

短 文

一种基于听觉掩蔽模型的语音增强算法

蔡汉添, 袁波涛

(华南理工大学 电子与通信工程系, 广东 广州 510640)

摘 要: 本文提出一种基于听觉掩蔽模型的语音增强算法。该算法对应用于语音编码中的听觉掩蔽模型进行了适当的修正, 动态地确定每一帧语音信号各个关键频率段的听觉掩蔽阈值, 有选择性地谱减。计算机仿真表明所提算法优于基本谱减法, 不仅信噪比有较大的提高而且有效地减少了主观听觉的失真和残留音乐噪声。

关键词: 语音增强; 谱减算法; 听觉掩蔽模型

中图分类号: TN912

文献标识码: A

文章编号: 1000-436X (2002)-08-0093-06

A speech enhancement algorithm based on masking properties of human auditory system

CAI Han-tian, YUAN Bo-tao

(Department of Electronic Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract: An approach to speech enhancement based on masking properties of human auditory system is presented. In this approach, the masking model of human auditory system used in speech coding is modified, the masking thresholds of the key frequency segments of each speech frame are determined dynamically, and the selective spectral subtraction is performed. The experiment results show that the proposed algorithm is superior to the general spectral subtraction methods, not only in the SNR improvement, but also in the reduction of audible distortion and music residual noise.

Key words: speech enhancement; spectral subtraction; masking model of auditory system

1 引言

有噪环境下的许多语音处理系统的性能急剧下降。语音增强是解决噪声污染的一种有效的预处理技术, 一直是语音信号处理研究领域中的热门课题之一, 广泛应用于语音识别和语

收稿日期: 2001-11-28; 修订日期: 2002-04-04

基金项目: 国家自然科学基金资助项目 (69896240)

作者简介: 蔡汉添 (1946-), 男, 广东广州人, 硕士, 华南理工大学副教授, 主要研究方向为信号处理新方法及其应用, 图像处理 and 图像压缩; 袁波涛 (1974-), 男, 广东东莞人, 华南理工大学硕士研究生, 主要研究方向为信号处理新方法及其应用。

音编码等数字语音系统中^[1-5]。语音增强的目的就是要从带噪语音信号中提取尽可能纯净的原始语音信号,提高信噪比,改善语音质量。目前常用的语音增强算法一般可分为两大类:基于语音生成模型参数和非基于语音生成模型参数。基于语音生成模型参数的语音增强算法主要有卡尔曼滤波算法和隐马尔科夫模型法。这类算法的特点是对语音生成模型参数进行估计和恢复。然而在低信噪比下很难对模型参数进行准确估计,并且此类算法往往需要迭代运算而增加算法的复杂度。非基于语音生成模型参数的语音增强方法主要有自适应噪声滤波法和基于语音短时谱估计的增强算法。其中,谱减法由于其简单有效性已为许多实际数字语音处理系统所采用。但是由于噪声与语音信号在频域中的重叠特性,在消除噪声,提高语音信号的信噪比的同时必然会或多或少地引起原始语音信号的失真。另外,常规谱减法还会产生较强的残留音乐噪声。如何兼顾去噪和减少语音失真是不少学者致力研究的课题。本文提出一种基于听觉掩蔽模型的语音增强算法,利用人类听觉系统的掩蔽特性,有选择性地对谱减,从而较好地兼顾了去噪和减少语音失真。

