文章编号:1671-4598(2018)12-0140-05 DOI:10.16526/j.cnki.11—4762/tp.2018.12.032 中图分类号:TN912 文献标识码:A

基于盲源分离和噪声抑制的语音信号识别

刘 晶
(南京理工大学 计算机科学与工程学院,南京 210094)

摘要:为了更准确地在噪声环境中对不同语音信号进行识别,提出了一种用于普适语音环境下的自优化语音活动检测(VAD)算法,该算法运用个性化语音命令自动识别系统的语音信号,并能够有效地从多个发声者的混合语音中分离出个体发声者的声音,通过跟踪语音功率谱的较高幅度部分和自适应地抑制噪声来检测发声者的语音信号;设计并实现了一种处理多个发声者任务的自动语音识别(ASR),免去了对干净的语音变化进行先验估计,直接利用噪声本身产生语音/非语音判决的阈值以完成自优化过程;使用语音数据库 NOIZEUS 进行了评价测试,实验结果表明,所提出的盲源分离和噪声抑制方法不需要任何额外的计算过程,有效地减少了计算负担。

关键词:语音恢复;时频分离;自适应噪声抑制;自动语音识别

Speech Signal Recognition Based on Blind Source Separation and Noise Suppression

Liu Jing
(School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing 210094, China)

Abstract: In order to more accurately identify different speech signals in a noisy environment, a self-optimizing speech activity detection (VAD) algorithm for universal speech environment is proposed, which uses personalized speech commands to automatically identify the speech signals of the system. And can effectively separate the sound of individual vocalists from the mixed speech of multiple utterers, and detect the utterer's speech signal by tracking the higher amplitude part of the speech power spectrum and adaptively suppressing the noise. Designed and implemented an automatic speech recognition (ASR) that handles multiple vocalist tasks, eliminating the need for a priori estimate of clean speech changes, and directly using the noise itself to generate speech/non-voice decision thresholds to complete the self-optimization process . The speech database NOIZEUS was used to evaluate the tests. The experimental results show that the proposed blind source separation and noise suppression methods do not require any additional calculation process and effectively reduce the computational burden.

Keywords: speech recovery; time-frequency separation; adaptive noise suppression; automatic speech recognition

0 引言

在多个发声者的普适环境中,正确地调整语音识别模型以提高语音识别精度一直是一项挑战。从噪声环境中恢复干净的语音对于语音增强、语音识别和许多其它语音相关应用具有重要意义。在现实生活中,有许多噪声源,如环境、信道失真和扬声器可变性[1]。因此,已经有许多算法用于消除语音中的噪声[2-4]。这些算法大多需要额外的噪声估计,并且只适用于听觉效果而不适用于自动语音识别(ASR)。

本文介绍了一种自优化语音活动检测(VAD)算法,以及信号分离后简单但有效的噪声消除过程,以提高语音识别率。所提出的 VAD 算法的关键是不需要对干净的语音变化进行先验估计。此外,用于语音/非语音判决的阈值是由噪声本身所产生,即自优化过程。对于噪声去除过程是基于广泛已知的频谱减法(SS)[5],而不需要任何额外的模型或训练过程。最后利用 NOIZEUS 数据库将 VAD 算法与 SS 方法、零交叉能量法(ZCE)[6]、熵权法[7]进行了性能比较。

收稿日期:2018-06-10; **修回日期**:2018-07-03。

作者简介:刘 晶(1981-),女,山西忻州人,硕士研究生,主要从事语音识别方向的研究。

1 基于 SSTFT 的盲源分离(BSS)

假设 $S_n(t), n = 1, \ldots, N$ 是未知的语音源,其中 N 是发声者的数目。M-传感器麦克风阵列的布置是线性的。输出向量 $x_m(t), m = 1, \ldots, M$ 可以建模为:

$$\boldsymbol{x}(t) = \boldsymbol{A}s(t) + \boldsymbol{n}(t) \tag{1}$$

其中:$\boldsymbol{A}$ 表示混合矩阵,$\boldsymbol{x}(t) = [x_1(t), \ldots, x_M(t)]^T$ 是接收混合的向量,$s(t)$ 包含多个语音源,$\boldsymbol{n}(t)$ 是加性高斯白噪声向量,T 是转置运算符。

