

基于样本的流行歌曲关键段分割方法

张一彬, 周 杰, 边肇祺

(清华大学自动化系, 北京 100084)

摘 要: 流行歌曲的关键段为歌曲中最能打动人、给人印象最深刻的一个完整片段. 将它分割出来可用于音乐试听和基于内容的音乐分类、检索、管理. 通过对人工截取的样本进行分析, 本文提出了一种流行歌曲关键段自动分割方法. 实验结果表明, 此方法可以比较准确和有效地分割出流行歌曲中的关键段.

关键词: 数字娱乐; 流行歌曲; 音乐自动摘要; 神经网络

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2006) 02-0220-06

Samples-Based Automatic Key Segment Extraction for Popular Song

ZHANG Yirbin, ZHOU Jie, BIAN Zhao-qi

(Department of Automation, Tsinghua University, Beijing 100084 China)

Abstract We define the key segment of a popular song as the most impressive part in this song. It can be regarded as a perfect index or summary for this song, therefore it is much helpful for popular songs' audition and management. We propose an effective algorithm for automatic key segment extraction based on the analysis of positive and negative samples of key segment extracted manually from a training database. An experiment is carried on a test database including 130 popular songs. Compared with manually extraction of key segments, the proposed algorithm has a rather satisfying performance.

Key words digital entertainment; popular song; automatic music summary; neural network

1 引言

音乐是人类共同语言, 它包含着复杂的结构和丰富的信息. 如何运用计算机手段自动地分析音乐是十分有趣同时也是十分困难的. 其结果有助于推动认知科学、机器听觉等领域的发展. 在音乐自动分析方向, 如何对音乐数据进行摘要描述(即对一段音乐生成一个高度简约又能反映出音乐特点的摘要信息)是一个重要的问题. 由于音乐所具有的高度复杂性、结构化、非语义性, 以及人们对音乐理解的主观性, 如何为音乐数据自动生成摘要已经成为一项具有特殊挑战性的工作.

另一方面, 随着娱乐业的快速发展大量流行歌曲被不断创作出来. 据统计^[1], 仅在西方国家每个月就新产出大约 4000 张音乐 CD, 其中流行歌曲占据了很大的比例. 对于普通消费者来说如何在浩如烟海的流行歌曲当中找到真正想听的歌曲? 对于商家来说如何快速有效地管理他们的海量音乐数据? 这些都已成为紧迫的问题. 为了满足用户的需要, 许多音乐销售商都允许顾客们试听他们打算

购买的 CD. 在日益流行的网络销售中, 为了防止免费下载, 商家通常仅允许顾客们试听一首音乐的起始部分. 显然一首音乐的开始部分往往不是这首音乐最吸引人的部分, 因而这种试听方式的效果不能令人满意. 理想的解决方式是由机器自动地从原版 CD 音乐中, 完整地截取出其中最有感染力的一段用作音乐试听.

本文把一首流行歌曲中最吸引人、最具感染力的那个片段称为这首流行歌曲的关键段. 流行歌曲关键段分割无疑是一种很好的音乐自动摘要方法, 而且关键段分割对音乐成份的进一步分析处理也有重要意义. 关键段分割还可用于音乐自动试听, 以方便销售商和顾客. 到目前为止, 在已发表的文献中还未见到类似工作. 与此相关的工作包括一些音乐自动摘要的方法. Kraft 等人提出了一种用于 MIDI (Musical Instrument Digital Interface) 格式的音乐自动摘要方法^[2]. 然而, 由于 MIDI 文件中包含了大量的乐谱信息, 基于 MIDI 格式的任何音乐分析方法其本质都是基于乐谱分析的, 都无法直接运用于其它音频格式的数据. 由于 MIDI 格式数据使用的有限性, 基于 MIDI 格式的音乐

分析方法其应用范围也是有限的. 在文献[3~9]中提出了一些基于一般音频格式的音乐自动摘要算法. 在该类算法中, 音乐摘要被定义为一些定长(如, 10秒、20秒或 30秒)的音频片段, 它们主要是通过相似性分析的方法被分割出来的. 所有方法的计算量都非常大, 并且其效果只在较小的测试集上通过一些主观的评价方法(如: 只对自动摘要结果直接评分)得到验证. 与上述的音乐自动摘要相比, 本文定义的流行歌曲关键段有以下不同之处: 流行歌曲关键段是着眼于一首流行歌曲中给人印象最深刻的那个片段, 而不仅仅是歌曲中重复性最高的部分; 在语义上它是一个完整的段落具有完整的含义, 因而是变长的.

