

# 波形变换编码语音分析和 合成开发系统研制\*

陈云凤 汪立全 陈志成 邵玉翠 伍康文

(中山大学无线电电子学系)

**摘要** 本文介绍一个适用于微型计算机的波形变换编码语音分析和合成开发系统。文中对波形变换编码方法作了概括,介绍语音分析和合成开发系统的硬件结构和软件组成,并介绍系统的使用情况和提出进一步改进意见。

**关键词** 语音分析, 语音合成, 波形变换编码

波形变换编码是八十年代初期由Mozer首先提出来的一种语音编码技术<sup>[1]</sup>,由于Mozer编码算法复杂,运算量大,通常需要在大型计算机上运行。本课题的目的是寻找适用于微型计算机的波形变换编码快速算法,并在此算法基础上建立一套微型机波形变换编码语音分析和合成开发系统。有关波形变换编码算法的研究,我们在文[2,3]中已经论述,本文的目的是介绍微型机波形变换编码语音分析和合成开发系统。

## 1 波形变换编码概述

人的听觉对声音的感知主要决定于语音信号的短时功率谱,而相位谱只起次要作用。以此为依据,可以对人们的语音信号进行某种波形变换,找出有利于量化和编码的新时域波形代替原语音波形进行编码以达到数据压缩的目的。图1示出了这种编码方法的数据处理过程。整个过程可分为3个步骤。

### 1.1 基音检测与分析帧划分和抽取

语音信号通过基音检测分出浊音段和清音段并检测出浊音段的基音周期,然后按浊音和清音两种类型划分为不同的分析帧。为保持波形变换后浊音语音的准周期性,浊音的分析帧长度取等于基音周期。清音语音的分析帧长度取128或256个样值。根据语音产生的物理过程是准稳态过程这一事实,对大多数语声,可以合理地认为短时功率谱在10~20ms或更长的时间内是保持不变的,因此邻近几个基音周期的功率谱可以用同一周期的功率谱得到。换句话说,邻近几个基音周期的波形用同一周期波形来代替,对人的听觉不引起多大差别。因此,在波形变换前,可以对分析帧进行抽取。例如每邻近3个或

本文1992年1月20日收到

\* 国家自然科学基金资助项目

4个基音周期选取一个分析帧。如采用四选一,则不但数码率缩减为1/4,而且波形变换编码的计算量也缩减至1/4。清音段也可同样抽取,但为使清音语音不因波形重构时出现附加的周期性,清音帧不能用浊音帧这种重复代替的方法。有关清音帧的重复方法详见文[2]。

1.2 波形变换

波形变换的目的是在保持短时功率谱不变的情况下寻找有利于量化和编码的波形。对以浊音语音,Mozer曾经提出通过相位调整找出能量集中在基音周期中心附近且具有中心对称和大部分样值群聚在16级电平附近的新波形进行量化的方法。因为这种波形具有中心对称,因此只要知道了新波形前半周期部分,就可通过对称性重构后半周期部分,这样需存储的比特率可减少一半,此外因能量集中在基音周期附近,前后1/4周期样值可用零值代替(叫半周期取零),数码率又减少近一半。最后因大部分样值群聚在16级电平附近,因此每个样值可只用4比特进行量化。这样对12比特的采样来说,数码率相当于减少了2/3。这种算法的缺点是运算繁杂费时,要在大型计算机上运行。

为了适应于微型计算机进行分析处理,我们摒弃了寻找大部分样值群聚在16级电平附近这种搜索算法,代之以寻找能量集中在基音周期中心附近,基音周期中心处有正最大样值且中心对称的某一新时域波形算法,这种算法的优点是每帧只需要作一次正、逆傅里叶变换,用近代的单片数字信号处理器容易做到实时处理,缺点是不保证每帧的新波形能适合于半周期取零技术,对新波形必须采用分段ADPCM等方法<sup>[3]</sup>进行4位量化。我们这种算法的相位谱取值是<sup>[2]</sup>

