

# 基于分数阶付立叶变换的谱减法及其语音增强应用

马多佳<sup>1</sup>, 谢湘<sup>2</sup>, 匡镜明<sup>3</sup>

(1. 北京理工大学 电子工程系, 北京 100081)

**文 摘:** 利用分数阶付立叶变换 (FrFT) 在时频分析上的特点, 提出一种新的谱减法-分数阶谱减法。首先, 对每一帧带噪语音进行的分数阶付立叶变换, 通过使得分数阶谱减法的信噪比达到最大从而确定最佳变换阶数。其中, 分数阶谱减法关键的噪声分数阶功率谱的估计采用最小统计估计, 带噪语音的分数阶功率谱减去噪声分数阶功率谱以得到增强的语音分数阶功率谱, 最后, 通过逆分数阶付立叶变换 (IFrFT) 得到增强后的语音信号。在实验中验证了这种新方法在信噪比和 Itakura-Saito LPC 距离上都优于传统的谱减法。

**关键词:** 噪声消除; 语音增强; 分数阶付立叶变换; 谱减法

**中图分类号:** TN912

谱减法作为经典语音增强方法已经发展出多种形式。总体来说, 谱减法就是从带噪语音功率谱中减去噪声功率谱得到干净的语音, 因此, 简单易于实现是谱减法的最大的优势。但是传统的谱减法由于采用付立叶变换而带来的信号失真又无法避免, 另外, 噪声功率谱的估计也是制约传统谱减法性能的主要方面, 一般的方法都是采用无声段进行噪声功率谱估计<sup>[2]</sup>。我们采用最小统计对噪声功率谱进行逐帧估计从而避免了判别有声-无声段的问题。

一般说来, 时域表示和频域表示是信号分析处理里面两种基本的物理表述。付立叶变换几乎是所有信号处理的基础: 它提供了信号时域与频域间的映射关系, 而信号在频域的表示可以揭示出信号一些重要的特性。但是付立叶变换是将信号作为平稳过程处理, 因此并不十分适合处理频率特性随着时间变化的信号。近些年来, 分数阶付立叶变换已经成功地应用于许多领域, 例如光学信号处理以及声纳信号处理等方面<sup>[3]</sup>。考虑到语音不仅是时变的而且也是频变的, 因此, 分数阶付立叶变换也同样适用于语音信号处理。

在论文的第一部分中, 我们将讨论分数阶付立叶变换的概念及其实现方法, 接下来描述了分数阶付立叶变换通过选取最佳变换阶数与传统谱减法的结合, 最后通过对传统谱减法和分数阶谱减法在信噪比和谱失真这两方面的比较得出对所采用的方法效果的评价。

## 1 关键算法描述

### 1.1 分数阶付立叶变换的基本概念

在许多情况下, 我们注意到信号混合后无论是

在时间域还是通过付立叶变换以频域表示都会有明显的重叠, 而在分数阶付立叶域在适当的变换角度下, 这种重叠可以减少甚至没有。因此分数阶付立叶变换作为一种新兴的时频分析方法在处理混叠信号方面有着其特有的优势<sup>[6-7]</sup>。

信号  $x(t)$  的分数阶付立叶变换表示为:

$$X_{\alpha}(u) = F_{\alpha}[x(t)] = \int_{-\infty}^{\infty} x(t) K_{\alpha}(t, u) dt \quad (1)$$

这里, 实数  $\alpha$  为分数阶付立叶变换阶数, 取值区间为  $[-2, 2]$ , 通过  $F_{\alpha+4l}[x(t)] = F_{\alpha}[x(t)]$  可以将变换阶数扩展至任意实数;  $F_{\alpha}[\bullet]$  为分数阶付立叶变换算子;  $K_{\alpha}(t, u)$  是分数阶付立叶变换的核函数:

$$K_{\alpha}(t, u) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}} \exp\left(j\frac{t^2+u^2}{2}\cot\alpha - j(u-t)\csc\alpha\right), & \alpha \neq n \\ \delta(t-u), & \alpha = 2n \\ \delta(t+u), & \alpha = 2n+1 \end{cases} \quad (2)$$

