

一种基于能零积的改进端点检测算法

韦金泉¹ 罗丽燕^{1,2*} 王玫³ 刘争红^{1,2} 何锐彬¹ 莫清麟¹

¹(桂林电子科技大学认知无线电与信息处理教育部重点实验室 广西 桂林 541004)

²(桂林电子科技大学广西无线宽带通信与信号处理重点实验室 广西 桂林 541004)

³(桂林理工大学信息科学与工程学院 广西 桂林 541007)

摘要 针对基于短时能量和短时过零率的传统端点检测算法在信噪比低于10 dB时其准确性明显下降的问题,提出一种基于声谱图特征增强的能零积端点检测改进算法。首先,对音频信号进行谱减法降噪,提升信号的信噪比;其次,依次对声谱图进行腐蚀、二值化和膨胀处理,以实现声谱图特征增强;最后,提取能零积特征,并利用双阈值端点检测算法对音频信号进行端点检测。实验结果表明,该算法在不同信噪比条件下可以有效捕获有环境异常音片段,具有良好的鲁棒性。

关键词 端点检测 能零积 低信噪比 声谱图 特征增强

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.07.032

AN IMPROVED ENDPOINT DETECTION ALGORITHM BASED ON ENERGY-ZERO-PRODUCT

Wei Jinquan¹ Luo Liyan^{1,2*} Wang Mei³ Liu Zhenghong^{1,2} He Ruibin¹ Mo Qinglin¹

¹(Provincial Ministry of Education Key Laboratory of Cognitive Radio and Signal Processing, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China)

²(Guangxi Key Laboratory of Wireless Broadband Communication and Signal Processing, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China)

³(College of Information Science and Engineering, Guilin University of Technology, Guilin 541007, Guangxi, China)

Abstract To tackle the problem of the significantly reduced accuracy of the traditional endpoint detection algorithm based on short-term energy and short-term zero-crossing rate when the signal-to-noise ratio is less than 10 dB, an improved energy-zero-product endpoint detection algorithm based on the enhancement of the Sonogram feature is proposed. The spectral subtraction method was utilized to reduce the noise of the audio signal and thus improving the signal-to-noise ratio. Perform erosion, binarization and dilation were processed on the sonogram to achieve the enhancement of the spectrogram characteristics. The energy-zero-product feature was extracted and the dual-threshold endpoint detection algorithm was used to conduct endpoint detection on the audio signal. Experimental results show that the algorithm can effectively capture the fragments with ambient abnormal sound under different signal-to-noise ratio conditions, and has good robustness.

Keywords Endpoint detection Energy-zero-product Low signal-to-noise ratio Sonogram Feature enhancement

收稿日期:2021-03-04。国家自然科学基金项目(61771151);广西重点研发计划项目(桂科AB17292058);广西科技基地和人才专项项目(桂科AD20159018);广西自然科学基金面上项目(2020GXNSFAA159004);桂林电子科技大学研究生教育创新计划项目(2021YCX024)。韦金泉,硕士生,主研领域:声音信号处理。罗丽燕,副教授。王玫,教授。刘争红,讲师。何锐彬,硕士生。莫清麟,本科生。

0 引言

在自然界和日常生活中,人类创造的语音和音乐之外的声音统称为一般音频或环境音,环境异常音通常指的是一段对人类具有警示作用或特定意义的连续声音事件,时间可长可短,例如尖叫声、鞭炮声、枪声、鸣笛声和玻璃破碎声等^[1]。端点检测(Endpoint Detection, ED)的作用则在于确定音频流中环境异常音的起止位置(Event onsets and offset)^[2]。在实际应用中,端点检测需要在各种背景声音的干扰下,在连续音频流中找到环境异常音的起始位置,然后再进行下一步的分析和处理,其常被用于公共场所安全监测、城市噪音监控^[3]和室内环境下辅助老年人生活等领域监测系统的音频信号处理中。