本文在关键段定义的基础上, 研究了基于一般音频格式的流行歌曲关键段自动分割方法. 我们收集了大量具有不同风格的流行歌曲, 并请志愿者们截取每首流行歌曲的关键段正样本和负样本; 然后从这些关键段正、负样本中提取音频特征, 组成特征向量, 并训练神经网络分类器. 利用训练好的分类器我们就可以在待处理的流行歌曲中搜索关键段的主体部分. 通过统计人工截取的关键段的长度分布确定关键段主体部分的搜索范围和步长, 在这个范围内的最有可能是关键段的音频片段将被确定为这首流行歌曲的关键段主体部分. 为了准确地确定关键段的起始点和结束点, 本文将在确定了关键段主体部分的基础上进一步确定关键段起始区域和结束区域, 并在这两个区域中利用能量最低原理分别搜索关键段起始点和结束点. 为了尽可能客观地评价算法的效果, 本文制定了统一的评分标准, 并请志愿者们对自动截取的关键段和相应的人工截取的关键段进行比较. 所有志愿者的评分的均值将作为最终的评价结果. 本文提出的算法在由 130 首流行歌曲所组成的测试数据库上进行了测试, 总体效果相当满意.

2 流行歌曲关键段

我们几乎每天都会听到音乐, 在各种各样的音乐类型当中最熟悉的莫过于流行歌曲. 一般来说, 一首流行歌曲可以大致被分成四个部分: 过门部分, 它通常是一段纯伴奏; 过渡部分, 这部分通常较长, 一般表现为歌手在乐队的伴奏下舒缓地歌唱; 高潮部分, 这是整首歌曲的精华, 在高潮部分中歌手们会在乐队的伴奏下激昂地歌唱; 尾声部分, 这一部分一般会是一段纯伴奏, 有时也会是高潮部分的自然延续, 但其音量会越来越弱直至结束^[10]. 这四个部分一般是按照上述顺序依次出现的, 但有时过渡部分和高潮部分会交替出现. 也就是说, 一首歌曲中可能会有一个以上的高潮段落, 但这些高潮段落的旋律通常是一样的, 它们给人的听觉感受也是一样的. 调查发现, 在大多数情况下, 人们更喜欢欣赏一首流行歌曲的高潮部分.

我们一般都会有这样的生活经验: 在一首流行歌曲当中总会有那么一个特殊的段落它令你最感动、最难忘. 一旦有人提起这首歌曲, 那个段落的旋律就会立即跃入你的脑中. 我们把这样的段落称为这首流行歌曲中的关键段.

例如, 在著名电影《泰坦尼克》的主题歌《my heart will go on》中, “you are here, there’s nothing I fear and I know that my heart will go on, we’ll stay forever this way, you are safe in my heart and my heart will go on and on”这一段被公认为是这首歌曲中最感人、最难忘的一段. 同样, 在著名歌手童安格演唱的流行歌曲《耶利亚女郎》中的关键段显然是“耶利亚, 神秘耶利亚, 耶利耶利亚, 耶利亚, 神秘耶利亚, 我一定要找到她”这一段. 调查发现, 在大多数流行歌曲中都有着被人们所普遍接受的关键段.

一般来说, 一首流行歌曲的关键段应该是在乐队伴奏下的一段优美、激昂的唱段, 而不应该是一段纯伴奏, 它通常就是歌曲中的高潮部分或高潮部分中的一个完整段落. 在关键段的歌词中, 很有可能包含着这首歌曲的名称. 不

同流行歌曲的关键段长度显然不尽相同, 但是通常不会太长或太短. 本文总共收集了 259 首不同风格的中文和英文流行歌曲, 并由志愿者们手工标出其中的关键段. 通过统计这些关键段的长度分布, 我们发现大多数流行歌曲的关键段长度集中在 20 秒至 40 秒之间. 人工截取的流行歌曲关键段的长度分布见图 1.

