

## 基于谱减法的藏语语音减噪研究\*

# Study on Noise Reduction of Tibetan Speech based on Spectral Subtraction

都格草,才让卓玛\*\*,南措吉

DOU Gecao, CAIRANG Zhuoma, NAN Cuoji

(青海师范大学计算机学院,青海西宁 810008)

(Computer School of Qinghai Normal University, Xining, Qinghai, 810008, China)

**摘要:**语音信息的采集过程中难免会有各种噪声的干扰,噪声降低了语音的可懂度及清晰度,甚至严重影响语音处理的准确性和可靠性。谱减法是语音减噪中最常用的方法,该方法基于人的感觉特性,对语音短时幅度谱进行估计,适用于受加性噪声污染的语音。本文通过统计分析谱减法的增益补偿因子和过减因子参数,研究基于谱减法的藏语语音减噪处理,并取得较为理想的实验结果。

**关键词:**语音信号处理 藏语语音分析 谱减法 语音减噪

**中图分类号:**TP391.1 **文献标识码:**A **文章编号:**1002-7378(2018)01-0078-05

**Abstract:** In the process of the acquisition of speech information, there will inevitably be all kinds of noise interference. The noise reduces the intelligibility and quality of speech and even seriously affects the accuracy of speech processing. Spectral subtraction is the most commonly used method for the noise reduction of speech, which estimates the short-time amplitude spectrum of the speech based on the characteristics of human consciousness. It is suitable for speech that is contaminated by additive noise pollution. In this paper, statistical analysis of spectral subtraction gain compensation factor and reduction factor parameter is used to study the noise reduction processing of Tibetan language speech based on spectral subtraction, and an ideal experimental result is obtained.

**Key words:** speech signal processing, Tibetan speech analysis, spectral subtraction, speech noise reduction

收稿日期:2017-11-23

修回日期:2017-12-21

作者简介:都格草(1992—),女,硕士研究生,主要从事藏文信息处理、藏语语音处理研究。

\* 国家自然科学基金项目(61262051,61163018),国家社科基金项目(16BYY167,15BYY132,14BYY132,13BYY141),教育部“春计划”(Z2016077)和青海省基础科研项目(2017-ZJ-767)资助。

\*\* 通信作者:才让卓玛(1970—),女,教授,硕士生导师,主要从事藏文信息处理、藏语语音处理研究,E-mail:cr-zhuoma@163.com。

## 0 引言

语音从人的口腔中发出时周围往往伴有各种各样的噪声,例如,马路上的汽车噪声、工厂和工地的机械轰鸣声,房屋中的空调声、风扇声、日光灯的蜂鸣声、打字声、谈话声和小孩的哭闹声等。在语音信息的采集过程中,难免会受到这些噪声的干扰。噪声不仅降低了语音的可懂度和清晰度,甚至严重地影响着语音处理的准确性。因此,减噪处理是语音处理的关键技术之一。

目前语音减噪方法主要有小波变换法<sup>[1]</sup>、

MMSE 法<sup>[2]</sup>、信号子空间法<sup>[3]</sup>、自适应滤波器降噪<sup>[4]</sup>、维纳滤波法<sup>[5]</sup>和谱减法<sup>[6]</sup>等。其中, 自适应滤波器降噪、维纳滤波法降噪和谱减法是使用最广泛的降噪算法。自适应滤波器降噪法有较快的收敛性和平稳性, 但计算量大, 需要不断的逐个点更新, 瞬时跟踪能力较弱; 维纳滤波法的适应面较广, 但不能用于噪声为非平稳的随机过程, 对于向量情况应用不方便; 相对于前两种方法, 由美国 Utah 大学的 Boll<sup>[7]</sup>提出的谱减法由于具有引入约束条件少和运算量小的特点, 而被广泛采用。谱减法的研究起源于改进噪声中的线性预测 (Linear Prediction, LP) 分析, 而今已被发展成估计噪声频谱并进行降噪的经典算法。随着降噪研究的不断深入, 人们陆续提出了一些谱减法的变形形式, 其中 1982 年由 Thomson 提出的多窗谱估计的谱减法<sup>[8]</sup>, 对 Boll 的谱减法进行了重要的改进<sup>[6]</sup>使得多窗谱估计的谱减法成为当前降噪处理的研究热点。因此, 本文通过分析谱减法及其改进方法中的过减因子和增益补偿因子, 研究了基于谱减法的藏语语音的降噪方法。

