

从存在伴奏的歌曲中提取歌声基音的时域算法

王佑民, 赵 杰, 江 城

(上海交通大学电子工程系, 上海市 200030)

摘 要:提取歌曲中的唱者基音拥有广泛的用途,如可用于基于内容的音频检索等。在歌曲中提取唱者基音存在许多与普通语音处理不同的问题,传统的时域算法在强音乐背景的干扰下很难正确提取唱者人声基音,近年来研究歌声特殊性的各系统则采用较为复杂的训练模型和频域算法。本系统为改进传统时域算法,提高歌声基音提取的准确度,同时为降低算法复杂度,利于用硬件实现基于内容的音频检索系统,提出一种以人声特征着眼、以开环-闭环基音提取为框架的时域算法。实验证明此算法在歌声基音提取上相对传统时域算法准确度有显著提高。

关键词:浊音能量判决法;开环-闭环基音提取;人声/乐音分离

中图分类号: TN912.34

0 引言

音乐伴奏是歌曲的重要组成部分,但要从歌曲中提取人声特征,却需排除它的巨大干扰。对数字语音处理的研究已历时多年,但至今仍少有系统能很好地解决这个问题。基音是重要的人声特征之一,歌曲中的唱者基音更有广泛的应用:歌词自动匹配将依赖于歌唱者基音的准确提取以分辨不同音乐片段,根据基音值的变化自动为歌曲记谱也将为音乐学习和创作提供有力工具。本算法提取出的基音将作为重要特征向量用于基于内容的音频检索,基音提取的准确性将直接关系到检索系统功能的实现。

20世纪60年代以来,多种基音检测算法被提出,如ACF(自相关函数)法^[1]、AMDF(平均幅度差函数)法^[2]、SIFT(简化逆滤波)法^[3]等,这些传统算法错误率较高、抗噪性较差。许多改进算法随后被提出,如修正的自相关算法^[4]、基于AMDF算法的循环AMDF法^[5]等,这些算法较传统算法有显著进步,但在强背景噪声情况下仍难以取得很好的性能,突出表现为倍频/半频等关系判断的错误。

唱者基音提取与普通的语音基音提取还有不同。首先,歌曲中歌声音调变化更为剧烈。其次,语音信号的基音限于80 Hz~400 Hz,而某些歌唱信号,如花腔女高音的基音最高可达1400 Hz,更多背景噪声处于与歌声基音相同的频段内。此外,歌曲中伴奏信号还与歌唱信号融成有机整体。以上特征使得用歌声信号的时域周期性受到很大干扰,上述传统时域算法更难以正确从歌曲中提取出唱者基音。

由于上述难点,至今尚未有人用时域算法实现从歌曲中提取歌声基音。目前也有学者针对歌曲的特殊性展开研究,有的着眼于人声和音乐伴奏分离,他们认为从歌曲中分离出人声将大大有利于准确地提取出歌声的基音。Meron 和 Hirose 曾设计出将人声和钢琴伴奏分离的系统^[6],但需预知包括大致基音在内的先验信息。Hu 和 Wang 提出用预测基音轮廓和声音信号幅度调制的方法分离歌声和音乐背景^[7]。这一系列研究的进展为基音提取提供了便利,但这些系统同时在相当程度上依赖于基音的准确提取。Li 和 Wang 采用了多分子带,利用自相关和 Mei 倒谱分析-综合,建立HMM(隐马尔可夫模型)、GMM(高斯混合模型)的方法分析歌曲中歌声的基音^[8],着重考虑了歌声基音提取的问题。但是,上述系统所采用的频域算法、训练模型等,存在运算复杂度较大等问题,难于在硬件上实现其功能。

本文提出一种针对歌曲提取唱者人声基音的时域算法,首先相对于传统的时域算法在提取歌声基音的准确性和抗噪性上有显著提高。同时,没有采用训练模型各频域参数,从而节约了计算复杂度。因此,既为音频检索系统提供了可靠的特征向量,又为整个系统的硬件实现提供了便利。实验证明此算法在歌声基音提取上相对传统时域算法准确度有显著提高。

