

文章编号: 1009-3443(2001)01-0041-04

一种改进的谱相减语音增强方法

蒋海霞, 成立新, 陈显治
(解放军理工大学 通信工程学院, 江苏 南京 210016)

摘要: 针对一般的谱相减语音增强算法所带来的“音乐”噪声较为严重的问题, 提出了一种改进的谱相减语音增强算法——噪声残差消除谱相减法。实验结果表明, 噪声残差消除谱相减法明显地削弱了一般谱相减法所带来的“音乐”噪声。对带噪语音质量的增强效果显著。

关键词: 语音增强; 谱相减; 噪声残差消除

中图分类号: TN912.3 **文献标识码:** A

An Improved Spectrum Subtraction Speech Enhancement Method

JIANG Hai-xia, CHENG Li-xin, CHEN Xian-zhi
(Institute of Communications Engineering PLAUST, Nanjing 210016, China)

Abstract: To solve the serious problem of the “music” noise brought by the generic spectrum subtraction speech enhancement method, we put forward an improved spectrum subtraction speech enhancement method: Residual Noise Removal Spectrum Subtraction. The experiment results show that this improved method weakens the “music” noise distinctly. What’s more, it notably improves the quality of noisy speech.

Key words: speech enhancement; spectrum subtraction; residual noise removal

人们在语音通信过程中不可避免地会受到各种各样的干扰, 这些干扰最终将使接收者接收到的语音已非纯净的原始语音信号, 而是受噪声污染的带噪语音信号。因此, 有必要进行语音增强。语音增强的一个主要目标是从带噪语音信号中提取尽可能纯净的原始语音。

在实际需求推动下, 早在 60 年代语音增强技术研究的课题就已引起人们的注意, 此后 30 多年间人们一直锲而不舍地进行这一方面的研究。随着数字信号处理理论的成熟, 70 年代曾形成一个研究高潮, 取得了一些基础性成果, 并使语音增强发展成为语音信号数字处理领域的一个重要分支。

目前, 大多语音增强算法都是建立在语音的生成模型及感知模型基础上的。由于噪声特性各异, 语音增强的方法也各不相同。在语音处理的实际应用

中, 各类宽带噪声的污染较其它噪声更为普遍, 也更难处理。因此, 设法增强带有宽带干扰的语音信号具有重要意义。30 多年来, 人们针对加性宽带噪声研究了各种语音增强算法。尽管目前语音增强在理论上并未完全解决, 还有待发展, 但某些增强算法已证明是有效果的。这些方法中有基于宽带噪声特性的自适应滤波噪声抵消法、频谱衰减法; 有基于语音产生模型的线性滤波法、梳状滤波法、自相关法等等。其中, 谱相减法是最早发展起来的语音增强方法之一。

一般谱相减法是将短时噪声语音谱与一个估计的噪声谱相减来达到压缩稳态噪声的目的。这种方法的最大优点在于其简单性。但这种方法有两个主要缺陷: 第一, 它仅提高了信噪比意义上语音“质量”, 而没有真正提高语音的“可懂度”; 第二, 这种方法引入了新的令人感到厌烦的“音乐噪声”。本文中采用了噪声残差消除的方法, 噪声残差消除谱相减

法是我们一般在一般谱相减法的基本原理上所提出的改进算法。这种改进的谱相减法在运算复杂度方面与一般谱相减算法相当,但它有效地削弱了“音乐”噪声,对带噪声语音质量的增强效果显著。

