

文章编号:1001-9081(2009)04-1188-05

## MP3 压缩域中语音分割的研究与实现

常辽豫,余小清,万旺根,李昌莲,许雪琼

(上海大学 通信与信息工程学院,上海 200072)

(yxq@staff.shu.edu.cn)

**摘要:**针对说话人改变点检测问题,在 MP3 格式下用改进型 BIC 算法实现了多话者改变点的检测。根据非压缩域中 MFCC 的求取过程,提出了一种在压缩域 MP3 格式下利用 MDCT 系数计算 MFCC 特征参数的新方法。在此基础上,使用改进型 BIC 改变点检测算法检测说话人改变点。实验表明在相同的 BIC 检测算法下,利用在压缩域中提取的 MFCC 特征参数进行语音分割,可以得到和非压缩域相似的分割精确度。

**关键词:**压缩域;BIC 检测;语音分割;MDCT 系数

**中图分类号:**TP274 **文献标志码:**A

## Research and realization of speech segmentation in MP3 compressed domain

CHANG Liao-yu, YU Xiao-qing, WAN Wang-gen, LI Chang-lian, XU Xue-qiong

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

**Abstract:** This article proposed an approach for detecting the voice change of speakers by employing improved Bayesian Information Criterion (BIC) algorithm in MPEG1-layer3 (MP3) compressed domain. According to the process of MFCC calculation in raw audio, a new Mel-Frequency Cepstral Coefficients (MFCC) algorithm by utilizing Modified Discrete Cosine Transform (MDCT) coefficient in MP3 domain was presented. Based on these coefficients, the improved BIC algorithm was employed to decide which point was the voice change point of speakers. The experimental results show that using the MFCC coefficients extracted in MP3 domain for speech segmentation, similar segmentation precision can be obtained as that in uncompressed domain.

**Key words:** compressed domain; Bayesian Information Criterion (BIC) detection; speech segmentation; Modified Discrete Cosine Transform (MDCT) coefficient

### 0 引言

音频分割是音频信号进一步应用的基础,其重要性是显而易见的。作为音频分割的一部分,话者改变点检测是音频检索及说话人识别的关键性工作;已有众多学者作了相关的研究。然而,传统音频分割算法大都是基于非压缩域格式,对于压缩域中音频分割的研究却相对较少;即使涉及到音频分割也大都是关于不同类别的划分。如文献[1-2]讨论了 MPEG 标准下无声、语音、音乐或其他声音划分;文献[3]对 MPEG 标准下语音识别进行了研究,但并未给出语音分割的具体方法。随着压缩音频数据增多尤其是 MPEG 凭借其易于传输存储、质量小的优点成为多媒体数据压缩通用标准,实现压缩域中的话者分割具有积极意义。

若以传统方式实现压缩域音频分割,必先进行音频解码,而后从解码后的数据中提取特征并对特征分析,这完全可以实现分割目的。这势必会造成计算量加大<sup>[3]</sup>,不能保证实时实现。由于 MPEG 音频编码时已考虑了人耳听觉感知特性,为利用这些感知信息和减少计算量,可以直接在 MPEG1-layer3 码流中以部分解码运算得到数据,从而实现 MFCC (Mel-Frequency Cepstral Coefficients) 特征参数的求取。同时,

由于不用考虑解码过程中的子带合成,用于解码部分的计算量是极小的。基本实现过程为:首先,从 MP3 (MPEG1-layer3) 码流中经部分解码,得到 MDCT (Modified Discrete Cosine Transform) 系数,随后,通过压缩域中 Mel 滤波处理每一帧的 MDCT 系数以形成每一帧的 15 维 MFCC 系数向量,最后,用改进型的 BIC (Bayesian Information Criterion) 改变点检测算法实现多个说话人的改变点检测。

### 1 压缩域 MFCC 特征提取

与普通实际频率倒谱分析不同,MFCC 特征的分析着眼于人耳听觉特性。类似于听觉临界带的划分它是在 Mel 频率域通过 Mel 三角滤波器对音频信号滤波得到的一种特征参数。