2 应用于语音增强的听觉掩蔽模型

人类听觉系统具有掩蔽效应。听觉掩蔽模型已成功地应用于音频信号的压缩编码。根据听觉掩蔽效应原理,低于听觉掩蔽阈值的语音信号频率成分可被高于听觉掩蔽阈值的语音信号相近的频率成分所掩蔽,因此这些被掩蔽的语音信号成分很难为人耳的听觉系统所响应,可以看作是冗余信号,从而为进一步提高语音信号的压缩比提供了可靠的依据。利用听觉掩蔽效应的语音增强技术也已引起学者们的兴趣。噪声信号的存在之所以影响到人耳对纯语音信号的接收,从本质上来讲就是因为噪声信号对纯语音信号产生了掩蔽。因此,语音增强就是要减少甚至消除噪声信号对于纯语音信号的掩蔽效应。然而,一般应用于语音压缩的听觉掩蔽模型是在无噪环境下建立的,因此不能直接用于带噪环境下的语音增强而需要对其进行修正。本文采用一个建立在 Bark(频率段)关键频率段基础上的常用的音频信号听觉掩蔽模型^[6,7]。我们得到估算带噪语音信号的各个 Bark 关键频率段的掩蔽阈值的算法如下:

(1) 求每个 Bark 关键频率段的能量,得到每帧语音信号在各个关键频率段的能量分布。
(2) 能量分布与一个听觉掩蔽模型函数卷积(该函数与一个听觉带通滤波器之幅度频率响应的倒置形状十分相似,如图 1(a)所示),求得语音信号的听觉掩蔽阈值 T_i' (这里的 i 指 Bark 编号,在 20Hz 到 16kHz 范围内的声音可以为 24 个 Bark)。

(3) 对(2)求得的语音信号的听觉掩蔽阈值 T_i' 进行修正,求得带噪语音信号各个 Bark 关键频率段的听觉掩蔽阈值 $T_i = T_i' - O_i$ 。

O_i 为每个 Bark 关键频率段听觉掩蔽阈值的调整值, $O_i = a(14.5 + i) + (1 - a)5$ 。

其中, $a = \min\left(\frac{SFM}{SFM_{\max}}, 1\right)$, $SFM = 10 \lg \left(\frac{G_m}{A_m}\right)$ 。SFM 反映信号频谱平滑程度,定义为信号频域能量谱的几何平均值和算术平均值之比。 G_m 是信号能量谱的几何平均值, A_m 是信号能量谱的算术平均值。 SFM_{\max} 为一个经验值,大小为 -60dB,表示只存在纯语音信号的掩蔽效应的情况。当 a 为零时,它表示只存在纯噪声信号的掩蔽效应的情况。

图 1 给出一帧语音信号的掩蔽阈值的计算结果。从图中可以看出,修正前的听觉掩蔽阈值分布比纯语音信号的听觉掩蔽阈值高一些,而修正后的听觉掩蔽阈值分布与纯语音信号的听觉掩蔽阈值分布十分相似。

