

基于MATLAB语谱图的声乐研究

白燕燕, 胡晓霞

(西安工业大学北方信息工程学院, 陕西 西安 710200)

摘 要: 语谱图是一种在语音分析以及语音合成中具有重要实用价值的时频图,能反映出语音信号动态频谱特征,被认定是语音信号的可视语言。横坐标代表时间大小,纵坐标代表频率大小,时间和频率所对应的像素点的值表示能量值的大小。本文对一小段语音信号和音乐信号进行处理,最终以语谱图的形式展现出来,并通过对语谱图的分析来判断语音信号和音乐信号,检测语音是否开始和结束,效果良好。这种方法容易区分语音信号中的有用声段和噪声声段。实验证明,纯语音信号频域能量和过零率变化要大于音乐信号。音乐信号较语音信号能量谱稳定。

关键词: 语谱图; 语音分析; 语音合成; 短时能量; 过零率

中图分类号: TP391 **文献标识码:** A

Research on Vocal Music Based on MATLAB Spectrogram

BAI Yanyan, HU Xiaoxia

(Xi'an University of Technological Information, Xi'an 710200, China)

Abstract: Spectrogram is a kind of time-frequency graph with important practical value in speech analysis and speech synthesis. It can reflect the dynamic spectrum characteristics of speech signals and is regarded as the visual language of speech signals. The horizontal coordinate represents time, the vertical coordinate represents frequency, and the value of pixels corresponding to time and frequency represents the magnitude of energy value. This paper deals with a small segment of speech signal and music signal, which are finally presented in the form of spectrogram. By analyzing the spectrogram, we can judge the speech signal and music signal, and detect whether the speech starts or ends. This method is easy to distinguish the useful sound segments from the noise segments in speech signals. It is proved that the change of frequency domain energy and zero crossing rate of pure speech signal are greater than that of music signal. The energy spectrum of music signal is more stable than that of speech signal.

Keywords: spectrogram; speech analysis; speech synthesis; short-term energy; zero-crossing rate

1 引言(Introduction)

语音信号数字处理是一门涉及诸多学科的交叉学科,它以生理学、心理学、语音学和声学等学科为基础^[1]。语谱图的广泛应用研究,从而推动了语谱图在生活中的利用,结合语谱图的技术分析,让其在声纹鉴定、聋人语训、数字音频水印、车型识别、语音合成,以及语音编码等方面也具有研究意义。语音输入人机交互手段,日渐成为智慧生活的一部分。

目前开设人工智能专业,研究语音交互的高校日益增

多。越来越多人工智能产品问世,有代表性的是腾讯公司的微信,小米公司的米聊、百度公司的小度智能音箱和小米公司的小爱智能音箱。本文针对自然语音和音乐从频域能量和过零率方面比较,并分析了语音和音乐语谱图特征。

2 音频信号处理(Audio signal processing)

语音信号采集处理:第一步将信号进行采样量化,第二步进行预处理。

预处理步骤包括:预加重、加窗、分帧等。

在此项研究中,音频是多媒体中的一种重要载体,声音经过模拟设备记录或再生,成为模拟音频,在经数字化成为数字音频。音频分析就是以数字音频信号为分析对象,以数字信号处理为分析手段,提取信号在时域、频域内一系列的过程。音频测量一般包括信号电压、频率、信噪比、谐波失真等基本参数。

语谱图的产生流程框图如图1所示。

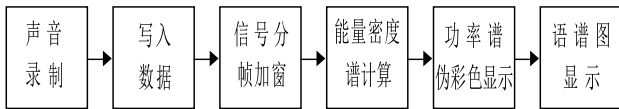


图1 语谱图流程框图

Fig.1 Flow chart of the implementation of the spectrum

2.1 音频信号的预加重

本设计利用录音软件进行录音,把录制好的声音信号和音乐信号以WAVE文件形式存在计算机硬盘上^[2]。预加重为了提升高频段、保持频谱平坦和频谱分析^[3]。预加重能提高信号信噪比,且信号动态范围缩小。预加重发生在A/D转换前后均可。常用一阶数字滤波器实现。这个数字滤波器有6dB/倍频程的提升高频特性,表达式为式(1):

$$H(z) = 1 - \mu z^{-1} \quad (1)$$

式(1)中, μ 值向1趋向,其取值范围是0.94—0.97。

2.2 音频信号的分帧加窗

预加重步骤之后是分帧加窗处理。为了得到10ms—30ms平稳短时音频信号,对录制的音频信号进行分帧加窗操作^[4]。窗函数平滑地滑动,将语音信号分成一帧一帧。

窗函数选择矩形窗和汉明窗^[5]。

矩形窗表达式:

$$\omega(n) = \begin{cases} 1, & 0 \leq n \leq (N-1) \\ 0, & n = \text{其他值} \end{cases} \quad (2)$$

