

Speech Enhancement Algorithm Combining Speech Absence Probability

Ruirui Han, Ying Gao, Chen Chen

School of Opto-Electronic Information, Yantai University, Yantai Shandong
Email: 506470634@qq.com

Received: Jul. 1st, 2018; accepted: Jul. 18th, 2018; published: Jul. 30th, 2018

Abstract

The research work of this paper is mainly on the basis of the amplitude squared spectrum least mean square estimator and proposes a new algorithm. Due to the uncertainty of the speech in the statistical model of noisy speech, the unified processing of speech signals will inevitably result in the loss of speech components, which will affect the performance of speech enhancement. Therefore, this paper mainly studies and estimates the frequency of each signal. The speech probability is then combined with the gain function of the squared spectrum least mean square error algorithm to derive a new gain function. Finally, we can see through the experimental simulation, the algorithm proposed in this paper can significantly improve the voice quality and improve the intelligibility of the voice.

Keywords

Speech Enhancement, Speech Absence Probability, Minimum Mean-Squared Error, Gain Function

基于无语音概率的语音增强算法

韩蕊蕊, 高颖, 陈晨

烟台大学, 光电信息科学技术学院, 山东 烟台
Email: 506470634@qq.com

收稿日期: 2018年7月1日; 录用日期: 2018年7月18日; 发布日期: 2018年7月30日

摘要

本文的研究工作主要是在幅度平方谱最小均方估计器的基础上提出了一种新的算法。由于带噪语音的统计模型中语音存在不确定性, 统一对语音信号进行处理必然会造成语音成分的丢失, 从而影响语音增强

的性能,因此,本文主要研究和估计出每个信号频点的无语音概率,然后与幅度平方谱最小均方差算法的增益函数相结合,推导出一个全新的增益函数。最后通过实验仿真可以看出,本文提出的算法能够明显的改善语音质量,提高语音的可懂度。

关键词

语音增强,无语音概率,最小均方误差,增益函数

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

语音通信和人机交互是现代通信中重要的通信方式。但是,在语音通信和语音识别的过程中,语音信号无法免除地会受到外界噪声的干扰,所以阻碍了人与人之间,人与机器之间的交流。因此从带噪声语音中提取出纯净语音信号是语音增强技术的目的所在。单信道语音增强是语音信号处理的关键分支,其应用技术的研究具有重要的适用价值,尤其在语音识别,医疗领域,军事通信,数字家电等领域已被广泛应用[1]。为了提高移动环境下的通信品质,实现实时通信,需要在传输到远端之前有效抑制背景噪声[2]。虽然语音增强技术看似只是一个恢复纯净语音的简易过程,但在学术领域涉及到的众多技术和方法是不容小觑的。随着人们对接收语音信号的质量要求不断提高,越来越多的语音增强算法被提出。其中比较经典和具有代表性的算法有谱减算法,维纳滤波算法,最小均方误差算法,小波变换算法等等[3]。由于基于谱估计统计模型的语音增强算法中的模型能够很好的适应信号的变化,因此这种算法受到了广大学者的青睐。基于谱估计统计模型算法的优点是适用范围广,易于理解,实施起来比较简单,且对于各种背景噪声抑制效果较好。

其中比较经典的谱估计统计模型算法是幅度平方谱最小均方估计算法,该算法在低先验信噪比和高后验信噪比的情况下能够有效地抑制背景噪声,减少语音失真。但是,基于谱估计统计模型的幅度平方谱最小均方算法并没有将语音信号的语音存在不确定性考虑在内[4],因此会不可避免的造成语音成分的丢失,影响了语音增强的效果。因此本文针对幅度平方谱最小均方估计器算法中没有考虑到语音信号存在不确定度的缺点提出了一种新的算法融合无语音概率的算法。本算法是在幅度平方谱最小均方估计器的基础上,推导出每一个频点的无语音概率,进而将幅度平方谱最小估计器的增益函数与无语音概率相结合成一个新的增益函数。