1 从存在伴奏的歌曲中提取歌声基音

从存在伴奏的歌曲中提取歌声的关键是利用人声与音乐背景根本性质上的区别,强化语音特征,以利于基音提取。我们研究了人的发声特征,重点利用浊音能量在频域上周期性分布的特性,提出了基于SIFT的浊音能量判决法,解决传统时域算法普遍存在的倍频/

半频判决错误等问题。仅一个判决法仍不足以保证基音提取的精确性,为此我们借鉴了 3GPP AMR-WB+标准^[9]的编码中基音提取部分,采用开环-闭环方式确定基音,即先确定基音大致取值范围,再做调整和精确取值,提高了基音提取的准确性和系统的抗噪性。此外,歌曲中有些时段是不存在歌声的,如果对歌曲每个时间点都提取基音,对这些纯背景音时段提取的歌声基音必然是错误的。针对这个问题,提出了能量比较法分辨歌声/乐音背景,除去了对只存在乐音背景的时段提取的基音。

1) 基于 SIFT 的浊音能量判决法

SIFT 法是较为经典的传统算法,其核心是采用 LPC(线性预测编码)抵消人的声道参数的调制效应。在语音信号的数字模型中,浊音的激励函数是周期性的,其周期为 $1/F_0$,即基音周期。语音可视为基音脉冲经声道调制后产生。

基于 LPC 的语音信号模型^[4],把唇辐射、声道、声门激励的组合效应用一个短时不变的全极点数字滤波器来表示:

$$H(z) = \frac{1}{A(z)}$$

式中: $A(z)$ 为逆滤波器:

$$A(z) = 1 - \sum_{m=1}^p a_m z^{-m}$$

p 为 LPC 分析的阶数。

让语音信号经过 $A(z)$ 就可得到滤波器 $H(z)$ 的输入信号,即浊音的激励函数,用于基音检测。

传统的 SIFT 算法中,数据经过逆滤波器滤波之后,通过计算自相关函数的极值点,作为最终的基音输出。但是,由于音乐中背景噪声干扰较为强烈,信号的周期性不能很好体现,因而使用自相关函数提取基音值错误率较大。我们利用浊音能量在频域上周期性分布的性质,用浊音能量判决法代替直接的自相关函数提取基音值。选取自相关谱上的多个峰值作为候选基音,并在候选基音各倍频点上用峰值滤波器滤波,得到候选基音对应的浊音脉冲序列信号能量。这里的峰值滤波器为一窄带通滤波器,带宽参照人声浊音脉冲带宽设置。选取对应能量最大的基音值作为输出基音。

2) 开环-闭环基音提取

利用前文所述的基于 SIFT 的浊音能量判决法,为了针对性解决传统基音提取算法普遍存在的倍频/半频关系判断错误,也为了降低浊音能量判决法判决的复杂度,先通过开环基音分析大致估算基音周期,记为 T_{op} 。开环基音分析每帧(39 ms)进行一次。

在此基础上,利用未减采样数据内该帧的前几帧、后几帧以及开环基音值周围数值的信息设定候选闭环基音集合 δ ,用 δ 中包括的 21 个基音候选值进一步确保基音轨迹的平滑性和正确性,这就提取出闭环基音,记为 T_{cl} 。定义某一帧闭环基音与开环基音之间的转化关系为:

$$T_{cl} = NT_{op}$$

式中: N 为减采样的倍数。

3) 能量比较法分辨歌声/乐音背景

将各帧分为歌声占主导的帧和乐音背景占主导的帧两种情况。首先通过构造基于 LPC 的逆滤波器,加强信号中的浊音成分,再用峰值滤波器在闭环基音所对应的频率及其各次谐波频率上进行滤波,得到仅包含基频及其谐波信息的信号,再求出此信号的短时能量,即得到浊音脉冲序列能量,记为 E_{vo} 。

