

◇ 研究报告 ◇

基于 ERB 尺度划分的多子带语声信号 抗噪谱减算法*

周挺挺 曾毓敏[†] 王蓉蓉 卞乐

(南京师范大学物理科学与技术学院 南京 210000)

摘要 为了研究心理声学在语声增强方面的应用,本文提出了一种基于等效矩阵带宽(ERB)尺度划分的多子带语声信号抗噪谱减算法。此算法根据 ERB 尺度将带噪信号的频谱划分成多个子带,然后再根据每个子带的分段信噪比以及心理声学掩蔽原则分别计算每个子带的谱减参数,最后在每个子带中分别进行谱减算法处理。实验结果表明,应用新算法所获得的语声增强结果在信噪比、IS 失真以及 PESQ 方面均优于之前提出的多子带语声信号抗噪谱减算法。

关键词 ERB 尺度,心理声学掩蔽,多子带谱减

中图分类号: TN912.35

文献标识码: A

文章编号: 1000-310X(2017)03-0212-08

DOI: 10.11684/j.issn.1000-310X.2017.03.004

A multi-band anti-noise spectral subtraction algorithm using ERB scale

ZHOU Tingting ZENG Yumin WANG Rongrong BIAN Le

(School of Physics and Technology, Nanjing Normal University, Nanjing 210000, China)

Abstract This paper addresses a multi-band spectral subtraction algorithm based on equivalent rectangular bandwidth (ERB) scale for applying psychoacoustics to speech enhancement. In the proposed algorithm, the whole spectrum of noisy speech is divided into multiple bands based on ERB scale. The subtraction parameters are then calculated according to the segment SNR of each band and psychoacoustics criteria. Finally, spectral subtraction with different subtraction parameters is executed in each band. The measurements of SNR improvement, IS distortion and PESQ show that the proposed algorithm outperforms the previous speech enhancement algorithms.

Key words ERB scale, Psychoacoustic masking, Multi-band spectral subtraction

2016-09-24 收稿; 2016-11-13 定稿

*江苏省科技项目 (BE2014139)

作者简介: 周挺挺 (1992-), 男, 江苏南通人, 硕士研究生, 研究方向: 语声增强, 心理声学, 声频处理。

[†] 通讯作者 E-mail: zengyumin@njnu.edu.cn

1 引言

语音增强的目的是为了提高在噪声环境中语音通信系统的性能。语音增强可以应用到很多领域,例如手机通信系统、语音识别、助听器等等。由于没有额外的信息来帮助区分语音和噪声信号,单通道的语音增强往往比多通道的语音增强难得多。目前语音增强的方法大致可以分为四种类型:谱减法、子空间方法、统计模型方法以及维纳滤波方法^[1]。

谱减法^[2]在这四类算法中是最常用的,这是因为实现谱减法非常的简单并且谱减法的计算量也很小。谱减法的实现基于对短时频谱幅度的直接估计。谱减法的基本原理是减去带噪信号中的噪声的幅度谱。然而,谱减法有一个不足之处在于处理后的语音会引入一种随着信号处理会改变韵调的有调噪声,这些有调噪声会随机分布在语音的不同频率处,带来一种很难受的听音感觉。随着心理声学的不断研究,许多基于心理声学准则的谱减法的变型已经被提出。本文就侧重在基于心理声学的单通道语音增强研究。

Tsoukalas等^[3]定义了一种可听噪声的心理声学量来描述被人的听觉机制感知的降低的噪声量。通过调节算法中的一些参数,可听噪声的心理声学量可以降低到0。可听噪声的抑制可以获得从客观和主观上的语音增强。可听噪声的心理声学量的计算是基于心理声学的同步掩蔽。Virag^[4]也将心理声学的同步掩蔽应用到了单通道语音增强上。在Virag的算法中,通过计算噪声掩蔽阈来给人的掩蔽效应建模。只要噪声的强度低于人的掩蔽阈,人就听不到该噪声。根据计算得到的噪声掩蔽阈来调节谱减参数,最终获得噪声抑制,语音失真以及残余噪声量之间的最好折中。

Gunawan和Ambikairajah^[5-6]将心理声学的时域掩蔽应用到了语音增强算法中。该算法根据时域掩蔽阈,计算得到了一个加权函数,用来放大在语音活动期间的信号。这样做的好处是保证了在噪声活动期间的信号被掩蔽,不再被人感知。在他们的文章中,只考虑了前掩蔽的作用,并没有考虑后掩蔽的作用。