#### 1.1 压缩域 MDCT 系数

特征参数选取的正确与否对检测结果至关重要。文献[4-5]在非压缩域中采用 MFCC、谱质心、短时傅立叶变换等特征实现了分割;与此类似,我们采用了压缩域中 MFCC 作为特征参数。在非压缩域中,特征参数可从原始数据直接结算得到;而压缩音频已经过滤波、声学模型处理和 MDCT 变换过程,显然不能采用直接求取的方法。因此,在 MP3 格式下,

**收稿日期:**2008-10-10;**修回日期:**2008-12-15。 **基金项目:**国家自然科学基金资助项目(60872115);上海市国际合作基金资助项目(075107035);上海市教委电路与系统重点学科项目(J50104)。

**作者简介:**常辽豫(1982-),男,河南漯河人,硕士研究生,主要研究方向:压缩域音频的分类检索;余小清(1958-),女,上海人,副教授,主要研究方向:语音信号处理、音频信息检索;万旺根(1961-),男,上海人,教授,主要研究方向:数字媒体、虚拟现实、音视频分类与检索;李昌莲(1985-),女,江苏南通人,硕士研究生,主要研究方向:音频场景分析;许雪琼(1985-),女,福建泉州人,硕士研究生,主要研究方向:音频分类检索。

我们根据部分解码后的 MDCT 系数求解特征参数。所需数据来源为图 1 中逆量化器后的输出。

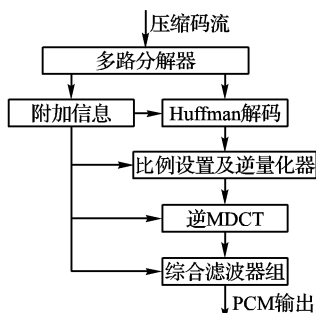


图1 ISO/MPEG 音频解码(第三层)的方框结构

MDCT 作为一种时频转化方式广泛用于各种先进编码中。一段完整的 MP3 音频由许多数据帧构成,帧是组成 MP3 数据的基本单位。一个数据帧有两个颗粒(granule)共包含 1152 个 PCM 值的变换信息,其中每个颗粒含有 576 个对应信息。数据帧经 Huffman 及其他解码处理后可得到 MDCT 系数;这一系数实质上是编码过程中 32 个子带的 PCM(Pulse Code Modulation)值经修正离散余弦变换后的变换参数,每个子带对应 18 个 MDCT 系数<sup>[6]</sup>。文献[7]证明了 MDCT 系数可由原始加窗信号及其混叠信号的漂移 DFT(Discrete Fourier Transformation)变换的线性叠加获得,当进一步假定频域时域漂移为零,频域漂移为 1/2 时可以发现 MDCT 系数实质是原始 DFT 经频移后的线性变换,因此,这些 MDCT 系数反映了音频数据经子带混合滤波后类似于频域的特征,可在其基础上进行特征的提取。在 MP3 标准定义的采样频率  $F_s$  下,每个子带带宽为  $F_s/64$  Hz。每个颗粒 MDCT 系数对应的实际频率值可通过式(1)计算。

$$F(f) = \begin{cases} (f+1) \times \frac{f_{BW}}{NoF}, & \text{无混合窗} \\ \left. \begin{aligned} (f+1) \times \frac{f_{BW}}{576} & f < 36 \\ \frac{f_{BW}}{16} + \frac{(f-35) \times f_{BW}}{192} & f \geq 36 \end{aligned} \right\} & \text{混合窗} \end{cases} \quad (1)$$

其中,  $f_{BW}$  为信号最高频率,根据 Nyquist 采样定理  $f_{BW} = F_s/2$ 。  $NoF$  为不同类型窗时所对应的频率线个数<sup>[2]</sup>。短窗作用是控制前向回声,从而增加时间分辨率;因此,实际 MP3 编码中只有很少一部分是短窗数据帧<sup>[8]</sup>。结合本文只是对语音信号分割,信号变换相对平缓,我们只考虑了长窗情况。每一个颗粒中 MDCT 系数在频域上等间隔分布,44.1 KHz 采样率下,每一子带带宽为 689.06 Hz,频率线间隔为 38.28 Hz。其频率线分布如图 2 所示。