## 1 谱减法

谱减法主要利用在无声期间统计得到的噪声方差来代替当前帧的噪声频谱, 其基本原理如图 1 所示:

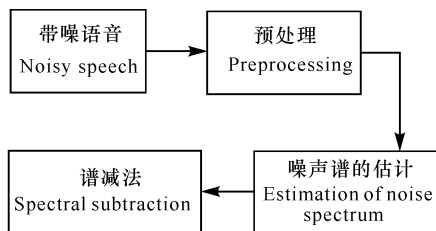


图 1 谱减法基本原理

Fig. 1 The basic principle of spectral subtraction

### 1.1 基本谱减法

设语音信号的时间序列为  $x(n)$ , 加窗分帧处理后得到第  $i$  帧语音信号为  $x_i(m)$ , 帧长为  $N$ 。任何一帧语音信号  $x_i(m)$  做 DFT (离散傅里叶变换) 后为

$$x_i(k) = \sum_{m=0}^{N-1} X_i(m) \exp(j \frac{2\pi mk}{N}) \quad k=0, 1, \dots, N-1. \quad (1)$$

要对  $x_i(k)$  求出每个分量的幅值和相角, 幅值是  $|x_i(k)|$ , 它的相角是

$$x_{\text{angle}}^i(k) = \arctan \left[ \frac{\text{Im}(x_i(k))}{\text{Re}(x_i(k))} \right]. \quad (2)$$

已知前导噪声段时长为  $IS$ , 对应的帧数为  $NIS$ , 可以

求出该噪声段的平均能量和谱减算法为

$$D(k) = \frac{1}{NIS} \sum_{i=1}^{NIS} |x_i(k)|^2. \quad (3)$$

$$\begin{aligned} & |\hat{x}_i(k)|^2 = \\ & \begin{cases} |x_i(k)|^2 - \alpha \times D(k) & |x_i(k)|^2 \geq \alpha \times D(k) \\ \beta \times D(k) & |x_i(k)|^2 < \alpha \times D(k) \end{cases} \end{aligned} \quad (4)$$

谱减算法中  $\alpha$  和  $\beta$  是两个常数,  $\alpha$  称为过减因子;  $\beta$  称为增益补偿因子。求出了谱减后幅值为  $|\hat{x}_i(k)|^2$  和相角  $X_{\text{angle}}(k)$  结合就能经快速傅里叶逆变换 (IFFT) 求出谱减后的语音序列  $\hat{x}_i(m)$ 。

### 1.2 多窗谱减法

传统的周期图法只用一个数据窗, 而 Thomson 提出的多窗谱估计的谱减法对同一数据序列用多个正交的数据窗分别求直接谱, 其次求平均得到谱估计, 因此可以得到较小的估计方差。Thomson 使用的数据窗是一组相互正交的离散球序列 (Discrete Prolate Spheroidal Sequences, DPSS), 也叫 Slepian 窗。多窗谱是一种比周期图法更准确的谱估计方法。多窗谱定义如下:

$$S^m(\omega) = \frac{1}{L} \sum_{k=0}^{L-1} S^m_k(\omega), \quad (5)$$

其中  $L$  为数据窗个数;  $S^m_k(\omega)$  为第  $k$  个数据窗的谱:

$$S^m_k(\omega) = \left| \sum_{n=0}^{N-1} a_k(n) x(n) e^{-jm\omega} \right|^2, \quad (6)$$

其中  $x(n)$  为数据序列;  $N$  为序列长度;  $a_k(n)$  为第  $k$  个数据窗, 满足多个数据窗之间相互正交:

$$\begin{aligned} \sum a_k(n) a_j(n) &= 0 \quad k \neq j, \\ \sum a_k(n) a_j(n) &= 1 \quad k = j. \end{aligned} \quad (7)$$

## 2 基于谱减法的藏语降噪算法

多窗谱估计中的过减因子和增益补偿因子两个参数对整个算法有着重要影响, 过减因子  $\alpha$  主要影响语音谱的失真参数, 增益补偿因子  $\beta$  可以控制残留噪声的多少以及音乐噪声的大小<sup>[9]</sup>。藏语言文字是一种以辅音字母和元音字母为构件的拼音文字, 以音节为单位, 一般一个字为一个音节, 各音节间用音节符“·”分隔。本文为了从字、词、句及篇章等各个层次上研究考察藏语语音的降噪效果, 将藏语语音降噪算法设计如下 (图 2):