同时,计算未经过逆滤波的歌曲原始信号的各帧短时能量,记为 E_{io} 。定义一帧中浊音能量占总能量的比重 $R = E_{vo}/E_{io}$ 。由于浊音能量占总能量的比重反映语音在信号中所占的比重,通过设定门限,将 R 值较小的帧判为乐音背景, R 值较大的帧判为歌声。

2 系统描述与效果

依照前文所述,本系统分为 3 个环节,如图 1 所示。

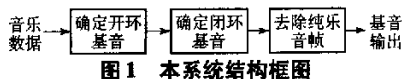


图1 本系统结构框图

本系统采用 20 首流行乐曲作为实验材料,手工截取其背景音比较强的高潮片段,歌曲采样率为 44.1 kHz,量化位数为 16 bit,格式为 .wav。

以下以背景音较为强烈的歌曲《暗香》片段为例,介绍各环节性能。

2.1 确定开环基音周期

该环节主要目的是在较粗糙范围内确定基音周期的大致取值。

步骤如下:

1) 预处理

依次经过 80 Hz ~ 4 410 Hz 滤波、5:1 减采样、分帧(39 ms)。滤波是为减采样时不产生频谱混叠,同时也为滤除人声频带外的噪声。分帧时,由于后续步骤采用的是修正的自相关函数(见步骤 3),实际每帧样本点共计 256 + 88 个。

2) 构造 LPC 逆滤波器对输入数据调制并在修正自相关法谱上待选基音值

这里采用了基于 SIFT 的浊音能量判决法后提取

的基音轮廓效果如图2(c)所示。

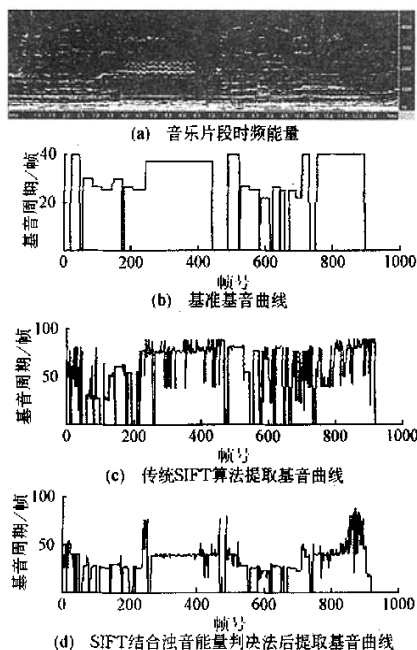


图2 开环基音提取环节效果

为了用数据说明各算法提取基音的准确程度,借用文献[8]中采用的计算误差的方式和参数,如果某帧的基音值超过或小于参考基音20%,该帧基音就判为检测错误,记歌曲总帧数为 N_{total} ,被判为检测错误的帧数为 N_{err} ,基音提取错误率定义为 $R_{err} = N_{err}/N_{total}$ 。目前对大部分歌曲不能找到不带伴奏的纯净歌声用于提取出基音轮廓作为参考基音,于是采用该歌曲的曲谱,配合观察该歌曲的时频图(如图2(a)所示)推断出每一帧的基音,得到较为粗糙的音阶轮廓作为参考基音,如图2(b)所示。

有了参考基音,用基音提取错误率的定义公式可以得到用传统SIFT法检测减采样数据该片段基音值的错误率为80.48%,其主要问题是倍频/半频关系处理错误。

本文采用前文所介绍的基于SIFT的浊音能量判决法,在白相关谱上选取7个峰值(非最大值)作为候选基音,供浊音峰值能量法判决。

引用文献[4],这里采用修正自相关函数,它是自相关法的改进算法,公式为:

$$R_n(k) = \sum_{m=0}^{N-1} x_n(m) x_n'(m+k) \quad 0 \leq k \leq K$$

3) 浊音能量判决法判决

采用浊音能量判决法,选取对应浊音能量值最大的候选基音作为开环基音 T_{op} 。经过判决后的基音轮廓如图2(d)所示。相对图2(c)所示传统SIFT算法

提取的基音轮廓有明显改善,错误率降至29.18%。

2.2 确定闭环基音周期

此环节主要目的是根据输入的开环基音 T_{op} 转换、修正得到为闭环基音 T_{cl} 。

1) 预处理

依次经过清音切除、静音切除、80 Hz ~ 800 Hz 带通滤波器、分帧(39 ms)。判别某帧数据为清音依据的是短时过零率。引用文献[4],浊音语音能量约集中在3 kHz以下,而清音能量多出现在较高频率上,即意味着高的平均过零率,因此具有较高过零率的帧判为清音,予以切除。短时能量代表该帧信息量的多少,短时能量过小的帧判为静音,予以切除。

带通滤波器的频率上下限依据歌曲中歌声基音的上下限而定。分帧时,因对候选基音值检验时采用修正的自相关函数,实际每帧样本点为 $1280 + 450$ 个。

2) 用修正自相关法确定闭环

先依据公式 $T_{cl} = NT_{op}$ 将开环基音进行转换。此处由于确定开环基音环节采用了5:1减采样,所以 $N = 5$ 。随后采用修正的自相关法确定闭环基音。图3(a)为未减采样数据下参考基音周期值图。图3(b)为经过闭环环节之后的基音轮廓。对比图2(d),闭环提取后,基音轮廓更为平滑,通过引入前后帧信息并在转换后的开环基音周围选取基音值进行检验。基音提取错误率降至26.25%。

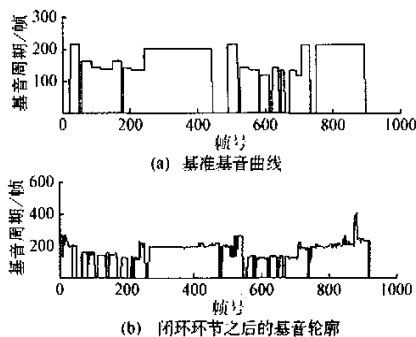


图3 闭环提取基音环节效果

2.3 依靠基音去除纯乐音帧

首先对歌曲原始数据分帧(39 ms),计算各帧短时平均能量。同时,用峰值滤波法对经过80 Hz ~ 4 410 Hz 滤波、减采样和LPC逆滤波后的数据提取出浊音脉冲序列能量。然后利用能量比较法,依据公式 $R = E_v/E_n$,分辨语音/乐音帧。此处 R 的门限值,设为0.0067,比值小于门限的帧判为乐音背景。

图4方框内所围的部分表示要去除的纯乐音帧。经过此环节削除这些帧,与参考基音相比,错误率降至24.08%。

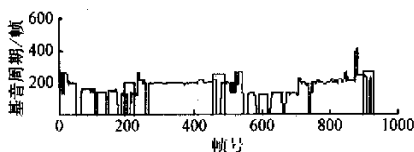


图4 去除纯乐音帧效果

3 实验结果与对比

本系统相比传统时域算法有了很大的改进,图5是各算法运算得到的歌曲《暗香》片段的基音轮廓。

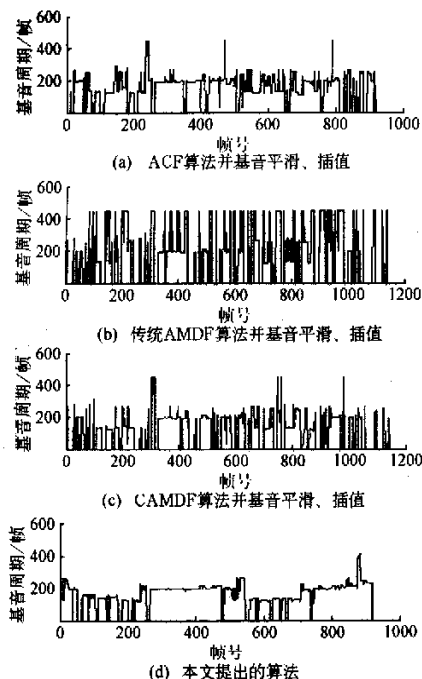


图5 各算法效果比较

图5(c)所示CAMDF(循环平均幅度差函数)法^[5]克服了平均幅度差函数的峰值幅度随着滞后时间的增加而逐渐下降的现象,相对传统AMDF法改进了倍频/半频等关系的处理问题。