$$S'_{\varphi}(k) = \pi \quad \text{当 } k = 2m - 1, m = 1, 2, \dots, N/2$$

$$S'_{\varphi}(k) = 0 \quad \text{当 } k = 2m, m = 0, 1, \dots, (N/2 - 1)$$

式中 $N$ 为分析帧样点数,即基音周期。在我们的分析中, $N$ 只取偶数。

对于清音语音的波形变换与浊音语音有类似之处,也有一些差别,详见文[2]。

1.3 对新时域波形的量化编码

波形变换后的新时域波形,经对称性删除和半周期取零后(半周期取零决定于前后1/4周期能量是否小于某一设定的门限)再提取公共增益因子 $G$ ,然后对规格化的新

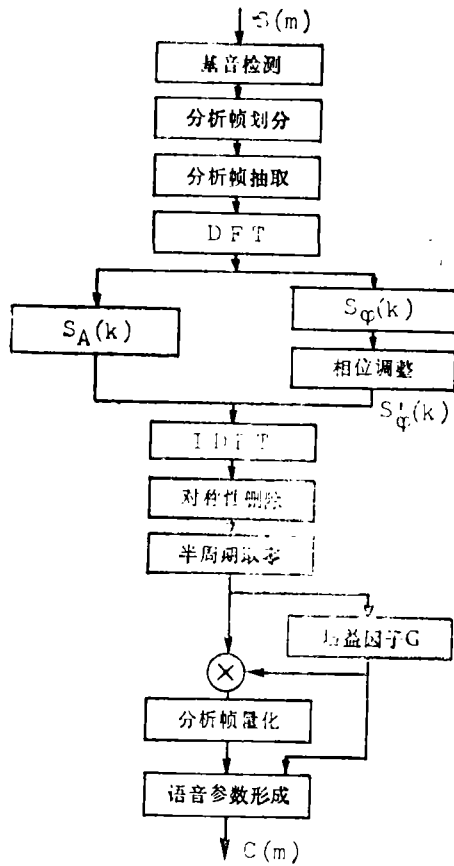


图1 波形变换编码的数据处理过程  
Fig.1 Data processing of the waveform transformation coding

波形进行量化和编码。最后将编码结果按一定的格式组成便于连续存储的数字语音参量<sup>(2)</sup>。曾经用3种方式对规格化波形进行量化和编码。

- A 直接用8比特进行量化和编码;
- B 用4比特分段ADPCM进行量化和编码<sup>(3)</sup>;
- C 经简单处理的4比特量化和编码。

第三种量化编码方式是:首先是因规格化后的最大正值其最高两位必为01,编码时可以删去,只记录其随后4位,译码时再在代码前加上两位01,组成6位有效位。其次最邻近最大正值的样值,必为正数,其符号位必为0,编码时亦可删去,只记录其随后4位,译码时加上符号0,组成5位有效位,最后其余样值乘以3/2后再用4比特量化,以增加其量化信噪比,译码时,再将译码结果乘以2/3,以恢复原值。3种量化编码方法均收到了良好效果,特别是第一种方法能得到数码率8000比特/s左右的高质量的语音。表1和表2分别列出了汉语普通话5个常见单元音用上述3种方法测得的分段信噪比和数码率。实验时采样频率为8kHz,原始语音数据由12位A/D转换器采样,实验材料为彩翎音像制作有限公司制作,广东省音像出版社出版发行的六年制小学语文课本第1册录音带中汉语拼音朗读中的男声,由录音机输入,分段信噪比的测定是以波形变换后的新波形在不进行半周期取零条件下(样值仍用12位表示)作为输入信号,编码后的数字语音参量经解码后的重建语言作为输出信号,分段长为32ms,即256个抽样。数码率的测定是在分析帧以四选一抽样,半周期取零门限为前后1/4周期样值能量小于整个周期内样值总能量的7%条件下测定。

表1 五个单元音的分段信噪比SNR seg (单位、dB)

Tab.1 Short-time quantization SNR (dB) of the five vowels for Mandarin Chinese

方式	1	2	3	4	5	6	7	8	9	10	平均
a	39.9	40.3	43.2	44.0	44.9	44.3	44.1	44.9	43.0	38.0	42.7
o	38.4	40.6	40.8	41.2	41.3	40.2	39.2	38.0	38.2	38.0	39.6
A e	38.4	37.1	39.5	37.8	37.1	35.5	36.5	38.0	40.0	40.8	38.1
i	37.4	40.4	38.1	39.8	40.5	40.5	40.9	40.0	40.3	40.9	39.9
u	39.0	41.4	40.8	42.2	41.2	43.7	44.5	44.1	43.8	44.2	42.5
a	23.0	21.9	25.2	23.0	23.9	23.3	23.9	24.0	24.7	25.9	23.9
o	19.7	17.5	18.8	21.7	21.4	18.7	18.9	23.0	17.7	18.8	19.6
Be	16.6	20.6	25.6	24.4	20.4	16.5	19.2	22.6	18.0	18.5	20.2
i	26.7	22.9	21.5	22.9	23.7	23.6	22.2	20.4	23.7	23.7	23.1
u	25.6	29.9	27.9	30.4	31.2	30.3	29.9	31.2	31.0	30.7	29.8
a	21.7	22.7	24.0	24.8	25.0	24.1	25.0	24.1	22.6	22.6	23.7
o	21.5	22.4	23.4	22.8	22.6	23.8	22.5	21.9	22.5	21.7	22.5
C e	20.3	20.8	21.2	20.8	19.7	19.6	19.1	20.7	21.2	22.1	20.6
i	20.1	21.7	21.8	22.1	22.2	23.2	22.3	22.3	22.8	22.5	22.1
u	22.3	22.6	23.0	22.7	22.6	22.7	24.1	23.3	23.0	22.7	22.9