Aicha和Jebara^[7]提出了一种后处理的算法来检测并且消除处理后语音中的有调噪声现象。该算法根据音调系数以及同步掩蔽,计算出音乐临界频段,进而将有调噪声中的感知部分消除。这个算法

的一个优点是可以用在任何一个谱减降噪技术的后处理部分。

Upadhyay和Karmakar^[8]在他们的算法中使用了多子带谱减技术。多子带谱减技术^[9]最初由Kamath等人提出。多子带谱减技术基于真实的噪声一般都是有色的这个事实,认为在整个语音谱上进行谱减操作是不合理的。因此将整个语音谱划分成不同的子带,分别在不同的子带上进行谱减操作。在Kamath等人的文章中,多子带谱减技术的应用得到了很好的效果。然而,该算法中多子带的划分是基于线性划分的,并没有考虑人的听觉特性。因此,Upadhyay和Karmakar在此基础上提出了基于临界频带划分的多子带谱减算法。从他们的文章中可以看出,基于临界频带划分的多子带谱减算法得到了更好的结果。

随着心理声学的发展,更加符合人的听觉特性的等效矩阵带宽(Equivalent rectangular bandwidth, ERB)划分方式已经被提出^[10]。本文就是将心理声学的ERB划分方式应用在多子带谱减技术中,替代已经提出的线性划分方式以及Bark划分方式。事实上,Zoghiami等人^[11]在2009年的信号处理会议上已经调查了基于ERB尺度的多子带谱减算法。在他们的算法中,每个子带的谱减算法是基于Ephraim-Malah规则的。EM规则中频谱加权系数的计算仅仅基于该子带中的分段信噪比进行的。然而,在本文的算法中,在确定每个子带中的谱减参数时,不仅考虑了分段信噪比的作用,而且考虑了心理声学的掩蔽效应,这样所得到的结果更加符合人的听觉特性。实验结果表明,本文的算法相比较之前的算法,在信噪比,IS失真以及PESQ方面,获得了更好的结果。

2 谱减法的基本原理

本章描述了基本的谱减算法以及该算法的一个改进版本,即过减法。

2.1 基本的谱减算法

假定语音信号 $s(t)$ 受到加性噪声 $d(t)$ 的污染,这样,带噪语音信号 $y(t)$ 可以表示为

$$y(t) = s(t) + d(t). \quad (1)$$

尽管语音信号是非稳态的,也就是频谱特性随时间不断变化的。STFT(短时傅立叶变换)将带噪

信号分帧后,进行进一步的处理。通过应用STFT,公式(1)可以被转换到频域:

$$Y(\omega) = S(\omega) + D(\omega). \quad (2)$$

因此,将带噪声语的短时幅度谱 $|Y(\omega)|$ 减去噪声频谱幅度估计 $|\hat{D}(\omega)|$,就可以得到加强语的短时幅度谱 $|\hat{S}(\omega)|$ 。这里,噪声频谱一般在语声间隙中进行估计。对于最常见的能量谱减,减去噪声估计可以写成如下形式:

$$|\hat{S}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - |\hat{D}(\omega)|^2, & \text{if } |Y(\omega)|^2 > |\hat{D}(\omega)|^2, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

带噪声语的相位信息保留了下来。最终,加强后的语声信号 $\hat{s}(t)$ 可以被描述成如下形式:

$$\hat{s}(t) = \text{IFFT}(|\hat{S}(\omega)| \cdot e^{j \cdot \arg(Y(\omega))}). \quad (4)$$

2.2 改进的谱减算法

Berouti等人^[12]提出了谱减算法的一个改进版本,又叫做频谱过减算法。该算法给出了两个额外的参数:过减因子 α 和谱下限 β 。频谱过减算法的公式如下:

$$|\hat{S}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha \cdot |\hat{D}(\omega)|^2, & \text{if } \frac{|\hat{D}(\omega)|^2}{|Y(\omega)|^2} < \frac{1}{\alpha + \beta}, \\ \beta \cdot |\hat{D}(\omega)|^2, & \text{otherwise,} \end{cases} \quad (5)$$

其中, $\alpha \geq 1, 0 < \beta \leq 1$ 。

过减因子和谱下限的不同取值决定了背景噪声抑制,残余音乐噪声以及语声失真之间的不同折中:

(1) 将要从带噪声语信号谱中减去的噪声能量受过减因子控制。一般来说,实际减去的噪声能量大于所估计的噪声能量,这是因为过减因子的取值一般大于1。这种情况下,残余的有调噪声的峰值会降低,但同时语声失真会提高。

