

# 基于短时傅里叶变换的敏感语音提取

张学敏

(长春工程学院 电气与信息学院,吉林 长春 130012)

**摘要:**在纯净语音条件下,提出了一种基于短时傅里叶变换的敏感语音提取,该提取利用3种FIR数字滤波器对纯净语音在不同频率范围的特征语音进行选取,从而产生不同的音效。待处理的语音文件可以是自己录制的一段语音,也可以是任意格式的音频文件,通过DSP和编译软件实现敏感语音提取,并在数字处理芯片TMS320C5402上实现。

**关键词:**短时傅里叶变换;数字滤波器;敏感语音提取;DSP

中图分类号:TP391.42

文献标识码:B

## Sensitive speech extraction based on short-time Fourier transform

ZHANG Xue Min

(School of Electrical Engineering & Information Technology, Changchun Institute of Technology, Changchun 130012, China)

**Abstract:**On condition of pure sound, this paper put forward a sensitive speech extraction based on short-time Fourier transform, which used three FIR digital filter extract sensitive speech in different frequency range. The sound that will be processed can be recorded or other sound with any format, then the extraction is compiled by CCS and realized by digital processing chip TMS320C5402.

**Key words:** short-time transform; digital filter; sensitive speech extraction; DSP

语音信号属于非平稳信号,它的一个重要特点是语音信号随时间变化而随机变化的,又由于人的声道形状及其变化规律具有一定惯性,因此在一段时间间隔内语音信号保持相对稳定,所以对语音信号的分析就必须建立在“短时”概念基础上。本文基于短时傅里叶变换方法,利用FIR数字低通、带通和高通滤波器对原始语音信号在不同频率范围进行提取,得到感兴趣的敏感语音。处理后的语音在低音、中音和高音段音效明显,达到了敏感语音提取的目的。由于所处理的音频文件不受格式限制,灵活性和操作性好,因此这种方法可应用于动漫音效合成和特殊音效生成场合,具有实际应用价值<sup>[1-2]</sup>。

### 1 短时傅里叶变换

#### 1.1 敏感语音提取流程

敏感语音提取框图如图1所示。



图1 敏感语音提取框图

敏感语音提取的傅里叶变换运算及滤波在TMS320C5402上实现。TMS320C5402是低功耗、16位定

点DSP芯片,它具有10ns的指令周期和100MIPS的运算能力。首先,将由话筒采集来的模拟声音信号通过A/D转换变成数字信号,然后对该数字信号做短时傅立叶变换,继而送入滤波器对变换后的频谱进行相应频段的滤波处理,再将处理后的数字信号通过D/A转换变成模拟信号,通过扬声器播放。这里A/D和D/A转换采用音频芯片TLC320AD50C,它集成了16位A/D、D/A转换器,采样频率最高达22.05kb/s。

#### 1.2 短时傅里叶变换

语音的感知过程与人类听觉系统的频谱分析功能是密切相关的,因此对语音信号进行频谱分析是认识语音信号和处理语音信号的重要方法。傅里叶变换可将时域上较复杂的运算转化为相对简单的频域运算,所以通过傅里叶变换可以使信号的某些感兴趣的特性变得明显,而在时域信号中这些特性是无法体现出来的。由于语音信号在短时间内是局部平稳的,所以可对某一帧语音进行傅里叶变换,即短时傅里叶变换,定义式如下:

$$X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)\omega(n-m)e^{-j\omega m}$$

式中 $\omega(n-m)$ 是窗函数序列,针对不同窗函数序列,可以

# 应用奇葩

## Example of Application

得到不同傅里叶变换结果。由上式可知,短时傅里叶变换有两个变量  $n$  和  $\omega$ ,所以它是时间  $n$  和频率  $\omega$  的时频函数。当  $n$  固定不变时,它仅是频率  $\omega$  的函数,即是序列  $x(n)$  的标准傅里叶变换。

由于语音信号是随时间缓慢变化的,在短时间间隔内语音信号的时间特性是固定不变的,因此可将语音信号进行短时分析,语音信号短时分析的一个重要问题是采样点数的选取。短时范围内,采样频率、采样点数和持续时间之间的对应关系见表 1。

表 1 采样频率与采样点数选择对应表

采样频率 $f_s$ (Hz)	采样点数 $N$ (个)	持续时间 (ms)
5k~10k	50~200	10~20
10k~20k	100~500	10~25
20k~50k	500~1000	20~25
50k~100k	1 000~2500	20~25
100k 以上	大于 2 500	大于 25