在嘈杂的环境中检测出环境异常音的起点和终点,不仅可以减少采集的数据量,消除背景噪声和静音段,从而减少音频信号的计算量和特征提取的时间^[4],还可以有效提高环境异常音识别分类的准确性^[5]。因此,在嘈杂的环境中准确检测环境异常音的开始和结束位置有利于提高音频信号处理系统的性能。基于时域参数短时能量和短时过零率的双门限判决方法^[6]是最早提出的经典端点检测算法,使用也最为广泛,当信噪比大于 10 dB 时,其具有较高的准确率,但当信噪比小于 10 dB 时其准确率会随着信噪比的下降而明显降低^[7],因此近年来许多人员对嘈杂环境中的端点检测进行了研究。为了获得更好的结果,人们开始关注频谱和倒谱中的特征。一些研究人员利用环境异常音与背景噪声在频域中的差异较大的特性,提出了基于子带频谱方差^[8-9]和子带谱熵^[10-11]的端点检测,并且频谱在经对数运算后再进行傅里叶逆变换即可转变到倒谱域,于是随后又提出基于倒谱距离^[12]、MFCC 欧氏距离^[13]和 MFCC 余弦相似^[14]的端点检测。上述基于频谱和倒谱特征的方法对实验后的各种噪声具有良好的鲁棒性,但是其误判率容易随着信噪比的降低而逐步提升,于是以提高音频信号信噪比的端点检测算法出现了。文献[15-16]针对信噪比低的问题,采用多窗口谱估计的谱减法对环境音频信号进行降噪处理,提升其信噪比,然后再利用传统的端点检测算法进行检测。实验结果表明,在不同信噪比的条件下,多窗谱估计谱减法都能较好地削弱音频信号中的噪声,从而使得基于提升信噪比的端点检测算法具有良好的鲁棒性。尽管多窗谱估计谱减法的降噪效果已经非常显著,但是由于谱减的效果是作用在全段音频信号的,故

而其在降噪的同时亦会削弱环境异常音的信号特征,进而影响到检测结果的判定,对此文献[17-18]从图像处理的角度出发,对音频信号进行傅里叶变换得到其声谱图,然后再利用图像处理中的腐蚀算法和膨胀算法相继对声谱图进行处理,不仅削弱了随机分布且能量较弱的残留噪声,而且还增强了连续分布的环境异常音。虽然腐蚀和膨胀算法可以抑制噪声并对环境异常音的特征进行增强,但其在低信噪比情况下的效果仍不能令人满意。

本文针对低信噪比环境下传统端点检测算法准确率明显下降的问题,提出了一种基于声谱图特征增强的能零积端点检测改进算法。其通过获取降噪后的声谱图并对其依次进行腐蚀、二值化和膨胀处理,以实现声谱图特征增强,最后,提取能零积特征进行端点检测。提出的改进算法在 -10 ~ 10 dB 的信噪比区间内,较之基于频带方差、能熵比和倒谱距离的传统端点检测算法有更高的命中率以及更低的过分割率,故而改进后的端点检测算法较之传统的端点检测算法更具优越性。

1 传统的能零积端点检测算法

传统能零积端点检测算法主要使用短时能量和短时过零率这两个参数的乘积来作出相应的判断。在室外相对安静环境下通过蓝牙音箱播放并录制尖叫声信号,如图 1 所示。可以看出,短时能量与短时过零率在有尖叫声出现的地方其对应的都是向上凸起的,而在只包含背景噪声的地方其对应的参数曲线则都是向下凹陷的。因此,为了使环境异常音和背景音之间的差异更加突出,可以将短时能量和短时过零率的乘积作为特征进行环境异常音起始端点的检测,这样更容易检测到环境异常音的起点和终点。

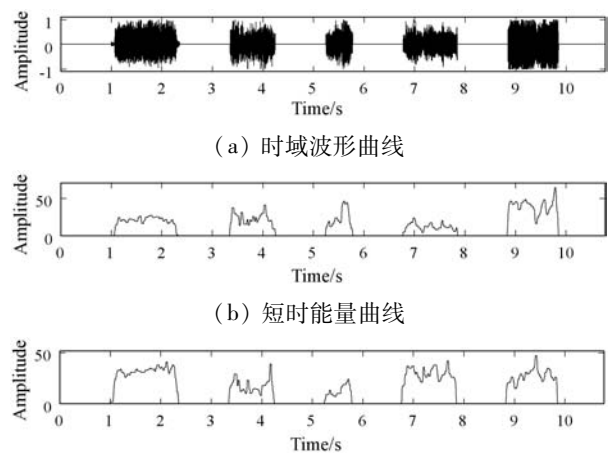


图 1 鸣笛声音频信号

设带噪的环境异常音信号为 $x(n)$, 对声音 $x(n)$ 进行分割处理得到 $x(i, m)$ (i 表示 i 帧, m 表示第 m 点), 之后对每帧声音信号加汉明窗来防止频谱泄漏和边界信息丢失。本文以 256 个采样点组成一帧音频帧, 并进行 $1/2$ 帧长度即 128 个采样点的帧重叠。信号中第 i 帧的短时对数能量 $E(i)$ 如式(1)所示。

$$E(i) = \log_{10} \left(1 + \sum_{m=1}^N |x(i, m)|^2 \right) \quad (1)$$

信号中第 i 帧的短时过零率 $Z(i)$ 如式(2)所示。

$$Z(i) = \sum_{m=1}^N |\text{sign}[x(i, m)] - \text{sign}[x(i, m-1)]| \quad (2)$$

