

基于实值离散 Gabor 变换的联合时频域语音增强

周健^{1,2} 赵力¹ 陶亮² 金赞¹

(1. 东南大学水声信号处理教育部重点实验室, 南京, 210096;

2. 安徽大学计算机科学与技术学院, 合肥, 230031)

摘要: 本文提出一种基于实值离散 Gabor 变换的联合时频域语音增强方法。算法采用具有最佳能量聚集特性的高斯函数作为综合窗, 采用双正交法计算其对应的分析窗函数, 利用能快速实现的实值离散 Gabor 变换 (RDGT) 将语音信号变换到联合时频域, 然后利用语音和噪声谱分别服从高斯分布的假设和无语音概率的思想进行基于最小均方误差的语音对数谱估计, 采用改进的最小受控递归平均算法 (IMCRA) 进行噪声时频谱估计, 在得到纯净语音的谱估计后利用实值离散 Gabor 逆变换获得纯净语音估计。实验表明, 该算法相比传统频域变换算法具有较好的语音去噪度和较低的语音失真度。

关键词: 实值离散 Gabor 变换; 联合时频域; 语音增强

中图分类号: TN912.34 **文献标识码:** A **文章编号:** 1003-0530(2010)12-1870-07

Speech Enhancement in joint time-frequency domain based on Real-Valued Discrete Gabor Transform

ZHOU Jian^{1,2} ZHAO Li¹ TAO Liang² JIN Yun¹

(1. Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education Southeast University,

Nanjing, China, 210096; 2. School of Computer Science and Technology Anhui University, Heifei, China, 230031)

Abstract: In this paper, we propose a novel speech enhancement method in the joint time-frequency domain. It is motivated by the fact that speech is actually a non-stationary signal and the joint time-frequency is the most important method when processing non-stationary signal. Noisy speech is first transformed to the joint time-frequency by fast Real-Valued Discrete Gabor Transform (RDGT) in which the Gaussian window is used as the transform kernel function since they assure the best utilization of the information area in the sense that they possess the smallest product of effective duration by effective width. The Minimum Mean-Square Error (MMSE) based log-amplitude estimator of speech is derived under speech presence uncertainty hypothesis as well as the assumption that speech and noise data are statistically independent Gaussian random variables. The noise spectral is estimated by Improved Minima Controlled Recursive Averaging (IMCRA) algorithm. The clean speech estimate is got by inverse transform of RDGT. Experimental results show that the proposed method is very effective in avoiding the musical residual noise and retaining weak speech components while retaining lower computation complexity.

Key words: Real-Valued Discrete Gabor Transform (RDGT); joint time-frequency domain; speech enhancement

1 引言

语音增强的目的在于从被噪声污染的语音中尽可能提取纯净的语音。语音增强技术在语音识别, 音

频编码及人机交互接口等应用领域都起着重要作用^[1]。当前的语音增强研究主要集中在如何更准确的估计噪声谱, 研究者们提出了很多改进方法且取得了一定的效果^[2-5]。然而, 这些变换域语音增强方法

收稿日期: 2009年11月27日; 修回日期: 2010年4月9日

基金项目: 本论文获国家自然科学基金(编号: 51075068, 61071169, 60975017, 60472058)以及安徽省教育厅优秀青年基金(编号: 2010SQRL018)、安徽大学青年科学研究基金(编号: 2009QN027B)资助。

在计算频谱时均采用短时傅里叶变换,而有关基于其他联合时频分析方法的语音增强算法的报道并不多见。

语音信号本质上是非平稳信号,Gabor 联合时频分析是处理非平稳信号的重要方法之一。本文提出一种基于实值离散 Gabor 变换 (RDGT) 的联合时频域语音增强算法,RDGT 中采用的高斯窗函数具有时宽带宽积最小的特性,相比其他窗函数能更好的描述信号的局部时频特征,对语音信号分析更加细腻,从而更好的分离语音和噪声谱。算法中 RDGT 正变换和反变换分别采用的两个窗函数满足双正交关系,使得重建信号与原信号误差更小。另外,在 RDGT 中,可以根据过抽样率参数控制频谱的粗细程度。在本文中,探讨了过抽样率参数取值大小对语音增强效果的影响;为了评价本文提出算法的有效性及其增强性能,我们将本文算法与其他经典算法进行了算法复杂性比较,并且在不同的噪声环境和不同原始信噪比条件下分别采用本文算法和其他算法进行语音增强实验,实验同时采用客观评价与主观评价相结合的方式衡量本文算法的性能。实验结果表明,本文提出的联合时频域增强算法相比传统频域方法在多种噪声背景下都获得了较好的语音增强性能。