(2) 谱下限的引入是为了使最终的频谱限定在一个预设的最小级之上,而不是直接设为0。同样地,这会降低残余有调噪声,但同时提高了背景噪声。

因此,过减因子和谱下限的选择是非常重要的。本文根据ERB尺度划分不同的子带,在不同的子带

中计算出相应的分段信噪比,并且考虑心理声学掩蔽作用,从而设定不同的过减因子和谱下限。

3 ERB尺度多子带谱减法

本章节详细描述了提出的基于ERB尺度划分的多子带语声信号抗噪谱减算法。图1给出了该算法的流程图。

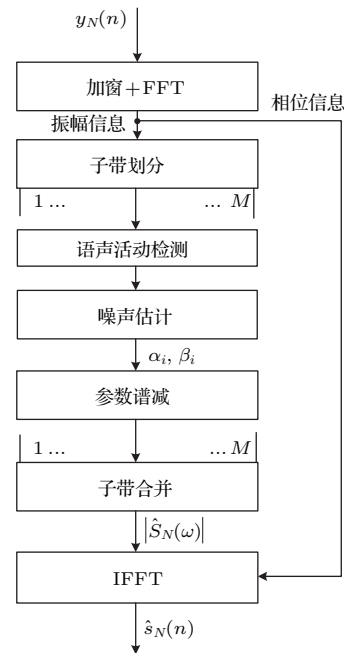


图1 多子带谱减算法流程图

Fig. 1 Structure of multi-band spectral subtraction

在图1中, $y_N(n)$ 表示第 N 帧带噪声语声信号, M 代表子带划分的个数, α_i 代表第 i 个子带的过减因子, β_i 代表第 i 个子带的谱下限, $|\hat{S}_N(\omega)|$ 代表第 N 帧估计的语声信号幅度, $\hat{s}_N(n)$ 代表第 N 帧估计的语声信号。

提出的算法包括如下几个步骤:

- (1) 频谱分解;
- (2) 根据ERB尺度进行子带划分;
- (3) 每个子带进行语声活动检测以及噪声估计;
- (4) 根据噪声估计结果并结合心理声学掩蔽效应确定每个子带的过减因子和谱下限;
- (5) 根据公式(5),对每个子带进行参数谱减;
- (6) 子带合并得到语声信号的幅度谱估计;
- (7) 结合初始的相位信息,根据公式(4),计算估计的语声信号。

本章剩余部分详细介绍其中的第2步,第3步,以及第4步。

3.1 子带划分

Upadhyay 和 Karmakar^[8] 的算法是根据 Bark 尺度进行子带划分的。然而, 现在有充分的证据可以表明, Bark 尺度所得到的临界频带宽度比人听觉掩蔽实际的临界频带宽度偏高^[13-14]。因此, 本文使用更加符合人的听觉特性的 ERB 尺度进行子带划分。公式 (6) 和公式 (7) 给出了 ERB 值与频率值之间的转换关系^[15]。

$$erb(f) = 21.366 \cdot \lg(0.004368f + 1), \quad (6)$$

$$f(erb) = \frac{10^{erb/21.366} - 1}{0.004368}, \quad (7)$$

其中, f 是频率值, erb 是 ERB 值。

子带划分分为如下 3 步:

$$|\hat{D}_{N,M}(\omega)|^2 = \begin{cases} \lambda_D \cdot |\hat{D}_{N-1,M}(\omega)|^2 + (1 - \lambda_D) \cdot W_{D,M}(\omega) \cdot |Y_{N,M}(\omega)|^2, & \text{if } W_{N,M}(\omega) > 0, \\ |\hat{D}_{N-1,M}(\omega)|^2, & \text{if } W_{N,M}(\omega) = 0, \end{cases} \quad (9)$$

其中, N 是帧计数, M 是子带计数, λ_D 是一个平滑因子满足 $0 < \lambda_D < 1$, $W_{N,M}(\omega)$ 是一个在噪声能量谱上的加权系数。

