

摘要

本文旨在以国际流行的低码率语音编码算法为背景，概要地说明了低码率语音编码算法的分类和大概的研究情况，并以波形内插语音编码算法为重点，系统而深入地介绍了有关波形内插语音编码算法的基本理论和主要研究方向。在对波形内插语音编码主要算法进行了深入研究的基础上，本文作者依据现在语音编码领域出现的新技术和作者本身对波形内插算法的一些新想法，对波形内插语音编码算法中的一个研究方向上的算法作出了一定程度上的改进，试图使得新算法在计算复杂度和语音合成质量方面都比原来的算法更有优势。通过仿真实验，改进算法的目的已被证实达到。但是改进算法仍是作为一个纯浊音的语音编码算法，与清音分开编码，使得它还保留有原来算法的局限性。

本文的研究工作主要包括以下几个方面：

1. 低码率语音编码的研究概况及其几种主要算法的介绍。首先给出了语音编码算法的分类和对应的国际标准，跟着分别对低码率语音编码各算法的起源，编解码的原理框图以及主要应用方向等进行了分析。

2. 波形内插语音编码算法的研究。详细地介绍了波形内插算法的基本原理和它的两个主流研究算法，并用 C 代码分别实现了这两个算法的浊音编码部分。在分别对两个研究方向的算法进行了功能分析和算法比较的基础上，指出了它们的不足，同时为后面即将提出的改进的波形内插语音编码算法指明了改进方向。

3. 改进的波形内插算法的研究。具体说明了改进算法的基本原理和实施流程，并对改进算法进行了仿真。仿真结果证明了改进算法的有效性，同时也指出了该改进算法由其实现原理的限制而存在的内在缺陷。

论文的最后对整个的研究工作进行了总结，并提出了波形内插语音编码算法将来的研究重点和发展方向。

关键词：波形内插； 原型波形内插； 特征波形内插； 原型周期波形； 特征波形

Abstract

This paper introduces the classification and the research condition of low-bit-rate speech coding algorithms briefly. With the emphasis upon the waveform interpolation speech coding algorithm, it elaborates the basic theory and the main research directions of the waveform interpolation speech coding algorithm. On the basis of the research on the waveform interpolation algorithm, the author proposed a modified algorithm according to some new ideas for the waveform interpolation algorithm and some new techniques in the speech coding. The paper strives to demonstrate the modified algorithm advantageous in the computation complexity and the speech quality. The modified algorithm has been proved effective by the simulation experiments. But the modified algorithm is still a pure-vocal speech coding algorithm, it must work without consonant. That made the modified algorithm retains the limitation of the original algorithm.

The research work of the paper is embodied as following:

1. The introduction about the general research condition and the main algorithms of the low-bit-rate speech coding. Firstly, the paper lists the classifications and the corresponding international standards of the speech coding algorithms. Secondly, it analyses the source, the coding theory and the main application directions of the low-bit-rate speech coding algorithms respectively.

2. Research on the waveform interpolation algorithm. The paper expands on the basic theory and the two main research directions of the waveform interpolation speech coding algorithm, and simulates their voicedness part with C code. Furthermore, it makes algorithm comparison between the two directions and conducts the functional analysis about the two corresponding coders respectively. Hence the paper points out the shortcomings of the two research directions and the improvement of the modified waveform interpolation algorithm.

3. Research on the modified waveform. The paper specifies the basic theory and the execution flow of the modified algorithm, and then carries out emulation on the

modified algorithm. The emulation result has proved the modified algorithm work while it shows the shortcomings of the modified algorithm as well.

Finally the paper summarizes the whole research work, and discusses the research focus and the development future of the waveform interpolation speech coding algorithm.

Keyword: waveform interpolation; PWI; CWI; PCW; CW

华南理工大学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：李巧玲

日期：2004年5月10日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权华南理工大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密口，在____年解密后适用本授权书。

本学位论文属于

不保密口。

(请在以上相应方框内打“√”)

作者签名：李巧玲

日期：2004年5月10日

导师签名：

韦岗

日期：2004年5月10日

第一章 绪论

1.1 语音编码算法的分类及其标准化

语音编码主要分为三类：波形编码、参数编码和混合编码。波形编码力图使重建的语音波形能够保持原语音波形的波形形状，具有适应能力强、语音质量高等优点，但也有致命的缺点就是比特率高，一般在 64~16kb/s 范围内，当码率降低时其性能会迅速下降，编码效率不高。参数编码则主要通过提取语音信号的特征参数，并对特征参数进行编码，目标在于保持原语音的语意，这使得重建语音的波形跟原始语音波形有较大的差别，直接导致合成语音的自然度低、质量差。但是参数编码与波形编码相比，其优势在于编码的速率可以很低，低至在 2.4~1.2kb/s 的范围。而混合编码则克服了波形编码和参数编码的缺点，同时结合了两者的优点，从而能在较低的码率上获得高自然度的合成语音。

由于人们强烈需要对通信方式实行共同化，从而语音编码算法的标准化受到了公众的关注。语音编码算法经过多年的研究与试用，其中很多的编码算法都已被标准化，并进入了实用性阶段。因为参数编码得到的音质难以达到通话要求，所以在当前来说，语音编码算法的标准主要是波形编码和混合编码方面的。已被标准化的语音编码算法按不同的算法类型分，主要有

1. 波形编码：G.711 64kb/s 的 PCM(Pulse-Code Modulation 脉冲编码调制)^[1]；G.721, G.723, G.726 和 G.727 的 ADPCM(Adaptive Differential PCM 自适应差分脉冲编码调制)^{[2][3]}；G.722 7kHz 音频编码^[4]。

2. 参数编码：FS1015 2.4kbit/s LPC(Linear Prediction Coding 线性预测编码)-10(e)^[5]。

3. 混合编码：GSM 13kb/s RPE-LTP(Regular Pulse Excited-long Term Prediction 长时预测的规则脉冲激励)编码器^[6]；FS1016 4.8kb/s CELP(Code-Excited Linear Prediction 码激励线性预测)^[7]；G.728 16kb/s LD-CELP(Low Delay CELP 低延时码激励线性预测)^[8]；IS54 8kb/s VSELP(Vector Sum Excited Linear Prediction 矢量和激励线性预测)^[9]；GSM 半速率语音编码器^[10]；JDC(Japanese Digital Cellular 日本数字蜂窝) 6.7kb/s VSELP^[11]；JDC 半速率 3.6kb/s PSI-CELP(Pitch Synchronous Innovation-Code Excited Linear Prediction 基音同步更新码激励线性预测)^[11]；G.729 8kb/s CS-ACELP(Conjugate Structure-Algebraic Code Excited Linear Prediction 共

结构一代数码本激励线性预测编码)^[12]；G.723.1 5.3/6.3kb/s ACELP^[13]；IS96 8.5kb/s QCELP(Qualcomm CELP-Qualcomm 通信公司的专利 CELP 语音编码标准)^[14]；3GPP 的第三代移动通信自适应多码率语音编码标准 AMR-WB (Adaptive Multi-Rate WideBand, 也即 G.722.2)^{[15][16]}。

1.2 低码率语音编码算法的主要研究方向

随着社会信息化的发展，有限的频率资源要用于无限的业务需求，使得频率资源愈加紧张，人们急切地盼望着更低码率的语音编码算法的出现，因此，低码率/较低码率的语音压缩算法正是当前语音编码算法的研究焦点。低码率语音编码算法主要有四个研究方向：

- 多带激励 MBE (Multiband Excitation) 编码
- 正弦变换编码 STC (Sinusoidal Transform Coding)
- 混合激励线性预测 MELP (Mixed-Excitation Linear Prediction) 编码
- 波形内插 WI (Waveform Interpolation) 编码