1 噪声残差消除谱相减增强法的基本原理与实现

1.1 一般谱相减法

图 1 为一般谱相减法实现的原理框图。 $\alpha=1$ 时,为短时功率谱的相减; $\alpha=2$ 时,为短时能量谱的相减。

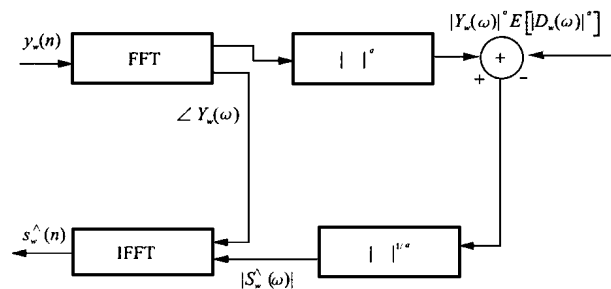


图 1 谱相减法原理框图

Fig.1 Block diagram of Spectrum subtraction principle

图中 $y_w(n)$ 为加窗后的带噪信号, $D_w(\omega)$ 为估测的噪声信号功率谱, $\hat{s}_w(n)$ 为增强后的语音信号。

谱相减法的应用前提是基于以下 5 点假设: (1) 噪声信号和语音信号是互不相关的,在谱域是加性的关系; (2) 背景噪声环境相对于语音活动区间来说是近似稳态的,这样就可以利用在无声段估测的平均噪声谱来逼近有声段的噪声谱; (3) 如果背景噪声环境变化到一个新的稳态,则应有足够的时间(约 300 ms 左右)以便于估计出新的背景噪声谱幅度估值; (4) 对于缓慢变化的非平稳噪声环境,谱相减算法中有语音激活检测环节以便适时地判断并进行调整; (5) 假设主要的噪声影响的消除可以通过仅仅从带噪语音谱幅度中减去噪声而实现。

由于人耳对语音相位不敏感,所以对语音的可懂度及质量起重要作用的是语音信号的短时谱幅度而不是相位。在谱减法中利用人耳的这种感知特性,以原始含噪语音的相位代替估计语音的相位。以下 3 式就是完整的谱相减法的公式。

$$|\hat{S}_w(\omega)|^2 = |Y_w(\omega)|^2 - E[|D_w(\omega)|^2] \quad (1)$$

$$\hat{S}_w(\omega) = |\hat{S}_w(\omega)| \exp(j \angle Y_w(\omega)) \quad (2)$$

$$\hat{s}_w(n) = F^{-1}[\hat{S}_w(\omega)] \quad (3)$$

由于我们假设噪声信号和语音信号是互不相关的,所以由式(1)得到增强后的语音信号的功率谱;最后利用原始含噪语音的相位 $\angle Y_w(\omega)$ 代替增强后语音的相位,从而恢复出增强后的语音信号 $\hat{s}_w(n)$ 。

1.2 噪声残差谱相减法

以上所述的谱相减技术存在两个明显的问题: ①相减谱中有负幅度出现; ②相减谱中残存有噪声信号的频谱峰,导致增强语音中产生“音乐”噪声。

在具体实现谱相减法时,为了尽量削弱这两个问题所带来的影响,我们进行了如下的改进:

(1) 输入/输出数据的加窗与缓存

对于待处理的输入语音数据,首先应该分帧处理。我们依据以下理论进行处理: 如果将一个序列分成半重迭的数据段,每一段数据乘以汉明窗,那么,由这些加窗的数据段迭加就可以恢复原始序列,如图 2 所示。

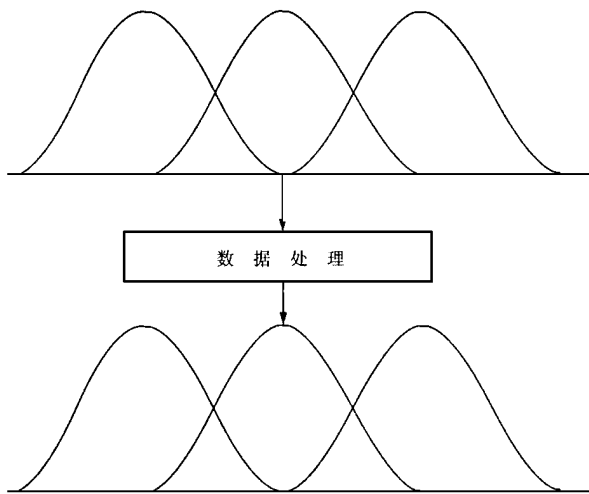


图 2 数据的分帧与恢复

Fig.2 Data segmentation and recovery

对于 8 kHz 抽样的语音信号,我们采用汉明窗长为 256 点,即帧长为 256 点(每帧长为 32 ms),帧移为 128 点。快速傅立叶变换(FFT)的长度为 256 点。