汉明窗表达式:

$$\omega(n) = \begin{cases} 0.54 - 0.46 \cos[2\pi n / (N-1)], & 0 \leq n \leq (N-1) \\ 0, & n = \text{其他值} \end{cases} \quad (3)$$

要选择一个好的窗函数利用其短时特性分析音频信号^[6]。定义好的窗函数须是:在时域时,减小两端时间窗的坡度,使窗口边缘慢慢平滑降为零,不引起急速变化,缓和两帧语音信号的截断效应;在频域时,其带宽要达到3dB以上。

选择矩形窗:

$$h(n) = \begin{cases} 1, & 0 \leq n \leq (N-1) \\ 0, & n = \text{其他值} \end{cases} \quad (4)$$

其数字滤波器的频率响应为:

$$H(e^{j\omega T}) = \sum_{n=0}^{N-1} e^{-j\omega n T} = \frac{\sin(N\omega T / 2)}{\sin(\omega T / 2)} e^{-j\omega T (N-1)/2} \quad (5)$$

矩形窗相位—频率特性呈现线性特性^[7],频率响应公式如

式(6):

$$f_{01} = f_s / N = 1 / NT_s \quad (6)$$

式(6)中, f_s 代表采样频率, $T_s = 1 / f_s$ 代表采样周期。

选择汉明窗:

$$h(n) = \begin{cases} 0.54 - 0.46 \cos[2\pi n / (N-1)], & 0 \leq n \leq N-1 \\ 0, & n = \text{其他值} \end{cases} \quad (7)$$

表1对比了3组矩形窗和汉明窗的参数数据。

表1 矩形窗和汉明窗

Tab.1 Rectangular window and hamming window

窗类型	旁瓣峰值	主瓣宽度	最小阻带衰减
矩形窗	-13	$4\pi/N$	-21
汉明窗	-41	$8\pi/N$	-53

从时频域来看,汉明窗比矩形窗带宽大两倍。矩形窗的主瓣宽度小于汉明窗,频谱分辨率高,然旁瓣峰值较大,容易频谱泄露,造成高频成分丢失。汉明窗比矩形窗显得更平滑些,其旁瓣衰减明显,更能反映信号时频特性。

往往选择窗函数还要看信号基音周期。一个语音帧正常包含1—7个基音周期。不同人的基音周期存在差异。基音周期变化范围通常是2—14ms,基音频率变化范围是500—700Hz,致使N的选择比较困难。采样频率是10kHz时, N点取100—200点比较合适(即10—20ms持续时间)。

2.3 音频信号的能量和过零率分析

短时能量用来分辨短时音频信号中的清音和浊音,浊音和清音差别很大,容易区分;还可以界定有声段和无声段,以及声母和韵母。语音识别中,短时能量也能代表语音信号的一些特征^[8]。

如果时间出现改变,那么信号的能量将随着时间的改变而变化。对于语音信号 $x(n)$ 短时平均能量的定义如下:

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)\omega(n-m)]^2 = \sum_{m=-\infty}^{\infty} x^2(m)h(n-m) = x^2(n) * h(n) \quad (8)$$

一段语音信号中,短时能量高低也是不一样的,无声时能量最低,浊音时能量最高,而清音时能量次之^[9]。图2为语音“西安工业大学”的短时能量图,图3为一小段音乐“darling”的短时能量图。

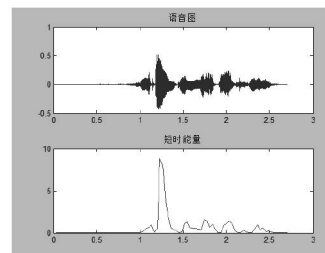


图2 语音“西安工业大学”短时能量图

Fig.2 Short-term energy diagram of Xi'an Technological University

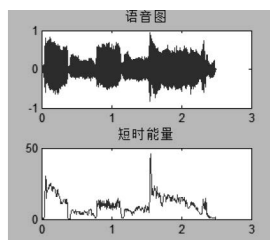


图3音乐信号“darling”的短时能量函数变化图

Fig.3 Short-term energy function of music signal *darling*

从图2和图3可以看出语音部分能量是整段语音中最高的部分。而语音部分又分为静音段、清音段和浊音段三部分,从图中可以明显看出 E_n 值小的是清音段, E_n 值大的是浊音段。短时能量可以用来区分语音的清音段和浊音段。语音中含有比音乐中更多的静音,语音的能量变化比音乐中的大的多。