基于上述信号模型的空间短时傅立叶变换(SSTFT)BSS 算法的过程如下:

1)计算公式(1)中混合 $x(t)$ 的 STFT,得到每个时频(TF)点 (t,f) 的 $M \times 1$ 向量 $\boldsymbol{S}_x(t,f)$:

$$\boldsymbol{S}_x(t,f) = AS_s(t,f) + S_n(t,f) \tag{2}$$

其中:下标 S 表示 STFT 运算符。

2)基于每个时刻的标准来检测自动源 TF 点,即 TF

域中语音源的自动定位：

$$\frac{\|S_x(t,f)\|}{\max_v \|S_x(t,v)\|} > \varepsilon_0 \tag{3}$$

其中：$\|\cdot\|$ 表示范数运算符，ε_0 是用于选择自动源 TF 点的经验阈值。满足公式（3）准则的所有 TF 点都包含在集合 Ω 中。

3）基于 SSTFT 的算法的前提是估计源 N 的数量以及混合矩阵 $\boldsymbol{A}$。本文应用文献［8］中提出的方法，尝试检测一些主导的 TF 点，即具有主要能量的源点与其他源和噪声能量的点相比较。应用均值漂移聚类方法[9]在不知道源的数量的情况下对主导 TF 点进行分类。通过对同一簇中的所有 TF 点的空间矢量求平均来估计混合矩阵 $\boldsymbol{A}$，并且通过计算得到合成簇的数量来估计 N。

4）基于检测到的自动源 TF 点集 Ω 和估计混合矩阵 $\boldsymbol{A}$，应用基于子空间的方法来估计每个源的 STFT 值[10]。假设在每个自动源 TF 点属于点集 Ω 处最多存在 $K < M$ 个源。因此，公式（2）中的表达式可简化为：

$$S_X(t,f) \approx \widetilde{A}\widetilde{S}_S(t,f),(t,f) \in \Omega \tag{4}$$

其中：$\widetilde{A}$ 表示每个点 $(t,f) \in \Omega$ 处 K 个源的导向矢量，并且 $\widetilde{S}_S(t,f)$ 包含这些 K 个源的 STFT 值。每个自动源 TF 点处的 $\widetilde{A}$ 可以通过以下最小化来确定：

$$\widetilde{A} = \arg_{\hat{a}_{m_1},\ldots,\hat{a}_{m_K}} \{PS_X(t,f)\} \tag{5}$$

其中：$P = I - \widetilde{A}_m(\widetilde{A}_m^H\widetilde{A}_m)^{-1}\widetilde{A}_m^H$ 表示将正交投影矩阵转换为噪声子空间，$\widetilde{A}_m = [\hat{a}_{m_1},\ldots,\hat{a}_{m_K}]$ 包含估计混合矩阵 $\boldsymbol{A}$ 的随机 K 列。每个自动源 TF 点处的 K 个 STFT 值可以表示为：

$$\widetilde{S}_S(t,f) \approx \widetilde{A}^+ S_X(t,f),(t,f) \in \Omega \tag{6}$$

其中：$+$ 表示摩尔－彭罗斯的伪逆运算符。Ω 中的每个自动源 TF 点处的能量分配给相应源的 K 个 STFT 值。

6）每个源通过逆 STFT[11]利用公式（6）估计的 STFT 值进行恢复。

2　噪声评估

噪声和语音通常在统计学上是相互独立并且具有不同的统计特性[12]。噪声更为对称分布且始终呈现，而语音由于其有效/无效周期，通常呈现出非平稳性。语音的主动/非主动转换使语音能量更集中在语音活动期。

2.1　噪声描述

噪声和语音的不同行为使得基于语音频谱的最小/最大值来跟踪语音或噪声成为可能。具有高能量的部分更可能是语音，而低能部分更可能是噪声，语音幅度大于噪声幅度，这使得可以通过分析有噪语音的最大值来检测语音。与噪声相比，清晰语音幅度的概率分布函数在“尾部”部分更平坦，这意味着清晰语音幅度更可能远离其平均值。即使对于信噪比 $SNR = 0$ dB，也可以证明信号的峰值部分更可能来自语音。