2 实值离散 Gabor 变换 (RDGT)

设 $x(k)$ 表示一个周期为 L 的有限长实序列,实值离散 Gabor 展开定义为^[6]:

$$x(k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} a(m,n) \tilde{h}_{m,n}(k) \quad (1)$$

展开系数 $a(m,n)$ 可从下式获得:

$$a(m,n) = \sum_{k=0}^{L-1} x(k) \tilde{\gamma}_{m,n}(k) \quad (2)$$

上式中

$$\tilde{h}_{m,n}(k) = \tilde{h}(k-m\bar{N}) \text{cas}(2nk/N) \quad (3)$$

$$\tilde{\gamma}_{m,n}(k) = \tilde{\gamma}(k-m\bar{N}) \text{cas}(2nk/N) \quad (4)$$

其中 $L = \bar{N}M = N\bar{M}$, $\text{cas}(2nk/N) = \cos(2nk/N) + \sin(2nk/N)$, M 和 N 分别为时域和频域中的抽样点数, \bar{M} 和 \bar{N} 分别为频域和时域的抽样间隔。 $\tilde{h}(k)$ 、 $\tilde{\gamma}(k)$ 分别是综合窗 $h(k)$ 、分析窗 $\gamma(k)$ 的周期延伸,即

$$\tilde{h}(k) = \sum_i h(k+iL) = \tilde{h}(k+L) \quad (5)$$

$$\tilde{\gamma}(k) = \sum_i \gamma(k+iL) = \tilde{\gamma}(k+L) \quad (6)$$

另外,设 $\tilde{h}(k)$ 具有单位能量,即

$$\sum_{k=0}^{L-1} |\tilde{h}(k)|^2 = 1 \quad (7)$$

(1)和(2)式分别称为实值离散 Gabor 展开和变换。

定义 $\beta = MN/L$ 为过抽样率。 $\beta=1$ 时称为临界抽样

(即: $\bar{N}\bar{M} = NM = L$, $a(m,n)$ 的个数等于 $x(k)$ 的样点数)。稳定的重建条件(完备性条件,即由系数能完全重建原信号)必须满足 $\beta > 1$, 此时的抽样称为过抽样。在欠抽样条件($\beta < 1$)下会丢失信息。

在 RDGT 中,系数 $a(m,n)$ 是实数并且是以 m 和 n 为变量的周期函数,即

$$a(m+iM, n+jN) = a(m,n) \quad i, j = 0, \pm 1, \pm 2, \pm 3, \dots \quad (8)$$

$\tilde{h}(k)$ 与 $\tilde{\gamma}(k)$ 满足如下双正交条件式:

$$\sum_{k=0}^{L-1} \tilde{h}(k+mN) \text{cas} \frac{2nk}{N} \tilde{\gamma}(k) = \frac{L}{NM} \delta(m)\delta(n) \quad (9)$$

其中, $0 \leq m \leq \bar{M}-1$, $0 \leq n \leq \bar{N}-1$, $\delta(m)$, $\delta(n)$ 表示 Kronecker delta 函数。

可以证明 $a(m,n)$ 和复值离散 Gabor 变换系数 $c(m,n)$ 之间有如下关系:

$$\text{Re}[c(m,n)] = \frac{a(m,n) + a(m, N-n)}{2} \quad (10)$$

$$\text{Im}[c(m,n)] = -\frac{a(m,n) - a(m, N-n)}{2} \quad (11)$$

实值离散 Gabor 变换不需要进行复数运算且可以采用快速 Hartley 变换实现,从而明显减少了 Gabor 变换的计算量^[7]。

3 过抽样条件下语音基于 Gabor 变换的时频谱最优估计

设被噪声污染的语音信号为 $y(n) = x(n) + d(n)$, 其中 $x(n)$ 和 $d(n)$ 分别代表纯净语音和噪声信号,假设纯净语音和噪声不相关。设 $Y_r(k,l)$, $X_r(k,l)$ 和 $D_r(k,l)$ 分别为根据(2)式获得的对应 $y(n)$, $x(n)$ 和 $d(n)$ 的实值离散 Gabor 变换系数,定义:

$$Y(k, l) = \sqrt{\frac{1}{2} [Y_r(k, l)^2 + Y_r(k, N-l)^2]} \quad (12)$$

为带噪声语音信号的第 k 时刻第 l 个频率点的联合时频谱, 根据噪声与语音的独立性假设, 有下式成立:

$$Y(k, l) = X(k, l) + D(k, l) \quad (13)$$

其中,

$$X(k, l) = \sqrt{\frac{1}{2} [X_r(k, l)^2 + X_r(k, N-l)^2]},$$

$$D(k, l) = \sqrt{\frac{1}{2} [D_r(k, l)^2 + D_r(k, N-l)^2]}$$

$X(k, l)$, $D(k, l)$ 分别定义为纯净语音和噪声第 k 时刻第 l 个频率点的时频谱。

根据基于 MMSE 的语音最优对数谱幅度估计的定义:

$$\tilde{X}(k, l) = \arg \min_{\tilde{X}} \{ E \{ [\log X(k, l) - \log \tilde{X}(k, l)]^2 | Y(k, l) \} \}$$

可得纯净语音谱估计为^[8]:

$$\tilde{X}(k, l) = \exp(E[\log X(k, l) | Y(k, l)]) \quad (14)$$

假设语音和噪声均服从高斯分布, 并且设 $H_0(k, l)$, $H_1(k, l)$ 分别代表第 k 时刻第 l 个频率点不出现和出现语音信号, 则根据上述假设可得如下语音信号条件概率密度函数^[9]:

$$P(Y(k, l) | H_0(k, l)) = \frac{1}{d_a(k, l)} \exp \left\{ -\frac{|Y(k, l)|^2}{d_a(k, l)} \right\} \quad (15)$$

$$P(Y(k, l) | H_1(k, l)) = \frac{1}{(d_a(k, l) + d_x(k, l))} \exp \left\{ -\frac{|Y(k, l)|^2}{d_a(k, l) + d_x(k, l)} \right\} \quad (16)$$

其中 $d_x(k, l) = E\{X(k, l)^2 | H_1(k, l)\}$, $d_a(k, l) = E\{D(k, l)^2\}$ 分别代表纯净语音和噪声的方差。 $\gamma(k, l) = \frac{|Y(k, l)|^2}{d_a(k, l)}$

和 $\xi(k, l) = \frac{d_x(k, l)}{d_a(k, l)}$ 分别代表语音出现的后验信噪比和先验信噪比, 后验信噪比的条件概率密度可以改写为如下形式:

$$f(\gamma(k, l) | H_0(k, l)) = e^{-\gamma(k, l)} \mu(\gamma(k, l)) \quad (17)$$

$$f(\gamma(k, l) | H_1(k, l)) = \frac{1}{1 + \xi(k, l)} \exp \left\{ -\frac{\gamma(k, l)}{1 + \xi(k, l)} \right\} \mu(\gamma(k, l)) \quad (18)$$

式(17), (18)中 $\mu(\cdot)$ 为单位阶跃函数。令 $p(k, l) = P(H_1(k, l) | \gamma(k, l))$, 根据贝叶斯准则有:

$$p(k, l) = \frac{1}{1 + \frac{q(k, l)}{1 - q(k, l)} (1 + \xi(k, l))} \exp(-\nu(k, l)) \quad (19)$$

其中 $q(k, l) = P(H_0(k, l))$ 代表无语音出现的概率, $\nu = \frac{\gamma(k, l) \xi(k, l)}{1 + \xi(k, l)}$ 。