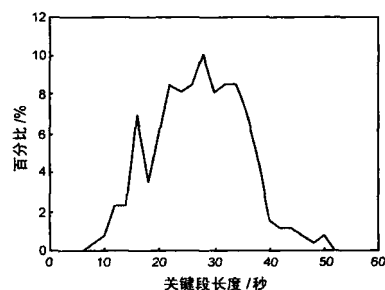


图 1 人工截取的关键段的长度分布

3 特征提取

本文中, 每首流行歌曲被转换成采样率为 11025 秒、量化位数为 8 位的混和单声道 WAV 文件. 并请志愿者们为其截取出一个关键段正样本和一个负样本. 一首流行歌曲当中除了关键段外的其它任何部分都可以被看作是负样本. 相对正样本来说, 负样本的截取比较随意. 只要其内容不与正样本有明显重复且长度与正样本相仿即可. 通过借鉴音频分析与识别领域中的前人成果, 在本文中所使用的基本音频特征为: 短时能量 (SE)、过零率 (ZCR)、Mel 频率 (MF)、和谐度 (HD)^[11]、低短时能量值比率 (LSTER)、高过零率比率 (HZCRR)、谱通量 (SF)^[12]、谱矩 (SC)、带宽 (BW) 和频谱滚动频率 (SRF)^[13]. 实验中求取基本音频特征时所使用的“帧”的长度为 512 个采样点, 约 46 毫秒; 相邻帧之间有 112 个采样点的重叠区域, 约 10 毫秒.

有了基本的音频特征后, 便可以从每个关键段正、负样本中提取出一个 17 维的特征向量, 并用这个特征向量来代表这个音频片段. 这个 17 维的特征向量中包括短时能量序列、过零率序列、Mel 频率序列、谱矩序列、带宽序列、谱通量序列以及频谱滚动频率序列的均值和标准差再加上低短时能量值比率、高过零率比率和和谐度. 值得注意的是在计算 Mel 频率序列的均值和标准差时将不考虑 Mel 频率值为 0

的那些帧;在计算短时能量序列的均值时需要对其做标准化处理,因为数据库中的歌曲来源不同,不同歌曲之间的平均能量水平有很大差异,用直接求得的片段平均短时能量作为特征是不合适的.本文将对音频片段的平均短时能量做如下的标准化处理:

$$\bar{E}_{seg} = \frac{E_{all}}{E_{song}} E_{seg} \quad (1)$$

其中 E_{all} 是数据库中所有歌曲的平均短时能量, E_{song} 是一首歌曲的全曲平均短时能量, E_{seg} 是从这首歌曲中截取的关键段正样本或负样本的平均短时能量, \bar{E}_{seg} 是最后经过标准化处理后的样本片段的平均短时能量值.有了关键段正、负样本的特征向量后就可以训练分类器,并利用这个分类器从待处理的流行歌曲中寻找合适的关键段.

4 分类器选择与特征有效性分析

4.1 分类器选择

本文将 259 首流行歌曲分为训练集和测试集两部分,其中训练集 129 首、测试集 130 首.从每首歌曲当中分别截取一个关键段正、负样本,并从每个样本当中提取出一个 17 维的特征向量.在过去的关于基于内容的音乐分类和音频分割的工作中^[11,14],我们比较了一些常见的分类器,包括欧氏距离下的 k 近邻分类器(K-NNC)、马氏距离下的最近邻分类器(MMDC)、帕森窗分类器(PC)、二次分类器(QDC)、Fisher 分类器(FC)、支持向量机(SVM)和 PB 神经网络分类器(BPNNC),并且发现在处理有关音频分类的问题时 BPNNC 的效果最好.本文中将从这些分类器中选择一个用于流行音乐关键段分割.不同分类器在测试集上得到的平均分类正确率见表 1.从表 1 可以看出在区分关键段样本与非关键段样本的两类分类问题中, BPNNC 的分类效果要优于其它分类器,这一点与我们在^[11,14]中的结论是一致的.本文所使用的

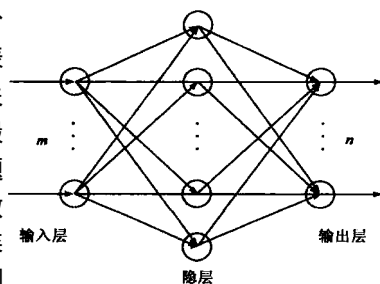


图 2 BPNNC 网络结构示意图

BPNNC 是一个 3 层的前馈神经网络,并用 BP 算法做为训练方法,其输入层的节点数为特征向量的维数,输出层的节点数为需要区分的类别数,中间层的节点数为 10.网络结构示意图见图 2. BPNNC 在测试集上得到的平均分类正确率为 92.7%,其详细分类结果见表 2.