加权系数 $W_{N,M}(\omega)$ 与估计的信噪比 $\bar{\gamma}_{N,M}(\omega)$ 基本成反比关系。公式 (10) 和公式 (11) 分别给出了 $\bar{\gamma}_{N,M}(\omega)$ 和 $W_{N,M}(\omega)$ 的计算。

$$\bar{\gamma}_{N,M}(\omega) = 10 \cdot \lg \left(\frac{|Y_{N,M}(\omega)|^2}{|\hat{D}_{N-1,M}(\omega)|^2} \right), \quad (10)$$

$$W_{N,M}(\omega) = \begin{cases} 1, & \text{if } \bar{\gamma}_{N,M}(\omega) \leq 0, \\ -\frac{1}{\tau} \cdot \bar{\gamma}_{N,M}(\omega) + 1, & \text{if } 0 < \bar{\gamma}_{N,M}(\omega) \leq \varepsilon, \\ 0, & \text{if } \bar{\gamma}_{N,M}(\omega) > \varepsilon, \end{cases} \quad (11)$$

其中, τ 是一个斜率常数, ε 是一个用来进行语音噪声检测的阈值。

在传统的噪声估计算法中, 只有平滑系数 λ_D 被用来更新噪声能量谱。这样会导致每一帧每一子带内所有频率成分的不同更新。这种噪声估计适用于稳态噪声。对于非稳态噪声, 每一帧每一子带内的不同频率成分的更新应该是不同的。因此, 引进了加权系数 $W_{N,M}(\omega)$ 来实现这个功能。加权系数 $W_{N,M}(\omega)$ 的使用保证了每一帧每一子带的频率成分的噪声估计更新是不一样的。Kato^[16] 等人在他们的论文中给出了详细的描述。

(1) 根据公式 (6), 将待分析的频率范围 $f_1 \sim f_2$ 转换到 ERB 范围 $erb_1 \sim erb_2$;

(2) 根据子带数 M , 将 ERB 范围平均划分成 M 等分, $erb_1 \sim erb_1 + \Delta$, $erb_1 + \Delta \sim erb_1 + 2\Delta, \dots$, $erb_2 - \Delta \sim erb_2$ 。其中, Δ 由公式 (8) 给出:

$$\Delta = \frac{erb_2 - erb_1}{M}. \quad (8)$$

(3) 根据公式 (7), 将 M 等分的 ERB 范围转换到频率范围。

3.2 语音活动检测以及噪声估计

在语音增强系统中, 噪声估计是一个重要的组成部分。本文利用公式 (9) 进行噪声能量谱的估计^[16]。

3.3 结合心理声学掩蔽效应调节过减因子 α 和谱下限 β

本文中, 每一帧每一子带的过减因子 α 和谱下限 β 是由该帧该子带的分段信噪比 $\text{SNR}_{N,M}$ 决定。分段信噪比 $\text{SNR}_{N,M}$ 的计算公式如下:

$$\text{SNR}_{N,M} = 10 \lg \left(\frac{\sum |Y_{N,M}(\omega)|^2}{\sum |\hat{D}_{N-1,M}(\omega)|^2} \right). \quad (12)$$

正如 2.2 节提到的, 谱减参数 α 和 β 的提高能够降低残余的有调噪声, 但同时会导致更多的语音失真以及残留的背景噪声。因此, 谱减参数 α 和 β 的调节基于如下准则:

(1) 如果信噪比较高, 根据心理声学掩蔽效应^[17], 残余的有调噪声能够被自然掩蔽并且听不见。因此就没有必要去降低残余音乐噪声, 这样保证了将语音失真和背景噪声降到最低。这种情况下, 谱减参数保持最小值。

(2) 如果信噪比较低, 残余有调噪声就不能被自然掩蔽, 这样就必须人为降低它。通过提高谱减参数的值, 虽然会造成一定的语音失真和背景噪声的增加, 但令人厌烦的残余有调噪声也能够大大减少。

根据上文的描述, 对于第 N 帧语音, 所有子带的分段信噪比中最小的一个对应着谱减参数的最大值, 所有子带的分段信噪比中最大的一个对应着

谱减参数的最小值。其他子带的分段信噪比所对应的谱减参数通过在这两个极端值间插值获取。这种谱减参数选取方式可由公式(13)和公式(14)表示。

$$\alpha_{N,M} = F_{\alpha}[\alpha_{\min}, \alpha_{\max}, \text{SNR}_{N,M}], \quad (13)$$

$$\beta_{N,M} = F_{\beta}[\beta_{\min}, \beta_{\max}, \text{SNR}_{N,M}], \quad (14)$$

其中, F_{α} 和 F_{β} 为对数插值函数。这里选择对数插值而不是线性插值的原因是人的听觉感知被认为是符合对数特性的^[17]。

谱减参数最小值以及最大值的选取是建立在大量的实验仿真基础上的。选取的最终目的是获取残余有调噪声, 语音失真以及背景噪声三者之间的最好折中。当然, 在不同的应用中, 改变谱减参数的最小值和最大值, 就能获得不同的折中结果。