MBE 和 STC 均是基于正弦合成分析模型 SASM(Sinusoidal Analysis/Synthesis Models) 的语音编码算法，MELP 则是基于二元激励模型的线性预测编码算法，WI 为基于 SASM 和线性预测模型的混合型编码算法。

1.3 低码率语音编码各个主要算法的基本原理

1.3.1 多带激励编码

MBE 算法是在 80 年代由 Griffin D.W. 博士提出。MBE 编码器是一个不用预测残差的完全的参数语音编码器，工作在频域里。与 LPC 算法所采用的二元激励模型不同，它并不是简单地将一帧语音判定为浊音或是清音，而是按基音各谐波频率，将一帧语音的频谱分成若干个谐波带，再以若干个谐波带为一组进行分带，分别对各带进行 V/U 判决，总的激励信号由各带激励信号相加构成。对于浊音带，用以基音周期为周期的脉冲序列谱作为激励信号频谱；对于清音带，则使用白噪声谱作为激励信号频谱。它考虑到了某段语音的激励信号是周期（浊音）和类噪声（清音）能量的混合情况^[17]。正是这种在激励模型中附加自由度的做法使得 MBE 语音模型的重建语音质量高于传统的语音模型，另外，还使得 MBE 语音模型在背景噪声面前有坚韧性。

在一个短时间隔内，一个加窗语音段 $s_w(n)$ 的傅立叶变换 $S_w(\omega)$ 被模型化为谱

包络 $H_w(\omega)$ 和激励谱 $E_w(\omega)$ 的乘积。谱包络是原语音谱的平滑部分，对于每组基频谐波，激励谱由基频和清/浊判决信息来确定，即把清音频率区域的随机噪声功率谱段与浊音频率区域的周期功率谱段相组合来获得激励谱。周期谱完全由基频 $P_w(\omega)$ 所决定。清浊音信息通过频率相关法使得具有随机功率谱的谐波谱最大来获得。假定人耳能够区分基频谐波支配的频率区域和其他噪声能量支配的频率区域，则可以从随机噪声中分离出浊音语音。另外，编码参数的确定使用了 ABS(Analysis-by-Synthesis 合成分析)法，即使原语音谱与合成语音谱的均方误差最小。数学处理则使用了 FFT(Fast Fourier Transform 傅立叶变换)或 DCT(Discrete Cosine Transform 离散余弦变换) 技术。