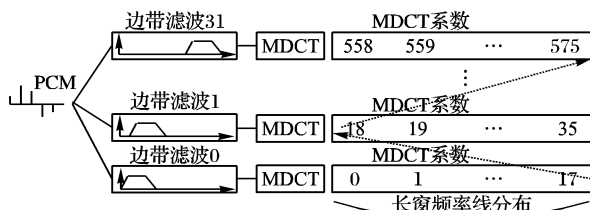


图2 MDCT 系数分布示意

## 1.2 压缩格式下 MFCC 提取

根据 1.1 节中所提到的文献[7]对 MDCT 系数的分析可知,MDCT 系数很大程度上是 DFT 频谱的一个线型近似,尤其是我们只考虑其能量值时。因此,在得到这些系数后,可利用

非压缩域中计算 MFCC 的方法来实现 MP3 域中参数的求取。经过 MP3 混合滤波器组,MDCT 系数在频率范围内以等间隔形式出现,只不过这些参数不是以  $2^n$  的形式对整个频域划分。由于 MDCT 系数包含了频域分量的丰富信息,因此,可以在压缩域中利用这些系数求解每帧的 MFCC 特征参数。之所以在一帧范围内计算参数,一方面是考虑了 MP3 数据格式,另一方面也考虑了语音信号处理时的短时要求。

根据实际临界带宽,文献[1]将每一帧的 MDCT 系数重新定义到六个相似临界带中,然后以 50% 重叠的三角滤波器组对这些系数滤波得到 10 维的 MFCC 特征向量。在本文计算过程中,并未考虑临界带的重新定义,而是通过使用 Mel 滤波器对这些系数滤波后再处理。(类似于文献[9])。根据将要求取的滤波器组数或特征向量维数,确定 Mel 三角滤波的中心频率并设计滤波器组(50% 重叠率)。使用 Mel 三角滤波器组对每一帧 MDCT 能量系数滤波,最后对滤波器输出结果做余弦变换即可得到 MFCC 特征参数。在本文中定义了滤波器的个数为 16,选取前除第一个以外的滤波器输出作为 MFCC 特征参数,形成 15 维特征向量。虽然相比于非压缩域中超过 576 点的 DFT 变换,其频谱分辨率较低,但在语音信号频率范围内,以 576 点等间隔划分频率域所得的分辨率可以满足实际需要。具体计算流程如图 3 所示。

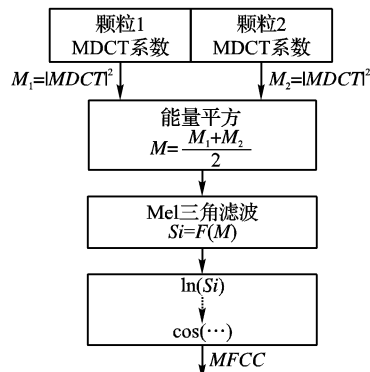


图3 MFCC 计算流程

具体的算法描述如下:

### 1) 帧内能量的求取。

为获得 MFCC 参数,文献[1]的作者采用了对 MDCT 系数直接滤波的方法,但考虑到噪声和谐估计误差有更好的鲁棒性,我们对 Mel 滤波后的 MDCT 频谱取对数能量。

实际语音信号虽然可以看作短时的近似稳定,但仍然存在时变的成分<sup>[10]</sup>,不同于文献[1]中在 15 帧范围内求取此参数,我们使用了基于一帧的求取方式。具体如下:

计算一帧中第一个颗粒的 MDCT 系数的平方,并将其对应的能量向量记为  $MDCT1^2$ ,计算第二个颗粒的 MDCT 系数平方,并将其对应的能量向量记为  $MDCT2^2$ ,求取两个向量的平均,重新得到含有 576 个元素的能量向量  $MDCT^2$ ,如式(2),且满足频域等的间隔分布。

$$MDCT^2 = (MDCT1^2 + MDCT2^2) / 2.0 \quad (2)$$

### 2) Mel 三角滤波器函数。

人耳对频率的感知是一个非线性过程,在 1000 Hz 以上感知能力与频率成对数关系。为实现对式(1)中得到的 MDCT 能量谱进行类似的处理,我们需要定义一个 Mel 三角滤波器组。考虑到计算的复杂性,本文所定义的滤波器个数为 16 个,则其对应于 Mel 频率下的 16 个中心可如下计算:

分别将滤波器应用范围的最低频率  $f_L$  和最高频率  $f_H$  映射为 Mel 频率并确定其 Mel 频率范围;

$$\begin{cases} B(f_L) = 1125 \ln(1 + f_L/700) \\ B(f_H) = \ln(1 + f_H/700) \end{cases} \quad (3)$$