从表1可见:①直接用8比特量化编码,量化信噪比平均高达40dB,影响语音质量的因素不在量化器而在半周期取零和分析帧抽取时所作的近似处理。试听结果表明,当分析帧抽取为四选一,半周期取零能量门限为7%时,仍能得到数码率为8 kbps的高质量语音(见表2)。②分段ADPCM编码与简单处理后4比特编码的分段信噪比均在22dB上下,其信噪比的总平均值前者稍高于后者,试听表明,两者的合成语音质量亦相当。

表2 五个单元音的数码率(单位: kbps)

Tab.2 Data rate(kbps) of the five vowels for Mandarin Chinese

方式	a	o	e	i	u
A	8.2	4.9	8.4	8.3	8.2
B,C	4.6	3.2	4.8	4.7	4.9

## 2 硬件结构和软件组成

本系统以IBM-PC/XT计算机为主机,配上ATD-20A型TMS32020高速数字信号处理器插板,构成主从式并行处理结构,主机负责数据采集,端点检测,显示修改,结果的打印或存储,放送原始语音和重构语音以及整个程序的管理调度等。TMS32020数字信号处理器负责基音检测,分析帧划分与抽取,波形变换及编码,通过解码重构语音信号等。由于TMS32020数字信号处理器运算速度快,功能强,有丰富的指令系统<sup>[4]</sup>。其整体运算速度比通用微机快100倍以上,这种结合使PC机运算速度提高到小型机水平。

系统采用自制的语音信号采样与语音放音插板。语音的输入可直接通过话筒或录放机作为媒介,经放大及低通抗混滤波后,再用12位A/D转换器得到离散的语音信号,语音的采样数据存储于PC机的内存存储器中,待端点确定后送TMS32020系统的全局存储器里,再由IBM-PC/XT机控制ATD-20A处理。语音放音电路主要包括12位D/A数模转换器,低通滤波器,功率放大器以及外接的小型扬声器,它用于原始语音和合成语音的监听和比较。系统的硬件框图如图2所示。

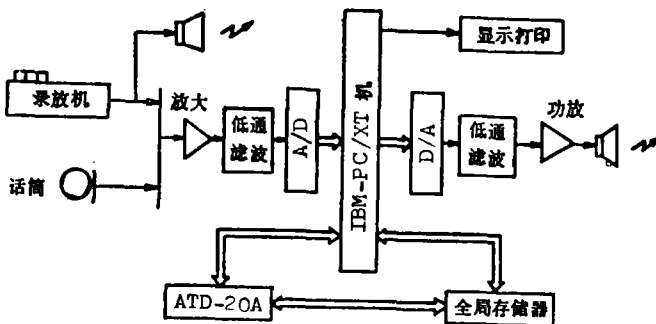


图2 波形变换编码语音分析和合成开发系统硬件框图

Fig.2 Block diagram of WTC Speech analysis-synthesis develop system

软件分别用8088汇编语言和TMS32020汇编语言编写,目标程序占PC机内

存2k字节和占ATD-20A程序存储器2k字。主要功能块有: ①背景噪音检测与语音信号端点门限设定程序; ②语音数据采集, 端点检测, 语音波形显示和端点修改程序; ③基音周期检测, 显示, 修改以及清浊音分类程序; ④分析帧划分与抽取, 波形变换与编码程序; ⑤代码编辑与重构语音(合成语音)信号的解码程序; ⑥原始语音与合成语音放送程序; ⑦数字语音参量存盘和显示打印程序。