MBE 的编译原理框图如下所示：

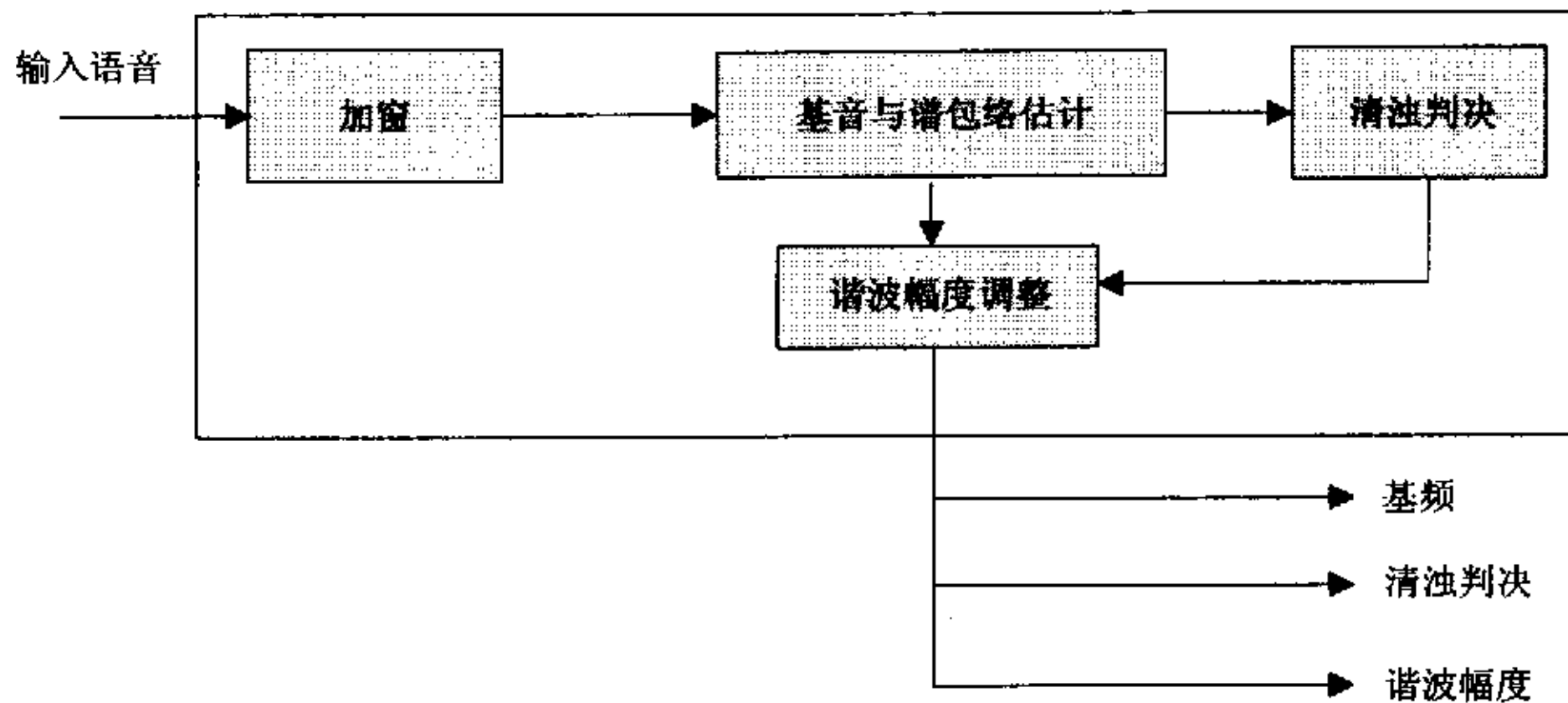


图 1-1 MBE 编码器
Fig.1-1 MBE Encoder

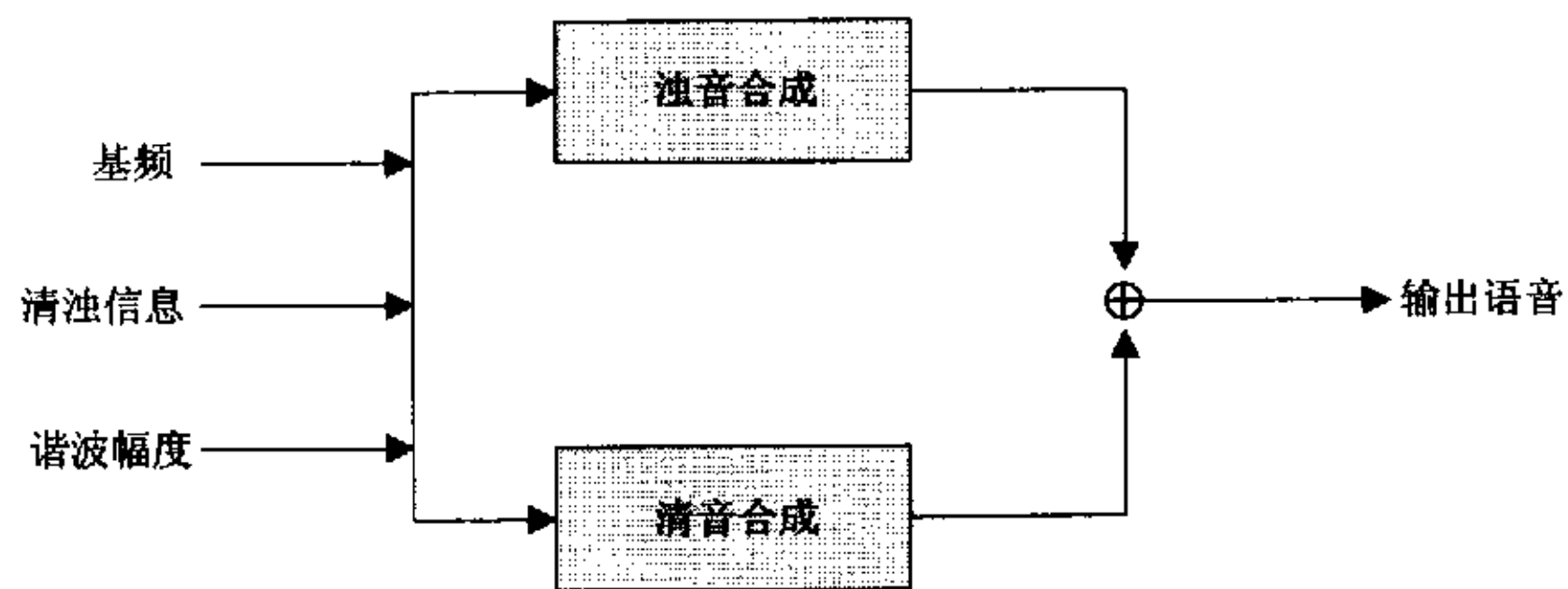


图 1-2 MBE 解码器
Fig.1-2 MBE Decoder

由于 MBE 编码器不需要残差信号，而且它也不需要码书，所以有低的计算复杂度。它在 2.4kb/s 的数码率时仍能保持语音的易懂度和自然度。由 MBE 编码器演变出的 IMBE(Improved MBE 改进的 MBE)和 AMBE(Advanced MBE 先进的 MBE)编码器已经被商业化，这些 AMBE 和 IMBE 硬件产品具有纠错、有声/无声检测、DTMF(Double Tone MultiFrequency 双音多频)检测/合成、回波抵消和软判决译码等功能，可用于蜂窝电话、PCS(Personal Communications Service 个人通信业务)、卫星通信、数字移动无线电、保密通信、话音多路复用、话音邮寄、多媒体应用和会议电视等。其中，IMBE 技术被选为几种应用的语音编码标准之一。

1.3.2 正弦变换编码

最早的 STC 编码器由美国的林肯实验室发明，码率范围为 9.6kb/s~1.2kb/s。STC 编码器不是波形编码器，但是采用了使用由正弦波的幅度、频率和相位表征的分析/合成技术的正弦模型。STC 编码器的合成语音是由简谐相关的正弦波组成，对于频率使用谐波模型，对于正弦波幅度使用基音自适应幅度包络，对于正弦波相位使用浊音度相关模型^[18]。下图给出了 STC 的编译原理图。

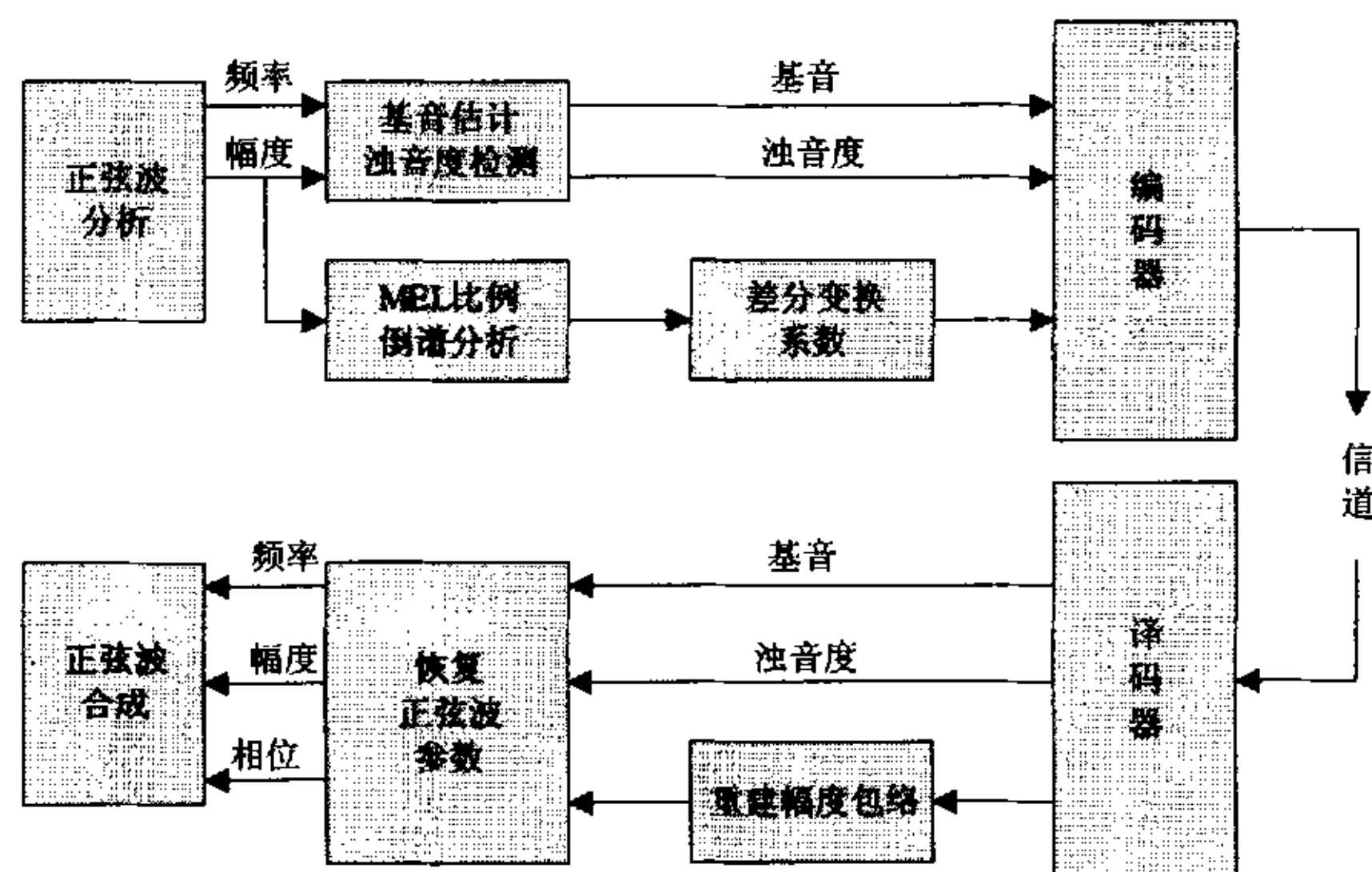


图 1-3 STC 编译原理图

Fig.1-3 STC Compilation Theory

由于基本的正弦波分析/合成系统能够产生高质量的语音信号，当给定足够高的数据速率时，可使编码器的性能任意接近于基本系统的要求。当使用基于正弦波的基音估计器时，用正弦波的谐波组产生非常高质量的合成语音是可能的，其编码语音的质量取决于对正弦波幅度和相位高保真编码的能力。正弦波参数提供

了语音信号的频域分解，更容易利用听觉机理的感性特征达到编码效率。另外，频域表达也考虑了后滤波器设计的可选择性方案。正如其他低码率编码器一样，后滤波器在 STC 编码器得到无抑制的合成正弦波语音是非常重要的。最后，频域表达还提供了将激励谱分带的方便基础，以至于可作多带浊音度判决，考虑了混合浊音度激励。这将改善合成语音的自然度，增加了语音信号在附加噪声中的坚韧性。

正弦模型除了提供开发参数声码器的基础外，也可以与波形编码方法相结合，得到一类波形内插声码器的产生。在满足低码率高质量合成语音要求的方面来说，正弦模型方法对于低速率语音编码是一个主要的趋势。其原理图如下：