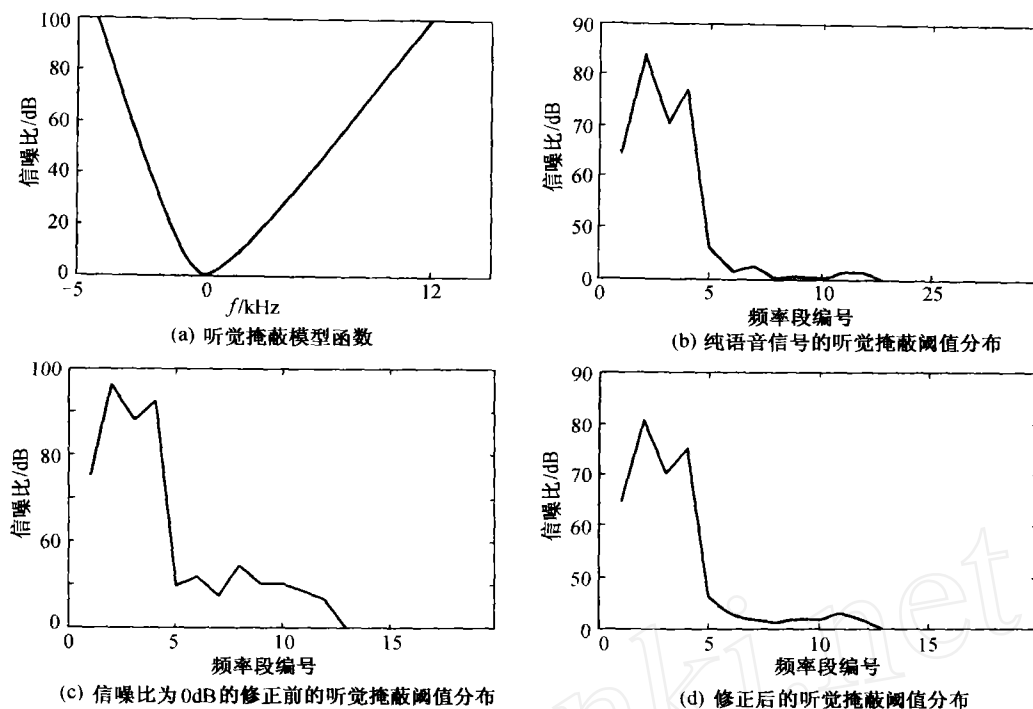


图1 一帧语音信号的听觉掩蔽阈值的计算结果

3 基于听觉掩蔽模型的语音增强谱减算法

一种常用的语音增强谱减法为

$$X'(k) = \begin{cases} \left(1 - \alpha \left[\frac{|N(k)|}{|Y(k)|}\right]^{\gamma_1}\right)^{\gamma_2} Y(k) & \left[\frac{|N(k)|}{|Y(k)|}\right]^{\gamma_1} < \frac{1}{\alpha + \beta} \\ \beta \left[\frac{|N(k)|}{|Y(k)|}\right]^{\gamma_1} Y(k) & \text{其它} \end{cases} \quad (1)$$

其中 $X'(k)$ 为纯语音信号频谱的估值, $Y(k)$ 和 $N(k)$ 分别为带噪语音信号和噪声的频谱。 α 为谱减阈值系数。加大这个系数, 将会提高信噪比, 但也将加大语音信号的失真。 β 为谱减噪声系数。它的作用是为了掩蔽残留的音乐噪声而加上一背景噪声。 β 的加大将会起到减少音乐噪声的作用, 但也将加大背景噪声的能量而使信噪比有所下降。

$\gamma_1=1/\gamma_2$ 决定了公式的尖锐程度。通常取 $\gamma_1=1/\gamma_2=2$ 。

事实证明在所有频段和所有语音帧内采取固定的谱减系数 α 和 β , 语音增强的效果并不理想。特别在低信噪比情况下, 利用上面公式很难在提高信噪比的同时又减少音乐噪声, 并保持原始语音信号质量没有明显下降。为此, 我们提出一种基于听觉掩蔽模型的谱减法, 在不同的频率段和不同的语音帧动态地调整谱减系数。听觉掩蔽阈值较大的 Bark 关键频率段取较小的谱减阈值系数 α 。听觉掩蔽阈值较大说明此 Bark 关键频率段中人耳对其它相近频率段的语音信号和噪声信号的抗干扰能力较强, 所以只需采用较小的谱减阈值系数。反之, 对于听觉掩蔽阈值较小的 Bark 关键频率段, 我们采用较大的谱减阈值系数。同样, 我们可以相应地调整谱减噪声系数 β 。对于各个 Bark 频率段的两个谱减系数 α 和 β 由如下公

式确定

$$\frac{T_{\max} - T}{\alpha(T) - \alpha_{\min}} = \frac{T - T_{\min}}{\alpha_{\max} - \alpha(T)} \quad (2)$$

$$\frac{T_{\max} - T}{\beta(T) - \beta_{\min}} = \frac{T - T_{\min}}{\beta_{\max} - \beta(T)} \quad (3)$$

其中, T 为各个 Bark 关键频率段的听觉掩蔽阈值。 T_{\max} 和 T_{\min} 是每一语音帧的听觉掩蔽阈值的最大和最小值。 $\alpha(T_{\max}) = \alpha_{\min}$, $\alpha(T_{\min}) = \alpha_{\max}$, $\beta(T_{\max}) = \beta_{\min}$, $\beta(T_{\min}) = \beta_{\max}$ 。通过大量的实验, 兼顾提高信噪比和保证原始语音信号质量达到一定的可懂度和清晰度, 减少音乐噪声, 选择 $\alpha_{\max} = 6$ 和 $\alpha_{\min} = 1$, $\beta_{\max} = 0.02$ 和 $\beta_{\min} = 0$ 。