根据上述语音出现概率假设和(14)式, 可以求得如下纯净语音最优谱增益^[10]:

$$G(k, l) = \{ G_{H_1}(k, l) \}^{p(k, l)} G_{\min}^{1-p(k, l)} \quad (20)$$

上式中 $G_{H_1}(k, l) = \frac{\xi(k, l)}{1 + \xi(k, l)} \exp\left(\frac{1}{2} \int_{\nu(k, l)}^{\infty} \frac{e^{-t}}{t} dt\right)$, G_{\min} 为无语音出现时增益的最小阈值。一旦获得语音谱增益可利用下式计算纯净语音谱估计:

$$\tilde{X}(k, l) = G(k, l) Y(k, l) \quad (21)$$

再利用实值离散 Gabor 逆变换可获得增强后的纯净语音:

$$\tilde{X}_r(k, l) = \text{Re}(\tilde{X}(k, l)) - \text{Im}(\tilde{X}(k, l)) \quad (22)$$

$$\tilde{x}(n) = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} X_r(k, l) \tilde{h}(n - k\bar{N}) \text{cas}(2 \ln/N) \quad (23)$$

其中 $\tilde{X}(k, l)$ 代表估计的纯净语音时频谱, $\tilde{x}(n)$ 为增强后的语音。

4 实验及性能评价

为了评价本文算法的有效性, 本文从两个方面进行实验, 第一, 在不同过抽样率下进行语音增强比较; 第二, 在不同的噪声环境下将本文算法(简称为 Gabor)与文献[8]提出的基于最小均方误差对数谱估计的语音增强算法、文献[9]提出的基于最小均方差的幅度谱估计语音增强算法以及文献[10]提出的最优对数谱估计算法进行比较(为表述简便, 后文将文献[8]、[9]、[10]及本文中的算法分别简称为 MMSELSA、MMSE、OMLSA、Gabor)。

实验采用来自 IEEE 语音数据库^[11]的纯净语音数据共 30 句, 分别由 3 男 3 女发音; 噪声数据采用 NOISEX-92 数据库的三类噪声信号: White 高斯白噪声、F16 飞机驾驶舱噪声和 M109 坦克噪声。纯净语音数据和噪声数据均为 8kHz 采样率, 混合产生信噪比为 $[-5, 15]$ 范围内带噪语音。实验采用文献[9]中的方

法进行先验信噪比估计,采用改进的最小受控递归平均法 IMCRA 进行噪声谱估计^[10]。

算法性能评价分别采用客观测量和主观测量两种方式。客观测量^[12]采用分段信噪比 (segSNR)、对数频谱距离 (LSD) 来衡量。分段信噪比计算公式如下:

$$SegSNR = \frac{1}{J} \sum_{l=0}^{J-1} 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} x^2(n + lN/2)}{\sum_{n=0}^{N-1} [x^2(n + lN/2) - \tilde{x}^2(n + lN/2)]} \right) \quad (24)$$

对数谱距离计算公式如下:

$$LSD = \frac{1}{J} \sum_{l=0}^{J-1} \left\{ \frac{1}{N/2 + 1} \sum_{k=0}^{N/2} [10 \log_{10} X(k, l) - 10 \log_{10} \tilde{X}(k, l)]^2 \right\}^{\frac{1}{2}} \quad (25)$$

(24), (25) 式中 J 代表语音帧数。

图 1 显示了过抽样率分别为 2, 4, 8, 16, 32 时本文算法的信噪比改进,从图 1 可以看出,过抽样率为由 2 增加到 8 的过程中,分段信噪比改进逐渐提高,当由 8 增加到 32 时,分段信噪比逐渐降低,这说明在用本文算法进行语音增强时,并不是过抽样率越高越好,另外过抽样率越小,CDGT 计算量越小,越易于系统实时实现。在本文后面实验中,过抽样率取为 8。

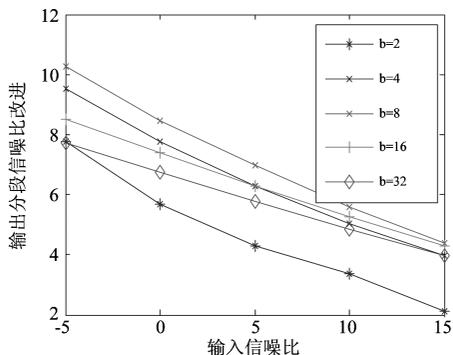


图 1 不同过抽样率下的本文算法信噪比改进

图 2 显示了在不同噪声环境下本文算法与待比较的算法的分段信噪比改进,从图 2 可以看出,本文提出的算法在三种噪声环境中的分段信噪比改进值均大于其他三类算法。这表示本文算法获得了较高的语音去噪度。