其中,  $B = 1125 \ln(1 + f/700)$  表示实际频率到 Mel 频率的映射关系,  $B_L$  和  $B_H$  表示 Mel 域的最高和最低频率。

由式(3)可得 Mel 频率范围为:

$$B_{\text{Mel}} = B_H - B_L \quad (4)$$

根据滤波器个数均匀等分式(4)中得到的 Mel 带宽  $B_{\text{Mel}}$ , 从而确定滤波器的 Mel 中心频率并将其映射为对应实际中心频率的频率线序号。

$$FC(m) = \frac{N}{F_s} B^{-1}(B(f_L) + m \times \frac{B(f_H) - B(f_L)}{M+1}); 1 \leq m \leq 16 \quad (5)$$

其中  $N = 576$ , 对应于一个颗粒中 MDCT 的个数,  $F_s$  为编码时的采样频率,  $M = 16$ ,  $B^{-1}$  为  $B$  的逆函数:  $B^{-1}(b) = 700(e^{b/1125} - 1)$ , 式(5)中乘以  $N/F_s$  的作用是把中心频率变换为对应的频率线序号。

Mel 频域的三角滤波其实是计算落入三角滤波器范围内的频域分量, 并将 MDCT 能量幅度乘上对应的因子。由式(5)的滤波器中心频率可得到三角滤波器传递函数如式(6)所示。

$$H_m(k) = \begin{cases} \frac{k - FC(m-1)}{FC(m) - FC(m-1)}, & FC(m-1) \leq k \leq FC(m) \\ 0, & k < FC(m-1) \text{ 或 } k > FC(m) \\ \frac{FC(m+1) - k}{FC(m+1) - FC(m)}, & FC(m) \leq k \leq FC(m+1) \end{cases} \quad (6)$$

其中  $1/(FC(m) - FC(m-1))$ ,  $1/(FC(m+1) - FC(m))$  可看作是三角滤波中心两侧的滤波因子, 不同的三角滤波器, 因其带宽的非线性, 对应因子也各不相同。  $m$  表示对应的滤波器序号, 这里  $m$  的范围是  $0 \leq m \leq 15$  且为整数。  $k$  表示频率线的序号, 这里  $k$  的范围为  $0 \leq k \leq 575$  分别和图 2 中 576 个 MDCT 系数对应。

3) 滤波后能量输出。

步骤 2) 中所得到的三角滤波函数体现了对 Mel 频率刻度的划分, 因此可以用其来实现对步骤 1) 中所得到的能量系数的滤波。考虑到频谱的动态范围, 环境噪声的剔除以及对数能量谱域分布适于建模等原因, 我们对滤波器组输出进行了对数处理, 如下式所示:

$$X(m) = \ln\left(\sum_{k=0}^{575} \text{MDCT}^2 \times H_m(k)\right); 0 \leq m \leq 15 \quad (7)$$

其中,  $m$  是滤波器序号,  $k$  是 MDCT<sup>2</sup> 序号, MDCT<sup>2</sup> 可由式(2)得出。在高频部分其值有可能为 0。

4) 利用离散 DCT 变换将滤波输出变化到倒谱域。

考虑到便于对不同频段的 MDCT 谱成分做解相关处理, 方便进一步建模和计算, 我们将滤波器输出  $X(m)$  做了 DCT 变换。

$$\text{Mef}(n) = \sum_{m=0}^{15} X(m) \times \cos[\pi n(m+0.5)/16]; 0 \leq n \leq 15 \quad (8)$$

由式(6)可得到一帧中的 MFCC 向量, 但向量中各成分对后续的话音人分割处理的贡献是不同的, 在说话人识别中

第一维或第二维的贡献是不大的, 仅是语音中的能量信息<sup>[11]</sup>, 考虑到本文的工作是话者的分割, 因此本文采用了除第一维之外的 15 维向量。