4 实验评估

本章评估了提出的算法与谱减法, 线性划分的多子带谱减以及Bark尺度划分的多子带谱减的

性能。

实验中, 声音信号的拾取在全消声室进行。全消声室的中间放置人工头播放纯净的语音信号, 周围安装的八个架子上的扬声器箱播放噪声信号。模拟的含噪环境就是通过人工头播放语音信号, 周围八个扬声器箱播放噪声信号构成的。采用标准的BK传声器来拾取实验中声音信号。关于实验环境更加详细的描述, 请参考本实验室刘敏等人^[18]的研究报告。

实验中语音信号采取完整出现全部汉语声母与韵母及五个声调、每个因素出现概率与国家标准公布数值的误差不超过 $\pm 5\%$ 、并且由国家文字改革委员会认可的标准普通话发声人方明、雅坤朗诵的“GSBM 6001-89”国家标准中的汉语普通话样件《美谈不美》男女声各一段。从NOIZEUS库^[1]中选择了四种类型的背景噪声进行性能测试, 包括: 高斯白噪声、飞机场噪声、babble噪声以及汽车噪声。

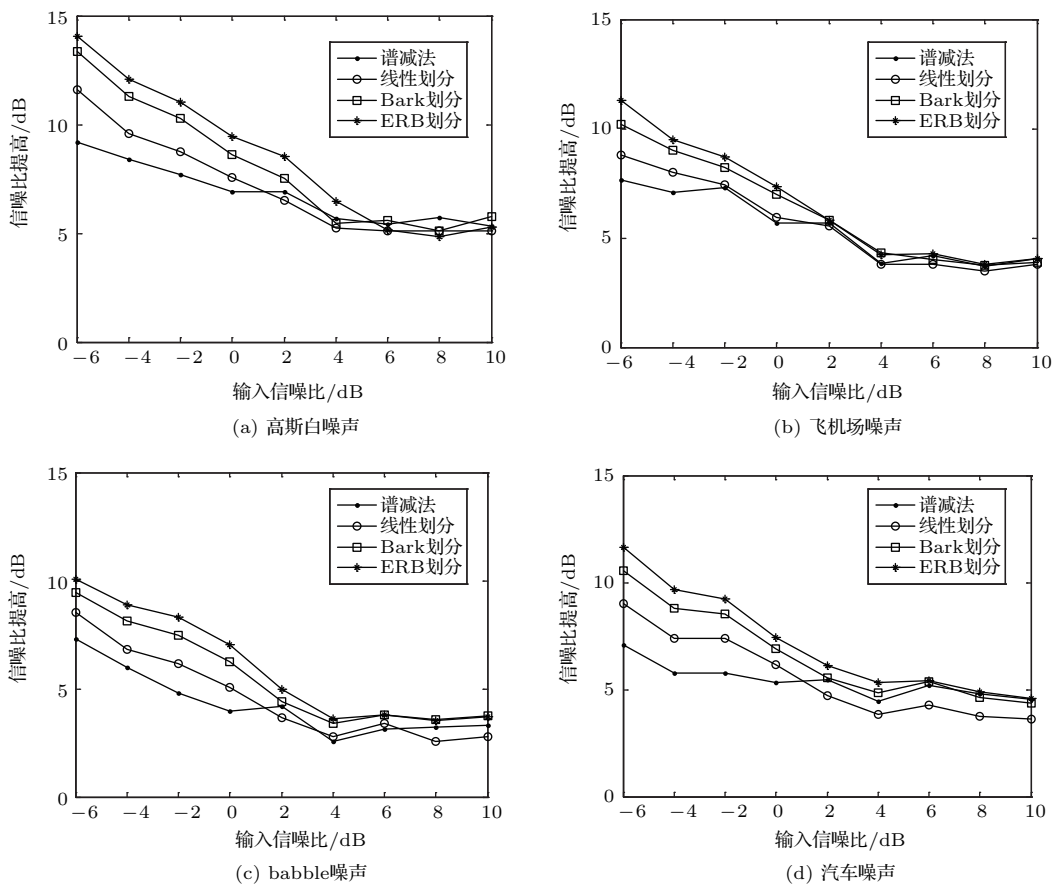


图2 不同噪声类型和噪声级的SNR提高

Fig. 2 SNR improvement for various noise and level

Yi和Loizou^[19]评估了不同的客观评价参数的性能,并且发现PESQ(语音质量的感知评估)在整体质量以及信号失真方面与人的主观感知有最高的相关性。因此,PESQ用来评估提出算法的性能。除了PESQ,SNR(信噪比)以及IS失真也被用来评估提出的算法性能。