核函数  $K_{\alpha}(t, u)$  具有下面属性:

$$K_{-\alpha}(t, u) = K_{\alpha}^{*}(t, u) \quad (3)$$

$$\int_{-\infty}^{\infty} K_{\alpha}(t, u) K_{\alpha}^{*}(t, u') dt = \delta(u - u') \quad (4)$$

因此, 分数阶付立叶变换的逆变换为:

$$x(t) = F_{-\alpha}[X_{\alpha}(u)] = \int_{-\infty}^{\infty} X_{\alpha}(u) K_{-\alpha}(t, u) du \quad (5)$$

### 1.2 分数阶付立叶变换的快速算法

在实际工程应用中, 需要快速算法来实现分数阶付立叶变换, 因此, 不断有人提出各种对分数阶付立叶变换的快速算法。根据 H.M.Ozaktas 等提出

快速算法的约束条件, 各种快速算法必须满足下面这些分数阶变换的基本性质<sup>[5]</sup>:

- 幺正性, 即:  $(F_\alpha)^H = F_{-\alpha}$ .
- 旋转可加性, 即:  $F_{\alpha_1} F_{\alpha_2} = F_{\alpha_1 + \alpha_2}$ .
- $F_1 = F$ ,  $F$  是 DFT 算子.
- 变换阶数的连续性
- 低复杂度

本文所采用的是文献<sup>[4]</sup>中描述的快速算法:

$$X_\alpha = F^\alpha x \quad (6)$$

$$F^\alpha = DK_\beta J \quad (7)$$

这里

$$F^\alpha = \begin{cases} DK_\beta J, & \text{对于 } 0.5 \leq \beta \leq 1.5 \\ DK_\beta JF, & \text{对于 } |\beta| \leq 0.5 \text{ 或 } 0.5 \leq \beta \leq 2 \end{cases} \quad (8)$$

式中,  $\beta$  是  $\alpha$  模 4 的值, 取值范围为  $[-2, 2]$ ;  $F$  是常规 DFT 矩阵。

## 2 基于最小统计估计方法的分数阶谱减法

分数阶谱减法的基础是基于对于语音信号的短时分数阶付立叶变换分析与传统的短时付立叶变换分析相比可以得到更好的谱图分辨率。

图 1 为分数阶谱减法的实现框图。分数阶谱减的算法与传统谱减法的主要不同点在于: 由在频率域减去噪声幅值改为在分数阶域的相减。假设  $x(i) = s(i) + n(i)$ , 其中  $s(i)$  为未失真的干净语音信号,  $n(i)$  为噪声信号, 这里我们假设噪声和语音是统计不相关的。因此就有  $E\{x^2(i)\} = E\{s^2(i)\} + E\{n^2(i)\}$ 。由于分数阶域谱图的性质和频谱的性质在幅度可加上有相似性, 可以得到:

$$X_\alpha(u_k) = S_\alpha(u_k) + N_\alpha(u_k) \quad (9)$$

这里,  $X_\alpha(u_k)$ ,  $S_\alpha(u_k)$  and  $N_\alpha(u_k)$  分别

$x(i)$ 、 $s(i)$  和  $n(i)$  的离散分数阶付立叶变换,

其中  $n(i)$  可能为非平稳的噪声序列。

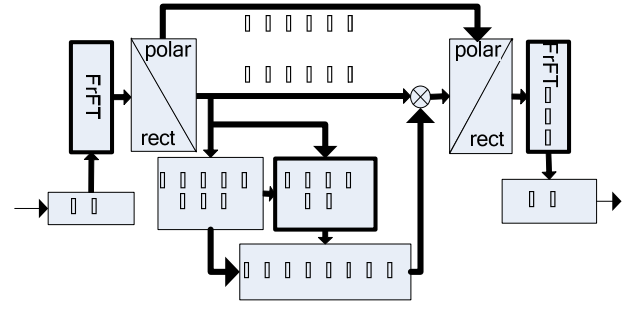


图 1 分数阶谱减法的实现框图

### 2.1 分数阶谱减法的描述

谱减法的基本算法以及噪声的最小统计估计方法都才有 R.Martin 的论文<sup>[1]</sup>所描述的方法。

首先, 对带噪声语音信号进行平滑:

$$|X_\alpha(\lambda, k)|^2 = \gamma |X_\alpha(\lambda-1, k)|^2 + (1-\gamma) |X_\alpha(\lambda, k)|^2 \quad (10)$$