通常童声高音频率范围 260Hz~880Hz,低音范围在 196Hz~700Hz;女生高音范围在 220Hz~1.1kHz,低音范围在 200Hz~700Hz;男生高音范围在 160Hz~523Hz,低音范围在 80Hz~358Hz。所以,按照奈奎斯特准则,采样频率 8kHz 就能满足要求,比如对声音的采样频率是 8kHz,持续时间 16ms,则采样点数  $N=16 \times 10^{-3} \times 8000 = 128$  点。

### 2 短时傅里叶变换的实现

敏感语音提取是在 DSP 编译软件 CCS 环境下实现的。CCS 是 DSP 芯片开发需要的一套完整的软、硬件开发工具。它具有强大的源代码编辑器,允许编辑 C 语言源代码和汇编源代码,能在 C 代码之后显示与之对应的汇编指令。利用 CCS 集成开发环境,用户可以在一个开发环境下完成工程定义、程序编辑、编译链接、调试和数据分析等工作环节。本文语音提取在 CCS 环境下的一个界面图如图 2 所示。

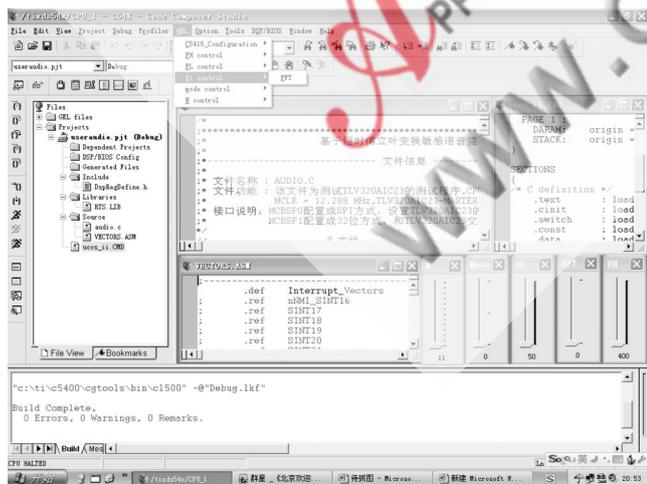


图 2 CCS 环境下敏感语音提取的界面图

由 C 语言实现的短时傅里叶变换的部分代码如下:

```
void kfft(double pr[],double pp[],int n,int k,double fr[],
double fi[],int l,int il)
```

```
{
int it,m,is,i,j,nv,l0;
double p,q,s,vr,vi,poddr,poddi;
for (it=0; it<=n-1; it++)
{ m=it; is=0;
for (i=0; i<=k-1; i++)
{ j=m/2; is=2*is+(m-2*j); m=j;}
fr[it]=pr[is]; fi[it]=pp[is];
}
pr[0]=1.0; pp[0]=0.0;
p=6.283185306/(1.0*n);
pr[1]=cos(p); pp[1]=-sin(p);
.....
if (il !=0)
for (i=0; i<=n-1; i++)
{ fr[i]=fr[i]/(1.0*n);
fi[i]=fi[i]/(1.0*n);
}
if (il !=0)
for (i=0; i<=n-1; i++)
{ pr[i]=sqrt(fr[i]*fr[i]+fi[i]*fi[i]);
if (fabs(fr[i])<0.000001*fabs(fi[i]))
{ if ((fi[i]*fr[i])>0) pp[i]=90.0;
else pp[i]=-90.0;
}
else
pp[i]=atan(fi[i]/fr[i])*360.0/6.283185306;
}
}
```

为不失一般性,自录一段语音,命名为 myvoice.wav。将其加载到硬件芯片上,采样频率 8kHz,采样点数 128 点,在 DSP 编译软件 CCS 环境下运行,该声音信号的时域和频域图形如图 3、图 4 所示。

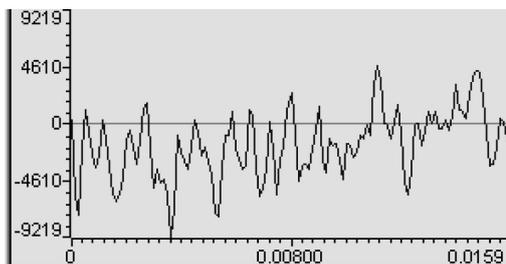


图 3 语音信号的时域图

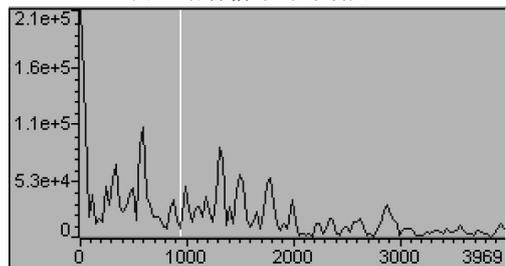


图 4 语音信号的频域图

# 应用奇葩

## Example of Application

由图 2 可见,短时范围为 15.9ms,近似 16ms。由图 3 可见,声音频率范围主要集中在 3 000Hz 以内。下面将频率范围分为 500 Hz 以内,500Hz~1 000Hz 和 1 000Hz 以上 3 种情况对语音进行滤波处理<sup>[3-5]</sup>。