实验中录取声音信号的采样率设定为8 kHz,帧长度设为256,50%的重叠率,选用Hanning窗。

4.1 信噪比

噪声降低的量通常由SNR的提高来评估,也就是输出的分段信噪比减去输入的分段信噪比。计算信噪比提高的公式在Virag的文章^[4]中给出。图2给出了不同的噪声类型以及不同的噪声级的SNR提高。

从图2可以看出,不同算法所获得的SNR的提高在高输入SNR时很接近。在低输入SNR的情况下,基于ERB尺度划分的多子带谱减算法获得了较

好的SNR提高。在高斯白噪声情形下,获得了最好的噪声抑制。而对于其它类型的噪声,SNR提高有所降低。

Yi和Loizou^[19]指出,SNR与主观听音结果的相关性是很差的。因此,需要研究其他的客观评价参数。

4.2 IS失真

IS失真测量对不同的频谱包络进行比较,获得一个比SNR相对高一点的与主观结果的相关性。IS失真测量结果一般在0~10之间,IS越小,相应的语音质量也越好。图3给出了不同的噪声类型以及不同的噪声级的IS失真。

从图3可以看出,相比于基本的谱减算法,多子带谱减算法能够得到一个很大的提升。这是因为子带划分实质上是符合人的多子带处理声音的心理声学特性的。同时可以看出,基于ERB尺度划分的