## 2 改进型 BIC 检测算法

在得到 MFCC 基础上, 使用改进型 BIC 检测算法对 MP3 数据做了分割。贝叶斯准则的基本思想是样本的极大似然减去模型复杂度。BIC 准则可由式(9)定义:

$$\text{BIC}(M) = \log L(X, M) - \frac{\lambda}{2} m \cdot \log(N) \quad (9)$$

其中,  $X = \{x_i, i = 1, 2, \dots, N\}$  是模型样本集合,  $M = \{M_i, i = 1, 2, \dots, K\}$  是候选参数模型,  $L(X, M)$  是样本数据  $X$  在模型  $M$  中的极大似然,  $\lambda$  是惩罚因子,  $m$  是模型  $M$  的参数数目。根据基本 BIC 准则, 可得到 BIC 改变点检测算法,  $X = \{x_i \in R^d, i = 1, 2, \dots, N\}$  是取自压缩域的 MFCC 特征序列,  $X$  满足多元的高斯分布。对一段语音特征向量的组合, 假设  $H_0$  代表无改变点高斯模型,  $H_1$  则代表有改变点高斯模型,  $i$  表示说话人发生的改变点,  $H_0$  和  $H_1$  的最大似然比定义为式(10):

$$R(i) = N \log |\Sigma| - N_1 \log |\Sigma_1| - N_2 \log |\Sigma_2| \quad (10)$$

其中  $N, N_1, N_2$  分别是整个语音段和改变点前后的帧数,  $\Sigma, \Sigma_1, \Sigma_2$  分别是整个段语音特征向量, 改变点前后特征向量的协方差矩阵。

若从  $H_0, H_1$  两个模型选择其一考虑, 它们之间的 BIC 值之差可表示为如下:

$$\text{BIC}(i) = R(i) - \frac{1}{2} \lambda \left( d + \frac{1}{2} d(d+1) \right) \log N \quad (11)$$

其中  $\lambda = 1$ ,  $d$  表示特征空间的维数,  $N$  表示整个语音段的帧数。如果式(11)大于零, 说明有改变点发生, 即  $H_1$  模型成立, 因此当

$$\{\max_i \text{BIC}(i)\} > 0 \quad (12)$$

时, 表示有改变点出现, 改变点出现的时刻为:

$$\hat{i} = \arg \max_i \text{BIC}(i) \quad (13)$$

为保证 BIC 检测的有效性, BIC 用于分割的音频段长度要大于 2 s。但上述的检测算法在说话人音调的突然变化、语气的加重情况下会出现多余的改变点, 为此改进型算法在上述算法基础上, 增加了对改变点的进一步判决。1) 它把上述算法得到的改变点作为备选改变点, 2) 分析备选改变点两端语音 BIC 值的变化趋势, 根据趋势变化作出有无改变点的判断, 实现方法是在改变点两侧判断上升或下降点所占有的比例, 若上升或下降 BIC 点占所有 BIC 点的比例大于 65%, 则判断为正确的改变点。

多点改变检测算法是在单改变点基础上的推广, 在当前语音段检测到一定长度后若无改变点, 则把检测点位置后移重新开始检测, 避免了数据量计算过大的问题。改进型多点检测算法流程如图 4 所示。

上述流程图中  $\text{MaxLength}$  长度是语音检测段长度的 4 倍为 1000 帧约 26.12 s, 当在  $\text{MaxLength}$  长度内无改变点时, 后移  $\text{MaxLength}$  长度重新开始检测。

## 3 实验结果分析

本文所采用的实验数据是某电视台的标准新闻节目, 语音效果为 CD 音质, 采样率为 44 100 Hz 的 MP3 音频格式。语音数据段均是几个男女播音员声音的组合。

每次改变点检测在 250 帧约 13.58 s 内进行,由于 MP3 数据格式,不考虑窗移情况,前后数据之间无重叠。判断改变点是否真实的比例因子设为 65%,多点检测时候最大无改变点语音长度为 1000 帧约 26.12 s。