本文的结构如下:第二部分简单描述了基于统计模型的语音增强算法基本理论,第三部分主要是对提出的新的算法——融合无语音概率的语音增强算法进行理论分析和公式推导,第四部分通过实验仿真的语谱图,以及四种客观评价标准:分段信噪比(Segmental SNR, SegSNR),短时客观可懂度(Short Time Objective Intelligibility, STOI),感知语音质量(Perceptual Evaluation of Speech, PESQ)和对数谱距离(Log-Spectral Distortion, LSD)结果进行定量分析,进一步比较了两种算法的优缺点,最后做出总结。

2. 语音增强算法基本理论

基于统计模型的语音增强算法,经过傅里叶变换后其纯净语音信号以及噪声语音信号通常都假设服从高斯分布,基于此模型 Loizou 等人提出了一种语音增强算法幅度平方谱最小均方误差估计器算法

(MMSE-MSS), 该算法假设带噪语音信号的频谱幅度的平方等于纯净语音频谱幅度的平方加上噪声频谱幅度的平方, 实际上传统的谱减算法以及谱估计统计模型都是采用这种假设。但是上述假设是在统计意义上成立的, 即假设 $X(k)$ 和 $D(k)$ 是两个不相关的随机变量, 根据上述假设可得[2] [3]

$$Y_k^2 = X_k^2 + D_k^2 \quad (1)$$

计算条件最小均方差由

$$\hat{X}_k^2 = E\{X_k^2 | Y_k^2\} = \int_0^{Y_k^2} X_k^2 f(X_k^2 | Y_k^2) dX_k^2 \quad (2)$$

其中

$$f(X_k^2 | Y_k^2) = \frac{f(Y_k^2 | X_k^2) f(X_k^2)}{f(Y_k^2)} = \begin{cases} \Psi_k \exp\left(-\frac{X_k^2}{\lambda(k)}\right) & \sigma_x^2 \neq \sigma_d^2 \\ \frac{1}{Y_k^2} & \sigma_x^2 = \sigma_d^2 \end{cases} \quad (3)$$

其中, 当 $\sigma_x^2 \neq \sigma_d^2$ 时,

$$\frac{1}{\lambda(k)} = \frac{1}{\sigma_x^2(k)} - \frac{1}{\sigma_x^2(k)} \quad (4)$$

$$\Psi_k = \frac{1}{\lambda(k) \left\{ 1 - \exp\left[-\frac{Y_k^2}{\lambda(k)}\right] \right\}} \quad (5)$$

很明显从可以看出(5)式之中为一正数。将式(3)带入式(2)经过计算可得

$$X_k^2 = \begin{cases} \left(\frac{1}{v_k} - \frac{1}{\exp(v_k) - 1} \right) Y_k^2, & \sigma_x^2 \neq \sigma_d^2 \\ \frac{1}{2} Y_k^2, & \sigma_x^2 = \sigma_d^2 \end{cases} \quad (6)$$

其中, v_k 定义为 $v_k = \frac{1 - \xi_k}{\xi_k} \gamma_k$ 。

因此式(6)可得幅度平方谱最小均方差估计器的增益函数为

$$G_{\text{MMSE-MSS}} = \begin{cases} \sqrt{\frac{1}{v_k} - \frac{1}{\exp(v_k) - 1}}, & \sigma_x^2 \neq \sigma_d^2 \\ \sqrt{\frac{1}{2}}, & \sigma_x^2 = \sigma_d^2 \end{cases} \quad (7)$$

上述算法在很大程度上消除了背景噪音并提高了语音质量, 不幸的是, 该算法由于没有将语音存在概率考虑在内, 因此导致了增强所需的语音成分损失严重影响了人类所感知的语音可懂度的下降。

3. 本文提出的融合无语音概率的语音增强算法

上述 Liozou 等人提出的算法并没有考虑到带噪语音中存在语音不确定度的问题, 因此本节主要是研究语音存在的不确定问题, 将无语音概率估计出来融合到增益函数中[4], 能够在一定环境中明显的改善语音质量, 提高语音可懂度。本文假设纯净语音信号和噪声语音信号统计独立, 且它们经过傅里叶变换

后的系数服从高斯分布，因此，它们的系数的平方服从指数分布[5]，即：

$$f_{X_k^2}(X_k^2) = \frac{1}{\sigma_x^2} \exp\left[-\frac{X_k^2}{\sigma_x^2}\right] \quad (8)$$

$$f_{D_k^2}(D_k^2) = \frac{1}{\sigma_d^2} \exp\left[-\frac{D_k^2}{\sigma_d^2}\right] \quad (9)$$