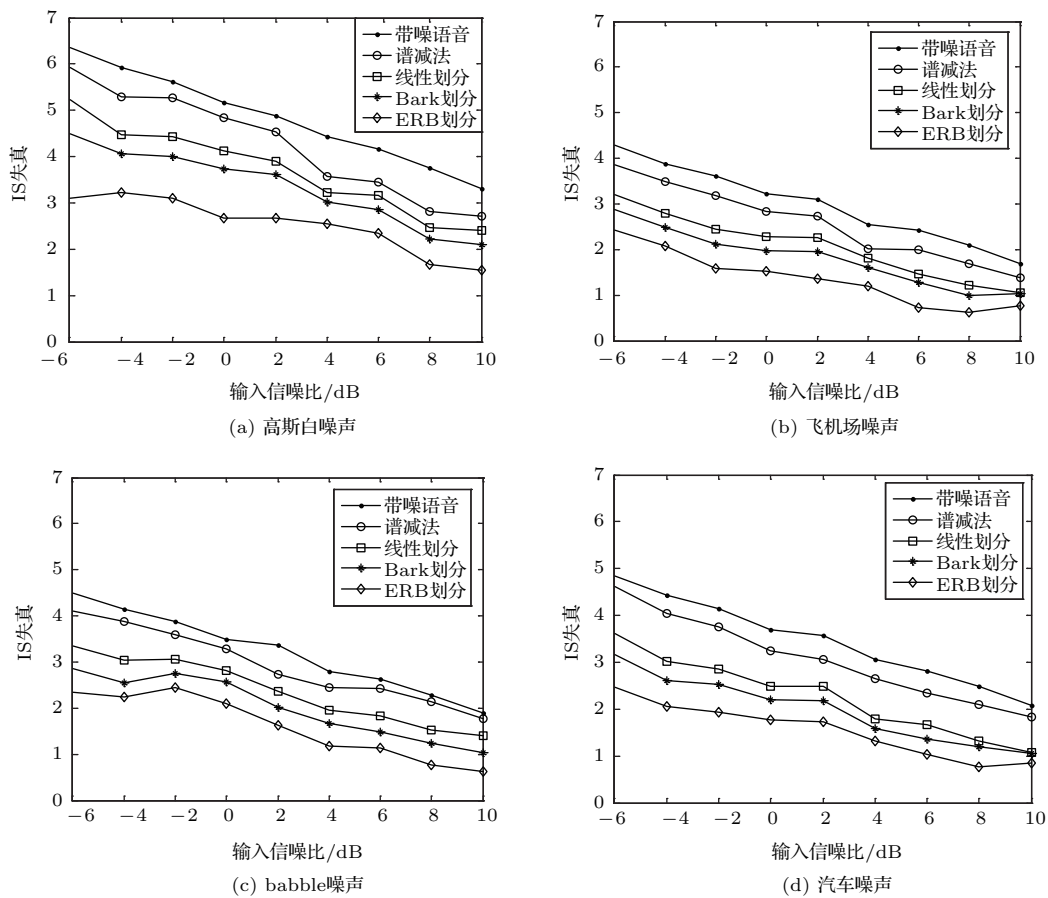


图3 不同噪声类型和噪声级的IS失真

Fig. 3 IS distortion for various noise and level

多子带谱减算法相比于线性尺度划分以及Bark尺度划分的多子带谱减算法得到了更好的结果。这是因为ERB尺度的划分目前是最符合人的听觉特性的。

4.3 PESQ

ITU-T推荐使用PESQ来进行客观评估^[20]。尽管PESQ的计算是所有客观评价参数中最复杂

的,然而PESQ能够在整体质量和信号失真方面与人的主观感知有很高的相关度。图4给出了不同的噪声类型以及不同的噪声级的PESQ值。

从图4可以看出,相比于其他算法,基于ERB尺度划分的多子带谱减算法得到更高的PESQ结果。这说明了基于ERB尺度划分的多子带谱减算法所得到的结果更加符合人的听觉特性。

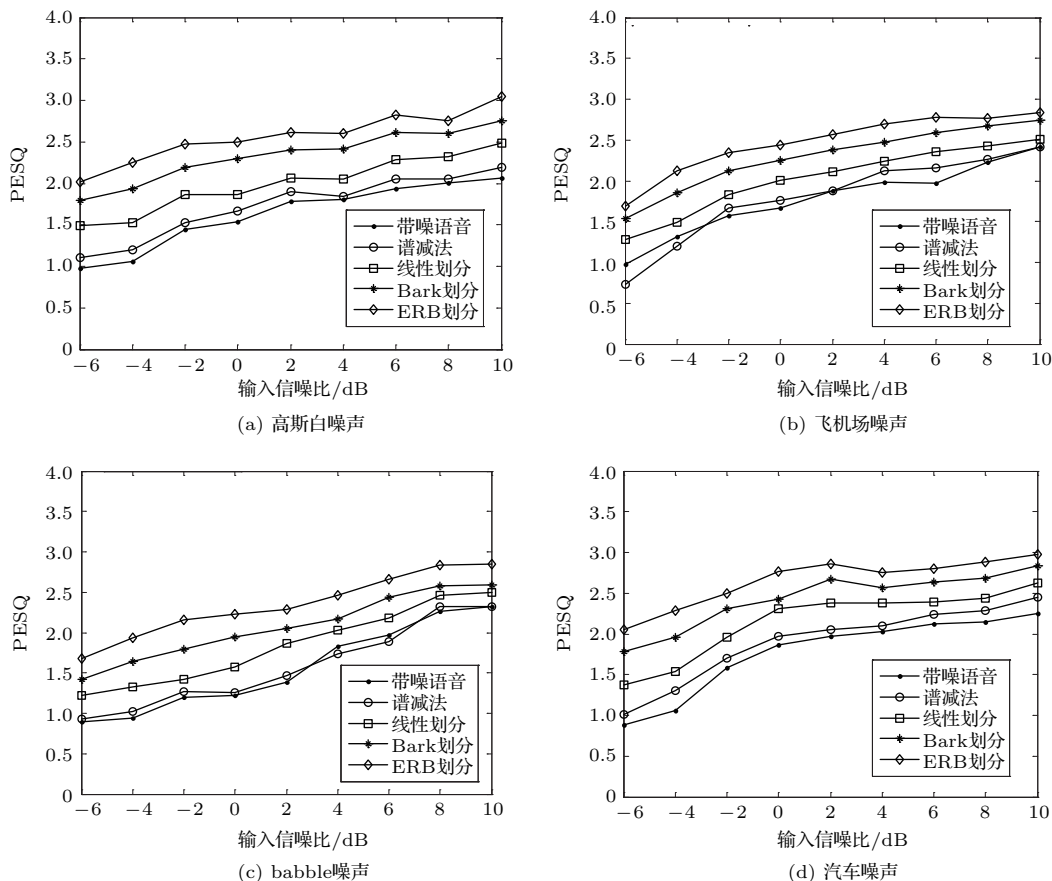


图4 不同噪声类型和噪声级的PESQ

Fig. 4 PESQ for various noise and level

5 结论

本文提出了一种基于ERB尺度划分的多子带语音信号抗噪谱减算法。根据ERB尺度,将带噪信号的频谱划分成多个子带,计算每个子带的分段信噪比,结合心理声学掩蔽效应,从而确定每个子带的谱减参数,最后在每个子带进行不同的谱减处理。ERB尺度相比较Bark尺度在低频处的临界频带宽度更低^[13-14],这样的划分更符合人处理声音的特性。同时,谱减参数的选取在考虑了分段信噪比的基础上,结合了心理声学的掩蔽效应,能够得到残余