图 2 是本文所提的基于听觉掩蔽模型的谱减算法流程图。噪声幅度谱通过语音端点检测在静音时段进行估计。经过一次基本谱减运算, 适当提高信噪比, 然后再确定听觉掩蔽阈值。根据所求的阈值由公式 (2)、(3) 调整谱减系数 α 和 β 。采用公式 (1) 求得语音信号的谱估计, 进行反傅里叶变换得到增强后的语音信号。

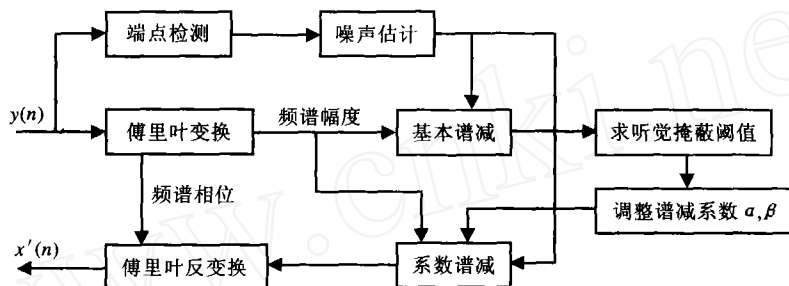


图 2 基于听觉掩蔽模型的谱减算法流程图

4 实验结果及分析

这里给出一段女播音员的录音信号的实验结果。采样率为 8kHz, 时间长度为 5s。背景噪声为高斯白噪声。图 3 为原始语音信号的波形图。对信噪比分别为 5dB、0dB 和 -5dB 的带噪语音信号采用基本谱减法和本文所提的基于听觉掩蔽模型谱减算法进行语音增强。实验结果如表 1 和图 4 所示。表 1 为两种谱减增强算法输出信噪比的比较。图 4 给出较低输入信噪比 (-5dB) 情况下两种谱减增强算法的输出信号波形, 图 4(a) 为输入带噪语音信号波形。从实验结果可以清楚地看到基于听觉掩蔽模型的谱减算法在提高信噪比方面比起传统的谱减算法有了较大的提高。当输入信噪比为 5dB、0dB 和 -5dB 时, 基本谱减法后的输出信噪比为 10dB、6.5dB 和 2.8dB, 而基于听觉掩蔽模型的谱减算法后的输出信噪比为 16dB、13dB 和 9.5dB。此外, 我们通过主观的试听, 可以明显地感觉到基于听觉掩蔽模型谱减法增强后的语音信号的音乐噪声大大地减小了, 同时语音质量没有明显的下降。特别是较低信噪比的情况下 (如 -5dB), 采用基本谱减法增强后的输出信号依然受到较强的背景噪声所掩蔽 (图 4(b)), 无法听清楚原语音的内容, 而采用基于听觉掩蔽模型谱减算法增强后的语音信号则较为清晰可听 (图 4(c))。本文的初步实验结果表明, 利用人类听觉系统的掩蔽效应, 有选择性地谱减, 能有效地兼顾去噪和保持语音信号的主观听觉质量。