表 1 不同分类器在测试集上的平均分类正确率

K-NNC	MMDC	PC	QDC
59.3%	86.9%	62.2%	89.6%
FC	SVM	BPNNC	—
91.9%	90.3%	92.7%	—

表 2 BPNNC 在测试集上的详细分类结果

	关键段	非关键段
关键段	93.8%	6.2%
非关键段	8.5%	91.5%

4.2 特征有效性分析

本文对不同特征对分类的贡献进行了分析.为了考察各维特征的有效性,首先将测试集中的特征向量在各维特征上分别做归一化,方法如表 3.

$$\bar{V}_i = \frac{V_i - V_{min,i}}{V_{max,i} - V_{min,i}} \quad (2)$$

其中, V_i 、 \bar{V}_i 分别为测试集中某样本的第 i 个特征的实际值和归一化后的值, $V_{min,i}$ 、 $V_{max,i}$ 分别为测试集中所有样本在第 i 个特征的最小值和最大值;然后按照 BPNNC 的分类结果将测试集分成两个子集,并分别计算这两个子集的中心特征向量 C_1 和 C_2 .我们只需考察 $|C_{1,i} - C_{2,i}|$,其中 $C_{j,i}$ 为第 j 个子集的中心特征向量中的第 i 个特征值.显然,对分类结果有较大贡献的特征其相应的 $|C_{1,i} - C_{2,i}|$ 值应该较大,详细的特征有效性分析结果见表 3.

表 3 特征有效性分析结果

特征名称	有效性指数
短时能量均值	0.365
低短时能量比率	0.201
高过零率比率	0.192
谱矩标准差	0.181
过零率均值	0.163
频谱滚动频率标准差	0.147
谱通量标准差	0.125
带宽标准差	0.116
频谱滚动频率均值	0.116
谱矩均值	0.114
短时能量标准差	0.101
和谐度	0.095
带宽均值	0.087
过零率标准差	0.084
Mel 频率标准差	0.052
Mel 频率均值	0.021
谱通量均值	0.001

分析结果表明在这 17 个特征中,对分类结果贡献最大的 3 个特征分别为:标准化后的短时能量均值、低短时能量比率和高过零率比率.这与人们对关键段的直观感受是相同的.因为关键段往往属于一首歌曲中的高潮部分,这部分的能量水平相对歌曲中的其它部分来说是较高的;另外关键段一般都是一些歌手正在演唱的片段,很少包含过多的纯伴奏部分.在语音分析中,人们常常用过零率水平的高低来

5 关键段分割算法

上一节中训练得到的分类器可用于在一首给定的流行歌曲当中搜索关键段的大概位置.根据前面得到的人工截取的关键段的长度分布,首先在全曲中搜索长度在 20 秒至 40 秒之间的有可能是关键段的音乐片段,并在这个片段的基础上进一步精确确定关键段的起止点,从而得到最终的流行歌曲关键段.算法包括“离线训练”和“在线处理”两部分,其流程图见图 3.具体的步骤如下:

(1)训练神经网络分类器(离线).这一步已经在前面的小节中做了详细的介绍.

(2)截取关键段主体候选片段集.本文将给定的流行

歌曲分割成一系列定长的音乐片段, 片段长度从 20 秒逐步增至 40 秒, 下一级音乐片段比上一级音乐片段增加 2 秒; 分割音乐片段的步长均为 2 秒. 假设一首流行歌曲的长度为 L 秒, 则分割得到的关键段主体候选片段总数为:

$$S = \sum_{i=0}^{10} L(L - 20 - 2i) / 2! \quad (3)$$

(3) 关键段主体候选片段特征提取. 本文将从每个关键段主体候选片段中提取一个 17 维的特征向量用来表示这个音频片段, 详见第 3 小节.