### 3 不同频率范围的敏感语音提取

#### 3.1 敏感语音提取流程图

敏感语音提取的程序流程框图如图 5 所示。

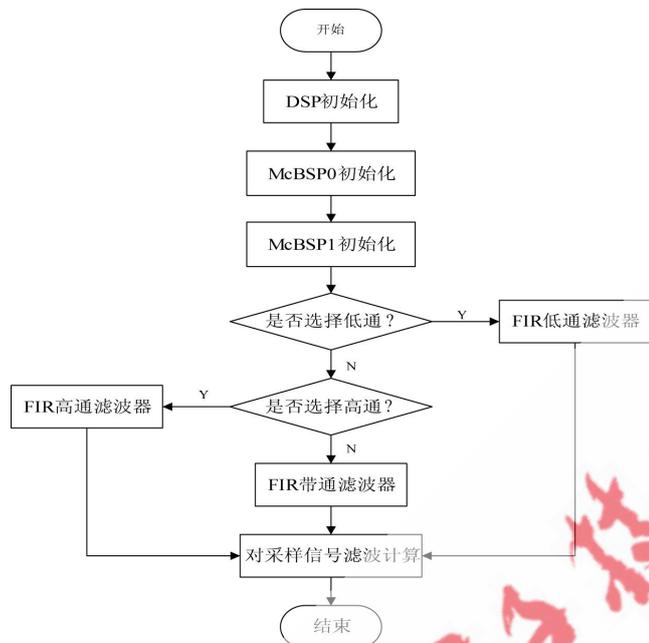


图 5 敏感语音提取流程图

#### 3.2 提取语音的低音部分

利用 FIR 滤波器提取 500 Hz 的低音部分,根据上面语音信号特点,FIR 数字低通滤波器的截止频率设为 500 Hz,采样频率设为 8 kHz,阶数选 20。FIR 低通滤波的 C 语言代码如下:

```
void firldi(double h[],double nlpass,int N)
{
    double wc;
    int a,i;
    wc=nlpass;
    for(i=0;i<N;i++)
        h[i]=0;
    a=(N-1)/2;
    h[a]=wc/pi;
    for(i=0;i<a;i++)
        h[i]=sin(wc*(i-a))/(pi*(i-a));
    for(i=a+1;i<N+1;i++)
        h[i]=sin(wc*(i-a))/(pi*(i-a));
}
```

原声音的频谱图和滤波后得到的低音部分的频谱图如图 6、图 7 所示。

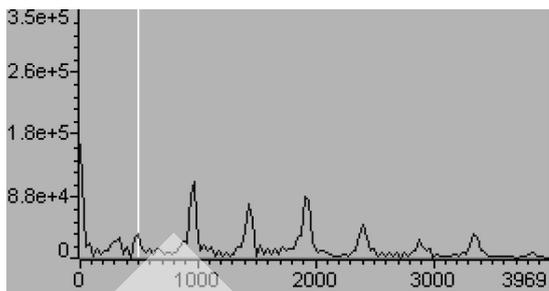


图 6 原语音信号的频谱图

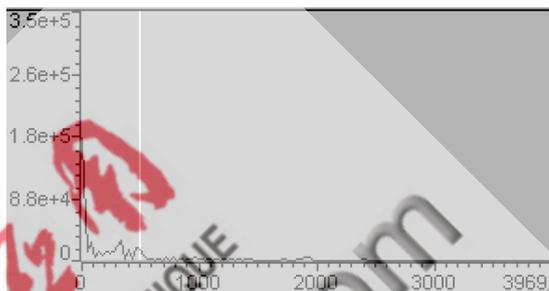


图 7 低音部分的频谱图

由图 7 可见,500 Hz 以上的频谱部分几乎完全被衰减,仅留下了 500 Hz 以内的低频频谱。该频段的声音播放后,较原声音低沉,带有低音炮效果<sup>[6-7]</sup>。