因此带噪语音信号 Y_k^2 的傅里叶变换系数则服从如下分布[6]：

$$f_{Y_k^2}(Y_k^2) = \frac{1}{\sigma_x^2 - \sigma_d^2} \left(\exp\left(-\frac{Y_k^2}{\sigma_x^2}\right) - \exp\left(-\frac{Y_k^2}{\sigma_d^2}\right) \right) \quad (10)$$

其中， k 代表频点， Y_k^2 ， X_k^2 ， D_k^2 分别是带噪语音频谱，纯净语音频谱以及噪声频谱，由于语音存在的不确定，将语音分为有语音和无语音两个状态，表示如下：

$$\begin{cases} H_0^k : Y_k^2 = D_k^2 \\ H_1^k : Y_k^2 = X_k^2 + D_k^2 \end{cases} \quad (11)$$

其中， H_0^k, H_1^k 分别表示无语音段和有语音段，则 $P(\cdot)$ 和 $P(\cdot|\cdot)$ 分别代表概率密度函数和条件概率密度函数。

即

$$P(Y_k^2 | H_1^k) = \frac{1}{\sigma_x^2 - \sigma_d^2} \left(\exp\left(-\frac{Y_k^2}{\sigma_x^2}\right) - \exp\left(-\frac{Y_k^2}{\sigma_d^2}\right) \right) \quad (12)$$

$$P(Y_k^2 | H_0^k) = \frac{1}{\sigma_d^2} \exp\left(-\frac{Y_k^2}{\sigma_d^2}\right) \quad (13)$$

利用贝叶斯公式可得：

$$P(H_1^k | Y_k^2) = \frac{P(Y_k^2 | H_1^k) P(H_1^k)}{P(Y_k^2 | H_0^k) P(H_0^k) + P(Y_k^2 | H_1^k) P(H_1^k)} = \frac{\Lambda(Y_k^2)}{1 + \Lambda(Y_k^2)} = G \quad (14)$$

其中， $P(H_1^k) = q, P(H_0^k) = 1 - q$ ，代表无语音概率， ξ_k 是先验信噪比用经典直接判决算法(DD)计算得出， γ_k 代表后验信噪比可由带噪语音功率比上噪声功率得到。

因此将无语音概率引入到幅度平方谱最小均方误差估计器中得到新的增益函数即：

$$G_{\text{new}} = G_{\text{MMSE-MSS}} \cdot \sqrt{G} \quad (15)$$

即融合了无语音概率后的幅度平方谱最小均方误差估计器为：

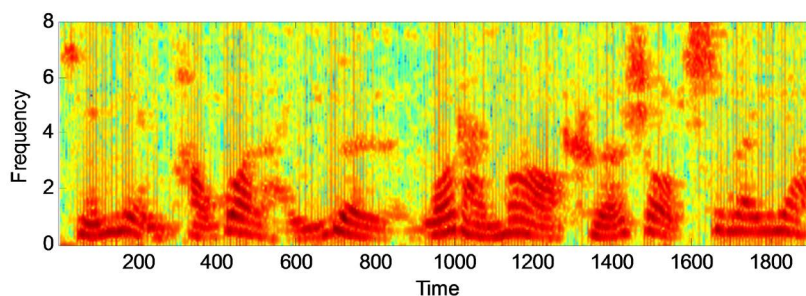
$$\hat{X}_k^2 = G_{\text{new}} \cdot Y_k^2 \quad (16)$$