式中: $\text{sign}[x(i, m)]$ 的定义如式(3)所示。

$$\text{sign}[x(i, m)] = \begin{cases} 1 & |x(i, m)| \geq 0 \\ -1 & |x(i, m)| < 0 \end{cases} \quad (3)$$

第 i 帧的能零积 $W_{EZ}(i)$ 如式(4)所示, 其是结合式(1)和式(2)得到的。

$$W_{EZ}(i) = E(i) \times Z(i) \quad (4)$$

在实际应用时为了提升信号的抗干扰能力, 常利用多窗谱估计谱减法对其进行降噪, 其是一种利用噪声与环境异常音在频域中的差异性进行降噪处理的方法, 是目前最常用的一种方法。谱减法的降噪效果好, 且其不涉及复杂的计算, 运算量较小, 适合于真实环境场景下的实时实现。对分帧后的信号 $x(i, m)$ 进行 FFT, 可求得其幅度谱 $|X(i, k)|$, 并通过增益因子 $g(i, k)$ 可求得如式(5)所示的谱减后的幅度谱 $|\hat{X}(i, k)|$ 。

$$|\hat{X}(i, k)| = g(i, k) \times |X(i, k)| \quad (5)$$

式中: i 表示第 i 帧; k 表示第 k 条谱线; $g(i, k)$ 为增益因子。设 $P_n(k)$ 为估计噪声的功率谱密度, $P_s(i, k)$ 为多窗谱估计的功率谱密度, 则增益因子的计算式为:

$$g(i, k) = \begin{cases} \frac{P_s(i, k) - \alpha P_n(k)}{P_s(i, k)} & P_s(i, k) \geq \alpha P_n(k) \\ \frac{\beta P_n(k)}{P_s(i, k)} & P_s(i, k) < \alpha P_n(k) \end{cases} \quad (6)$$

式中: α 为过减因子, 其值的大小对谱减结果有最直接的影响, 如果其值过大, 则可能会引起音频信号的失真, 在实验中的令 $\alpha = 3$; β 为增益补偿因子, 且 $\beta = 0.001$ 。

2 改进端点检测算法

2.1 算法概述

大多数音频信号包含噪声, 当信噪比 SNR (信噪比) 大于 10 dB 时, 传统的能零积端点检测方法具有较

好的检测效果; 然而当信噪比降低时会容易发生误判, 将一些能量较大的非平稳噪声当作环境异常音, 又或者同一个异常音被分割为多份, 导致端点检测算法的性能下降。针对低信噪比或非平稳噪声干扰下端点检测性能低的问题, 本文提出了一种改进的能零积端点检测, 其算法框图如图 2 所示。首先使用多窗函数谱估计谱减法对包含噪声的信号进行降噪, 在很大程度上减少了环境噪声对有用信号的干扰; 其次, 为了加强噪声与环境异常音之间的特征差异性, 依次对声谱图进行腐蚀、二值化和膨胀处理; 最后, 将增强后的声谱图转换回时域信号, 提取时域上的短时过零率特征与短时能量特征并计算能零积, 进行端点检测算法判决。

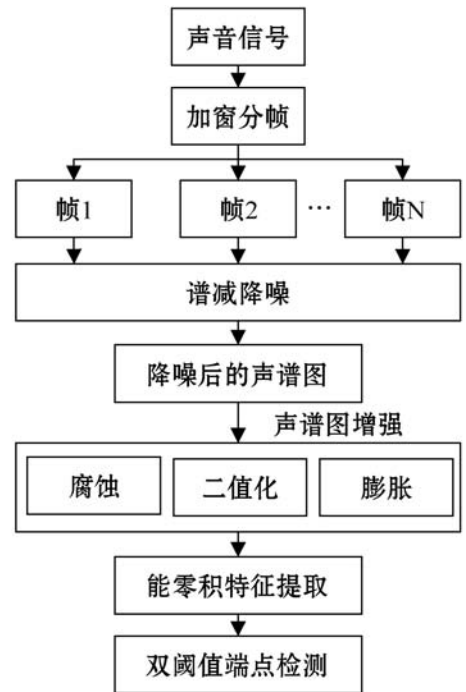


图2 改进的端点检测算法

2.2 声谱图特征增强

尽管在经过多窗谱估计谱减降噪后极大地削弱了所录制信号的噪声, 但是在声谱图上仍然残留着结构微小、或强或弱的噪声, 这将极大地影响到环境异常音起始边界的确定; 此外由于谱相减是作用在整个声谱图上的, 故而在削弱噪声结构信息的同时, 有用声音的结构信息也遭受到一定程度的衰减, 这将会影响端点检测结果的准确性。因此, 本研究依次使用腐蚀、二值化和膨胀对声谱图进行处理, 从而实现声谱图的特征增强。下面分别对腐蚀、二值化和膨胀算法进行介绍。