短时平均过零率代表一帧内信号过零次数。连续信号,观察其时域波形通过X轴次数即可。离散信号,观察其信号采样点符号的变化次数即可。

图4为语音“西安工业大学”的短时平均过零率图,图5为小段音乐“darling”的短时平均过零率图。

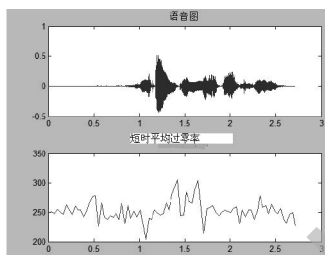


图4 语音“西安工业大学”的短时平均过零率图

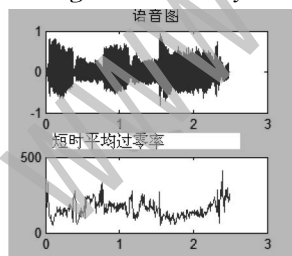
Fig.4 Short-term average zero-crossing rate of the voice *Xi'an Technological University*

图5 音乐“darling”的短时平均过零率图

Fig.5 Short-term average zero-crossing rate of music *darling*

从图4和图5可以看出,语音信号和音乐信号不一样的地方,其由发音的音节和不发音的音节交替组成。语音由于清浊音交替出现,过零率变化明显比音乐激烈。

3 语谱图生成和分析(Spectrogram generation and analysis)

本次设计利用Cool Edit Pro, 普通立体声麦克风和笔记

本电脑进行录音。设定采样频率8000Hz, 帧长取100点, 两帧间距取100点, 进行16比特量化。经过采样之后, 得到标准化数字语音信号。

此次录入的语音信号“西安工业大学”文件名为xg.wav, 截取音乐“darling”, 将音乐保存文件名为darling.wav。

将音频信号经过前面一系列的处理之后, 我们得到语音信号和音乐信号的语谱图, 分别如图6和图7所示。

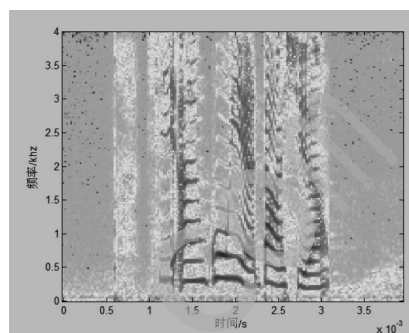


图6 语音“西安工业大学”语谱图

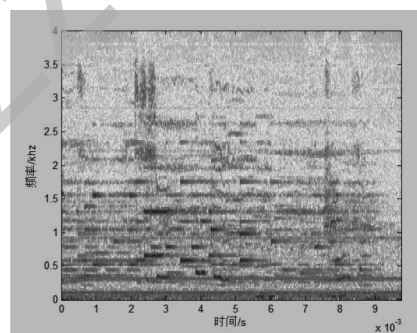
Fig.6 The spectrum of *Xi'an Technological University*

图7 音乐“darling”语谱图

Fig.7 Music *darling* spectrum

从图6和图7得到, 语音信号和纯音乐歌曲信号相比, 语音频谱中峰值变化快, 不稳定。语谱图横轴代表时间, 纵轴代表频率, 一个个像素点的值对应于语音信号的能量密度值。采用二维平面示意三维信息, 其能量值的大小用颜色来表示, 颜色深, 代表这个点的语音能量越强。渲图效果越亮即此处能量越大。

观察图6可发现在彩色语谱图中会间断的出现空隙, 这是因为人在说话的时候会有呼吸的交替, 从而出现空隙, 在图7中, 间隙很少出现, 在音乐播放同时乐谱的伴奏是连续且不间断地演奏从而很少会出现图像中间隙的部分。

观察图6和图7可以看出横杠清楚的出现在图中的摩擦乱纹, 说明了此处语音的录入出现了浊音。还有一些没有规则的乱纹, 这些条纹代表了语谱图中具有和时间轴平行的横条和与时间轴垂直的竖直条。其中, 横条的出现表示语音信号中的共振峰, 对于竖直条而言, 代表了语音信号基音的周期以及基音频率等信息。

如果在连续的若干帧中,能量谱中的峰值^[10]出现在频域中相对稳定的位置,则认为这些音频信号含有音乐成分。

4 结论(Conclusion)

语谱图是一种可观察信号特性的平面图,通过分析音频的产生原理可以观察到声源点的共振属性和声乐在自然界的特性。语谱图用颜色深浅表示声纹强度大小。颜色深在语谱图中所占比重较大,那么相应影响人感知的效果要强烈得多。通过实验分析,可以有效区分纯音乐、背景音乐语音、纯语音。这一过程在场景识别、语音识别、声纹识别中起着关键性的作用。

参考文献(References)