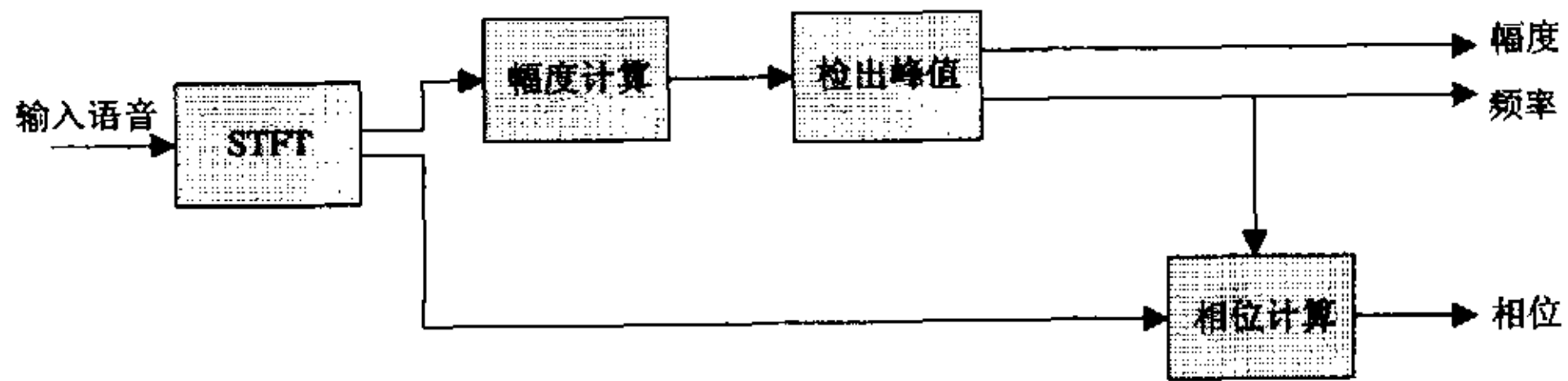


图 1-4 正弦模型分析端

Fig.1-4 Sinusoidal Model Analyzer

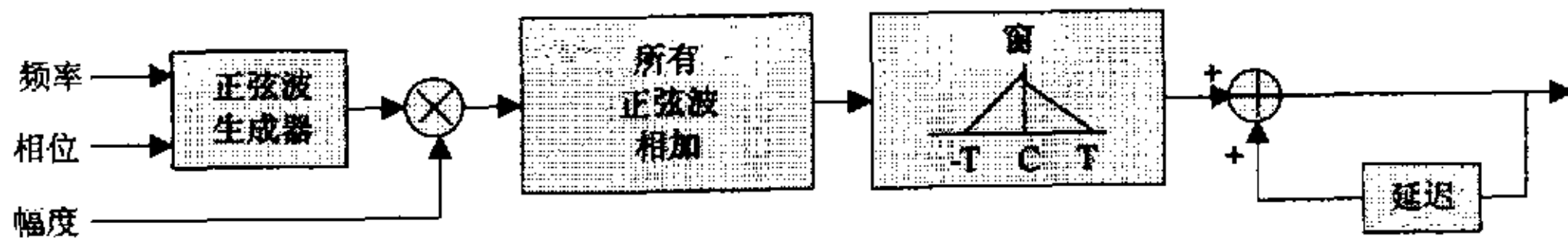


图 1-5 正弦模型合成端

Fig.1-5 Sinusoidal Model Synthesizer

1.3.3 混合激励线性预测编码

MELP 编码器是由美国的 McCree A.V.博士提出的，目的是针对传统的基音激励 LPC 声码器中，简单的脉冲串没有能力产生各种浊音语音，导致合成语音带有机械性或蜂鸣性的现象。

MELP 编码器使用了混合激励模型，能模仿较丰富的可能语音特征集合，从而产生了更自然的语音，即使在是严重的背景噪声环境里也是坚韧的。MELP 编码器采用传统的 LPC 参数模型为基础，但也引入了四个附加特征：混合激励、非

周期脉冲、脉冲散布、自适应谱滤波^[19]。混合激励用多带混合模型实现，主要作用是减少与 LPC 声码器相联系的嗡嗡声，尤其是在宽频声学噪声里。非周期脉冲常用于语音信号的浊音和清音间的过渡区域，主要作用是排除浊音度过强而引起的偶然的音调噪声。脉冲散布通过基于谱平整的三角脉冲的固定脉冲散布滤波器来实现的，从而在一个周期内起到了展开激励能量的作用，减少了合成语音的粗糙质量。自适应谱滤波器以 LPC 声道滤波器的极点为基础，用于增强合成语音的共振峰结构，改善了合成波形和自然带通波形间的匹配，产生更自然的输出语音。其合成语音模型的原理图如下：