1) 声谱图腐蚀。腐蚀算法常被用于图像的数据降噪, 其可以通过减弱或消除特定的二维结构以实现二维数据的降噪。在声音信号经过谱减法降噪之后的声谱图中, 会残存着一些随机噪声, 如图 3(a) 所示, 虽然这些噪声结构的尺寸普遍较小, 但能量通常较大,

会在声谱图膨胀处理的时候被放大进而对端点检测的结果产生影响。为了削减甚至消除这些具备特定结构的噪声,采用二维噪声腐蚀算法对声谱图进行处理,其腐蚀公式如式(7)所示。

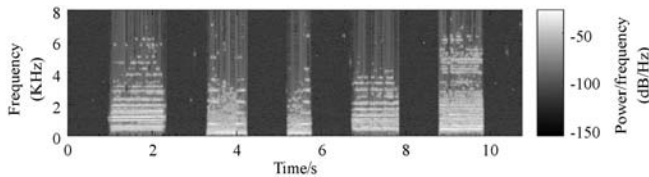
$$Y_e(i, k) = \min \{ \hat{X}(i+i', k+k') - a(i', k') \} \quad (7)$$

式中: $a(i', k')$ 是结构元素,且 $(i+i'), (k+k') \in D_{\hat{X}}$; $i', k' \in D_a$ 。 $D_{\hat{X}}$ 是 $\hat{X}(i+i', k+k')$ 的定义域, D_a 是 $a(i', k')$ 的定义域。平移参数 $(i+i'), (k+k')$ 必须在 $\hat{X}(i+i', k+k')$ 的定义域内,且 i', k' 必须在 $a(i', k')$ 的定义域之内。

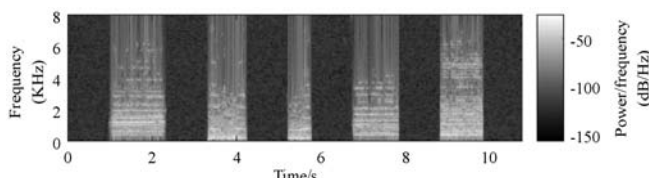
在形态学图像处理上,用于降噪(即腐蚀)的结构元素通常需要满足一定的条件。对于声谱图中的噪声结构,当其与腐蚀处理中所用到的结构元素相似或者一致时,噪声可以很好地被削弱甚至消除,所以在选择结构元素类型时,应尽可能让其与声谱图噪声结构的形状保持一致;此处腐蚀的作用是在保留下环境异常音结构的同时尽可能地削弱噪声结构,因此结构元素的尺寸应小于非噪声结构的尺寸,但同时又应大于噪声结构的尺寸^[20]。经过查阅相关文献资料与验证,包含白噪声的音频信号经过降噪后的残留背景噪声为雪花状^[18],此外圆形的结构元素也已被证明能较好地滤除噪声^[21],故而本研究选择了具备平面圆盘形的元素结构,通过测试分析将其半径指定为2,二维噪声腐蚀算法的结构元素 $a(i', k')$ 被定义为式(8)。

$$a(i', k') = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (8)$$

式中: $a(i', k')$ 是由 MATLAB 中的 $strel('disk', R)$ 生成的,其可以创建一个指定半径 R (非负整数)的平面圆盘形的结构元素,即结构元素的所有像素是由到中心像素距离小于等于 R 的像素组成。如图3所示。



(a) 腐蚀前的声谱图



(b) 腐蚀后的声谱图

图3 腐蚀仿真结果

2) 声谱图二值化。经过谱减法降噪和腐蚀后的声谱图在非异常音段留下了许多的微小杂波,其能量通常较小,且容易在声谱图膨胀处理过程中被放大从而影响到端点检测的结果;此外,在实验的过程中发现,声谱图中各像素点间值差异性过大从而容易导致端点检测对阈值的大小较为敏感,因此在形态学图像处理开运算中加入了声谱图二值化,以降低上述两种因素对端点检测结果的影响。

为了凸显声谱图中感兴趣目标的轮廓,即凸显环境异常音的二维结构,本研究对腐蚀过后的声谱图进行二值化。所谓图像二值化,即通过设置一个局部或者全局阈值 T ,对于灰度图像像素矩阵 I ,将 I 中所有高于阈值 T 的灰度值替换为1,并将其剩余部分灰度值替换为0。在对图像的数字化处理过程中,因其可以让整幅图像呈现出清晰黑白的效果,有助于对图像的下一步处理,所以其在图像处理中一直以来都占据着非常重要的位置。