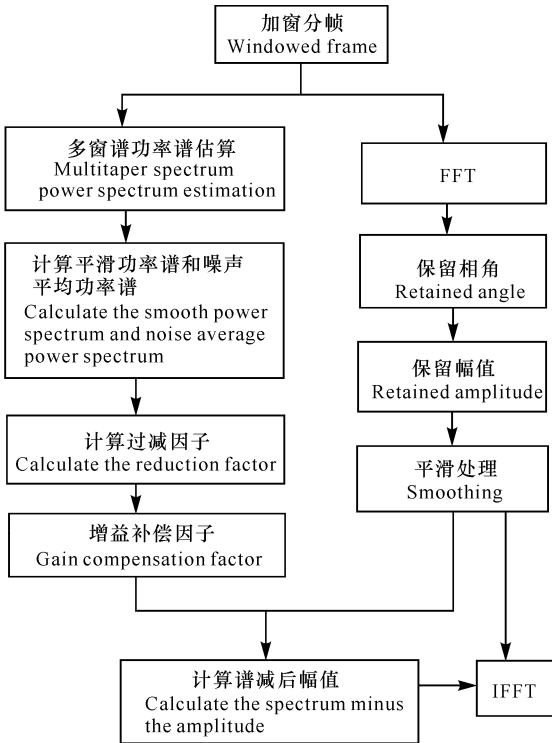


图2 基于谱减法的藏语语音减噪算法流程

Fig. 2 The flow chart of Tibetan speech noise reduction algorithm based on spectral subtraction

基于谱减法的藏语减噪算法具体步骤如下:

步骤 1: 读入语音文件, 消除直流分量, 幅值做归一化。

步骤 2: 设置初始噪声比后生成叠加高斯白噪声的带噪语音, 计算叠加噪后的信噪比。

步骤 3: 设置参数, 比如前导无话段长度、帧长、帧移、过减因子和增益补偿因子。

步骤 4: 计算多窗谱改进谱减法:

(1) 设置窗后开始分帧, 求帧数, 本实验中同样使用了汉明窗。

(2) 每帧语音信号计算快速傅里叶变换(FFT)。

(3) 求出正频率部分的幅值和相位角。

(4) 相邻帧之间做平滑处理, 然后计算平均幅值谱。

(5) 多窗谱法对每一帧数据进行功率谱估计, 对功率谱的相邻帧之间进行平滑处理。

(6) 前导无话段中求出噪声平均功率谱, 利用谱减关系计算出增益因子。

(7) 用增益补偿因子计算谱减后幅值, 幅值经快速傅里叶逆变换(IFT), 求出谱减后语音序列。

步骤 5: 计算谱减后的信噪比, 输出信噪比与信噪增益, 合成减噪后语音。

### 3 结果与分析

本文实验语料均来自藏语网站, 从大到小分为 4 个层次, 即包含 631 音节的篇章、平均包含 11 个音节的 10 个句子、20 个多音节的词和短语、100 单