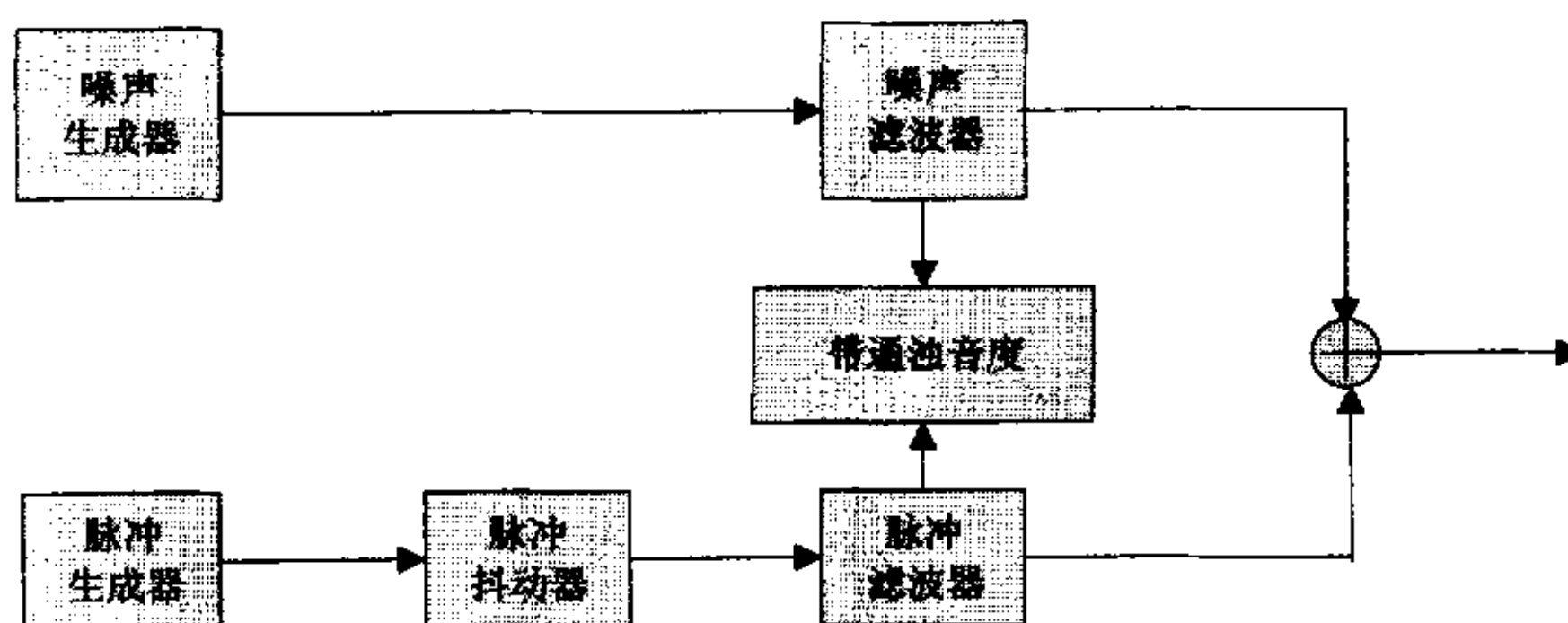


图 1-6 MELP 分析端

Fig.1-6 MELP Analyzer

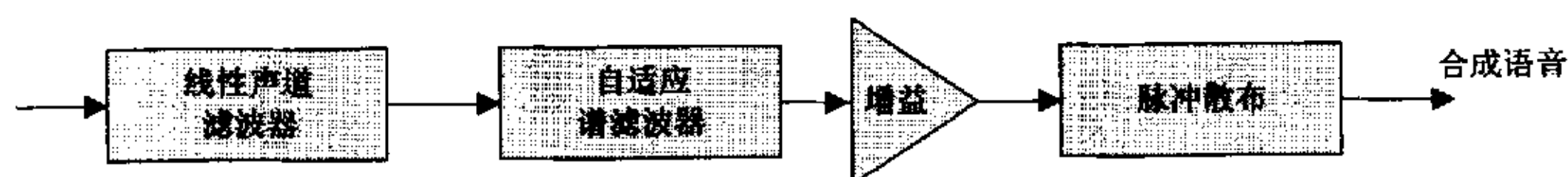


图 1-7 MELP 合成端

Fig.1-7 MELP Synthesizer

MELP 声码器已被美国国防部数字声音处理协会选择为 2.4kb/s 语音编码的联邦标准。MELP 声码器还可用于民用，如在无线通信、Internet 电话，以及语音邮寄系统等，但其语音质量有待进一步的改善。

1.3.4 波形内插编码

1991 年美国的 AT&T 贝尔实验室的 Kleijn W.B. 博士提出了原型波形内插 PWI (Prototype Waveform Interpolation) 语音编码算法。该算法原理是在一定的时间

范围之内（如 20~30ms）以慢渐变的基音周期波形为基础，在浊音帧里提取和传送一段原型波形，用线性插值恢复未传送的信号^[20]。由于 PWI 有着明显的局限性，那就是需要和 CELP 编码器组合才能实现一个完整的编码系统，导致清浊音不同编码模式转换处附近产生较大的听觉失真；并且 PWI 算法复杂，在硬件上难以实时实现。所以 Kleijn W.B. 博士又提出了另一种波形内插算法—特征波形内插 CWI（Characteristic Waveform Interpolation）语音编码算法。该方法将语音信号视为渐变的特征波形，当语音信号为明显的浊音时波形慢渐变，当语音信号为明显的清音时波形快渐变，借助简单的非自适应滤波器将特征波形分解为慢渐变和快渐变波形两个部分，分别对它们进行编码，既有效地提高了编码效率，又避免了清浊判决带来的影响^[21]。PWI 和 CWI 算法都是目前很有潜力的低速率语音编码算法，正为国际上的许多研究机构所集中研究开发着，期望能在低速率上产生通信质量的重建语音。

1.4 波形内插语音编码算法

波形内插 WI（Waveform Interpolation）语音编码是由 Kleijn W.B. 博士提出，经近十年研究发展起来的一种低速率语音编码算法，具有在低码率却具备高自然度质量的语音的特点。它利用了语音特征的感觉重要性，允许信号的有效压缩，促进了低速率语音编码算法的发展，已成为语音编码领域研究的热点。

传统的合成分析语音编码算法在 4.8kb/s 以上已获得好的语音质量，如 CELP 算法。这些方法由波形的匹配过程重建语音信号，即在逐帧分析的基础上变得谱加权 SNR（Ratio of Signal and Noise）最大。当比特率下降时，由于没有足够多的比特数来精确地描述一个波形，波形匹配的准确性下降，CELP 算法会带来很大的量化噪声，致使重建信号有噪声特征，重建语音的质量将严重下降。

要在低码率提供高质量的语音信号压缩，必须利用语音信号的内在周期性和信号的展开本质，而不是依靠 CELP 的互相关技术来产生周期。WI 算法提供了这两种属性，通过表示信号或更经常表示的是 LP 残差信号作为一个基音周期的展开式（众所周知的是原型波形或是特征波形）。这个描述语音信号基音周期的明确方法使人联想到第一代的话音编码算法。但是，尽管 WI 利用了许多熟悉的概念，例如 LP 编码和序列 LSF 量化，大部分概念对语音编码来说都是新的。原信号或残差到一个相位对准的特征波形的展开表面的转换，以及为了量化，该表面到近乎独立的缓变和快变表面的序列分解也许是 WI 编码的最大不同特征。另外，这种技术在低码率通过平滑内插几乎它的所有参数获得了高的质量。同样地这种技术需要仔细地考虑各种事件例如基音加倍。这个技术是一个真正的混合语音编码

算法，在时域和离散频域都进行分析。尤其是原型波形傅立叶分解的使用允许了有效的相位对准，和不同基音周期的特征波形之间的内插^[22]。正是波形内插语音编码算法的这些新颖特点使它能在众多的语音压缩编码算法中脱颖而出，成为国际上语音压缩编码的研究热点。

1.5 本文的主要研究内容和章节安排

本文对两种主要的波形内插语音编码算法和它们的实施流程分别进行了详细的说明，重点研究了波形内插编码算法中的原型波形内插算法，并就其不足之处提出了改进方案。全文共分四章，主要结构如下：

第一章为绪论，对低速率语音编码算法的研究概况、国际标准以及主流算法作了简要的介绍，并独立一节特别介绍波形内插语音编码算法。

第二章详细分析了波形内插语音编码算法之一——原型波形内插算法的工作原理和实现过程，对原型波形内插算法所得到的编码器做了功能分析，并在此基础上指出了该算法的不足之处。

第三章具体说明了波形内插语音编码算法之一——特征波形内插算法的基本思想和实现过程，同样地对由特征波形内插算法所实现的编码器做了功能分析，并在前一章所做的原型波形内插编码器的性能分析的基础上，对两个不同的算法进行了比较。

第四章重点研究了基于原型波形内插算法的改进算法，依次给出了它的原理框图、实现流程、以及仿真结果，并对改进算法的仿真结果作出了分析。仿真结果证明了改进算法在保持计算复杂度基本不变的情况下，所合成的浊音语音的波形形状准确度和周期性准确度均明显优于原来的 PWI 算法。

最后是全文的结论，主要总结了作者的研究工作，并提出了今后进一步在本研究方向进行研究工作的展望。

1.6 本章小结

本章主要概述了语音编码算法，尤其是低速率语音编码算法当前在国际上的研究概况，以及它们主流的研究方向和算法原理，并重点介绍了波形内插语音编码算法的算法特点。

第二章 原型波形内插语音编码算法

2.1 基本思想

浊音语音具有高水平的周期性特征,可以看作是慢变化的基音周期波形 PCW (Pitch Cycle Waveform)的连接。原型波形内插方法就利用了浊音语音的这种周期性,试图在低比特速率重现浊音语音信号的周期性水平。它每隔 20~30ms 提取一单个有代表性的基音周期波形,并称这个基音周期波形为原型波形 PW (Prototype Waveform), 传送描述这个 PW 的信息,然后在更新点处对 PW 的长度和形状进行线性内插得到重建的语音信号^[23]。所以,语音信号是由减样的 PCW 序列重建的,内插过程得到了平滑渐变的 PCW,重建语音保持了浊音语音的周期性水平。对提取的原型波形进行编码,再利用原型波形内插方法重建未经编码的信号,将大大降低编码率。PWI 方法能在 4kb/s 以下的编码速率上获得优良的浊音语音质量,并能与 CELP 算法相组合对清音语音编码。