式中  $\gamma$ ——为平滑因子, 取为 0.82

$|X_\alpha(\lambda, k)|^2$ ——带噪语音  $\alpha$  阶分数阶付立叶变换的能量

$|X_\alpha(\lambda-1, k)|^2$ ——带噪语音平滑后的  $\alpha$  阶分数阶付立叶变换的能量。

然后, 利用回归平滑周期谱计算子带信号  $\alpha$  阶分数阶付立叶变换的能量:

$$P_{x_\alpha}(\lambda, k) = \beta P_{x_\alpha}(\lambda-1, k) + (1-\beta) |X_\alpha(\lambda, k)|^2 \quad (11)$$

式中  $\beta$ ——平滑常量, 取值为 0.9~0.95。

通过

$$P_{n_\alpha}(\lambda, k) = \text{omin} \bullet P_{\min_\alpha}(\lambda, k) \quad (12)$$

对噪声能量进行估计, 即为一定窗长范围内的最小分数阶能量的加权,  $\text{omin}$  即权重变量。

### 2.2 分数阶变换阶数的选取

最佳变换阶数可以通过各种最佳准则进行选取, 按照不同的应用主要有最小均方误差准则(MMSE)、最大信噪比和最小方差准则等。每种判断准则都有其优点和局限性<sup>[8]</sup>。如果信号的特性是已知的, 那么最佳变换阶数  $\alpha$  就可以提前确定。同样的, 可以对信号的全局进行分析后得到全局最佳变换阶数也可以对加窗的信号进行分析得到局部最佳变换阶数。虽然不断调整最佳变换阶数可能需要的计算量比较大, 但是同时它也不需要先验知识的优点。

图 2 所表示的是变换阶数  $\alpha$  与分数阶信噪比  $SNR_\alpha$  的关系。这里定义分数阶信噪比为:

$$SNR_\alpha(\lambda) = \frac{E[P_{x_\alpha}(\lambda, k) - \min(P_{n_\alpha}(\lambda, k), P_{x_\alpha}(\lambda, k))]}{E[P_{n_\alpha}(\lambda, k)]} \quad (13)$$

对每一帧的语音数据进行求取  $SNR_\alpha$ , 一般情

况下变换阶数为接近 1（传统付立叶变换）的值，又由于分数阶变换的对称性<sup>[7]</sup>，为了减少计算量因此可以只计算变换阶数 $\alpha$ 取值为 0.9~1 的  $SNR_{\alpha}$ 。变换阶数 $\alpha$ 就选择使得  $SNR_{\alpha}$  最大的值。

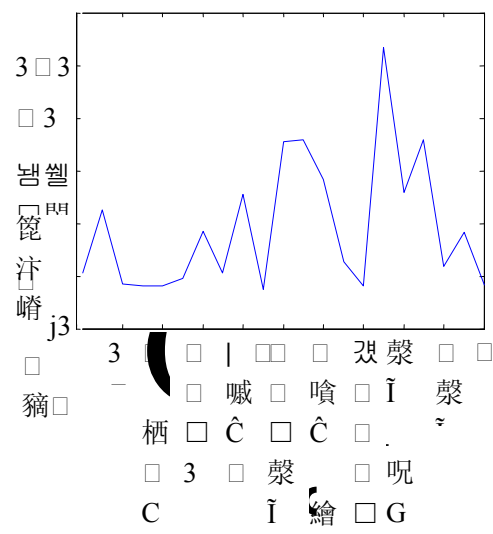


图 2 变换阶数 $\alpha$ 与分数阶信噪比  $SNR_{\alpha}$  的关系举例（背景噪声为咖啡馆噪声信噪比为-5db，横轴为变换阶数 $\alpha$ ，纵轴为信噪比  $SNR_{\alpha}$ ）