对于经过腐蚀处理后的二维声谱矩阵 $Y_e(i, k)$,在 MATLAB 中调用 $graystresh(Y_e(i, k))$ 来计算全局阈值 T ,阈值函数采用 Otsu 方法,通过选择阈值来最小化黑白像素的类内方差,该阈值 T 可用于将强度图像转换为二值图像,其级别是位于范围 $[0, 1]$ 内的标准化强度值。拿到全局阈值 T 后,在 MATLAB 中利用 $imbinarize(Y_e(i, k), T)$ 函数来获取二值转化的二维声谱矩阵 $Y_b(i, k)$ 。

3) 声谱图膨胀。经过谱减法降噪与二维噪声腐蚀后的声谱图,干扰噪声已经被很好地抑制,然而谱减法降噪与声谱图腐蚀的作用是双重的,即噪声和环境异常音信号被同时削弱,为了对有用的声音信号进行补偿,利用膨胀算法对二维声谱图进行增强。针对二值化后的声谱图矩阵 $Y_b(i, k)$ 进行膨胀处理,设膨胀后的声谱图矩阵为 $Y_d(i, k)$,则膨胀公式如式(9)所示。

$$Y_d(i, k) = \max \{ Y_b(i+i'', k+k'') + b(i'', k'') \} \quad (9)$$

式中: $b(i'', k'')$ 是结构元素, $(i+i''), (k+k'') \in D_{Y_b}$; $i'', k'' \in D_b$ 。 D_{Y_b} 是 $Y_b(i+i'', k+k'')$ 的定义域, D_b 是 $b(i'', k'')$ 的定义域。

此处膨胀算法的目的是,在相对抑制噪音结构的前提下,尽可能地增强环境异常音结构。在本研究的膨胀处理过程中使用的仍是具备平面圆盘形的元素结构,并将其半径指定为10,则结构元素 $b(i'', k'')$ 如式(10)所示,这里的 $b(i'', k'')$ 也是由 MATLAB 中的 $strel('disk', R)$ 生成的。

$$b(i'', k'') = \begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 1 & \cdots & 1 & \cdots & 1 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \end{pmatrix} \quad (10)$$

2.3 评价指标

在一些文献中常使用命中率(Hit Rate, HR)、过分割率(Over Segmentation, OS)来评价端点检测的性能,对应计算公式分别如式(11)和式(12)所示。这两种评价中前者的使用频次是最高的,然而仅仅使用命中率来评判端点检测算法的性能是不够严谨的。由于有噪声信号的干扰,容易造成端点检测的误判、漏判,故而除了命中率外还需要考虑其对应的过分割率。当过分割率为正的时候,证明算法检测到了比预期更多的环境异常音;而当过分割率为负的时候,则证明了端点检测算法检测到了比预期更少的环境异常音。

$$R_{HR} = \frac{N_{hit}}{N_{ref}} \quad (11)$$

$$R_{OS} = N_{alg}/N_{ref} - 1 \quad (12)$$

式中: N_{ref} 表示在实验音频信号中实际包含并被标注的环境异常音的个数,且实验中的 $N_{ref} = 113$, N_{hit} 表示算法检测结果与手动标记结果一致的个数, N_{alg} 则表示算法所检测出环境异常音的个数。

2.4 偏移量

声谱图在经过腐蚀、二值化和膨胀处理后,环境异常音的特征已经得到了增强而噪声也得到了削弱,将特征增强后的声谱图由频域转换为音频时域信号,并分别计算其短时能量和短时过零率,进而得到新的能零积,最后通过单参数双阈值的端点检测算法进行检测并得出结果。为了方便实验,通常需要手动标记研究对象在音频信号中的起始位置,然后通过比对算法标记与手动标记的结果,验证算法的性能。手动给音频信号中环境异常音起始端点标记的过程并不是绝对严谨的,会产生人为误差;端点检测算法在对音频信号进行降噪与增强处理的过程也会产生一定的误差。如果直接以算法标记的位置边界与手动标记的位置边界进行严格比对,上述存在的两种误差可能会使算法失效,故而与端点检测算法所得边界位置进行判决比较的对象应为一个范围值。当前,大多数方法使用的是通过人为确定环境异常音起始边界位置的具体所在并打上对应的标记 ϕ ,然后向前或者向后移动一个固定偏移量 Δ ,则落入范围 $(\phi - \Delta, \phi + \Delta)$ 内的分割边界被认为是正确的分割边界^[22],偏移量的大小将直接影响端点检测的命中率。