在低速率语音编码中,对语音信号周期性特征的不适当修正会导致合成语音信号中常见的三类失真:噪声、回声、音调的人工产物。音调的人工产物指的是重建语音信号的周期性水平高于原始语音信号的周期性水平。回声的引起是由于重建语音中相继的 PCW 缺少相位一致性。噪声则大部分是由于量化失真而引起。这些失真强调了 PCW 动态性的重要性^[24]。波形动态意味着寻找一种 PCW 随时间渐变的规则或模式。在波形编码器中,比特率的减少将导致波形的不准确描述,重建语音中的波形动态变成了依赖于声码器中所使用的语音模型。基于这个理由,PWI 方法则是提供平滑的 PCW,因为平滑的 PCW 表达了浊音语音中最常见的波形动态。PWI 的附加特征使波形动态自适应于原始语音信号,在目前的方法中,这些动态由两个频率相关参数表达,一个是介于相邻 PCW 间的相关,另一个是间隔为 20~30ms 的 PCW 间的相关。通过强制重建信号,使重建信号中的这两个相关值类似于原信号中获得的相关值,可以保持波形的动态。

2.2 基本原理

2.2.1 瞬时激励波形及其内插

PW 的表达及内插必须以浊音语音信号的基本性质为基础。浊音语音的短时谱可以粗略地分为谱包络和谱的精细结构两部分,谱包络由声道形状确定,谱的精细结构由来自声带的周期激励信号确定。PWI 方法假定声道的形状相对于声带

产生的激励信号独立变化，所以在进行 PW 内插前，要分离共振峰结构和激励信号，其中，激励波形和它的基音周期为一方面，共振峰结构为另一个方面，这两个方面独立插值，然后利用共振峰结构和内插的激励信号产生语音信号。

为了获得高质量的浊音语音，内插方法应使激励信号保持连续性，并有一个平滑的基音轮廓。因为浊音语音的激励信号近似是周期的，所以数学上它能近似为一个具有时变参数（傅立叶级数系数和基音周期）的周期函数（傅立叶级数）。对于一个长度为 P 的单个一维原型波形，其离散时间傅立叶级数表示为

$$s(m) = \sum_{k=0}^{\lfloor P/2 \rfloor} [A_k \cos(\frac{2\pi km}{P}) + B_k \sin(\frac{2\pi km}{P})] \quad 0 \leq m \leq P \quad (2-1)$$

式中， P 为基音周期， $\{A_k\}$ 和 $\{B_k\}$ 为傅立叶级数系数。

有了上面单个原型波形表达式后，只要对公式 (2-1) 中的参数附加一个时间标号 t ，就可以得到原型波形序列的二维表达。这时公式 (2-1) 可重写为

$$s(m, t) = \sum_{k=0}^{\lfloor p(t)/2 \rfloor} [A_k(t) \cos(\frac{2\pi km}{p(t)}) + B_k(t) \sin(\frac{2\pi km}{p(t)})] \quad 0 \leq m \leq P \quad (2-2)$$

式中，系数 $\{A_k(t)\}$ $\{B_k(t)\}$ 和基音周期 $p(t)$ 是时变的。

现在式 (2-2) 是一个二维信号表示，其中 t 和 m 是运动的变量，沿 m 轴显示的是单个的原型波形，而原型波形的形状沿时间轴 t 渐变。然而式 (2-2) 的长度取决于 $p(t)$ ，不同时刻的原型波形可能有不同的长度。为了方便，所有的原型波形要归一化为同样的长度。

$$\varphi = \varphi(m) = \frac{2\pi m}{p(t)} \quad (2-3)$$

此时式 (2-2) 变为

$$s(t, \varphi) = \sum_{k=0}^{\lfloor p(t)/2 \rfloor} [A_k(t) \cos(k\varphi) + B_k(t) \sin(k\varphi)] \quad 0 \leq \varphi(\bullet) \leq 2\pi \quad (2-4)$$

以这种方法，所有的原型波形有同样的长度 2π 。基音周期的归一化使得 φ 与周期函数 $s(t, \varphi)$ 的最低谐波相位一致，所以称 φ 为瞬时激励波形的相位。

对于浊音语音，瞬时激励波形 $s(t, \varphi)$ 和基音周期 $p(t)$ 在时间轴上慢渐变，所以它们能在间隔 20~30ms 的范围内插值。PWI 编码器利用了这个事实，在每一更新区间仅仅传送一个瞬时激励波形和它的基音周期，所传送的瞬时激励波形构成了重建语音的原型激励波形 PEW (Prototype Excitation Waveform)。设当前插值区间的分界修正时刻为 t_m 和 t_{m+1} 。为了使内插简便，2 个原型波形应该具有相近的相位，则原型波形 $s(t_m, \varphi)$ 和 $s(t_{m+1}, \varphi)$ 的线性内插得到重建的瞬时波形为

$$s(t, \varphi) = [1 - \alpha(t)]s(t_m, \varphi) + \alpha(t)s(t_{m+1}, \varphi) \quad t_m \leq t \leq t_{m+1} \quad (2-5)$$

式中， $\alpha(t)$ 为从 $\alpha(t_m) = 0$ 到 $\alpha(t_{m+1}) = 1$ 的单调递增内插函数。值得注意的是，2 个原型波形之所以能以这种简单方式内插，是因为它们有同样的归一化基音周期。

基音周期 $p(t)$ 的内插方法和瞬时波形的内插方法相同：

$$p(t) = [1 - \beta(t)]p(t_m) + \beta(t)p(t_{m+1}) \quad t_m \leq t \leq t_{m+1} \quad (2-6)$$

此处的 $\beta(t)$ 为从 $\beta(t_m) = 0$ 到 $\beta(t_{m+1}) = 1$ 的单调递增内插函数。为了方便起见，对 $p(t)$ 和瞬时波形 $s(t, \varphi)$ 内插时可以使用相同的插值函数 $\alpha(t)$ 。

使每一时刻的瞬时波形有正确的基音周期，然后连接这些无穷小的瞬时波形，可以重建信号 $s(t)$ ，即

$$s(t) = s[t, \varphi(t)] = s[t, \varphi(t_m) + \int_m^t \frac{2\pi}{p(t')} dt'] \quad (2-7)$$

公式 (2-7) 描述了连续插值方法的基本原理。

2.2.2 PCW 波形动态的自适应

按式 (2-7) 进行的 PW 内插产生了平滑渐变的 PCW，它表现了浊音语音信号中最普通形式的 PCW 波形动态。然而，如果缺少波形动态自适应，则将产生两类失真：第一类失真是由低比特率对 PW 的不准确量化而引起的，这时所传送的原型激励波形的形状将产生大的波动，重建信号听起来有回声；当原语音信号中包含大量的送气噪声时，会有第二类失真出现，即用太多的周期性恢复这种语音信号时，在重建语音信号中会出现音调的人工产物。

在 PWI 算法中，为了说明信号的改变是由瞬时波形的渐变引起，而不是由噪声引起，采用信号改变比 SCR (Signal-to-Change Ratio) 来作为不同时刻激励波

形的相似性测度。为了保证波形动态类似于原语音信号，PWI 方法要求在不同的时间分隔处，要保持瞬时波形间的 SCR 不变。为此，PWI 算法引入了长时 SCR 和短时 SCR。长时 SCR 对应于 PW 间的信号改变比，而短时 SCR 对应于相邻 PCW 间的信号改变比。

定义 SCR：两个瞬时波形 $v(t_l, \varphi)$ 和 $w(t_m, \varphi)$ ，设定前者为原信号，两个瞬时波形 $v(t_l, \varphi)$ 和 $w(t_m, \varphi)$ 之间的内积为

$$\langle v(t_l, \varphi) | w(t_m, \varphi) \rangle = \int_0^{2\pi} v(t_l, \varphi) w(t_m, \varphi) d\varphi \quad (2-8)$$