### 3 实验结果及分析

实验采用男声发音，发音内容为汉语数字，8kHz 采样，帧长 256 采样点，窗移为 128 采样点。分数阶付立叶变换谱减法与传统谱减法在 SNR 的比较通过输入信噪比和输出信噪比进行，定义如下：

$$SNR_{in} = 10 \log_{10} \frac{\sum_{m=1}^M s(m)^2}{\sum_{m=1}^M n(m)^2} \quad (14)$$

$$SNR_{out} = 10 \log_{10} \frac{\sum_{m=1}^M s(m)^2}{\sum_{m=1}^M [\hat{s}(m) - s(m)]^2} \quad (15)$$

$s(m)$  和  $\hat{s}(m)$  分别表示干净的语音信号和经过谱减法处理过的语音信号。 $n(m)$  为噪声信号，实验中平稳噪声采用高斯白噪声 Gaussian White 以及 Hoth 谱噪声非平稳噪声采用了 cafeteria 噪声，输入噪声的范围从 0db 至 20db。

图 3 和图 4 分别从时域波形和频域语谱图上对传统谱减法和分数阶付立叶变换谱减法进行了比较。可以看到在输入信噪比较低的情况下，分数阶付立叶变换谱减法的效果明显比传统付立叶变换有改进。

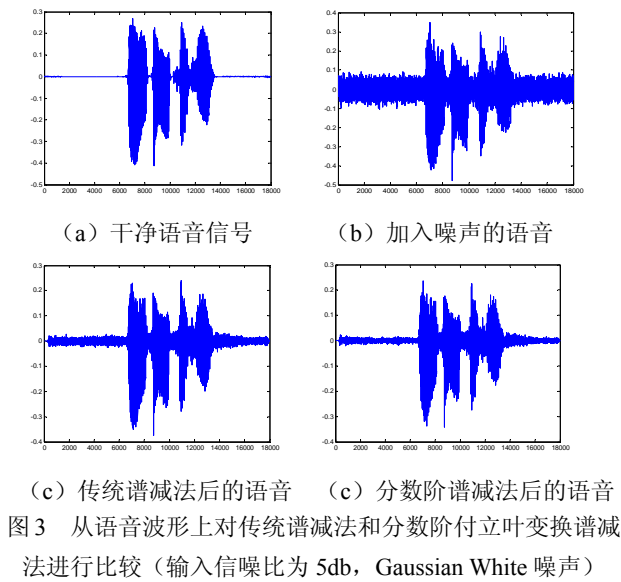


图 3 从语音波形上对传统谱减法和分数阶付立叶变换谱减法进行比较（输入信噪比为 5db，Gaussian White 噪声）

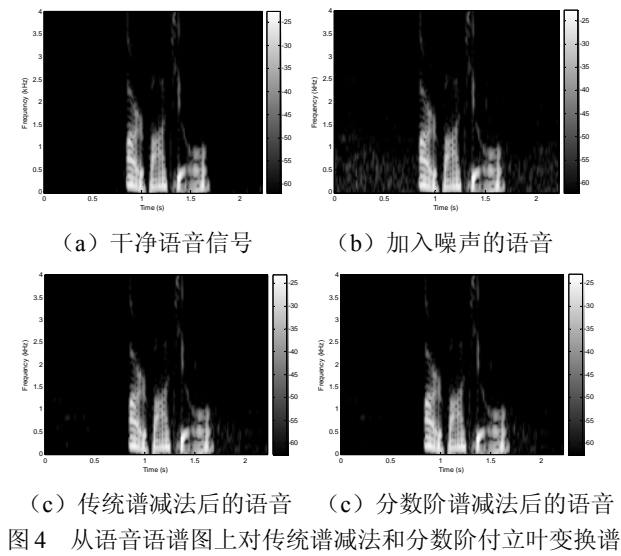


图 4 从语音语谱图上对传统谱减法和分数阶付立叶变换谱减法进行比较（输入信噪比为 20db，cafeteria 噪声）

由于所采用的传统谱减法和分数阶谱减法都采用最小统计估计的方法进行噪声能量估计，因此都不需进行语音激活检测。

表 1 为传统谱减法和所提出的分数阶谱减法的信噪比方面的比较。

表 1 传统谱减法和分数阶谱减法的信噪比比较表			
输入噪声 (dB)	噪声类型	$SNR_{out}$ (dB)	
		传统谱减法	分数阶谱减法
0	Gaussian White	4.29	5.35
	Hoth noise	2.60	3.80
	cafeteria	2.70	3.64
5	Gaussian White	5.53	6.74
	Hoth noise	8.99	9.13
	cafeteria	9.41	9.54
10	Gaussian White	15.89	16.00
	Hoth noise	13.08	13.17
	cafeteria	13.47	13.57
15	Gaussian White	20.01	20.11