由数字信号处理的知识可知: 由于圆周卷积的时域混叠效应,输入时间序列 FFT 后,经过谱修正,然后再进行反傅立叶变换(IFFT)恢复的输出时域波形将会有所失真;而如果 FFT 的变换点数足够长,并在输入时间序列的后面补零,则可以减小这种失真。在实验中,我们对此进行了研究并发现: 是否在输入信号后补零,对增强后语音质量没有明显的差别。因此,我们采用如上所述的分帧处理方法。

(2) 频谱整形

在式(1)中,当带噪信号的能量小于估测的噪声信号的能量时,就会产生负的相减谱幅度值。为了保证相减谱幅度值为正,我们采用了简单的半波整形的方法,即当相减谱幅度出现负值时,令其为零。当然这样将引入宽带噪声。

(3) 噪声残差消除

谱相减技术虽然有效地削弱了加性背景噪声的影响,但由其所带来的“音乐”噪声对语音质量也有很大的影响。“音乐”噪声产生的原因主要是:背景噪声的短时谱中存在着峰值和谷值,它们所在的频率点与幅度都是随机变化的。经过频谱整形后的残留噪声谱经 IFFT 后恢复的时域信号就会产生如同经过随机的基本频率调制的音调,故名“音乐”噪声。在有声段,噪声残差在那些语音信号无法掩蔽的频率上也能被感知。

为了削弱“音乐”噪声的影响,利用噪声残差信号的帧到帧间的随机性,并考虑到以下 3 个方面:①如果估测语音谱 $|S^{\wedge}(\omega)|$ 比噪声残差谱的最大值小,并且帧间的变化比较剧烈,这说明该频率点的谱极可能为噪声谱,用前后 3 帧的对应点的最小值来表示;②如果估测信号谱虽比噪声残差谱的最大值小,但前后帧间的变化较小,这说明该频率点的谱极可能为信号能量较低的语音信号谱。在这种情况下,用前后 3 帧的对应点的最小值来表示,可以保存语音信号信息;③如果估测信号谱比噪声残差谱的最大值大,则说明该频率点的谱为语音信号谱,且说明噪声的消除已经比较充分了。

以上所阐述的噪声残差消除思想可以用公式(4)、公式(5)表示:

$$\begin{aligned} &|S_i^{\wedge}(\omega)| = |S_i^{\wedge}(\omega)|, \\ \text{当 } &|S_i^{\wedge}(\omega)| \leq \max(|N_R(\omega)|) \end{aligned} \tag{4}$$
$$|S_i^{\wedge}(\omega)| = \min\{|S_j^{\wedge}(\omega)| \mid j = i - 1, i, i + 1\},$$
$$\text{当 } |S_i^{\wedge}(\omega)| < \max(|N_R(\omega)|) \tag{5}$$

其中, $|S_i^{\wedge}(\omega)|$ 表示第 i 帧的估测信号谱, $\max(|N_R(\omega)|)$ 为在非语音帧所得到的噪声残差信号谱的最大值。以上处理有效地平滑了相减谱中残存的噪声信号的频谱峰,“音乐”噪声明显削弱。

2 实验结果比较

经过反复多次的实验证明,采用了噪声残差消除的谱相减法比一般的谱相减法效果明显要好。同一般谱相减法一样,噪声残差消除的谱相减法适用于平稳的加性噪声。实验中我们选取了具有代表性的高斯白噪声。用噪声残差消除谱减法与一般谱减