系统采用会话方式工作, 系统工作时, 首先检测背景噪声并定出端点检测门限, 使系统处于待输入语音状态, 语音输入以后进行端点检测, 然后根据“菜单”提示或者选择将采集的数据送全局存储器交TMS32020进行基音检测和进一步处理, 或者选择显示语音波形并对端点进行修改后再送到全局存储器供TMS32020处理。每执行一步都有多种选择, 操作者可根据菜单提示决定选择程序的进程。直到得到自己认为满意的语音参数并存入磁盘为止, 然后进行下一步语音输入。由于系统备有完善的修改程序和显示监听程序, 系统除可以一次分析处理不长于3s的语音外还可以对句子中各音节分别进行语音参数提取, 甚至可以对音节中各音素单独进行分析处理。

### 3 系统的使用情况和进一步改进意见

利用本开发系统提供的单词和词组的语音参量构成的词汇表, 设计了两个单片机控制的语音合成系统, 一个系统已在实验室条件下进行了“汽车运行状况检测与语音预警”试验, 另一个系统已构成“WTC-1型键盘语音响应系统”。后一个系统已作为产品出售。此外, 这种合成器在学龄前儿童教育应用方面也作了初步探讨。

波形变换编码语音合成器以其高质量的合成语音和价格的低廉受到用户的欢迎。存在的问题是所需存储器的比特率仍较高, 只能用以有限词汇场合。作者曾经对新时域波形作2比特ADPCM量化和编码试验, 发现语音质量并不理想。作者认为, 如果利用归一化新时域波形相邻帧间的相似性(对应样值间的相关性)将一维语音信号作二维处理, 象图像信号的压缩编码那样采用帧间差值编码。可望在保持一定语音质量前提下大大降低所需存储器的比特率。最近关于对矢量量化后的LPC代码进行帧间内插的报导<sup>[5]</sup>亦从侧面说明我们这种提法是可行的。

### 参 考 文 献

- 1 Bristow G. Electronic Speech Synthesis. Ganada Publishing Ltd, 1984: 135~147
- 2 陈云凤, 汪立全, 伍康文. 数据采集与处理, 1992, 7(2): 86~93
- 3 陈云凤, 伍康文. 数据采集与处理. 1992, 7(3): 169~175
- 4 TMS32020 User's Guide. Texas Instrumente Corp, 1986
- 5 Martins J A, Violaro F. ICASSP, 1991(1), S1: 597~600

## Design and Implementation of Waveform Transformation Coding Speech Analysis-Synthesis System

Chen Yunfeng\* Wang Liquan Chen Zhicheng Shao Yuzhui Wu Kangwen

**Abstract** A Waveform Transformation Coding (WTC) speech analysis-synthesis system adapted to a microcomputer is developed. General introduction of the system as well as hardware and software development are presented. Suggestions on the system improvement are proposed.

**Keywords** speech analysis, speech synthesis, waveform transformation coding

· 简 讯 ·

### 由对碰脉冲锁模激光器获得18fs光脉冲

中山大学物理学系林位株、莫党等最近成功地由对碰脉冲锁模环型染料激光器(简称CPM)直接获得18fs( $18 \times 10^{-15}$ s)激光脉冲。据了解,这是目前世界上由此类激光器直接产生的最短光脉冲。该CPM激光器是一个改进型的六镜四棱镜系统。与典型的CPM系统比较,它采用了新的谐振腔设计和较短焦距的泵浦聚焦镜,使该系统具有较低连续波泵浦阈值和较宽的频带。该激光器输出的飞秒脉冲宽度由计算机控制的精密非共线式自相关器测量。输出脉冲的功率频谱用扫描式光栅单色仪记录。

该激光器输出2束飞秒脉冲列,较短的脉冲宽度一般在20~35fs,最短的达到18fs,激光脉冲的中心波长为630nm,脉冲列中脉冲的重复率为100MHz,平均功率为每束20mW。输出激光脉冲的宽度非常灵敏地依赖于激光器的工作状态,特别是激光腔内群速度色散和自相位调制以及增益和损耗的平衡情况。飞秒脉冲的强度自相关曲线显示出孤子脉冲特性,脉冲的两翼各有1个小脉座。当脉冲被压缩至18fs时,脉座降低,出现负的脉座,自相关曲线轮廓和 $\text{Sech}^2$ 的脉冲光强度形状的自相关函数很好符合。这是首次观察到的新特性。该激光器输出脉冲宽度为21fs时对应的功率频谱半宽度为17.8nm,脉冲的时间-带宽积 $\Delta t \Delta \nu \approx 0.29$ ,比理想的双曲正割平方型脉冲的时间-带宽积 $\Delta t \Delta \nu = 0.315$ 略小。这显示,此脉冲形状可能是非对称的。脉冲功率频谱的非对称性和理论计算亦说明这一点。

(朱文)

\* Department of Radio and Electronics, Zhongshan University