2.2　算法推导

假设语音被不相关的加性高斯噪声所扭曲，VAD 的两个假设是：

H_0：语音缺失：$Y = N + R$；

H_1：语音呈现：$Y = S + N + R$。

其中：Y，N，S 和 R 分别表示来自盲源分离过程的频域噪声话音，噪声，干净话音和残余话音。H_0 和 H_1 的概率密度函数由下式给出：

$$H_0 : P_N(Y) = f(Y \mid H_0) = \frac{\exp\left(-\frac{|Y|^2}{2\sigma_N^2}\right)}{2\pi(\sigma_N^2 + \sigma_R^2)} \tag{7}$$

$$H_1 : P(Y) = f(Y \mid H_1) = \frac{\exp\left(-\frac{|Y|^2}{2(\sigma_N^2 + \sigma_R^2 + \sigma_S^2)}\right)}{2\pi(\sigma_N^2 + \sigma_R^2 + \sigma_S^2)} \tag{8}$$

其中：σ_N^2，σ_R^2 和 σ_S^2 表示噪声，残差和干净的语音变化。

本文还需要假设两个条件 $P_N(Y)/P(Y) < \varepsilon$ 和 $P_S \geqslant P_N$，其中 ε 是 0.01 和 0.2 之间的启发式参数。定义 $\sigma_S^2/\sigma_N^2 + \sigma_R^2 = k$，则第一个条件可以简化为：

$$|Y|^2 > -\frac{2(\sigma_S^2 + \sigma_N^2 + \sigma_R^2)}{k+2}\ln\left(\varepsilon\sqrt{\frac{1}{1+k}}\right) \tag{9}$$

因此，可以定义：

$$|Y_\varepsilon|^2 = -\frac{2(\sigma_S^2 + \sigma_N^2 + \sigma_R^2)}{k+2}\ln\left(\varepsilon\sqrt{\frac{1}{1+k}}\right) \tag{10}$$

其中：Y_ε 可以作为更直接的阈值。频率等级 VAD 标志可以表示如下：

$$flag = \begin{cases} 1, |Y|^2 > |Y_\varepsilon|^2 \\ 0, |Y|^2 \leqslant |Y_\varepsilon|^2 \end{cases} \tag{11}$$

计算语音概率密度函数，可以得到 $|Y_\varepsilon|^2$。使用公式（11）实现二进制 VAD 标志。VAD 算法适用于抑制噪声，并且可以有效区分噪声和浊音。为了提高自动语音识别率，本文仍然需要跟踪噪声能量的变化，并更新包括语音帧在内的所有帧的噪声能量。

3　噪声抑制

在 VAD 算法的设计中，由于语音信号是高度非平稳，软判决算法优于二进制判决。没有明确的边界标记发音的开始或结束，因此，判别信息被用作软判决阈值。

3.1　子带能量计算

能量计算在逐帧的基础上进行，每个帧乘以适当的窗口以减少来自快速傅立叶变换（FFT）的频率混叠。其中，50％重叠意味着产生帧长度一半的初始延迟，应仔细选择框架尺寸。假设采样率为 F_S，帧大小为 $N = 2^m$，时间分辨率为 N/F_S，频率分辨率为 F_S/N。显然，较大的帧尺寸可以提供更好的频率分辨率，但时间分辨率较差。通常，对于 F_S 为 8 000 和 16 000 Hz，对应合适的帧大小 N 分别为 256 和 512。

信号被分为 16 个子频带。当帧的大小为 256 时，第 i 个子频带的能量为：

$$S_i = \sum_{k=16i-15}^{16i} R_{i,k}^2 \tag{12}$$

其中：$R_{i,k}$ R 是第 i 个频带的第 k 个傅立叶变换系数的绝对值。整个能量中的子带由下式计算：

$$P_i = \frac{S_i}{\sum_{j=1}^{16} S_i} \tag{13}$$

利用帧能量和子带能量计算基于当前帧和噪声帧的子带能量分布概率来计算识别信息。假设随机变量 Y 可能是 $a_1,\ldots,a_k$ 的值。Y 的概率分布与假设 H_0 和 H_1 有关。设 $P_0(a_k)=P(a_k \mid H_0)$，$P_1(a_k)=P(a_k \mid H_1)$，判别信息定义如下：