法相比,其增强后语音所含的宽带噪声和“音乐”噪声都要小得多,语音质量得到明显的改善,主观试听也证明了这一点。

2.1 语谱图比较

从处理后的语音的语谱图比较可以直观地看出各种语音增强算法对语音质量的改善效果(如图 3



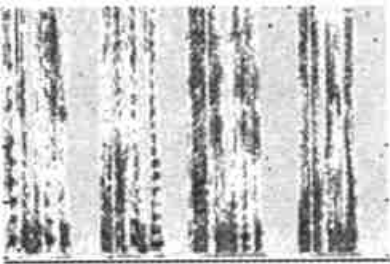
(a) 干净语音语谱图



(b) SNR= 0 dB 带噪语音语谱图



(c) 一般谱减法处理(α= 2)



(d) 噪声残差谱减法处理

图 3 语谱图比较

Fig. 3 Spectrograms of the speech in different environments

所示)。语音材料为 8 kHz 采样, 16 位表示的长度约为 8.2 s 的一段语音“我去无锡市, 他到黑龙江”。图

3(a) 为原始的干净语音的语谱图, 图 3(b) 为加了白噪声的噪比为 0 dB 的带噪语音的语谱图。从图 3(b) 可以看出, 由于受到噪声的影响, 语谱图中语音的谐波结构、共振峰等信息量都受到了很大的影响。图 3(c) 为一般的谱相减语音增强算法作用于 0 dB 的带噪语音后的语谱图。图 3(d) 为改进的谱相减法即噪声残差消除的谱相减语音增强算法作用于 0 dB 的带噪语音后的语谱图。不难看出, 经过增强后的语音的质量和信噪比在一定程度上都有所改善; 噪声残差消除的谱相减法的效果明显优于一般谱相减法。

2.2 语音质量客观评价结果比较

语音质量的客观评价方法是指用客观测量的手段来评价语音编码质量。它采用某个特定的参数去表征语音通过处理系统后的失真程度, 并以此来评估处理系统的性能优劣。实验中, 我们采用了传统的 LPC 倒谱距离客观评价方法(LPC-CD) 以及较新的巴克谱距离客观评价方法(BSD) , 如表 1 所示。语音材料同上。

LPC-CD 与 BSD 的数值越小, 说明处理后的语音与干净语音的距离越小, 即语音质量越好。这与非正式主观试听的结论是一致的。

表 1 客观评价实验数据

Tab. 1 Experiment data of the objective measurements

客观评价方法	带噪语音 ($SNR=0\text{dB}$)	一般谱相减法 ($\alpha=2$)	噪声残差消除谱相减法
LPC-CD	2.412	2.329	2.270
BSD	5.794	2.587	1.779

3 小结

本文提出了一种改进的谱相减语音增强算法——噪声残差消除谱相减语音增强算法。实验结果表明: 改进后的语音增强算法具有与一般谱相减法相当的运算量, 但更好地提高了带噪语音的信噪比, 明显地削弱了一般谱相减法所带来的“音乐”噪声, 对带噪语音质量的增强效果显著。

参考文献:

[1] 杨行峻, 迟惠生, 唐 昆, 等. 语音信号数字处理[M]. 北京: 电子工业出版社, 1995.

[2] Gray A, Markel J. Distance measures for speech processing[J]. IEEE Trans on ASSP, 1976, 24(5): 380-391.

[3] Wang S H, Sekey A, Gersho A. An objective measure for predicting subjective quality of speech coders [J]. IEEE Journal on Sel areas in Communication, 1992, 10(5): 819-829.

[4] Steven F B. Suppression of noise in speech using the SABER method[C]. ICASSP' 78. 606-609.

[5] Steven F B. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans on ASSP, 1979, 27(2): 113-120.

[6] Flanagan J L. Spectrum analysis in speech coding[J]. IEEE Trans on Audio and Electroacoustics, 1967, 15 (2): 64-69.

(责任编辑: 徐金龙)