3 实验与结果分析

实验中,本课题所提出的改进端点检测算法对比其他三种端点检测算法,并比较它们的命中率以及过分割率。这三种方法分别是频带方差端点检测、能熵比端点检测和倒谱距离端点检测。实验从 Google 数据集中拿到鸣笛声数据,在没有额外添加高斯噪声的前提下,对不同偏移量情况下的端点检测结果进行了仿真,用于确定最优的偏移量大小并将其用于后续其他实验;随后,对其叠加不同信噪比的高斯白噪声,测试不同端点检测算法在不同信噪比条件下的性能;最后通过分析端点检测结果,得出所提出的改进算法在低信噪比条件下比其他三种端点检测算法更具鲁棒性。

3.1 偏移量选择

实验从 Google 数据集中拿到鸣笛声数据,手动标记了实验对象中所有鸣笛声的起始点位置,在没有额外添加高斯噪声的前提下,然后分别用频带方差、能熵比、倒谱距离以及提出的改进算法对音频信号进行处理,它们对应偏移量和命中率的检测结果见图4。容易看出,在偏移量 $\Delta = 15$ 的时候,能熵比和所提出算法的曲线已经趋于平稳,倒谱距离的曲线在偏移量 $\Delta = 25$ 的时候趋于平稳,而频带宽度的曲线则在 $\Delta = 30$ 附近趋于平稳,此时再增加偏移量其对应的命中率也不会出现太大的变化。理论上偏移量的值越小,对应的端点检测结果越精确,即过大的偏移量会影响端点检测的精准度,故而综合考虑偏移量 $\Delta = 30$ 。

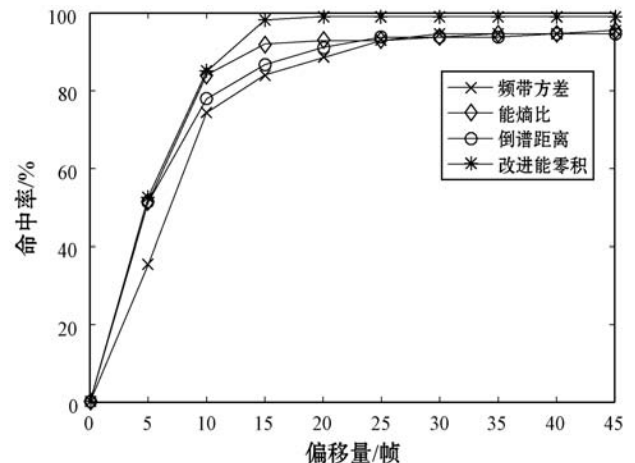


图4 偏移量与命中率的关系曲线

3.2 与传统算法的对比实验

为了验证在低信噪比情况下所提出的改进端点检测算法的鲁棒性,对 Google 音频数据集中的音频信号叠加不同信噪比的高斯白噪声,得到不同信噪比情况

下的命中率和过分割率,并与其他三种端点检测算法进行比较。本研究对传统算法和改进算法的不同信噪比情况下的命中率关系曲线和过分割率关系曲线分别见图5和图6。从图5可以看出,在SNR小于5的低信噪比情况下,本课题所提出的改进算法较之其他三种算法有更高的命中率,故而所提出算法更具在低信噪比情况下更具优越性。从图6中则可以知道,所提出算法的过分割率要远远低于其余三种算法,其在 $-10 \sim 5$ dB低信噪比环境下的最高过分割率也仅为23.9%,不足四分之一;反观其余三种算法,在低信噪比条件下最低的过分割率也高达38.9%,由此可见本课题所提改进算法的可靠性。