表1 基本谱减法法和基于听觉掩蔽模型谱减算法进行语音增强的输出信噪比比较

带噪语音信号输入信噪比/dB	5	0	-5
基本谱减法输出信噪比/dB	10	6.5	2.8
基于听觉掩蔽模型谱减法输出信噪比/dB	16	13	9.5

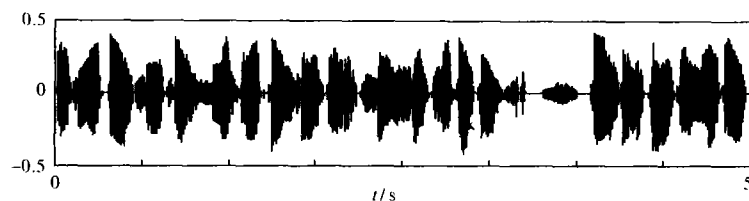
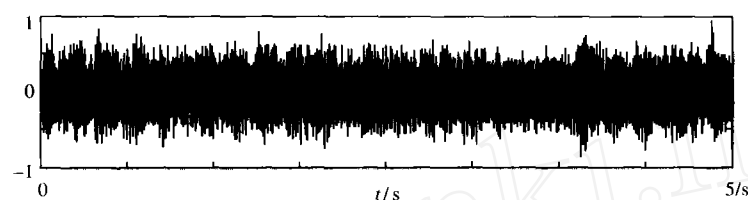
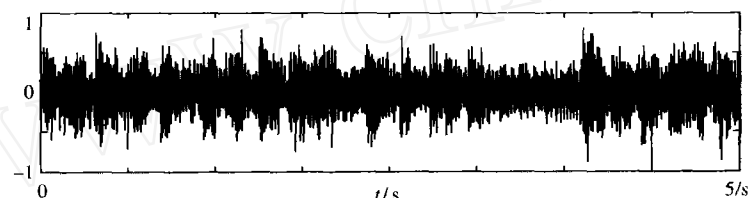


图3 原始语音信号



(a) 信噪比为-5dB的带噪语音信号



(b) 基本谱减法增强后的语音信号



(c) 基于听觉掩蔽模型谱减法增强后的语音信号

图4 信噪比为-5dB的带噪信号进行语音增强的实验结果

参考文献:

- [1] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans Acoustic, Speech, Signal Processing, 1979, 27(2):113-120.
- [2] EPHRAIM Y, MALAH D. Speech enhancement using a minimum mean square error short time spectral amplitude estimator[J]. IEEE Trans Acoustic, Speech, Signal Processing, 1984, 32(6):1109-1121.
- [3] TSOUKALAS D E, MOURJOPOULOS J N, KOKKINAKIS G. Speech enhancement based on audible noise suppression[J]. IEEE Trans Speech and Audio Process, 1999, 7(3):497-514.
- [4] SOON L Y, KOH S N. Low distortion speech enhancement[J]. IEEE Proc Vis Image Signal Processing, 2000, 147(3):247-250.
- [5] TSOUKALAS D, PARASKEVAS M, MOURJOPOULOS J. Speech enhancement using psycho-acoustic criteria[A]. Proc IEEE

ICASSP 1993[C].Minneapolis, 1993.359-361.

- [6] SINHA D, TEWFIK A H. Low bit rate transparent audio compression using adapted wavelets [J]. IEEE Trans Signal Processing, 1993, 41(12):3463-3479.
- [7] JOHNSTON J D. Transform coding of audio signal using perceptual noise criteria[J]. IEEE J Select Areas Commun, 1988, 6(2): 314-323.

《中国优秀博硕士论文全文数据库》(CDMD) 总体介绍

CDMD 由中国学术期刊(光盘版)电子杂志社与清华同方光盘股份有限公司共同研制, 得到了国务院学位办与全国近 300 家博士培养单位的大力支持与协助。CDMD 具有覆盖学科广、文献量大、收录质量高、全文收录、每日更新、使用方式灵活等特点, 是我国最具权威的优秀学位论文全文数据库。

1. 简介

CDMD 覆盖理工、农林、医卫、社会科学各学科, 精选收录全国近 300 家博士授予单位, 2000-2001 年的论文全文近 30000 册, 其中“211 工程”高校的收录率达 80%。CDMD 按学科划分为 9 大专辑出版, 今后, 每年精选增加论文全文 20000 册。