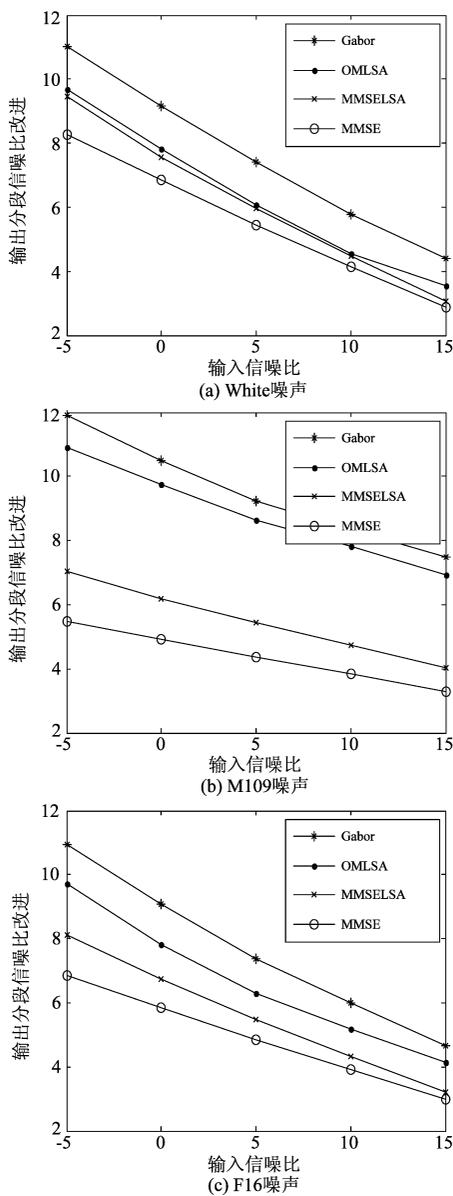


图 2 不同噪声下不同语音增强算法分段信噪比改进

表 1 列出了不同噪声环境下本文算法与其他待比较算法的对数谱距离。从实验结果可以看出,本文提出的算法具有更小的对数谱距离,这进一步证实本文提出的算法具有较大的语音去噪度。

以上客观测试的缺点是不能很好地反映语音信号的感知质量,同时也不能反应语音信号和噪声信号在频率上的分布情况。进行增强时引入的音乐噪声和语音失真会引起人们听觉的不适,客观性指标不能刻画音乐噪声的大小和语音失真的大小。因此,在本文的实验中,我们采用非正式听音测试和分析增强前后语音信号的时域波形图和语谱图进行主观测试,从而在直观上了解音乐噪声和语音失真大小。在进行

非正式的听音测试时,我们主要采用主观平均意见分(MOS)法,实验随机从不同信噪比下的30句增强语音中抽出10句进行分别播放给3位测试者,每人在听完给定的某个语音后给出打分值(5分制),然后对每类信噪比下的语音得分值取平均,得到的MOS分

如表2所示。从表2可以看出,经过本文算法增强后的语音在0dB到15dB范围内均获得了较高的得分,这表明这些语音的感知质量有了明显的改善,但在信噪比很低(比如-5dB)的时候这种语音感知质量并没有明显提高。

表1 不同算法增强语音对数谱距离比较

噪声类型	算法	输入信噪比(dB)				
		-5	0	5	10	15
WGN	Gabor	2.2115	2.0319	1.9115	1.8055	1.7064
	OMLSA	3.2416	2.0866	2.0036	1.9318	1.8197
	MMSELSA	3.2239	2.8318	2.4554	2.1254	1.8297
	MMSE	3.2611	3.1293	2.757	2.3981	2.0584
M109	Gabor	2.1054	1.9022	1.7147	1.5257	1.5257
	OMLSA	2.0342	2.0189	1.795	1.5488	1.3147
	MMSELSA	2.2809	2.0608	1.7988	1.5416	1.3112
	MMSE	2.2293	2.1895	1.9333	1.6606	1.4058
F16 cockpit	Gabor	2.1478	2.0107	1.8509	1.7127	1.5766
	OMLSA	2.6115	2.0711	1.9685	1.8232	1.6229
	MMSELSA	2.7503	2.4712	2.1463	1.8398	1.5673
	MMSE	2.7254	2.7049	2.3841	2.0503	1.7385