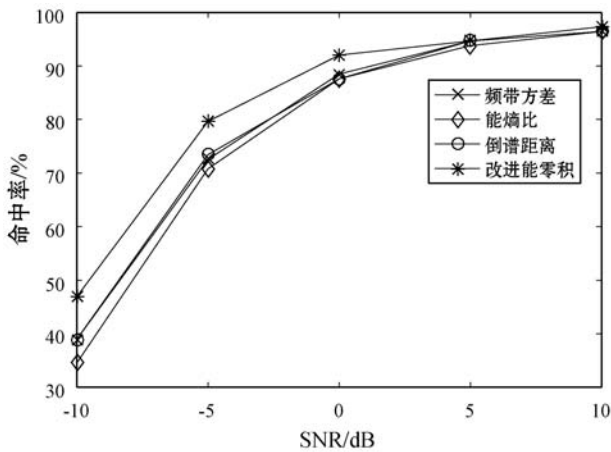


图5 命中率曲线

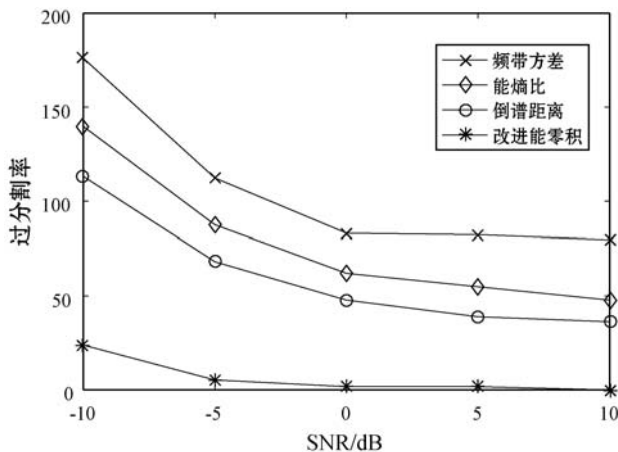


图6 过分割率曲线

为了对四种算法在不同信噪比下命中率和过分割率有一个整体的认识,实验中计算了四种算法在不同信噪比下的平均命中率和平均过分割率见表1。从表1中可以看到在同等条件下,能熵比的平均命中率最低,为76.6%;而频带方差和倒谱距离的平均命中率是一样的,均为78.2%;本文所提出改进算法的平均命中率最高,为82.1%,是四种算法中唯一平均命中率达到80%以上的,由此可见所提改进算法的优越性。从过分割率来看,从频带方差至所提改进算法,呈

现依次下降的趋势;并且除改进算法外的三种算法的平均过分割率均在50%以上,而频带方差算法的结果更是高达106.7%之上,而改进算法的结果仅为6.7%,是四者之中最低且唯一平均过分割率在个位数的,由此验证了改进算法的可靠性。

表1 平均命中率与过分割率(%)

指标	频带方差	能熵比	倒谱距离	改进算法
平均命中率	78.2	76.6	78.2	82.1
过分割率	116.7	78.4	60.9	6.7

为了探索所提改进算法与其他三种传统算法在过分割率上巨大的差异性,截取了在SNR=5条件下改进算法和传统算法的部分检测结果进行研究,如图7所示。可以看出,图7(c)-(e)的传统端点检测结果中,皆出现了将噪声判决为环境异常音的现象,唯独(f)中没有。在前面声谱图腐蚀部分有说到,在声音信号经过谱减法降噪之后的声谱图中,会残存着一些随机噪声,这些噪声结构尺寸通常较小,但其能量往往较大,具体图形见图3(a)。正是因为这些噪声干扰到了端点检测,为了削减甚至消除这些具备特定结构的噪声,本研究采用了二维噪声腐蚀算法对声谱图进行处理。通过分析端点检测结果,验证所以提出的改进端点检测算法要比传统的端点检测算法更具可靠性。

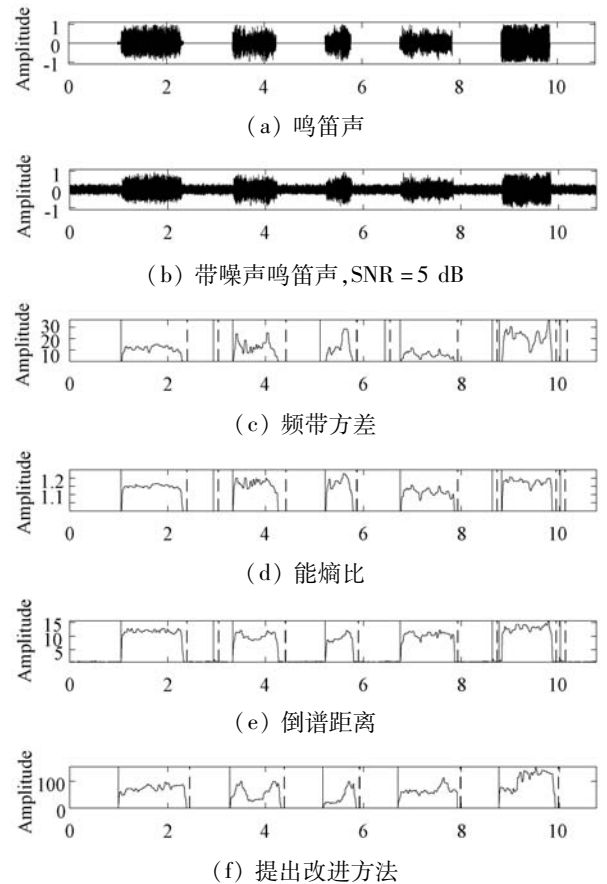


图7 端点检测结果

4 结 语

大多数音频信号包含噪声,当信噪比 SNR(信噪比)大于 10 dB 时,传统的能零积端点检测方法具有较好的检测效果;然而当信噪比降低时会容易发生误判,将一些能量较大的非平稳噪声当作环境异常音,又或者同一个异常音被分割为多份,导致端点检测算法的性能下降。对此,本研究在传统能零积端点检测算法上进行了改进,利用形态学图像处理的方法,依次对声谱图进行腐蚀、二值化和膨胀处理,实现了声谱图的特征增强。从实验结果来看,本研究提出的基于声谱图特征增强的改进能零积端点检测算法,其在信噪比低于 10 dB 时较之传统算法更具优越性。在 -10 ~ 10 dB 信噪比区间内的平均命中率分别比传统频带方差、能熵比和倒谱距离的端点检测算法前提升了 3.9%、5.6% 和 3.9%;平均过分割率则分别比传统频带方差、能熵比和倒谱距离的端点检测算法降低了 100%、71.35% 和 54.2%。综上所述,改进后的端点检测算法较之传统的端点检测算法有良好的优越性。