(4) 选择关键段主体. 利用上一节中训练得到的 BPNN 分类器, 从关键段主体候选片段集中选出一个与人工截取的关键段最为相似的片段, 并将其确定为这首歌曲中关键段的主体部分. 在本文中 BPNN 分类器的输出层有两个节点, 它们的输出值分别对应着“关键段”和“非关键段”这两个类别, 其和等于 1 将代表关键段主体候选片段的 17 维特征向量送入分类器中, 如果对应着“关键段”类别的输出节点上的值大于 0.5, 该片段将被分到“关键段”类别中. 将对应着“关键段”类别的 BPNN 输出值记为 λ_i 这个值越大分类结果的可信度就越高. 则在关键段主体候选片段集中, 被选为关键段主体部分的音乐片段可以表示为:

$$\text{seg}_{key} = \arg \max_{\text{seg}_i \in \Psi} (\lambda_i) \quad (4)$$

其中 Ψ 为关键段主体候选片段集, seg_i 为集合 Ψ 中的第 i 个候选片段, seg_{key} 为被确定为关键段主体部分的音乐片段.

(5) 确定关键段边界区域. 在确定了关键段的大致位置后, 需要进一步精确确定其起止点. 为此, 本文将分别建立关键段的起始点区域和结束点区域, 并在这两个区域中寻找最终的关键段起始点和结束点. 这两个区域都是长度为 6 秒的音乐片段, 其中心点分别为 seg_{key} 的起始点和结束点. 因此, 最后得到的关键段长度应在 14 秒至 46 秒之间.

(6) 关键段边界区域滤波. 我们知道一个好的关键段应该包括一个或几个完整的乐句. 为了避免生硬截断乐句的现象, 关键段起止点的合理位置应该在歌手演唱的间隙处. 这些地方所对应的短时能量值会明显低于其附近的地方. 研究表明人的噪音的能量主要集中在 100Hz 到 350Hz 之间^[15]. 为了突出这两个区域中的人的噪音成份, 本文将对其进行带通滤波以压制伴奏音乐.

(7) 确定关键段起止点. 经过滤波后, 在起始点区域中能量最小的音频帧的起始点被确认为这首歌曲的关键段的起始点; 同样在结束点区域中能量最小的音频帧的结束点被确认为这首歌曲的关键段的结束点. 这里, 音频帧的定义与特征提取时相同, 每个帧包含 512 个采样点, 约 46 毫秒; 相邻帧之间有 112 个采样点的重叠区域, 约 10 毫秒.

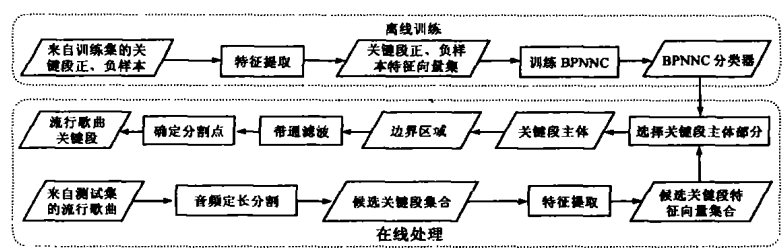


图 3 流行歌曲关键段分割算法流程图

6 实验结果

本文用测试集中的 130 首流行歌曲对关键段分割算法进行了测试, 这些风格各异的流行歌曲来自许多不同的歌手. 在本文中, 用于测试的流行歌曲数目远远多于相关文献中用于实验的歌曲数目 [2~ 9], 它们在 7 首至 30 首之间.

在对算法效果的评价方面, 本文采用了一种相对客观和严格的方法. 首先将程序自动分割的关键段保存下来, 然后请志愿者们用统一的评价方法对来自同一首流行歌曲的人工关键段和自动关键段进行比较. 这些志愿者都是流行歌曲爱好者, 并且与上文中截取人工关键段的志愿者们不同. 他们得出的结论具有较强的可信度和代表性. 显然, 这种与人工截取的关键段进行对照评分的实验结果评价方法和直接评价自动分割效果的方法^[2~ 9]相比更加严格、合理.

具体的评价方法如下: 首先考察关键段的总体质量, 每个志愿者给关键段 (包括人工关键段) 打两个分数, 分数 1 用于描述这个片段的内容作为关键段是否合适, 分别用 3~ 0 来表示“优”、“良”、“中”、“差”; 分数 2 用来描述这个片段的边界合理性, 分别用 2~ 0 来表示“合理”、“一般”、“不合理”. 分数 1 与分数 2 的和则用来表示这个片段的总体质量.