输入 噪声 (dB)	噪声类型	$SNR_{out}$ (dB)	
		传统谱减 法	分数阶谱减 法
20	Hoth noise	17.37	17.44
	cafeteria	17.74	17.81
	Gaussian White	24.09	24.15
	Hoth noise	21.78	21.87
	cafeteria	22.13	22.22

由于传统谱减法主要的缺点就是对信号造成了谱失真, 因此选择 Itakura-Saito 距离对传统谱减法和分数阶谱减法进行比较, 这里选择了 LPC 参数的 Itakura-Saito 距离 (ISD)。表 2 就对传统谱减法和所提出的分数阶谱减法在谱失真方面进行了比较。可以看到在输入信噪相同的情况下, 分数阶付立叶变换谱减法的效果明显比传统付立叶变换有改进。

表 2 传统谱减法和分数阶谱减法的 ISD 比较表

输入 噪声 (dB)	噪声类型 /ISD	ISD	
		传统谱减 法	分数阶谱减 法
0	Gaussian White	16.11	10.37
	/19.0467		
	Hoth noise	6.27	3.99
	/5.7686		
5	Cafeteria	5.32	3.78
	/4.1726		
	Gaussian White	6.09	4.29
	/13.9717		
10	Hoth noise	2.46	1.70
	/4.034)		
	cafeteria	2.30	1.83
	/3.0063		
10	Gaussian White	3.17	2.2
	/9.4813		
	Hoth noise	1.09	0.89
	/2.5113		
	cafeteria	0.95	0.83
	/1.9472		

#### 4 结论及以后的工作

我们比较了传统谱减法和分数阶谱减法对语

音增强的效果。在对噪声没有先验知识的情况下, 无论对平稳噪声还是非平稳噪声, 所提出的具有自适应变换阶数的分数阶付立叶变换的谱减法在信噪比和保持谱特性方面都表现出明显的优势。

但是分数阶谱减法在进行自适应选取最佳变换阶数时的计算量比较大, 这作为一个可以考虑改进的方向仍然需要进一步的工作。

#### 5 课题背景

该课题由国家自然科学基金 Grant NSFC 60605015 支持。

#### 参 考 文 献

- [1] R. Martin, "Spectral subtraction based on minimum statistics", in Proc. Eur. Signal Processing Conf., 1182-1185, 1994.
- [2] Zhenli. Wang, Xiongwei. Zhang, "On the application of fractional Fourier transform for enhancing noisy speech", in MAPE 2005, 289-292, 2005.
- [3] Soo-Chang Pei, Jian-Jiun Ding, "Relations between fractional operations and time-frequency distributions, and their applications", IEEE Trans. Signal Processing 1638-1655, 2001.
- [4] H.M.Ozaktas, O. Arikan, M. A. Kutay and G. Bozdog, "Digital computation of the fractional Fourier transforms", in IEEE Trans. Signal Processing, 44, 2141-2150, 1996.
- [5] C.Candan, M.A.Kutay, and H.M.Ozaktas. The discrete fractional Fourier transform. IEEE Trans. Signal Processing, vol.48, No.5, 1329-1337, May. 2000.
- [6] L.B. Almeida. "The fractional Fourier transform and time-frequency representation" in IEEE Trans. Signal Processing, 44, 2141-2150, 1994.
- [7] 陶然, 齐林, 王越, 《分数阶 Fourier 变换的原理与应用》清华大学出版社, 2004
- [8] Rajiv Saxena, Kulbir Singh. "Fractional Fourier transform: A novel tool for signal processing" in J. Indian Inst. Sci., Jan.-Feb. 2005, 85, 11-26