtraction. *IEICE Trans Fundamentals*, 2002, E85-A(3): 723-727
- [6] Hartmann U, Hermansen K. Model Based Spectral Subtraction Used for Noise Suppression in Speech with Low SNR. http://www.es.isy.liu.se/norsig2000/publ/page129_id101.pdf
- [7] Ghoreish M H, Sheikzadeh H. A Hybrid Speech Enhancement System Based on HMM and Spectral Subtraction. In: *Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 2000, 1855-1858
- [8] Kotnik B, Kacic Z, Horvat B. A Computational Efficient Real Time Robust Speech Recognition Based on Improved Spectral Subtraction Method. In: *Proc of the 7th European Conference on Speech Communication and Technology*, Scandinavia, Denmark, 2001, 1123-1126
- [9] Bhatnagar M. A Modified Spectral Subtraction Method Combined with Perceptual Weighting for Speech Enhancement. <http://www.utdallas.edu/~loizou/thesis/mukul.ms.thesis.pdf>
- [10] Kuo S M, Huang Y C. Implementation of Optimized Spectral Subtraction Techniques on TMS320C5x and TMS320C3x. <http://www.icspat.com/papers/470mfi.pdf>
- [11] Yao K, Shi B E, et al. Residual Noise Compensation for Robust Speech Recognition in Nonstationary Noise. In: *Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 2000, 1125-1128
- [12] Cho Y D, Al-Naimi K, Kondoz A. Improved Voice Activity Detection Based on a Smoothed Statistical Likelihood Ratio. In: *Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, 2001, 675-678
- [13] Hu Y, Bhatnagar M, Loizou P C. A Cross-Correlation Technique for Enhancing Speech Corrupted with Correlated Noise. In: *Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, 2001, 1939-1942

A METHOD FOR SPEECH ENHANCEMENT BASED ON IMPROVED SPECTRAL SUBTRACTION

Wang Rangding

(*AI Lab, Tongji University, Shanghai 200092*)

(*CKC Software Lab, Ningbo University, Ningbo 315211*)

Chai Peiqi

(*AI Lab, Tongji University, Shanghai 200092*)

ABSTRACT

The spectral subtraction method is the most suitable technique for the speech enhancement. It is of utmost its low computation complexity, real-time capability, and easy implementation. The idea of spectral subtraction is to obtain an estimate of the speech spectral level by subtracting the noise estimation from noisy speech. In this case, the phase of the speech is not processed. The main objective of speech enhancement is to reduce noise and to improve one or more perceptual aspects of speech, such as the speech quality.

In this paper, a new speech enhancement method is presented, which uses spectral and time domain processing and achieves a trade-off between effective and low computational load for real-time operations. In the frequency domain, high frequency components of a noisy speech spectrum are averaged and the spectrum of noise is estimated. In the time domain, a weighting function is used to reduce a residual "musical noise". Experiment results show the proposed method reduces greatly musical noise, and improves efficiently speech quality.

Key Words Speech Enhancement, Noise Spectrum, Spectral Subtraction, Weighting Function