$$I(P_1,P_0;Y) = \sum_{k=1}^{K} P_1(a_k)\log[P_1(a_k)/P_0(a_k)] \tag{14}$$

可以利用子带能量分布来计算当前帧和噪声帧的相似性：

$$P_0(a_k) = \frac{N_k}{\sum_{i=1}^{8} N_i} \tag{15}$$

$$P_1(a_k) = \frac{S_k}{\sum_{i=1}^{8} S_i} \tag{16}$$

3.2 阈值更新

阈值通过以下方式更新：

1）选择前 5 帧为噪声/非语音帧。

2）语音信号周期的前一帧认为是噪声帧。

3）当前一帧确定为是噪声帧时，如果当前帧满足 $|Y|^2 \leqslant |Y_\varepsilon|^2$，则当前帧将视为噪声帧。如果当前帧满足 $|Y|^2 > |Y_\varepsilon|^2$ 和 $d > T_r$，则当前帧将视为起始位置帧，并与接下来的 6 帧进行比较。如果 6 帧也满足 $|Y|^2 > |Y_\varepsilon|^2$ 和 $d > T_r$，则可以将起始位置帧作为语音周期的起始位置。否则，当前帧仍然认为是噪声帧。

4）当前一帧是语音帧时，如果当前帧满足 $|Y|^2 > |Y_\varepsilon|^2$，则仍然是语音帧。如果当前帧满足 $|Y|^2 \leqslant |Y_\varepsilon|^2$ 和 $d < T_r$，则将其归类为结束位置帧，然后与接下来的 6 帧进行比较。如果 6 个帧也满足 $|Y|^2 \leqslant |Y_\varepsilon|^2$ 和 $d < T_r$，则可以将结束位置帧作为语音周期的结束位置（也是语音的起始点）；否则，当前帧仍然是语音帧。T_r 是与最近的 5 个噪声帧的平均识别值相等的识别信息边缘值。

5）在上述确定的每个步骤中，噪声阈值将进行更新：

$$TH_n = TH_{n-1}(1-\lambda) + |Y|^2\lambda \tag{17}$$

其中：TH_n 表示第 n 帧的更新后的噪声阈值，$|Y|^2$ 为当前语音的概率分布函数值，λ 为噪声更新因子，该噪声更新因子由判别信息计算。

6）如果所有数据都已处理完毕，则自适应调整结束。

3.3 改进的 VAD 和噪声抑制

语音信号 $Y(w)$ 通常被加性高斯噪声 $N(w)$ 所破坏。理论上，可以通过估计其功率并使用以下滤波器对噪声信号进行滤波来实现最佳地消除噪声：

$$H(w) = (|Y(w)| - |N(w)|)/Y(w) \tag{19}$$

本文所提出的 VAD 将检测噪声帧，并从语音信号中减去噪声谱，试图在 ASR 的特征提取过程中保留更多的信息，并消除在特征提取和模板匹配期间提供错误信息的噪声。由于语音信号总是非平稳的，所以作出语音或噪声的二元决策变得相当困难。因此，本文通过计算语音活动评分（VAS）来估计语音，当导出的 VAS 表示语音和噪声的混合时，可以实现平滑的处理转换。

框架下的 VAS 是由两个方面决定：第一个涉及语音的可理解性，通过计数语音频带中的 Bark 频带的数量来近似量化，该频带的功率超过估计噪声的相应 Bark 频带的数量。语音频带范围从第 4 到第 14 个 Bark 频带。第二个是当前帧相对于估计噪声功率的相对功率，帧的相对功率越高，其包含语音的可能性就越大。最后的 VAS 仅仅是这两个方面的分数之和。将参数 ε 设置为 VAS 的倒数并对每个帧进行更新。连续 VAS 比固定参数提供更大的灵活性。即使需要对帧是否为纯噪声帧进行二元判决，仍然可以在一定的值上处理改变和收敛。

4 发声者和语音识别

本文将从前端特征提取，由词单元和词模板组成的训练过程以及最终识别过程阐述了整个系统。在 VAD 和噪声抑制之后，将在 ASR 系统中对处理后的语音信号进行评估。

4.1 前端特征提取

用于此识别任务的特征向量是 24 MFCC。帧窗口大小为 20 ms，语音在 16 kHz 下采样并具有 16 bit 分辨率。

4.2 分词单元生成

训练过程的第一部分要求用户记录他们大约两分钟的演讲。建议阅读语音丰富的句子，以获得更全面的分词单元。在这个实验中，用户被要求阅读一系列哈佛大学经典语录。通过使用 C 均值算法得到的 MFCC 聚类为 64 个不同的单元，大致对应于分词的集合。然后使用 4 种高斯混合模型对这些聚类中的每一个进行建模。在这个实验中，重新聚类不会在分词模板生成过程中完成。为了简化模型，进一步计算生成 64×64 的 Bhattacharyya 距离矩阵。这个过程如图 1 所示。