首先定义函数 $\Xi[\bullet, \bullet]$ 如下：

$$\Xi[v(t_l, \varphi), w(t_m, \varphi)] = \left[1 - \frac{\langle v(t_l, \varphi) | w(t_m, \varphi) \rangle^2}{\langle v(t_l, \varphi) | v(t_l, \varphi) \rangle \langle w(t_m, \varphi) | w(t_m, \varphi) \rangle} \right]^{-1} \quad (2-9)$$

$\Xi[\bullet, \bullet]$ 关于两个瞬时波形对称，可以将它看成是相关系数的单调函数。

则可得到 SCR 的定义：两个瞬时波形中的一个进行相位移位而得到的 $\Xi[\bullet, \bullet]$ 的最大值，即

$$SCR[v(t_l, \varphi), v(t_m, \varphi)] = \arg_{\varphi'} \max \Xi[v(t_l, \varphi), v(t_m, \varphi + \varphi')] \quad 0 \leq \varphi' \leq 2\pi \quad (2-10)$$

长时 SCR 定义为间隔 20~30ms 的 PCW 间的 SCR，这个间隔与编码系统的修正速率相同。经实验证明，制约长时 SCR 虽然使量化的原型波形的 SNR 降低，但抑制了回声，增强了重建信号的感性质量。

长时 SCR 的约束过程如下：首先，测量相继的、未量化的原型波形的 SCR；其次，强制量化的原型波形的 SCR 与这个值接近。为了减少可察觉的回声，在测量长时 SCR 之前，原型激励中必须加入原始语音信号的平坦共振峰结构（谱包络）。

令 $h(\tau)$ 为谱加权滤波器的脉冲响应， $H[\bullet]$ 为与 $h(\tau)$ 相联系的滤波运算，则对于瞬时激励波形 $s(t_l, \varphi)$ ，其谱加权的瞬时激励波形为

$$v(t_l, \varphi) = H[s(t_l, \varphi)] = \int h(\tau) s[t_l, \varphi - \frac{2\pi}{p} \tau] d\tau \quad (2-11)$$

需要注意的是当估计 SCR 时,两个瞬时波形所用的谱加权和基音周期 p 必须一致。

令 $v(t_{m+1}, \varphi)$ 为当前已经量化和谱加权的原型激励波形。一般地,这个波形能被分解成两种成分,一种成分为与原来量化和谱加权的原型激励波形 $v(t_m, \varphi)$ 平行,另一种成分与 $v(t_m, \varphi)$ 正交,即

$$v(t_{m+1}, \varphi) = y_p + y_o, \quad \langle y_o | v(t_m, \varphi) \rangle = 0 \quad \langle y_p | y_o \rangle = 0 \quad (2-12)$$

则根据 y_o 和 y_p 的相对比例调整长时 SCR, 这可以借助一个比例因数 μ 来完成, 即

$$v'(t_{m+1}, \varphi) = \mu y_p + [(1 - \mu^2) \frac{\langle y_p | y_p \rangle}{\langle y_o | y_o \rangle} + 1]^{\frac{1}{2}} y_o \quad (2-13)$$

当 $\mu = 1$ 时表示没有调整长时 SCR。如果原信号的长时 SCR 值为 Λ , 则当 $v(t_{m+1}, \varphi)$ 使用如下 μ 值时, 重建信号的长时 SCR 等于 Λ :

$$\mu = [1 - \Lambda^{-1}]^{\frac{1}{2}} [1 + \frac{\langle y_o | y_o \rangle}{\langle y_p | y_p \rangle}] \quad (2-14)$$

上式的长时 SCR 是针对整个信号频带的, 而公式 (2-11) 能被用于选择特定频带, 在每个所要求的频带内, 增加比特率可保持重建信号的长时 SCR 与原信号的长时 SCR 一致。

对于基频非常低的讲话者, 若在 PW 的描述中加入长时 SCR, 可以得到非常好质量的合成语音。然而, 对短基音周期的讲话者, 由于插值方法在某些语音段加入了太多的短时周期性, 在描述波形动态的时候, 必须考虑瞬时波形间的短时相关性。这种相关可以用短时 SCR 确定。由于送气音噪声的数量是频率的函数, 有必要对频率划分子频带后再计算各频带自身的短时 SCR。而短时 SCR 测量的是各频带自身的噪声, 把频带数限制为最小, 当测量短时 SCR 时, 在内插区间进行时间平均是有帮助的。测量短时 SCR, 可由式 (2-11) 选择一适当的频带, 然后用式 (2-9) 获得实际的短时 SCR, 在插值区间重复这个过程, 并求平均可获得一个更具坚韧性的短时 SCR 估计。