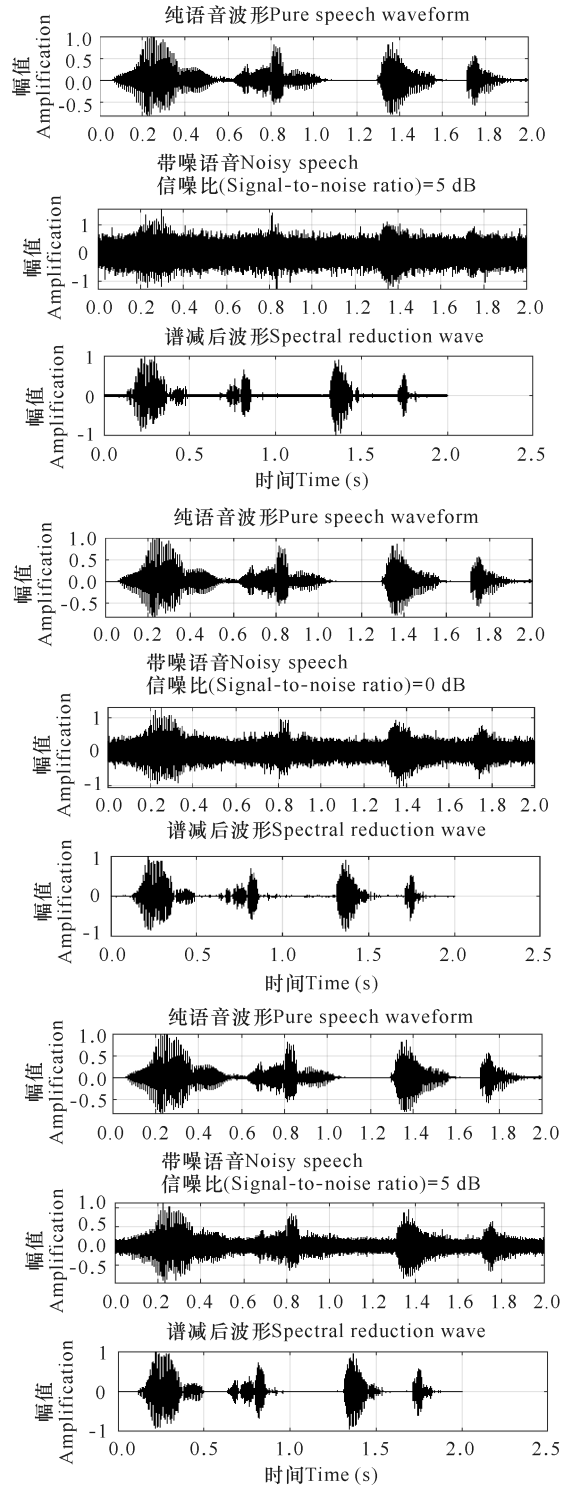


图3 藏语多音节使用多窗谱估计的改进谱减法效果

Fig. 3 Improved spectral subtraction effect of multisyllable using multitaper spectrum estimation in Tibetan

音节。语音的信噪比分别为-5 dB、0 dB、5 dB,采样率为8 000 Hz、16 bit量化的纯净语音信号,分别叠加高斯白噪声。

### 3.1 主观评测

算法对-5 dB、0 dB、5 dB带噪语音处理后的结果、纯净语音以及谱减后的效果如图3所示。不同

表1 MOS值

Table 1 MOS value

语料 Corpus	MOS值 MOS value					平均值 Average
	测听员 1 Interrogator 1	测听员 2 Interrogator 2	测听员 3 Interrogator 3	测听员 4 Interrogator 4	测听员 5 Interrogator 5	
篇章 Chapter	2.5	2.8	2.4	2.6	2.4	2.54
句子 Sentence	2.6	2.9	2.5	2.7	2.6	2.66
多音节 Multi-syllable	2.5	2.9	2.5	2.8	2.5	2.64
单音节 Monosyllable	2.4	2.7	2.4	2.5	2.4	2.48

### 3.2 客观评测

适当地选择过减因子 $\alpha$ 和增益补偿因子 $\beta$ ,可以有效除掉噪声或控制残留噪声以及音乐噪声的大小。反之将导致语音失真<sup>[9]</sup>。通过实验发现,藏语语音减噪中, $\alpha$ 取值为1.8~3.0, $\beta$ 取值为0.001时较为合适。当过减因子 $\alpha$ 取1.7或者1.8时信噪比最高, $\alpha$ 的值小于1.7或大于1.8时信噪比逐步减小; $\beta=0.001$ 时信噪比最高, $\beta$ 的值小于0.001或大于0.001时同样信噪比逐步减小。同时,语料越大,输出的信噪比就越提高。相对来说,句子和多音节的减噪效果优于篇章和单音节的减噪效果。具体实验数据见表2~4。

表2 不同信噪比下调整过减因子

Table 2 Adjusting the reduction factor under different signal to noise ratio

输入过减因子 Enter the reduction factor $\alpha$	改进谱减法的测算结果 The calculation results of the improved spectral subtraction		
	-5*	0*	5*
1	2.099 8	5.975 2	9.604 5
1.6	4.294 8	7.485 1	9.569 2
1.7	4.475 2	7.694 8	10.027 4
1.8	5.069 5	7.587 2	9.375 0
1.9	4.915 4	7.520 3	9.391 6
2.4	4.820 1	7.354 0	8.574 2
3	4.671 1	6.594 1	7.895 5

注: \* 为输入的信噪比(dB)

Note: \* means signal to noise ratio of input (dB)

