

基于语音帧检测和子带谱跟踪的噪声估计算法

贺慧芳, 马建芬

(太原理工大学计算机与软件学院, 太原 030024)

摘要: 针对传统基于最小统计和递归平均的噪声估计算法存在较长时间延迟的问题, 提出一种快速跟踪的噪声估计算法, 将语音活性检测应用于子带谱跟踪来估计非稳定噪声的功率谱。仿真结果表明, 该算法保证了噪声估计的精确性, 当噪声为非稳定时, 相比其他算法其跟踪速度更快。

关键词: 语音帧; 噪声估计; 子带

Noise Estimation Algorithm Based on Speech Frame Detection and Subband Spectral Tracking

HE Hui-fang, MA Jian-fen

(College of Computer and Software, Taiyuan University of Technology, Taiyuan 030024)

【Abstract】 Aiming at the problem of traditional noise estimation algorithms based on minimum statistics and recursive averaging existing long time delay, this paper proposes a quick tracking noise estimation algorithm, it applies voice activity detection into subband spectral tracking to estimate the power spectrum of non-stationary noise. Simulation results show that the algorithm ensures the accuracy of noise estimation, it tracts faster than other algorithm when the noise is highly non-stationary.

【Key words】 speech frame; noise estimation; subband

1 概述

目前, 噪声估计已成为语音增强技术中很关键的环节。在很多单通道语音增强算法中, 尤其在噪声源不可知的情况下, 实时噪声功率谱估计是很重要的。噪声估计的准确性会直接影响最终效果: 若噪声估计过高, 则微弱的语音将被去掉, 增强语音会产生失真; 若估计过低, 则会有较多的背景残留噪声。

近年来, 非稳定环境下噪声功率谱估计得到了重视。传统的噪声估计方法实现纯粹基于语音活动的检测, 制约了噪声在语音存在时的更新, 而且在弱语音信号的低输入信噪比时, 其可靠性性能恶化较明显。

文献[1]提出的最小值统计(Minima Statistical, MS)方法, 通过跟踪特定窗内语音功率谱的最小值来估计噪声, 乘以一个系数来补偿偏差。该方法所得噪声估计的方差是传统方法的2倍, 同时它可能偶尔会削弱低能量音素, 尤其当最小观察窗很小时, 只有以很高的计算复杂度为代价来克服其局限性。

文献[2-3]提出最小值控制递归平均(Minima Controlled Recursive Averaging, MCRA)法以及改进的MCRA(Improved Minima Controlled Recursive Averaging, IMCRA), 它们虽然保证了噪声谱估计的准确性, 但在追踪带噪声语音平滑功率谱最小值时采用了固定时间窗, 因此, 在噪声突变的情况下, 估计的噪声谱存在很长的延时。

文献[4]提出一种快速估计方法, 其噪声谱的更新在连续时间内不依赖固定时间的窗长, 但在语音或噪声能量过高时噪声估计会慢下来, 而且当时间大于0.5 s时, 会削弱一些语音能量。

本文提出的噪声估计方法, 能使估计的噪声谱很快适应背景噪声的变化。

2 语音帧检测

采用结合短时能量和短时过零率来进行语音帧检测^[5]。将语音帧分为3个可互相转化的状态, 即非语音帧、可能语音帧、语音帧。具体检测步骤如下:

(1) 设定门限

$energy(n)$ 代表第 n 帧的短时能量, 设定2个短时能量判决门限 $energy1$ 和 $energy2$ 。 $zcr(n)$ 代表第 n 帧的短时过零率, 同样设定2个短时过零率判决门限 $zcr1$ 和 $zcr2$ 。短时过零率用来检测清音的存在, 短时能量用来检测浊音的存在。

(2) 判断语音帧

从最开始进行检测, 一直处于非语音状态。如果短时能量 $energy(n)$ 一直低于 $energy2$, 且 $zcr(n)$ 低于 $zcr2$, 那么表示此刻既没有清音成分, 也没有浊音成分, 语音一直处于非语音帧状态。当检测到短时过零率 $zcr(n)$ 大于 $zcr2$, 或短时能量 $energy(n)$ 大于 $energy2$, 代表语音可能已出现, 但暂时不能判断其为语音的起始, 其原因是它可能是语音清音部分的开始, 也可能是某些噪声的干扰。只有短时能量 $energy(n)$ 大于 $energy1$ 时, 才代表语音浊音部分出现, 判定为语音帧。