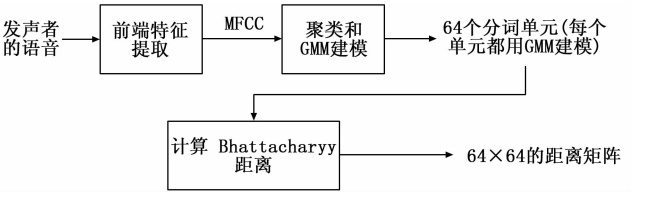

图 1 分词单元生成

4.3 语言模板生成

在这一步中，要识别的单词被记录。如图 2 所示，要求用户对单词进行发音，并且基于最大似然估计方式，模板生成将这些单词转换为子单词单元索引序列。为了避免

对单词进行过分割，只有当与相邻状态存在显着的似然差异余量时才允许改变子单词索引，从而采用过渡启发式。用户想要向系统录入每个单词都必须重复该过程。

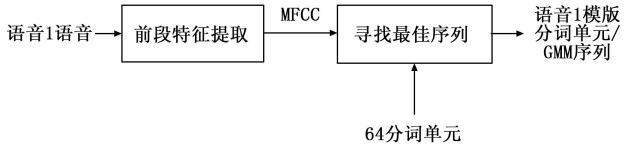

图 2　语言模板生成

4.4　匹配过程识别

假设系统中有 M 个词模板，识别过程计算由模板生成的用户输入特征向量 X 输入的概率，则选择的词是最大似然的词：

$$m^* = \arg\max p_m(X_{input}) \tag{20}$$

模板可以看作是高斯混合模型（GMM）序列，这使得随着词表模板数目的增加，$p_m(X_{input})$ 计算越来越复杂，并且很难观察所提出的 VAD 算法[13]的效果。本文使用 Bhattacharyya 距离矩阵[14]将输入特征转换为分词单元索引序列。两个概率分布 p_1 和 p_2 之间的 Bhattacharyya 距离矩阵为：

$$Bhatt(p_1 \mid p_2) = -\ln\left(\int \sqrt{p_1(x)}\ \sqrt{p_2(x)}\mathrm{d}x\right) \tag{21}$$

测试实验中的每个分词单元使用 4 个混合 GMM 建模，因此它们之间的距离为：

$$Bhatt(p_1 \mid p_2) = -\ln\left(\int \sqrt{\sum_{mix=0}^{3} p_{1,mix}(x)}\ \sqrt{\sum_{mix=0}^{3} p_{2,mix}(x)}\mathrm{d}x\right) \tag{22}$$

使用 Levenshtein 距离法计算所有 64 个分词单元的距离。通过原始模式匹配算法的识别任务的平均运行时间与模板的数目成比例地增加。对于 Bhattacharyya 编辑距离方法，当模板数量增加时，运行时间非常稳定，特别适用于实时识别系统，图 3 给出了其匹配过程。

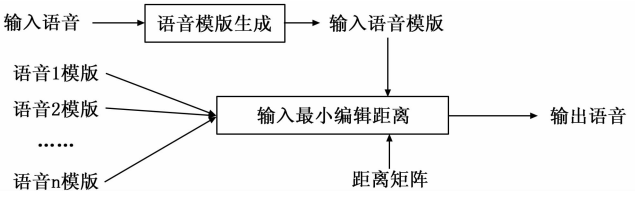

图 3　识别匹配过程

发声者识别过程与匹配过程相似，有两个主要的区别：(1) 只有在发声者识别过程中才加载所选择的说话者配置文件，由于发声者的身份已知。在发声者识别中，对发声者配置文件进行轮询，并将输入与每个发声者的相应激活关键字注册进行比较。(2) 不考虑编辑距离，发声者识别过程在给定模板中的 GMM 分布序列的情况下使用输入的后验概率。这种方法使在设定接受阈值时具有更大的灵活性。