表2 不同噪声不同信噪比下四类算法得到的增强语音MOS得分

噪声类型	算法	输入信噪比(dB)				
		-5	0	5	10	15
WGN	Gabor	0.8496	1.5813	2.1538	2.6013	2.9176
	OMLSA	0.9558	1.4709	2.0011	2.4918	2.7893
	MMSELSA	1.3912	1.5413	2.2842	2.5742	2.9144
	MMSE	1.3726	1.5640	2.1331	2.5231	2.8771
M109	Gabor	1.7166	2.3918	2.8982	3.3100	3.6336
	OMLSA	1.9464	2.2767	2.7887	3.2013	3.5644
	MMSELSA	2.1190	2.3540	2.8724	3.2756	3.5718
	MMSE	2.0692	2.3747	2.8838	3.1509	3.4111
F16 cockpit	Gabor	1.0894	1.7993	2.3156	2.7242	3.1576
	OMLSA	1.3674	1.6358	2.1907	2.6142	2.9080
	MMSELSA	1.5858	1.6322	2.2238	2.6864	3.0676
	MMSE	1.5708	1.6267	2.1296	2.7224	3.0780

另外,我们也可以从增强前后语音信号的时域波形图和语谱图分析语音失真度。图 3 和图 4 分别给出了在噪声类型为高斯 White 噪声,输入信噪比为 5dB 情况下语音增强的时域波形和语谱图。从图 3 及图 4 的 (c) 和 (d) 可以发现,本文算法增强后保留了更多的语音内容,而 (e), (f) 则含有更多的残留噪声,这些保留的语音内容对语音感知质量的提高有着重要作用,这也说明本文算法提高了增强后语音的感知质量。

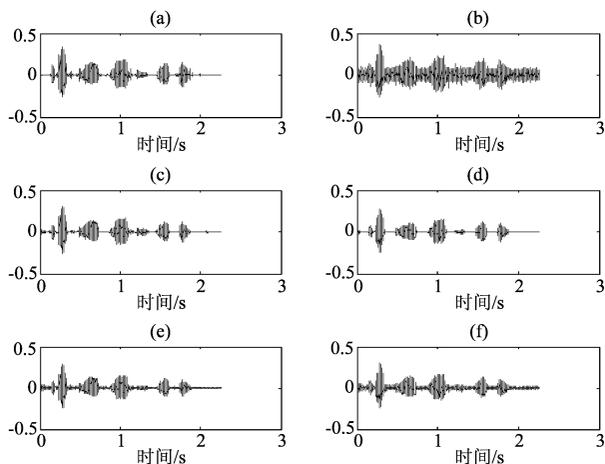


图 3 不同算法进行语音增强比较。(a) 纯净语音;(b) 含噪声语音,信噪比为 5dB;(c) 应用本文算法增强后的语音;(d), (e), (f) 分别应用 OMLSA, MMSELSA, MMSE 算法进行增强后的语音

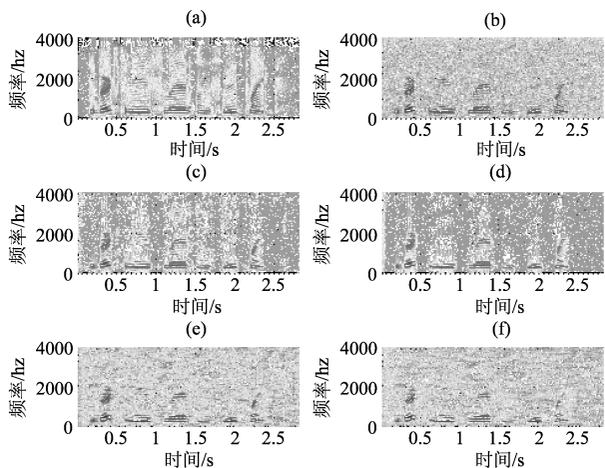


图 4 图 3 对应的语谱图。(a) 纯净语音语谱图;(b) 含噪声语音语谱图,信噪比为 5dB;(c) 应用本文算法增强后的语谱图;(d), (e), (f) 分别应用 OMLSA, MMSELSA, MMSE 算法进行增强后的语谱图