- [1] Zhen Huang,Sabato Marco Siniscalchi,Chin-Hui Lee,et al.A unified approach to transfer learning of deep neural networks with applications to speaker adaptation in automatic speech recognition[J].Neurocomputing,2016(218):448-459.
- [2] Ivan Himawan,Petr Motlicek,David Imseng ,Sridha Sridharan,et al.Feature mapping using far-field microphones for distant speech recognition[J].Speech Communication,2016(83):1-9.
- [3] Foster R.Goss,Li Zhou,Scott G.Weiner,et al.Incidence of speech recognition errors in the emergency department[J].International Journal of Medical Informatics,2016(93):70-73.
- [4] 韩纪庆,张磊,郑铁然,等.语音信号处理[M].北京:清华大学出版社,2004:10-19.
- [5] 白燕燕,胡晓霞,郑三婷,等.基于听觉特性的声纹识别系统的研究[J].电子设计工程,2015,2(4):86-91.
- [6] 赵力.语音信号处理[M].北京:机械工业出版社,2003:43-51.
- [7] 张峰,石现峰,张学智,等.数字信号处理原理及应用[M].北京:电子工业出版社,2010:43-55.
- [8] 李富强,万红,黄俊杰,等.基于MATLAB语谱图的显示与分析[J].微计算机信息,2005,21(103):71-76.
- [9] 白燕燕.基于声纹识别的身份确认系统的研究[D].西安工业大学,2012:16-21.
- [10] 陈青,龚乾,张鸣,等.基于语谱图的声乐分析[J].微计算机信息(管控一体化),2010,26(73):6-8.

作者简介:

白燕燕(1983-),女,硕士,讲师.研究领域:语音信号处理,语音识别.
胡晓霞(1986-),女,硕士,讲师.研究领域:图像处理,图像识别.

(上接第7页)

延的方向、烟雾扩散的范围后,工程机器人用机械臂夹取小型灭火球帮助消防员扑灭火源,消防员则利用破拆、喷雾水流、移动排烟设备等方法进行人工排烟。

4.3 清理工作

火灾扑灭后的清理工作,防止二次爆炸或燃烧。火灾扑灭后利用消防救援机器人全面、细致地检查火场,彻底消灭余火;对石油化工生产装置、储存设备的温度及其周围可燃气体、易燃可燃液体蒸汽的浓度进行检测,并进行相应的处理,防止复燃。

5 结论(Conclusion)

消防辅助救援机器人是由一台领航机器人和两台跟随机器人组成的多智能体编队。本产品的主要功能有火灾紧急辅助救援、动态监视、灾情探测、异常报警、危险品抓取、紧急物资递送等,可提高救援效率、保障消防员安全,为紧急情况救援、环境探测、灾后巡检等任务提供解决方案。这个设计未来有三个主要发展方向:一是将三辆车进行放大化,往人形机器人方向发展;二是提高系统语音识别的智能程度,向智能交互发展;三是利用基站定位或者雷达定位对阵型的变换进行加强。最终目的是希望它能升级应用到更大的工业活动中去。

参考文献(References)

- [1] Ando H,Ambe Y,Ishii A,et al.Aerial Hose Type Robot by Water Jet for Fire Fighting[J].IEEE Robotics & Automation Letters,2018,3(2):1128-1135.

- [2] Shamsudin A U,Ohno K,Hamada R,et al.Consistent map building in petrochemical complexes for firefighter robots using SLAM based on GPS and LIDAR[J].Robomech Journal,2018,5(1):7.
- [3] Lu Y,Zhang L,Wei J V D,et al.Beyond Eleven Color Names for Image Understanding[J].Machine Vision & Applications,2018,29(2):361-373.
- [4] 陶霖密,徐光祐.机器视觉中的颜色问题及应用[J].科学通报,2001,46(3):178-190.
- [5] 郭豪,李宝慧,赵树忠.基于模糊PID控制的步进电机建模与仿真[J].机械工程与自动化,2018,48(2):167-168.
- [6] 陈建斌.消防机器人在我国灭火救援中的应用现状和前景分析[J].地球,2017(10):98.
- [7] 郑鹏.基于消防机器人的救援辅助系统的研究[J].山东工业技术,2014(10):57-58.

作者简介:

陈伟哲(1998-),男,本科生.研究领域:智能机器人.
黄梦醒(1973-),男,博士,教授.研究领域:人工智能大数据海洋通信.本文通讯作者.
吴迪(1991-),男,硕士.研究领域:人工智能大数据海洋通信.
肖瑾(1998-),女,本科生.研究领域:智能机器人.
丁一寅(1998-),男,本科生.研究领域:智能机器人.
张晨宇(1998-),男,本科生.研究领域:智能机器人.