5　实验研究

5.1　算法应用与分析

在本文的案例中，使用 2 个麦克风接收来自 4 个发声者的混合声音。在分离的信号中，通过使用自动发声者识别选择一个发声者的声音，然后进行隔离单词识别测试。

图 4 给出了使用上述过程从 2 个麦克风接收到的信号中分离出 4 个发声者声音的结果，两个麦克风的噪声混合如图 5 所示。

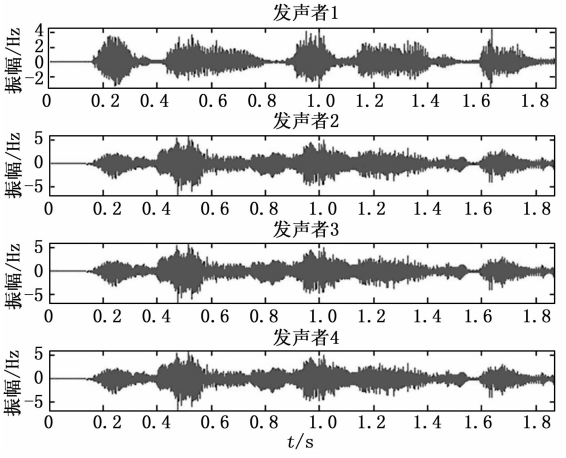

图 4　从 4 个发声者中分离出来的声音

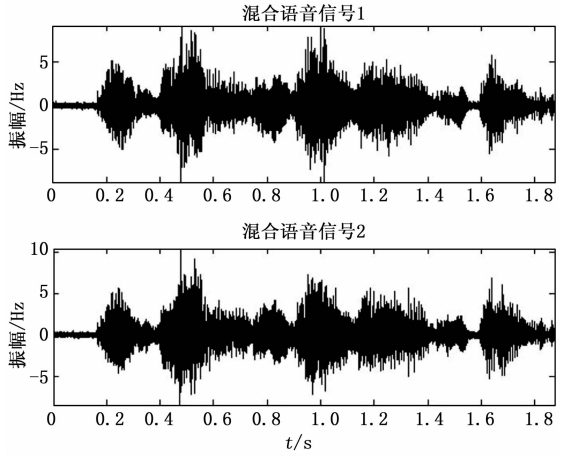

图 5　由 2 个麦克风接收的混合语音信号

由图 4 可见，4 个发声者的振幅随时间的变化趋势相似，仅在振幅的数值上有微小差异，说明 4 个发声者在信号分离过程中有语言重叠部分，但通过分词单元匹配识别，仍然能够分辨出不同发声者的语音信号。图 5 可见，2 个麦克风接收的混合语音信号随时间的变化趋势也相似，也仅在振幅的数值上有微小差异，说明 2 个麦克风接收混合语音信号过程中可以有效地对噪声进行抑制，更好的接收不同发声者的语音信号。因此，图 4 和图 5 分别从盲源分离和噪声抑制两个方面验证了所提出的语音信号识别方法的有效性。

5.2 算法评价

本文将给出 ASR 系统的结果和客观评价。首先定义信噪比：

$$SNR = 10\log \frac{\sum_{k=0}^{n-1} S^2(k)}{\sum_{k=0}^{n-1} N^2(k)} \tag{24}$$

其中：$S(k)$ 是语音信号能量，$N(k)$ 是噪声能量。在这个 ASR 实验中，车辆和餐厅的噪声来自 NOIZEUS 噪声数据库。

使用 ASR 系统进行语音识别测试，提出的盲盲源分离算法是在 VAD 算法之前实现。文献 [15] 指出，与麦克风相比，语音源更少的系统可以获得更好的分离效果。将 SS 方法、ZCE 方法和基于熵的方法的性能与在车辆和餐馆噪声环境中提出的 VAD 噪声抑制方法进行了比较。对于信噪比 SNR 为 0，5 和 10 dB 的情况，在表 1 中给出了车辆噪声与餐厅噪声内的语音识别精度实验结果，其中，括号内为餐厅噪声环境下的识别率。

表 1 车辆噪声与餐厅噪声内的语音识别精度

SNR	0 dB	5 dB	10 dB
SS	60.43(58.33)	79.86(75.94)	92.23(88.01)
ZCE	76.77(70.52)	85.42(79.41)	93.91(87.68)
信息熵	84.09(83.56)	87.18(85.09)	93.95(91.82)
VAD	86.59(84.72)	88.53(85.42)	93.98(92.08)