信噪比时多窗谱减后的波形和纯净语音波形大同小异(图3)。由6位测听员对4个语料在信噪比为0 dB下的减噪效果进行了测听,Mean Opinion Score (MOS)分值如表1所示。句子减噪效果最好,其次是多音节,说明句子和多音节的减噪效果优于篇章和单音节的减噪效果(表1)。

表3 不同信噪比下调整增益补偿因子

Table 3 Adjusting the gain compensation factor under different signal to noise ratio

输入增益补偿因子 Input gain compensation factor $\beta$	改进谱减法的测算结果 Improved spectral subtraction results		
	-5*	0*	5*
0	5.398 4	4.349 8	7.134 5
0.000 01	5.358 4	4.636 6	7.238 4
0.001	6.013 3	6.812 2	7.405 6
0.01	5.124 9	6.719 7	7.414 0
0.1	2.532 5	5.461 1	7.669 5
1	-3.662 6	-0.326 3	3.282 3

注: \* 为输入的信噪比(dB)

Note: \* means signal to noise ratio of input (dB)

表4 相同信噪比下4部分藏语语料测算结果

Table 4 The calculation results of four parts of Tibetan language data under the same signal to noise ratio

语料 Corpus	测算结果 Calculation results		
	$\alpha=1.8,$ $\beta=0.001$	$\alpha=2,$ $\beta=0.001$	$\alpha=3,$ $\beta=0.001$
篇章 Chapter	11.320 0	12.033 3	12.117 3
句子 Sentence	7.865 2	7.689 8	6.513 0
多音节 Multi-syllable	7.670 4	6.890 0	5.497 0
单音节 Monosyllable	5.560 8	5.406 0	4.406 3

注:输入的带噪语音信噪比为0 dB

Note: The input noisy speech signal to noise ratio is 0 dB

## 4 结论

过减因子和增益补偿因子是决定谱减法减噪效果的重要参数,本文通过统计分析藏语语音在不同信噪比下的过减因子 $\alpha$ 和增益补偿因子 $\beta$ ,研究了基于谱减法的藏语减噪算法,得出较为理想的 $\alpha$ 和 $\beta$ 的取值范围,并发现句子和多音节的减噪效果优

于篇章和单音节的减噪效果。

#### 参考文献:

- [1] 姚学锋,吴震,金观昌. 数字散斑相关法中的小波减噪分析[J]. 清华大学学报:自然科学版,2001,41(4/5):108-111.  
YAO X F, WU Z, JIN G C. Wavelet denoising using the digital speckle correlation method[J]. J Tsinghua Univ:Sci & Tech,2001,41(4/5):108-111.
- [2] EPHRAIM Y, MALAH D. Speech enhancement using a mini-mum mean-square error short-time spectral amplitude estimator[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(6):1109-1121.
- [3] EPHRAIM Y, VAN TREES H L. A signal subspace approach for speech enhancement[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1995, 3(4):251-266.
- [4] 宋知用. MATLAB在语音信号分析与语音合成中的应用[M]. 北京:北京航空航天大学出版社,2013:184-194.  
SONG Z Y, The application of MATLAB in speech signal analysis and synthesis applications[M]. Beijing: Beihang University Press,2013:184-194.
- [5] 张青,吴进. 基于多窗谱估计的改进维纳滤波语音增强[J]. 计算机应用与软件,2017,34(3):67-70,118.  
ZHANG Q, WU J. Improved wiener filter speech enhancement based on multi-taper spectrum estimation [J]. Computer Applications and Software, 2017, 34(3):67-70,118.
- [6] 武鹏鹏,赵刚,邹明. 基于多窗谱估计的改进谱减法[J]. 现代电子技术,2008(12):150-152.  
WU P P, ZHAO G, ZOU M. An improved spectral subtraction method based on multitaper estimation[J]. Modern Electronic Technology,2008(12):150-152.
- [7] BOLL S. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1979, 27(2):113-120.
- [8] HU Y, LOIZOU P C. Speech enhancement based on wavelet thresholding the multitaper spectrum [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing,2004,12(1):59-67.
- [9] [投稿]谱减法语音降噪原理[EB/OL]. [2017-0-11]. <http://blog.csdn.net/leixiaohua1020/article/details/47276353>.  
[Submission] The principle of noise reduction in spectral subtraction[EB/OL]. [2017-0-11]. <http://blog.csdn.net/leixiaohua1020/article/details/47276353>.

(责任编辑:米慧芝)