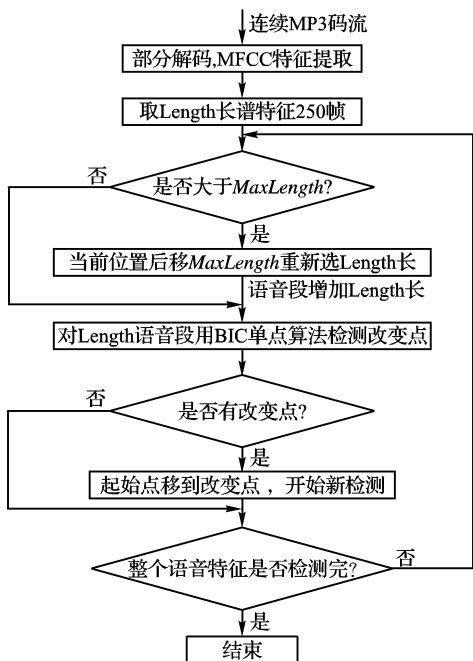
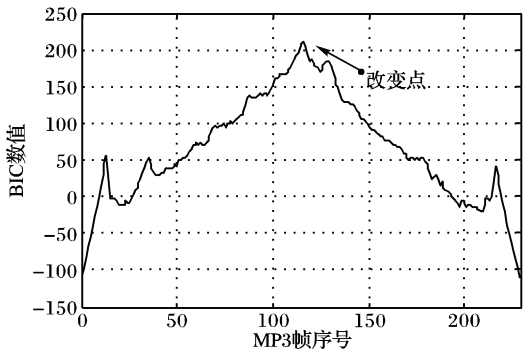


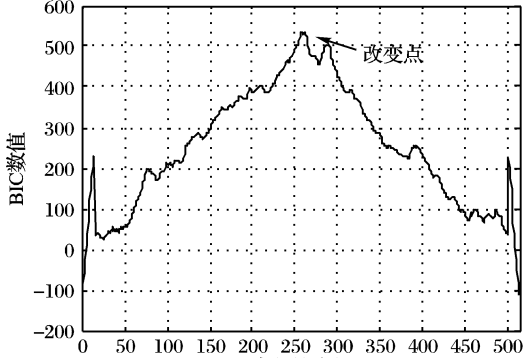
图 4 改进型多点 BIC 检测算法流程

3.1 压缩与非压缩域单改变点检测对比

大量实验证明,对含有单改变点的压缩语音段,用改进型 BIC 检测算法能准确地检测出话者的改变点。相比非压缩中的检测结果,并没出现误检的情况。两种情况下的 BIC 曲线图极其相似,且所检测到的改变点时间偏差很小。



(a) 压缩域单改变点BIC曲线



(b) 非压缩域单改变点BIC曲线

图 5 非压缩域与压缩域中单改变点 BIC 曲线对比

图 5 是同一段语音在非压缩与压缩形式下单改变点 BIC 曲线对比图。语音段的长度为 6 s,实际改变时刻点在 3 s。

非压缩域中采样率为 11 025 Hz,帧长为 256 点,50%窗移时共有 515 帧。压缩域中采样率为 44 100 Hz,帧长 1 152 点(约 26 ms)无重叠时共有 229 帧。使用改进型 BIC 检测算法,非压缩域中改变点发生第 258.743 帧约 3.004 s 时刻,压缩域中改变点发生第 119.935 帧约 3.133 s。单改变点时两者检测时刻相差 0.129 s,与真实值相差 0.133 s,在可接受的范围内。

3.2 压缩多改变点检测结果

语音分割的性能指标一般有:准确率(Precision, PRC)、召回率(Recall, RCL)和综合性能( $F_1$ -measure,  $F_1$ )。准确率是所有检测出的语音改变点中,正确改变点所占的百分比;召回率是所检测出的正确改变点占真实改变点的百分比,综合性能如下式:

$$F_1 = \frac{2PRC \times RCL}{PRC + RCL} \quad (14)$$

其中  $F_1$  介于 0 和 1 之间,值越大表明性能越好。

本文使用改进型 BIC 检测算法对六段压缩格式混合语音段进行了分割;结果表明在压缩情况下。总体准确率达到 93.1%、召回率达到 89.3% 以及综合性能达到 91.1%。表 1 列出了各个语音分割时的性能指标。

表 1 各音频段分割性能指标

音频段	实际改变点个数	音频段长/s	PRC/%	RCL/%	$F_1$ /%
文件 1	9	135.7	88.9	100.00	94.1
文件 2	19	150.2	94.4	89.47	89.8
文件 3	12	168.0	92.3	85.71	88.8
文件 4	10	148.9	90.0	90.00	90.0
文件 5	6	168.3	100.0	83.30	90.9
文件 6	16	150.8	93.3	87.50	93.1
平均	12	153.7	93.1	89.30	91.1