有调噪声,语音失真以及背景噪声三者之间符合人的听觉特性的更好折中。实验结果表明,提出的算法在信噪比,IS失真以及PESQ方面优于现有的多子带谱减算法。因此,本文提出的算法是有一定意义的。

参考文献

- [1] LOIZOU P. Speech enhancement: Theory and practice[M]. Boca Raton, FL: CRC Press, 2007.
- [2] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans. on Acoust., Speech, Signal Proc., 1979, 27(2): 113-120.

- [3] TSOUKALAS D, MOURJOPOULOS J, KOKKINAKIS G. Speech enhancement based on audible noise suppression[J]. *IEEE Transactions on Speech and Audio Processing*, 1997, 5(6): 497–514.
- [4] VIRAG N. Single channel speech enhancement based on masking properties of the human auditory system[J]. *IEEE Transactions on Speech and Audio Processing*, 1996, 7(2): 126–137.
- [5] GUNAWAN T S, AMBIKAI RAJAH E. Speech enhancement using temporal masking and fractional bark gammatone filters[C]. *Proc. of. 10th International Conference on Speech Science & Technology*, Sydney, 2004: 420–425.
- [6] GUNAWAN T S, AMBIKAI RAJAH E. A new forward masking model and its application to speech enhancement[C]. *IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP*, 2006.
- [7] AICHA A B, JEBARA S B. Perceptual musical noise reduction using critical bands tonality coefficients and masking thresholds[C]. *INTERSPEECH Conference*, Anvers, Belgium, 2007.
- [8] UPADHYAY N, KARMAKAR A. Single-channel speech enhancement using critical-band rate scale based improved multi-band spectral subtraction[J]. *Journal of Signal and Information Processing*, 2013, 4(3): 314–326.
- [9] KAMATH S, LOIZOU P. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise[C]. *ICASSP Conference*, 2002.
- [10] MOORE B, GLASBERG B. Suggested formulae for calculating auditory-filter bandwidths and excitation pattern[J]. *Journal of Acoustic Society of America*, 1983, 74(3): 750–753.
- [11] ZOGHLAMI N, LACHIRI Z, ELLOUZE N. Speech enhancement using auditory spectral attention[C]. *EU-SIPCO*, 2009.
- [12] BEROUTI M, SCHWARTZ R, MAKHOUL J. Enhancement of speech corrupted by acoustic noise[C]. *Proceedings Int. Conf. Acoustic, Speech, Signal Processing*, Washington DC, 1979: 208–211.
- [13] JURADO C, MOORE B J. Frequency selectivity for frequencies below 100 Hz: Comparisons with mid-frequencies[J]. *J. Acoust. Soc. Am.*, 2010, 128(6): 3585–3596.
- [14] JURADO C, PEDERSEN C S, MOORE B J. Psychophysical tuning curves for frequencies below 100 Hz[J]. *J. Acoust. Soc. Am.*, 2011, 129(5): 3166–3180.
- [15] ANSI S3.4-2007, Procedure for the computation of loudness of steady sounds[S].
- [16] KATO M, SUGIYAMA A, SERIZAWA M. Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA[J]. *IEICE Trans. Fundamentals*, 2002, E85-A(7): 1710–1718.
- [17] FASTL H, ZWICKER E. *Psychoacoustics: Fact and models*[M]. 3rd ed. US: Springer, 2006.
- [18] 刘敏, 曾毓敏, 张铭, 等. 基于二次相关的语音信号时延估计改进算法[J]. *应用声学*, 2016, 35(3): 255–264.
LIU Min, ZENG Yumin, ZHANG Ming, et al. A proposed time delay estimation in speech signal based on second correlation[J]. *Journal of Applied Acoustics*, 2016, 35(3): 255–264.
- [19] HU Y, LOIZOU P C. Evaluation of objective quality measures for speech enhancement[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2008, 16(1): 229–238.
- [20] ITU-T Rec. P. 862, Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephones networks and speech codecs[S].