当前研究对象所添加的噪声为白噪声,然而在实际场景中通常存在的都是非平稳噪声,故而下一步的研究方向为非平稳噪声低信噪比环境下的端点检测。

参 考 文 献

- [1] 李伟,李硕. 理解数字声音—基于一般音频/环境声的计算机听觉综述[J]. 复旦学报(自然科学版),2019,58(3):269-313.
- [2] Zhang T, Shao Y, Wu Y Q, et al. An overview of speech endpoint detection algorithms[J]. Applied Acoustics,2020,160:107133.
- [3] Foggia P, Petkov N, Saggese A, et al. Audio surveillance of roads: A system for detecting anomalous sounds[J]. IEEE Transactions on Intelligent Transportation Systems,2015,17(1):279-288.
- [4] 熊梓奥,苍岩. 育肥猪舍中的猪叫声端点检测算法[J]. 应用科技,2020,47(5):79-85.
- [5] 孙一鸣,吴杨扬,李平. 基于改进双门限法的语音端点检测研究[J]. 长春理工大学学报(自然科学版),2016,39(1):92-95.
- [6] 鲁运耀,周妮,肖珂,等. 强噪声环境下改进的语音端点检测算法[J]. 计算机应用,2014,34(5):1386-1390.
- [7] 曾剑飞. 低信噪比条件下的语音端点检测算法研究[D]. 广州:华南理工大学,2019.
- [8] 王威,胡桂明,杨丽,等. 基于谱减法和均匀子带频带方差法的端点检测[J]. 电声技术,2016,40(5):40-43,66.
- [9] Zhang C, Dong M. An improved speech endpoint detection based on adaptive sub-band selection spectral variance[C]//35th Chinese Control Conference,2016:5033-5037.
- [10] Pang J. Spectrum energy based voice activity detection [C]//7th Annual Computing and Communication Workshop and Conference,2017:1-5.
- [11] Zhang Y, Wang K J, Yan B. Speech endpoint detection algorithm with low signal-to-noise based on improved conventional spectral entropy[C]//12th World Congress on Intelligent Control and Automation,2016:3307-3311.
- [12] 卜玉婷,曾庆宁,郑展恒. 一种低信噪比环境下的语音端点检测算法[J]. 声学技术,2020,39(5):592-602.
- [13] 韩云霄,邵清,符玉襄,等. 复杂噪声中基于 MFCC 距离的语音端点检测算法[J]. 计算机工程,2020,46(3):309-314.
- [14] Cao D Y, Gao X, Gao L. An improved endpoint detection algorithm based on MFCC cosine value[J]. Wireless Personal Communications,2017,95(3):2073-2090.
- [15] Bao T, Li Y X, Xu K, et al. An improved endpoint detection algorithm based on improved spectral subtraction with multi-taper spectrum and energy-zero ratio[C]//14th International Conference on Intelligent Computing,2018:266-275.
- [16] 张涛,章小兵,朱明星. 低信噪比环境下改进的倒谱距离语音端点检测算法[J]. 电声技术,2017,41(7):108-112,125.
- [17] 吴迪,赵鹤鸣,陶智,等. 低信噪比下采用感知语谱结构边界参数的语音端点检测算法[J]. 声学学报,2014,39(3):392-399.
- [18] Wu D, Tao Z, Wu Y B, et al. Speech endpoint detection in noisy environment using spectrogram boundary factor[C]//9th International Congress on Image and Signal Processing,2016:964-968.
- [19] 张黄群,于盛林,白银刚. 形态学图像去噪中结构元素选取原则[J]. 数据采集与处理,2008(S1):81-83.
- [20] 狄红卫,许瑶. 数学形态学在图象滤波中的应用[J]. 暨南大学学报(自然科学与医学版),2003,24(3):42-45.
- [21] 杨健,李振鹏,苏鹏. 语音分割与端点检测研究综述[J]. 计算机应用,2020,40(1):1-7.
- [22] Gemmeke J F, Ellis D P, Freedman D, et al. Audio set: An ontology and human-labeled dataset for audio events [C]//IEEE International Conference on Acoustics, Speech and Signal Processing,2017:776-780.