从实验结果可以看出,在压缩域中使用改进型 BIC 检测算法进行改变点检测,其平均性能指标能接近 90%;准确分割率超过 90%;这说明了我们在压缩域使用 MFCC 作为分割参数的正确性。同时,也验证了所采用的改进型 BIC 检测算法分割的有效性。

结合以前的研究工作,我们将压缩域中分割指标与使用该算法时所获得非压缩分割性能指标<sup>[12]</sup>做了对比;同时,为便于说明加上判别改变点情况后该算法的精确度,我们也将此算法和一般的 BIC 分割结果做了对比。具体比较结果参考如图 6。

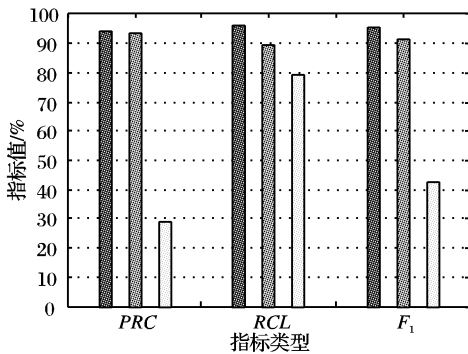


图 6 改进 BIC 算法对比

图 6 中,颜色最深条形的为改进算法在非压缩域中多次分割实验时三个指标情况,颜色次之的条形图为改进算法在压缩域多次分割时的指标,数值较小的条形图未改进算法在非压缩域中的多次分割情况。

## 4 结语

说话人分割是语音识别,音频检索等语音处理的基础。在非压缩域中,已经有许多学者使用 BIC 检测算法进行了语音分割研究。压缩格式音频的出现,使得这些研究已逐渐不能满足实际需要。鉴于此,本文在压缩域中,使用改进型 BIC 检测算法对 MP3 格式音频分割做了相关研究;实验结果证明在压缩域中进行语音分割的可行性和有效性,这对在压缩域中进行语音识别,音频检索等研究具积极的现实意义和指导意义。

### 参考文献:

- [1] SHAO XI, XU CHANG-SHENG, WANG YE, *et al.* Automatic music summarization in compressed domain[C]// Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing: ICASSP04. Singapore: [s. n.], 2004: 261 - 264.
- [2] KIRANYAZ S, AUBAZAC M, GABBOUJ M. Unsupervised segmentation and classification over mp3 and AAC audio bitstreams [C]// Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services. London: [s. n.], 2003: 338 - 344.
- [3] YAPP L, ZICK G. Speech recognition on MPEG/Audio encoded files[C]// Proceedings of the 1997 International Conference on Multimedia Computing and Systems. Ottawa: [s. n.], 1997: 624 - 625.
- [4] LU LIE, ZHANG HONG-JIANG, LI S Z. Content-based audio classification and segmentation by using support vector machines[J]. Multimedia Systems, 2003, 8(6): 482 - 492.
- [5] YASLAN Y, CATALTEPE Z. Audio music genre classification using different classifiers and feature selection methods[C]// Proceedings of the 18th International Conference on Pattern Recognition. Washington, D C: IEEE Computer Society, 2006: 373 - 376.
- [6] ISO/IEC 11172-3, Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s, Part 3: Audio[S]. ISO/IEC, 1992.
- [7] WANG Y, YAROSLAVSKY L, VILERMO M. On the relationship between MDCT, SDFT and DFT[C]// Proceedings of the 5th International Conference on Signal Processing. Beijing: [s. n.], 2000: 44 - 47.
- [8] BRANDENBURG K. MP 3 and AAC explained[C]// Proceedings of the AES 17th International Conference on High-quality Audio Coding. Florence, Italy: [s. n.], 1999: 99 - 111.
- [9] HAN WEI, CHAN C F, CHOY C S, *et al.* An efficient MFCC extraction method in speech recognition[C]// Proceedings of the 2006 IEEE International Symposium on Circuits and Systems: ISCAS 2006. Hong Kong: [s. n.], 2006: 145 - 148.
- [10] KIM S, ERIKSSON T, KANG H G. A pitch synchronous feature extraction method for speaker recognition[C]// Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. Piscataway, NJ: IEEE Computer Society, 2004: 405 - 408.
- [11] 甄斌, 吴玺宏, 刘志敏, 等. 语音识别和说话人识别中各倒谱分量的相对重要性[J]. 北京大学学报: 自然科学版, 2001, 37(3): 371 - 378.
- [12] 余小清, 谭海英, 万旺根. 一种改进型 BIC 话者改变检测算法[J]. 上海大学学报, 2007, 13(14): 403 - 408.