由图5(d)可见,本文提出的算法所提取的基音轮廓更为平滑,与其他传统算法相比,与图3(a)所示的参考基音相比最为接近。

统计对于20首歌曲的基音提取值,根据前文所述对于基音提取错误率的规定,各算法提取错误率平均值比较如下:ACF法为42.80%;AMDF法为71.84%;CAMDF法为53.21%;SIFT法为79.34%;新算法为22.79%。由此可见,新算法对于传统算法改进效果明显。

4 结束语

本文针对传统时域算法处理歌声基音提取的不足,提出了用时域算法在歌曲中提取唱者基音。实验证明,本系统采用的开环-闭环基音提取流程、能量判决法等一系列创新算法对于提高基音提取正确率效果显著,最终的提取结果相对传统算法有明显改进。进一步可以利用基音以硬件实现基于内容的音频检索系统。仍然存在的问题是在对于去除乐音背景干扰还有待进一步深入研究,以促进在歌曲中更为准确地提取基音。

参考文献

- [1] RABINER L R, CHENG M J, ROSENBERG A E, et al. A comparative performance study of several pitch detection algorithms[J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1976, 24(5): 399-418.
- [2] ROSS M J, SHAFFER H L, COHEN A, et al. Average magnitude difference function pitch extractor[J]. IEEE Trans on Acoustics, Speech, and Signal Processing, 1974, 22(5): 353-362.
- [3] MARKEL J D. The SIFT algorithm for fundamental frequency estimation[J]. IEEE Trans on Audio Electroacoust, 1972, 20(5): 367-377.
- [4] 赵力. 语音信号处理[M]. 北京: 机械工业出版社, 2003: 38-40.
- [5] 张文耀, 许刚, 王裕国. 循环AMDF及其语音基音周期估计计算[J]. 电子学报, 2003, 31(6): 886-890.
- [6] MERON Y, HIROSE K. Separation of singing and piano sounds[C]//Proceedings of 5th International Conference on Spoken Language Processing (ICSLP'98): Vol 3, Nov 30-Dec 4, 1998, Sydney, Australia. 1998: 1059-1062.
- [7] HU G, WANG D L. Monaural speech segregation based on pitch tracking and amplitude modulation[J]. IEEE Trans on Neural Networks 2004, 15(5): 1135-1150.
- [8] LI Y, WANG D L. Detecting pitch of singing voice in polyphonic audio[C]//Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05): Vol 3, Mar 18-23, 2005, Philadelphia, PA, USA. Piscataway, NJ, USA: IEEE, 2005: 17-20.
- [9] 3GPP TS 26.190 V 6.0.0. AMR wideband speech codec: transcoding function[S]. 2004.

王佑民(1986-),男,上海交通大学学生,主要研究方向为计算机视觉、听觉,生物医学信息处理。

(下转第61页)

3 结束语

采用上述光纤接口电路,输入和输出光信号能满足要求的通信速率,实现了CAN通信的光纤传输。使用光纤作为传输介质,各节点间有良好的电气隔离,也避免了高功率脉冲电源放电时产生的强电磁场对控制电路的影响。CAN总线网络不但具有相当高的抗干扰性、实时性和可扩展性,而且能够满足各个电源模块不同优先级的需要。

参 考 文 献

[1] 王莹,肖峰. 电驱动原理[M]. 北京:国防工业出版社,

1995.