对于每一对来自同一首流行歌曲的自动关键段和人工关键段, 志愿者们将给出一个更为重要的评价指标——旋律一致性. 也就是说, 要求志愿者们考察自动关键段在旋律上与其相应的人工关键段是否相似, 而不必考虑这两个片段中的歌词差异以及它们在歌曲中的绝对位置上的差异. 我们分别用 5~ 1 来表示“优”、“良”、“中”、“差”和“很差”. 所有志愿者给出的分数的平均值将作为最终的评价结果, 见表 4.

表 4 关键段分割实验评分结果

	人工关键段	自动关键段
分数 1 (0~ 3)	2.9	2.7
分数 2 (0~ 2)	1.9	1.4
总体质量 (0~ 5)	4.8	4.1
旋律一致性 (1~ 5)	3.9	

从实验结果中可以看到本文提出的流行歌曲关键段分割算法是有效的, 大多数自动关键段在旋律上与其相应

的人工关键段高度一致. 同时我们也发现在边界合理性方面, 自动关键段明显弱于人工关键段. 也就是说想十分精确地将关键段从全曲中分割出来是非常困难的. 当志愿者在手工截取关键段时, 他可以考虑到歌词的含义、前后乐句之间的关系等等非常重要的一些因素, 而这些在算法中是无法顾及到的. 这也为我们将来进一步改进工作指明了方向.

图 4和图 5给出了两个具体的例子, 将分别从两首著名的流行歌曲中分割关键段. 一首是电影《泰坦尼克》的主题歌——《My heart will go on》它的演唱者是 Celine Dion, 另一首是《耶利亚女郎》它的演唱者是童安格. 在这两个例子当中, 自动关键段与人工关键段非常一致. 在第一个例子当中, 事实上自动关键段要稍好一些, 因为它的歌词更加完整. 也许是由于志愿者工作时的粗心, 这首歌曲的人工关键段在边界处理上并不十分完美; 在第二个例子当中, 人工关键段与自动

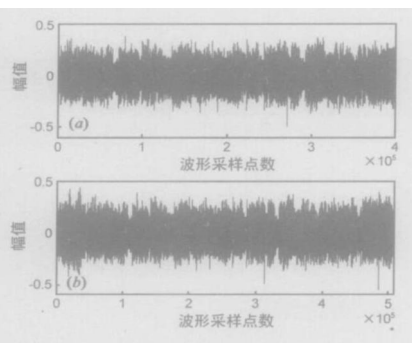


图 4 流行歌曲《My heart will go on》的关键段波形图. (a) 为人工截取的关键段的波形图, 长度为 33.1 秒, 相应的歌词 “are here there’s nothing I fear and I know that my heart will go on, we’ll stay forever this way, you are safe in my heart and my heart will go on”; (b) 为自动截取的关键段的波形图, 长度为 43.2 秒, 相应的歌词 “you are here, there’s nothing I fear and I know that my heart will go on, we’ll stay forever this way, you are safe in my heart, and my heart will go on and on”.

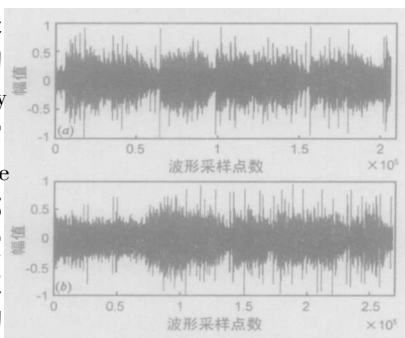


图 5 流行歌曲《耶利亚女郎》的关键段波形图. (a) 为人工截取的关键段的波形图, 长度为 18.7 秒, 相应的歌词为 “耶利亚, 神秘耶利亚, 耶利耶利亚, 耶利亚, 神秘耶利亚, 我一定要找到她”; (b) 为自动截取的关键段的波形图, 长度为 24.3 秒, 相应的歌词与人工关键段完全相同. 由于在开始部分包含了一小段伴奏音乐, 所以比人工截取的关键段要长一些.

段. 由于自动分割出来的关键段的前部包含了一小段伴奏音乐, 所以比人工截取的关键段略长一些.

7 结论

本文研究了流行歌曲中的关键段分割问题, 将流行歌曲中的关键段定义为一首歌曲当中给人印象最深刻的那个片段, 并且设计了一种基于样本的流行歌曲关键段分割方法. 算法在一个包含了 130 首流行歌曲的数据库上进行了测试, 通过与人工截取的关键段进行比较证明了其有效性.