与在 VAD 算法中实现精度最高的（基于熵的方法）相比，信噪比 $SNR=0$ dB 的情况下的相对改善达到 2.5%(1.2%)，而在信噪比为 5 dB 的情况下，改善率为 1.4%(0.33%)。整个 ASR 系统以逐帧的方式工作，满足大多数嵌入式电子应用的实时操作。除了在实验中使用的噪声，使用 NOIZEUS 的街道噪音也可以获得相似的结果。

6 结论

本文提出并实现了一种用于普适语音环境的完整语音恢复算法。它能有效地从多个发声者的混合语音中恢复个体发声者的声音。所提出的算法的关键特征是不需要先验信息的来源数量和干净语音方差的估计。用于抑制噪声的阈值是从语音本身生成，从而导致适应变化环境的理想能力。此外，所提出的盲源分离和噪声抑制方法不需要任何额外的计算过程，有效地减少了计算负担。最后，所提出的系统可以容易地在普适语音环境中实现。

参考文献：

[1] 王山海，景新幸，杨海燕. 基于深度学习神经网络的孤立词语音识别的研究 [J]. 计算机应用研究，2015，32 (8)：2289－2291，2298.

[2] Yugandhar D，Nayak S K. A heuristic speech de－noising with the aid of dual tree complex wavelet transform using teaching－Learning based optimization [J]. International Journal of Engineering and Technology (IJET)，2016，8 (5)：1967－1980.

[3] 李 洋，景新幸，杨海燕. 基于改进小波阈值和 EMD 的语音去噪方法 [J]. 计算机工程与设计，2014，35 (7)：2462－2466.

[4] Narayanam R. Efficient de－noising performance of a combined algorithm of translation invariant (TI) wavelets and independent component analysis over TI wavelets for speech－auditory brainstem responses [J]. Procedia Computer Science，2015，54：829－837.

[5] 邓利娜，黄晓革. 基于频谱减法的语音去噪算法研究 [J]. 电子设计工程，2011，19 (8)：113－115.

[6] Wang Z，Bi G. A Voice Activity Detector Based on Noise Spectrum Adaptation and Discrimination Information for Automatic Speech Recognition System [A]. 2014 5th International Conference on Intelligent Systems，Modelling and Simulation (ISMS) [C]. IEEE，2014：301－305.

[7] 宋 静，张雪英，孙 颖，等. 基于模糊综合评价法的情感语音数据库的建立 [J]. 现代电子技术，2016，39 (13)：51－54，58.

[8] May T. Robust speech dereverberation with a neural network－based post－filter that exploits multi－conditional training of binaural cues [J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2018，26 (2)：406－414.

[9] 张 睿，熊金虎，汪东兴，等. 基于体素邻域信息的均值漂移聚类算法检测 fMRI 激活区 [J]. 江苏大学学报（自然科学版），2016，37 (5)：556－561.

[10] Laljahan L，Iqbal S，Sarwar F，et al. The relationship of generalized fractional Hilbert transform with fractional Mellin and fractional Laplace transforms [J]. American Scientific Research Journal for Engineering，Technology，and Sciences (ASRJETS)，2017，38 (2)：90－97.

[11] Jaganathan K，Eldar Y C，Hassibi B. STFT phase retrieval：Uniqueness guarantees and recovery algorithms [J]. IEEE Journal of selected topics in signal processing，2016，10 (4)：770－781.

[12] 陈振锋，吴蔚澜，刘 加，等. 基于 Mel 倒谱特征顺序统计滤波的语音端点检测算法 [J]. 中国科学院大学学报，2014，31 (4)：524－529.

[13] 晏光华. 一种基于 MMSE－LSA 和 VAD 的语音增强算法 [J]. 移动通信，2014，38 (10)：59－62，66.

[14] 邹焕新，秦先祥，周石琳，等. 基于区域 Bhattacharyya 相似度的 SAR 图像地物分类方法 [J]. 系统工程与电子技术，2016，38 (12)：2752－2757.

[15] Williamson D S，Wang Y，Wang D L. Complex ratio masking for monaural speech separation [J]. IEEE/ACM Transactions on Audio，Speech and Language Processing (TASLP)，2016，24 (3)：483－492.