4. 仿真实验结果分析

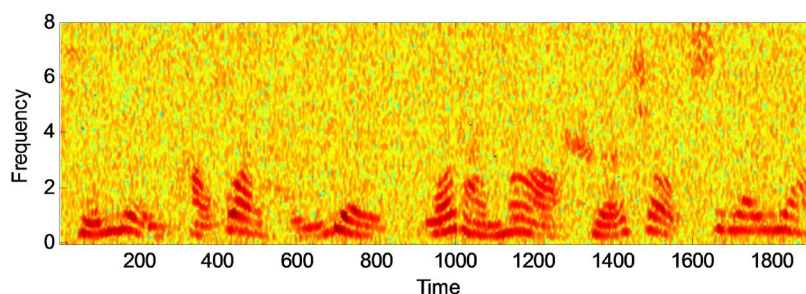
为了对比幅度平方谱最小均方误差算法与本文算法的性能，利用 MATLAB 为仿真实验平台，以此评比算法的优劣。实验数据主要来自于语音库 10 段纯净语音信号(5 段男声，5 段女声)作为测试。噪声来自于噪声库比较有代表性的 White 噪声，输入信噪比分别取 5 dB，10 dB，15 dB。噪声功率谱估计采用最小值控制递归平均算法(MCRA) [7]。所有信号的取样频率均为 8 kHz/s，仿真实验中语音信号帧长 $K = 256$ ，窗函数采用正交汉明窗，参数 q 取 0.8 [8]。

通过图 1 语音增强后语谱图的对比可以看出, 两种算法都能够有效的去除背景噪声, 但是本文提出的算法对去除背景噪声能力要优于幅度平方谱最小均方估计算法, 且语音失真程度方面本文算法失真较小, 具体的去噪能力, 失真程度, 以及语音可懂度性能方面还要从上述表格数据中对比分析。通过对数谱距离(LSD), 分段信噪比(segSNR), 语音的可懂度(STOI)以及感知语音质量(PESQ)等四种评价标准进行对两种算法增强后的语音性能进行评估[9] [10], 其中 LSD 表示增强后的语音信号与纯净语音之间的接近程度, 其值越小, 说明增强后的语音信号越贴近原始纯净语音信号, 其失真程度越小, 增强效果越显著。其中 segSNR 是表征带噪语音信号抑制噪声性能优劣的重要标准, 数值越大, 表明去除背景噪声的能力越强, 增强效果越好。STOI 和 PESQ 都是符合人类听觉系统的语音质量评价标准, 其数值越大, 代表增强后的语音可懂度越高。

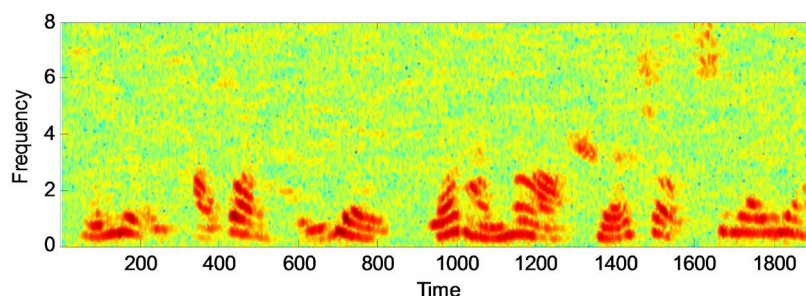
从表 1 中, 我们可以看出, 本文提出的算法在 white 输入信噪比分别为 5 dB, 10 dB, 15 dB 时, 其本文提出的算法其 LSD 数值更小, 说明其对原始纯净语音信号的语音失真比较小, 因此本文算法在语音失真方面有了明显改善。从表格的数据中可以看出, 本文提出的算法其 SegSNR 的数值要明显大于幅度平方谱最小均方估计算法, 说明本文提出的算法其抑制噪声的能力有了明显提升。进一步从语音质量和语音可懂度方面分析, 其本文提出的算法 STOI 值和 PESQ 值更大, 说明本文提出的算法对语音质量和语音可懂度方面有了明显提高。总体来说, 语音增强性能要明显优于 MMSE-MSS 算法。



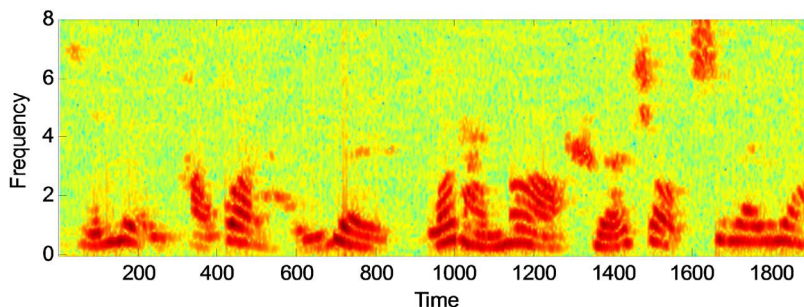
(a) 纯净语音信号语谱图



(b) 带噪语音信号语谱图



(c) 基于 MMSE-MSS 算法的增强语音语谱图



(d) 基于本文算法的增强语音信号语谱图

Figure 1. The spectrum of speech signal of different algorithms under white noise. (SNR = 10 dB)