算法时间复杂度方面,采用并行格型结构实现实值离散 Gabor 时频变换所需的时间复杂度为 $L+0.5N$

$\log_2 N$,由 Gabor 变换系数重建原信号时的时间复杂度为 $L+0.5L \log_2 N$,而其他三类算法均采用短时傅里叶变换获得频谱,故其正变换和反变化算法的时间复杂度均为 $M * (0.5N \log N)$,很显然,本文算法具有更小的时间复杂度。

5 结论

语音信号本质上是非平稳的,本文提出了一种基于联合时频域的语音增强方法,通过对带噪语音做实值离散 Gabor 变换以获取语音时频谱,并对时频谱进行增强后再重建为干净语音信号。实值离散 Gabor 变换相对于短时傅里叶变换具有一些特点,比如综合窗和分析窗满足双正交关系;实值离散 Gabor 变换对过抽样条件进行了更加细致的讨论等。本文利用实值离散 Gabor 变换这些特点,依托实值离散 Gabor 变换快速算法,在过抽样条件下进行语音增强并取得了较好的效果。

参考文献

- [1] You, C. H., Rahardja, S. and Li, H. Speech enhancement for telephony name speech recognition [C] // IEEE International Conference on Multimedia and Expo. GERMANY: Institute of Electrical and Electronics Engineers, 2008. 973-976.
- [2] I. COHEN. Relaxed statistical model for speech enhancement and a priori SNR estimation [J]. IEEE Transactions on Speech and Audio Processing, 2005, 13 (5): 870-881.
- [3] Zhao, D. Y., Kleijn, W. B., Ypma, A. and de Vries, B. Online Noise Estimation Using Stochastic-Gain HMM for Speech Enhancement [J]. IEEE Transactions on Audio Speech and Language Processing, 2008, 16(4): 835-846.
- [4] CHOI M S, KANG H G. An improved estimation of a priori speech absence probability for speech enhancement; in perspective of speech absence probability [C] // IEEE International Conference on ICASSP. USA: Institute of Electrical and Electronics Engineers, 2005. 1117-1120.
- [5] I. COHEN. Speech enhancement using a noncasual a priori SNR estimator [J]. IEEE Signal Processing Letters, 2004, 11(9): 725-728.
- [6] J. Wexler and S. Raz. Discrete Gabor expansions [J].

Speech Processing, 1990, 21(3): 207-220.

- [7] L. Tao, H. K. Kwan. Block time-recursive real-valued discrete Gabor transform implemented by unified parallel lattice structures [J]. IEICE Trans. Inform. Systems, 2005, 88(7), 1472-1478.
- [8] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1985, 33(2): 443-445.
- [9] Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator [J]. IEEE Trans Acoustics Speech and Signal Processing, 1984, 32(6):1109-1121.
- [10] I. Cohen. Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging [J]. IEEE Transactions on speech and audio processing, 2005, 11(5):466-475.
- [11] E. H. Rothauser, W. D. Chapman, et al. IEEE Recommended Practice for Speech Quality Measurements [J]. IEEE Trans. Audio and Electroacoustics, 1969, 17(3): 225-246.
- [12] S. R. Quackenbush. Objective Measures of Speech Quality [M]. USA: Prentice Hall, 1988.

作者简介



周健(1981-),男,安徽太湖县人,东南大学水声信号处理教育部重点实验室博士生,研究方向为信号与信息处理。

E-mail: ahjzhou@gmail.com



赵力(1958-),男,江苏南京人,教授,博士,博士生导师,主要从事语音信号处理、自然语言处理、声频和视频信号处理、情感信息处理等方面的研究工作。



陶亮(1963-),男,安徽合肥人,教授,博士,博士生导师,主要从事模式识别与信息处理。



金贇(1979-),男,江苏南京人,东南大学水声信号教育部重点实验室博士生,研究方向为语音信号处理。