2. 专辑清单

代码	专辑名称	ISSN/CN	学 科 范 围
M-A	理工辑 A (数理科学)	1007-8010/11-9101/N	数学 力学 物理 生物 天文 地理、测绘、资源 气象、水文、海洋地质 地球物理学
M-B	理工辑 B (化学化工能源与材料)	1007-8029/11-9102/T	化学 化工 矿冶 石油 天然气 金属及金属工艺 煤炭 轻工 劳动保护 环境 材料
M-C	理工辑 C (工业技术)	1007-8037/11-9103/N	工业通用技术及设备 机械 仪表 航空 航天 交通运输 水利工程 农业工程 建筑 动力 原子能技术 电工技术
M-D	农业辑	1007-8045/11-9104/S	农业基础科学 农艺学 植保 农作物 园艺 林业 畜牧、动物医学 狩猎、蚕蜂 水产、渔业
M-E	医药卫生辑	1007-8053/11-9105/R	预防医学与卫生学 基础医学 临床医学 中医、中药 药学 生物医学工程
M-F	文史哲辑	1007-8061/11-9106/C	文学 艺术 旅游 历史 哲学 宗教 体育 人物传记
M-G	经济政治与法律辑	1007-807X/11-9107/C	经济学 商贸 金融 保险 政论 党建 外交 军事 法律
M-H	教育与社会科学综合辑	1007-8088/11-9108/C	社会科学研究方法 社会学 民族学 人口学 人才学 各级各类教育
J-I	电子技术与信息科学辑	1008-6293	无线电 计算机 自动化 新闻与传媒 图书情报 档案

3. 出版背景

学位论文与期刊、图书、报纸等文献资料一样, 是记载人类创造的知识信息的一种重要文献类型。世界各国的文献信息机构都很重视对它的收藏与开发利用。我国在博硕士论文的收集、整理、开发方面已取得了积极的成果, 但远不能满足国家信息化建设的要求。CDMD 的建设是在国家信息化发展战略方针的大框架下进行的, 实行全面规划、统一标准、规模建库、资源共享。

4. 用途

- 1) 发展高等教育、特别是研究生教育事业, 培养高层次科技人才;
- 2) 推进科技工作人员之间的学术交流, 促进学科建设和科研工作;
- 3) 建设国家基础知识设施和国家创新体系。

5. 检索系统

- 1) 提供 CNKI 知识仓库分类导航与学科专业导航两套导航检索系统;
- 2) 提供关键词、中文题名、副题名、中文摘要、作者姓名、导师、全文、引文等基本检索功能;
- 3) 提供初级检索与高级检索两套检索界面, 支持二次检索、多种逻辑组合检索等专业检索功能;
- 4) 提供中文简体、中文繁体 and 英文检索三种检索界面, 支持中英文对照和中文简繁体对照检索;
- 5) 提供论文全文的在线浏览、全文下载、保存、打印等功能, 提供摘录功能。

6. 使用方式

- 1) 网上包库服务 (WEB 方式): 读者直接登录 CNKI 数据库交换服务中心网站 (全国共有 10 个) 进行检索;
- 2) 镜像站点方式: 将 CDMD 数据库系统安装到用户单位的内部网络服务器上, 读者在内部网上进行检索;
- 3) 全文光盘方式: 将 CDMD 全文光盘 (DVD 格式) 安装在本单位的计算机或局域网上使用。

7. 更新周期

CNKI 数据库交换服务中心网站数据每日更新, 镜像站点通过互联网或卫星每日更新, 光盘每半年出版一期。

8. 软件环境

用户端: Window 95/98/ME/2000/NT/XP 服务器端: Window 2000/NT/XP

全国免费咨询热线: 8008100946

地址: 北京清华大学学研大厦 1300 室 通信地址: 北京清华大学 84-48 信箱 邮编: 100084

联系人: 张莉 联系电话: 010-62791829/30/31

E-mail: qklw@cnki.net

详情请访问: CNKI 电信全国中心 <http://www.cnki.net/CNKI> 教育全国中心 <http://www.edu.cnki.net/>