图 1. 白色噪声下不同算法的语音信号语谱图(SNR = 10 dB)

Table 1. The data comparison table of the two algorithms

表 1. 两种算法的 LSD 数据对比表

客观评价测度	输入信噪比	算法	
		MMSE-MSS算法	本文算法
LSD	5 dB	5.7190	5.6961
	10 dB	5.8803	4.7070
	15 dB	6.0238	4.2512
SegSNR	5 dB	6.1954	6.5140
	10 dB	6.5744	7.3483
	15 dB	7.6732	8.3765
STOI	5 dB	0.6625	0.6770
	10 dB	8.7320	8.7570
	15 dB	9.0451	9.1620
Pesq	5 dB	1.6917	1.7928
	10 dB	1.9712	2.2366
	15 dB	2.0039	2.6425

5. 小结

本文主要研究了语音增强算法中基于谱估计模型中的语音存在不确定度问题，介绍了经典的幅度平方谱最小均方误差算法，以及概述了该算法在实现过程中没有对语音信号是否存在进行检测的缺点并对其缺点进行改进，提出了一种新的语音增强算法——通过融入语言存在概率的估计来去除背景噪声，改善语音质量，提高语音可懂度。最后通过使用 Matlab 软件进行实验仿真，仿真结果表明本文提出的算法在输入信噪比较高的情况下，其增强后的语音性能明显优于幅度平方谱最小均方估计算法。在输入信噪比较低的情况下，此算法也有相对来说较高的优越性。所以未来的研究工作是让本文提出的融合无语音概率的算法在任何噪声环境下增强后的语音性能都比较优越。

参考文献

[1] Lu, Y. and Loizou, P.C. (2011) Estimators of the Magnitude-Squared Spectrum and Methods for Incorporating SNR Uncertainty. *IEEE Transactions on Audio, Speech, and Language Processing*, **19**, 1123-1137.

- <https://doi.org/10.1109/TASL.2010.2082531>
- [2] Diethom, E.J. (2000) Subband Noise Reduction Methods for Speech Enhancement. *Acoustic Signal Processing for Telecommunication*. Springer US, New York, 155-178.
- [3] Xia, B. and Bao, C. (2014) Wiener Filtering Based Speech Enhancement with Weighted Denoising Auto-Encoder and Noise Classification. *Speech Communication*, **60**, 13-29. <https://doi.org/10.1016/j.specom.2014.02.001>
- [4] 卢志强. 基于新型先验信噪比估计的语音增强算法的对比研究[D]: [硕士学位论文]. 长沙: 湖南大学.
- [5] Papoulis, A. and Pillai, S.U. (2002) *Probability, Random Variables and Stochastic Processes with Errata Sheet*. McGraw-Hill Education, New York, 31.
- [6] Huan, Z. (2014) A New Soft Masking Method for Speech Enhancement in the Frequency Domain. *Elektronika ir Elektrotechnika*, **20**, 1392-1215.
- [7] Cohen, I. (2003) Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging. *IEEE Transaction on Speech and Audio Processing*, **11**, 466-475. <https://doi.org/10.1109/TSA.2003.811544>
- [8] Ephraim, Y. and Malah, D. (1984) Speech Enhancement Using a Minimum-Mean Squared Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **32**, 1109-1121. <https://doi.org/10.1109/TASSP.1984.1164453>
- [9] Taal, C.H., Hendriks, R.C., Heusdens, R., et al. (2011) An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech. *IEEE Transactions on Audio Speech & Language Processing*, **19**, 2125-2136. <https://doi.org/10.1109/TASL.2011.2114881>
- [10] Abramson, A. and Cohen, I. (2010) Simultaneous Detection and Estimation Approach for Speech Enhancement. *IEEE Transactions on Audio Speech & Language Processing*, **15**, 2348-2359. <https://doi.org/10.1109/TASL.2007.904231>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-3983, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: hjwc@hanspub.org