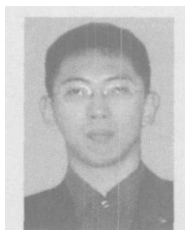
将来我们会继续研究流行歌曲关键段所具有的特征, 希望通过学习与规则相结合的方法来进一步提高算法的效果. 另外, 我们也将探索其它类别音乐中的典型片段的自动分割问题, 并将研究成果应用于基于内容的多媒体数据自动分析和管理.

参考文献:

- [1] Pachet F, et al A combinatorial approach to content-based music selection [A]. Proc of 1999 IEEE Inter Conf on Multimedia Computing and Systems [C]. Florence Italy: IEEE, 1999. 1: 457–1: 462
- [2] Kraft R, et al Method and apparatus for music summarization and creation of audio summaries [P]. US Patent 6225546 2001.
- [3] Logan B, et al Music summarization using key phrases [A]. Proc of 2000 IEEE Inter Conf On Acoustics, Speech, and Signal Processing [C]. Istanbul Turkey: IEEE, 2000. 2: 749–2: 752
- [4] Xu C S, et al Automatic music summarization based on temporal spectral and cepstral features [A]. Proc of 2002 IEEE Inter Conf on Multimedia and Expo [C]. Singapore: IEEE, 2002. 1: 117–1: 120
- [5] Cooper M, et al Automatic music summarization via similarity analysis [A]. Proc of 2002 Inter Symp on Music Information Retrieval [C]. Paris France: Springer-Verlag, 2002. 81–85
- [6] Cooper M, et al Summarizing popular music via structural similarity analysis [A]. Proc of 2003 IEEE Workshop On Applications of Signal Processing to Audio and Acoustics [C]. New York USA: IEEE, 2003. 127–130
- [7] Wei C, et al Music thumbnailing via structural analysis [A]. Proc of 2003 11th ACM Inter Conf on Multimedia [C]. Berkeley USA: Association for Computing Machinery, 2003. 223–226
- [8] Hinata K, et al Interactive music summarization based on generative theory of tonal music [J]. Journal of New Music Research, 2003. 165–177.

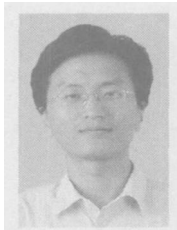
- [9] Shao X, et al Automatically generating summaries for musical video[A]. Proc of 2003 IEEE Inter Conf on Image Processing[C]. Barcelona Spain: IEEE, 2003. 2: 547- 2 550
- [10] Torgue H S Popular Song, 3th ed[M]. Paris France: University Press, 1984. 59- 84
- [11] Zhang Y B, et al A study on content-based music classification[A]. Proc of 2003 IEEE Seventh Inter Symp on Signal Processing and Its Applications [C]. Paris France: IEEE, 2003. 2: 113- 2 116
- [12] Lu L, et al Content analysis for audio classification and segmentation[J]. IEEE Trans Speech and Audio Processing 2002, 10(7): 504- 516
- [13] Li D G, et al Classification of general audio data for content-based retrieval[J]. Pattern Recognition Letters 2001, 22(5): 533- 544
- [14] Zhang Y B, et al Audio segmentation based on multi-scale audio classification[A]. Proc of 2004 IEEE Inter Conf on Acoustics Speech, and Signal Processing[C]. Philadelphia USA: IEEE, 2004. 4: 349- 4 352
- [15] Owens F G. Signal Processing of Speech[M]. London: Macmillan, 1993. 3- 7.

作者简介:



张一彬 男, 1974年10月出生, 分别于1997年和2000年在东北大学自动控制系统获得工学学士学位和硕士学位, 现在在清华大学自动化系攻读工学博士学位. 目前研究兴趣包括模式识别、基于内容的音频及音乐分析、图象分析和信息挖掘等, 已发表相关论文8篇.

E-mail: zybo@ mails. tsinghua. edu. cn



周杰 男, 1968年11月出生, 清华大学自动化系主任助理、教授、博士生导师, 国际刊物《Int J Robotics and Automation》编委, IEEE高级会员, 研究领域包括模式识别、信息融合、图像处理 and 计算机视觉. 近几年, 负责完成十多项科研项目, 在著名国际刊物 (IEEE Trans PAMI IEEE Trans Image Processing Pattern Recognition等) 上发表论文十多篇, 在国际会议上发表论文四十余篇.

E-mail: jzhou@ tsinghua. edu. cn