如果连续的可能语音帧持续超过300 ms, 那么将当前帧判定为非语音帧。此时, 用变量 T 来计算连续可能语音存在信号的持续时间。

作者简介: 贺慧芳(1983-), 女, 硕士, 主研方向: 语音信号处理, 谱分析; 马建芬, 副教授

收稿日期: 2009-05-23 **E-mail:** hehuifang789@163.com

3 子带谱跟踪的噪声谱估计

3.1 平滑功率谱

假设带噪语音信号为 $y(i) = s(i) + n(i)$, $0 \leq i \leq N-1$, 且 $s(i)$ 和 $n(i)$ 是相互独立的, 则 $|Y(\lambda, k)|^2 = |S(\lambda, k)|^2 + |N(\lambda, k)|^2$ 。其中, $|Y(\lambda, k)|^2$, $|S(\lambda, k)|^2$, $|N(\lambda, k)|^2$ 分别是带噪语音, 纯净语音和加性噪声的功率谱, λ 是时间, k 是频率。

为了方便计算, 令 $P_Y(\lambda, k) = |Y(\lambda, k)|^2$, 根据文献[2-3]的方法, 得到平滑功率谱:

$$p_Y(\lambda, k) = \sum_{v=D}^{\lambda} b(v) P_Y(\lambda, ((k-v))_K) \quad (1)$$

其中, K 是频谱的长度; $b(v)$ 为窗函数, 且 $\sum_{v=D}^{\lambda} b(v) = 1$ 。

3.2 短时平滑功率谱

用一阶递归公式计算短时平滑功率谱:

$$\overline{S}(\lambda, k) = \alpha_l \overline{S}(\lambda-1, k) + (1-\alpha_l) p_Y(\lambda, k) \quad (2)$$

其中, α_l 的经验值为 0.8。

3.3 语音存在检测^[6]

语音存在是根据特定窗长内, 短时平滑功率谱和其最小值比率决定的。在特定的窗长 L 内, L 为 1 s~2 s。跟踪 $\overline{S}(\lambda, k)$ 的最小值 $S_{\min}(\lambda, k)$ 。

若 λ 能被 L 整除, 则

$$\begin{cases} S_{\min}(\lambda, k) = \min\{S_{\min}(\lambda-1, k), \overline{S}(\lambda, k)\} \\ S_{\min}(\lambda, k) = \overline{S}(\lambda, k) \end{cases} \quad (3)$$

否则

$$\begin{cases} S_{\min}(\lambda, k) = \min\{S_{\min}(\lambda-1, k), \overline{S}(\lambda, k)\} \\ S_{\min}(\lambda, k) = \min\{S_{\min}(\lambda-1, k), \overline{S}(\lambda, k)\} \end{cases} \quad (4)$$

语音存在概率用局部能量和谱最小值的比率计算:

$$S_r(\lambda, k) = \frac{\overline{S}(\lambda, k)}{S_{\min}(\lambda, k)} \quad (5)$$

计算概率, 并将其与经验频率值 δ_b 比较, 若大于 δ_b , 则

$$I_{\text{subband}}(\lambda, k) = 1, \text{ 否则, } I_{\text{subband}}(\lambda, k) = 0。$$

对于所有帧, 若 $\sum_{\lambda} I_{\text{subband}}(\lambda, k) > B_{th}$, 则 $I_{\text{frame}}(\lambda) = 1$, 否则, $I_{\text{frame}}(\lambda) = 0$ 。其中, δ_b 的值为 8; B_{th} 的值为 4。

3.4 更新噪声功率谱

根据上述内容, 令 $I(\lambda, k) = I_{\text{frame}}(\lambda) \times I_{\text{subband}}(\lambda, k)$ 。若 $I(\lambda, k) = 0$, 则更新噪声功率谱为 $D(\lambda, k) = \alpha_d D(\lambda-1, k) + (1-\alpha_d) |Y(\lambda, k)|^2$ 。