#### 3.3 提取语音的中音部分

FIR 数字带通滤波器的截止频率设为[500 1000] Hz,采样频率设为 8 kHz,阶数选 20。FIR 带通滤波的 C 语言代码略。原声音的频谱图和中音部分的频谱图如图 8、图 9 所示。

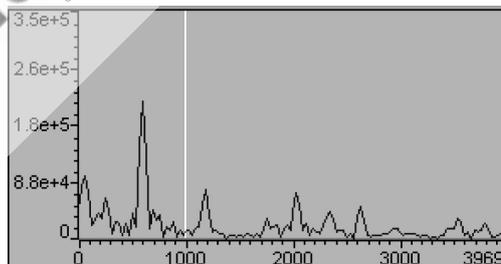


图 8 原语音信号的频谱图

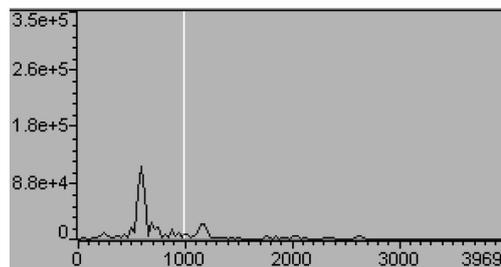


图 9 中音部分的频谱图

由图 9 可见,[500 1000] Hz 以外的频谱部分被衰减,仅留下了[500 1000] Hz 这段频谱。该频段的声音播放后,音调具有中音效果。

#### 3.4 提取语音的高音部分

FIR 数字高通滤波器的截止频率设为 1 000 Hz,采样  
《微型机与应用》2009 年第 24 期

## 应用奇葩

## Example of Application

频率设为 8 kHz, 阶数选 20。FIR 高通滤波的 C 语言代码略。原声音的频谱图和高音部分的频谱图如图 10、图 11 所示。

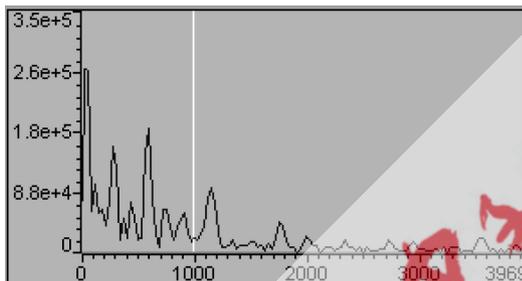


图 10 原语音信号的频谱图

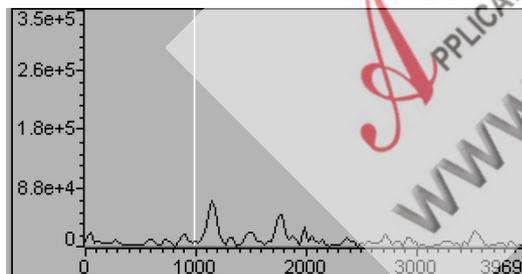


图 11 中音部分的频谱图

由图 11 可见, 1 000 Hz 以内的频谱被衰减, 高频分量被保留。该频段的声音播放后, 音调具有高音效果。

语音信号是一个非平稳过程, 因此适用于周期、瞬变或平稳随机信号的标准傅里叶变换不能直接用来表示语音信号, 本文利用短时分析, 对语音信号处理前后的频谱进行比较, 得出结论: 经过短时分析, 语音信号中不感兴趣的部分得到了抑制, 感兴趣的部分得到保留, 有效实现了敏感语音提取。

## 参考文献

- [1] 陈永彬. 语音信号处理[M]. 上海: 上海交通大学出版社, 2000.
- [2] 胡航. 语音信号处理[M]. 哈尔滨: 哈尔滨工业大学出版社, 2001.
- [3] 邹彦. DSP 原理及应用[M]. 北京: 电子工业出版社, 2007.
- [4] 张雄伟, 陈亮, 徐光辉. DSP 芯片的原理与开发应用(3 版)[M]. 北京: 电子工业出版社, 2003.
- [5] 乔瑞萍, 崔涛, 张芳娟. TMS320C54x 原理及应用[M]. 西安: 西安电子科技大学出版社, 2005.
- [6] 刘益成. TMS320C54x DSP 应用程序设计与开发[M]. 北京: 北京航空航天大学出版社, 2002.
- [7] 桑迪. Rulph DSP applications using C and the TMS320C6x DSK [M]. 北京: 电子工业出版社, 2007.

(收稿日期: 2009-08-28)