(上接第 1179 页)

虽然在实验过程中,在 CPU 上使用的算法时间复杂度为  $O(n^3)$ ,实验数据表明,GPU 上运算速度快得多,远高于时间复杂度为  $O(n^{2.376})$  的串行算法。而同在 GPU 上,本文改进算法比 SDK 例程算法有很大的速度提升,大约提升 60% 以上,很好的弥补了为提升精度使用 Kahan 求和公式增加的计算量,实现了精度提高的同时,速度也大大提高。

## 4 结语

GPU 是为了加速图形图像处理而出现的,它适用于计算密集型和数据密集型运算,适用于高性能并行计算,在算法相同的情况下,数据量越大,处理效率越高。针对如何提升在 GPU 上实现大型矩阵运算的计算精度和运算速度,本文提出了使用 Kahan 求和公式来提升大型矩阵乘法的计算精度,根据 GPU 多线程数来划分矩阵分块,以及利用显存的节距对齐特点完成内存分配,最终在保证计算精度的同时实现大型矩阵的快速运算。最后实验结果表明,该措施能够取得较好的效果,大型矩阵乘法的运算速度和精度都可以得到大的提升。

### 参考文献:

- [1] JOWENS J D, LUEKE D, GOVINDARAJU N, *et al.* A survey of general-purpose computation on graphics hardware[C]// Proceedings of the 2005 Annual Conference of the European Association for Computer Graphics: Euro-Graphics 2005. Dublin, Ireland: [s. n.], 2005: 21 - 51.
- [2] 张庆丹, 戴正华, 冯圣中, 等. 基于 GPU 的串匹配算法的研究[J]. 计算机应用, 2006, 26(7): 1735 - 1737.
- [3] 吴恩华. 图形处理器用于通用计算的技术、现状及其挑战[J]. 软件学报, 2004, 15(10): 1493 - 1540.
- [4] 吴恩华, 柳有权. 基于图形处理器(GPU)的通用计算[J]. 计算机辅助设计与图形学报, 2004, 16(5): 602 - 612.
- [5] 王吉春, 王顺绪. 大规模矩阵乘积的几个并行算法[J]. 淮海工学院学报, 2007, 16(2): 32 - 35.
- [6] 张浩, 李利军, 林岚. GPU 的通用计算应用研究[J]. 计算机与数字工程, 2005, 33(12): 60 - 62.
- [7] 张扬, 诸昌铃, 何太军. 图形硬件通用计算技术的应用研究[J]. 计算机应用, 2005, 25(9): 2192 - 2194.
- [8] NVIDIA. NVIDIA CUDA compute unified device architecture reference manual[EB/OL]. [2008 - 08 - 20]. [http://www.nvidia.com/object/cuda\\_develop.html](http://www.nvidia.com/object/cuda_develop.html).
- [9] 谭福平, 刘洪刚. Winograd 矩阵乘法算法用于任意阶矩阵时的一种新处理方法[J]. 应用数学与计算数学学报, 2004, 18(1): 92 - 95.
- [10] (美) MANBER U. 算法引论——一种创造性方法[M]. 黄林鹏, 谢瑾奎, 陆首博, 译. 北京: 电子工业出版社, 2005.
- [11] BRASSARD G, BRATLEY P. 算法基础[M]. 邱仲潘, 柯渝, 徐锋, 译. 北京: 清华大学出版社, 2005.
- [12] Sun Studio11. 数值计算指南[EB/OL]. [2008 - 08 - 25]. [http://developers.sun.com/export/sites/default/solaris/sun\\_studio.html](http://developers.sun.com/export/sites/default/solaris/sun_studio.html).
- [13] LEFOHN A, KNISS J, OWENS J. Implementing efficient parallel data structures on GPUs[C]// GPU gems 2: Programming techniques for high performance graphics and general purpose computation. Boston, MA: Addison-Wesley, 2005: 521 - 545.