- [2] 郭宽明. 现场总线技术应用选编1[M]. 北京:北京航空航天大学出版社,2003.
- [3] 饶运涛,邹继军,郑勇芸. 现场总线CAN原理与应用技术[M]. 北京:北京航空航天大学出版社,2003.
- [4] 史久根,张培仁,陈真勇. CAN现场总线系统设计技术[M]. 北京:国防工业出版社,2004.
- [5] 李正军. 现场总线及其应用技术[M]. 北京:机械工业出版社,2005.

支彬安(1984-),男,硕士研究生,主要研究方向为机械电子工程。

Realization of CAN Bus Communication in Electromagnetic Coil Drive

ZHI Bin'an¹, LEI Bin¹, LI Zhiyuan¹, ZHANG Hongyan²

(1. Ordnance Engineering College, Shijiazhuang 050003, China;

2. Military Representative office of Chongqing Military Representative Department
in the 919th Factory, Chongqing 630000, China)

Abstract: In order to gain high muzzle velocity and drive heavier object, electromagnetic coil driving needs several even hundreds of electrical source modules. Therefore the timely touch off of the high-pulsed power electrical source modules and the controlling and measuring of the charge and discharge parameters have become the critical components for Electromagnetic coil driving. The system brings forward CAN as the field control bus, uses fiber as the communication medium, thus, it fulfills the need of the electrical source modules and avoids the strong electromagnetic disturb in the process of driving.

Keywords: fiber; electromagnetic coil driving; CAN; node; CAN2.0B; frame

(上接第36页)

A Time-domain Algorithm of Extraction Singing Voice's Pitch with Musical Instrument Accompaniment

WANG Youmin, ZHAO Jie, JIANG Cheng

(Shanghai Jiao Tong University, Shanghai 2000030, China)

Abstract: Extraction singing voice from music with music instrument accompaniment can be applied in many areas such as content-based music retrieval. Conventional time-domain algorithms perform poor in strong noise circumstance due to its remarkable diversity from ordinary speech processing. Newly-emerging systems dealing with singing voice and background music employ complex training models and realize it in frequency domain. With the consideration of hardware implementation and resource saving, our system proposes a new time-domain algorithm, which is based on human voice feature. The system adopted an open-loop and close-loop framework. Simulations on MATLAB show that accuracy of pitch extraction is superior to conventional time-domain algorithms and some other modified ones.

Keywords: voiced sound energy decision; open-loop and close-loop pitch extraction; speech/music discrimination

从存在伴奏的歌曲中提取歌声基音的时域算法

作者: [王佑民](#), [赵杰](#), [江城](#), [WANG Youmin](#), [ZHAO Jie](#), [JIANG Cheng](#)
作者单位: [上海交通大学电子工程系, 上海市, 200030](#)
刊名: [电子工程师](#)
英文刊名: [ELECTRONIC ENGINEER](#)
年, 卷(期): 2007, 33(11)

参考文献(9条)

1. [MERON Y;HIROSE K](#) [Separation of singing and piano sounds](#) 1998
2. [张文耀](#); [许刚](#); [王裕国](#) [循环AMDF及其语音基因周期估计计算](#)[期刊论文]-[电子学报](#) 2003(06)
3. [赵力](#) [语音信号处理](#) 2003
4. [3GPP TS 26.190 V 6.0.0](#) [AMR wideband speech codec:transcoding function](#) 2004
5. [LI Y;WANG D L](#) [Detecting pitch of singing voice in polyphonic audio](#)[外文会议] 2005
6. [HU G;WANG D L](#) [Monaural speech segregation based on pitch tracking and amplitude modulation](#)[外文期刊] 2004(05)
7. [MARKEL J D](#) [The SIFT algorithm for fundamental frequency estimation](#) 1972(05)
8. [ROSS M J;SHAFFER H L;COHEN A](#) [Average magnitude difference function pitch extractor](#) 1974(05)
9. [RABINER L R;CHENG M J;ROSENBERG A E](#) [A comparative performance study of several pitch detection algorithms](#) 1976(05)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_dzgcs200711010.aspx