4 基于语音帧检测和子带谱跟踪的噪声估计算法

在上述内容的基础上, 利用语音帧检测方法^[5]检测出非语音帧、可能语音帧和语音帧。当确定帧的类别后, 对不同类型的帧分别进行噪声谱更新。

为了解决时间窗的延迟, 采用 2 个并行的最小跟踪器: 主跟踪器是在较大窗长为 L , 输出为 $S_{\min}(\lambda, k)$ 和 $S_{\text{mp}}(\lambda, k)$, 计算方法如式(3)与式(4)所示。次跟踪器^[6]是在较小窗长为 L' 下, 输出分别为 $S'_{\min}(\lambda, k)$ 和 $S'_{\text{mp}}(\lambda, k)$, 计算方法如式(3)与式(4)。当检测到当前帧为可能语音帧时, 更新 $S'_{\min}(\lambda, k)$ 和 $S'_{\text{mp}}(\lambda, k)$ 。其中, L' 为 200 ms~300 ms。可见, 次跟踪器的跟踪速度较主跟踪器提高很多。

对于语音帧, 当 $I(\lambda, k) = 1$ 时, 更新噪声功率谱为 $D(\lambda, k) = D(\lambda-1, k)$ 。对于非语音帧, 当 $I(\lambda, k) = 0$ 时, 噪声功率谱更新为 $D(\lambda, k) = \alpha_d D(\lambda-1, k) + (1-\alpha_d) |Y(\lambda, k)|^2$, 其中, α_d 的取值为 0.95。

5 实验仿真

实验采用的纯净语音来源于 HINT 数据库, 噪声语料来源于 NOISEX-92 数据库。语音信号和噪声信号均是 8 000 Hz 采样, 16 bit 量化。将纯净语音和噪声按比例线性相加产生不同信噪比的带噪语音, 采用汉明窗, 帧长 20 ms, 帧间重叠 50%。

对本文的噪声估计方法进行测试, 并与文献[4]方法进行比较。图 1 为噪声突变环境下噪声谱估计方法比较的仿真结果, 采用的前 3.2 s 的语音为 10 dB 的高斯白噪声, 后 3.4 s 为 -5 dB 的 F16 战斗机噪声。由此可见, 本算法噪声估计的跟踪速度较文献[4]有很大提高, 且大大缩减了时间延迟。

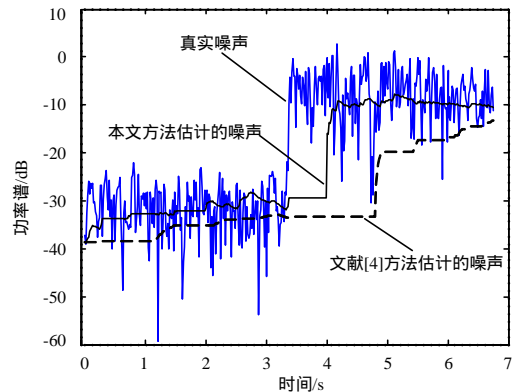


图 1 噪声突变环境下 2 种不同噪声谱估计算法对比

6 结束语

实验结果表明, 在输入信噪比较低和噪声突变时, 本文算法能准确估计语音信号的噪声功率谱, 且相比文献[4]方法, 能更好地跟踪源噪声, 缩短时间延迟。

参考文献

- [1] Martin R. Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics[J]. IEEE Trans. on Speech and Audio Processing, 2001, 9(5): 504-512.
- [2] Cohen I, Bergdugo B. Noise Estimation by Minima Controlled Recursive for Robust Speech Enhancement[J]. IEEE Signal Processing Letters, 2002, 9(1): 12-15.
- [3] Cohen I. Noise Spectral Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging[J]. IEEE Transactions on Speech and Audio Processing, 2003, 11(5): 466-475.
- [4] Rangachari S, Loizou P C. A Noise-estimation Algorithm for Highly Non-stationary Environments[C]//Proc. of ICASSP'04. Dallas, USA: [s. n.], 2004: 220-231.
- [5] 胡光锐. 语音处理与识别[M]. 上海: 上海科学技术文献出版社, 1994.
- [6] Zhong Lin, Goubran R A, Danserean R M. Noise Estimation Using Speech/Non-speech Frame Decision and Subband Spectral Tracking[J]. Speech Communication, 2007, 49(7/8): 542-557.

编辑 陆燕菲