2.2.3 算法框图

图 2-1 为 PWI 浊音编解码的流程图。

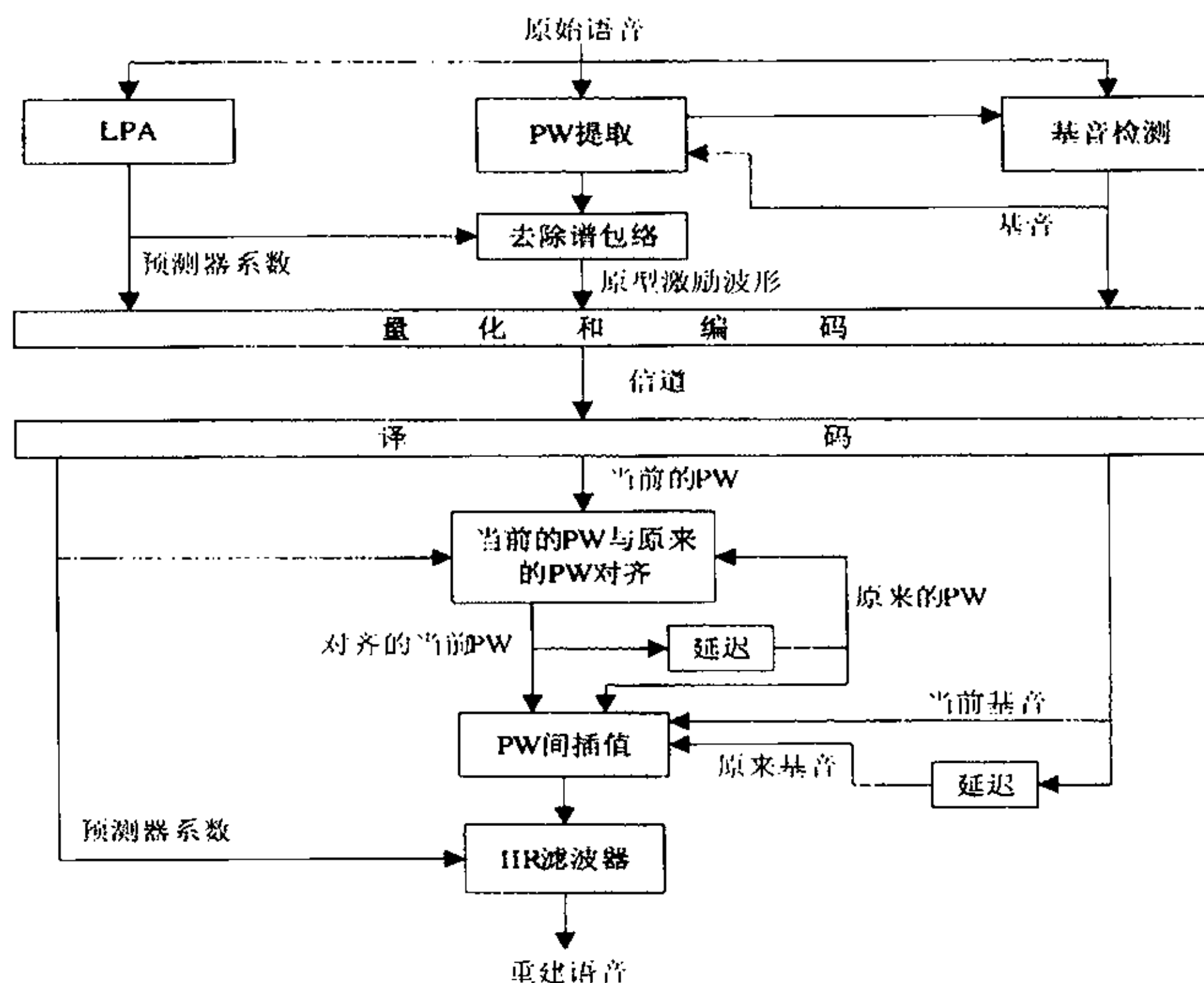


图 2-1 PWI 算法流程

Fig.2-1 PWI Algorithm Flowchart

在编码器里，首先对输入信号进行基音检测提取基音周期，以获得的基音周期为基础提取原型波形 PW，同时对原信号进行 LP 分析。跟着，利用获得的 LP 系数实现 PW 的谱包络修正。在量化前，还要先将由谱包络修正而得到的原型激励波形 PEW 对齐。对得到的参数：LP 系数、基音周期、PEW 的傅立叶级数系数分别量化编码。

解码是编码的逆过程。由译码得到的 PEW 的傅立叶级数系数重建单个的 PEW，再由单个 PEW 通过线性内插和 IDTFS 重建被谱包络修正的原型波形。将重建的原型波形通过由 LP 系数组成的 IIR(Infinite Impulse Response 无限冲激响应) 滤波器即可得到最终的重建浊音信号。

2.3 实现过程

2.3.1 PW 的提取

PW 的提取可有两种方法，一种是在语音域中进行，另一种是在 LPA 残差域中进行。这两个情况下，都是要先进行基音周期的估计。

2.3.1.1 在语音域中提取 PW

在语音信号中提取原型语音波形利用的是预测增益最大准则。首先，对 PCW 定义一个任意开始时刻，一般将起点定在过零处或其附近，可以使信号能量的变化对准则的影响最小。然后，搜索长度近似等于基音周期的那段区间，对于重复这段区间而得到的周期信号，可得到最佳线性预测器的最大短时增益，最后的区间长度对应于局部基音周期。如果原语音信号是周期的，当区间等于准确的基音周期长度时，预测增益最大；若语音信号不是准确的周期信号时，这个方法对于求得一个合理的 PW 表达式也是有效的。但是需要注意的是，最大预测增益准则仅仅使用了一个 PCW 的数据，这样产生的原型语音波形受语音信号的非稳态性影响最小。

通过计算每个候选区间的带限傅立叶级数来计算最大预测增益，可以避免增加采样率。令 M 为以采样周期为单位的基音周期（一般 M 不为整数），对于一个带限为一半采样率的系统，有 j 个谐波，这里 $j \leq M/2 \leq j+1$ 。用 $2j+1$ 个参数表示瞬时语音波形的傅立叶级数：

$$u(t, \varphi) = \sum_{k=0}^j [C_k(t) \cos(k\varphi) + D_k(t) \sin(k\varphi)] \quad (2-15)$$

如果长度为 M 的试探区间中心位于已有的样点处，那么傅立叶级数能够准确地拟合语音样点。求得 j 个谐波的傅立叶级数后，对所要求的延迟，计算自相关函数：

$$r(t) = \frac{1}{2} \sum_{k=0}^j [C_k^2 + D_k^2] \cos\left(\frac{2\pi kt}{M}\right) \quad (2-16)$$

从上式得到一系列的自相关值后，得到长度为 M 的区间上产生的周期信号的线性预测增益 $G(M)$ 为：

$$G(M) = \frac{r(0)}{r(0) - \mathbf{a}^T \mathbf{R} \mathbf{a}} \quad (2-17)$$

这里的 $\mathbf{a} = [\alpha_1, \alpha_2, \dots, \alpha_p]^T$ 为短时预测器系数，可用自相关法对以原采样率的语音信号计算得到， p 为预测器阶数； \mathbf{R} 为自相关法矩阵，是对称的 Toeplitz 矩阵，它的第一行为 $r(0), r(T), r(2T), \dots, r[(p-1)T]$ ， T 为采样周期。

在 M 值的适当范围内重复式 (2-17)，具有最大预测增益的区间选择为 PW。

然而，如果用的是为了获得语音合成所用的 LP (Linear Prediction 线性预测) 系数来提取原型波形，式 (2-17) 不再成立。因为此时的 LP 系数需在较长的区间上进行 LP 分析，导致预测器不是最优的。预测增益改用下式计算：

$$G(M) = \frac{r(0)}{r(0) + \mathbf{a}^T \mathbf{R} \mathbf{a} - 2 \sum_{n=1}^p \alpha_n r(nT)} \quad (2-18)$$

2.3.1.2 在 LPA 残差域中提取 PW

由于 LP 残差缺少短时相关，所以最大预测增益准则不适用于残差域提取 PW。在残差域中提取 PW，利用的是基音脉冲标志，一般来说，从一个增样的残差信号里提取基音脉冲标志相对于直接在原残差信号里要容易。在一个增样 10 倍的信号里，在某一帧里简单搜索绝对值最大点作为第一个基音脉冲标志点，跟着在距已确定的标志点大约一个基音周期处搜索其他绝对值最大点，以期求得其他基音标志点。

在残差域中，仅仅在基音脉冲附近（大部分能量集中处）的 PCW 和 PWI 方法的性能至关重要，而分解的准确位置并不重要，只要将原型激励波形所包含的基音脉冲标志点置于分解位置的中心就足够了。这也导致了在原型激励波形的基音周期里产生了小的误差，使得该方法相对于语音域提取原型激励波形方法来说，计算量减少，但是激励基音标志点的可信度下降。

2.3.2 原型波形谱包络的修正

因为瞬时激励波形 $s(t, \varphi)$ 是周期的，可用式 (2-4) 的傅立叶级数表示。在 PWI 编码器的各级中，必须修正 PW 的谱包络。在传统的合成分析编码器中，用 FIR 和 IIR 滤波器都可以改变谱包络。对于一个语音域内的 PW，当 FIR 滤波运算在

这个 PW 上时, 滤波器的初始状态是通过滤波当前 PW 的周期扩展部分而得到, 即使用循环卷积得到滤波器的初始状态。令时刻 t 处的瞬时语音波形为

$$u(t, \varphi) = \sum_{k=0}^{\lfloor p(t)/2 \rfloor} [C_k(t) \cos(k\varphi) + D_k(t) \sin(k\varphi)] \quad (2-19)$$

用系数为 $\alpha_0, \alpha_1, \dots, \alpha_p$ 的 LPA 滤波器对瞬时语音波形的采样信号进行数字滤波, 能得到一带限的采样的周期残差信号。在奈奎斯特频率处进行理想低通滤波, 可得到连续的残差信号, 瞬时残差波形可表示为

$$s(t, \varphi) = \sum_{k=0}^{\lfloor p(t)/2 \rfloor} \sum_{i=0}^p \alpha_i \left\{ C_k(t) \cos\left(k\varphi - \frac{2\pi k iT}{p(t)}\right) + D_k(t) \sin\left(k\varphi - \frac{2\pi k iT}{p(t)}\right) \right\} \quad (2-20)$$

式中, T 是采样间隔。式 (2-20) 意味着瞬时语音波形的傅立叶系数和瞬时激励波形的傅立叶系数存在如下关系 (由式 (2-4) 和式 (2-20) 相等求得):

$$\left. \begin{aligned} A_k(t) &= C_k(t) \sum_{n=0}^p \alpha_n \cos\left(\frac{2\pi knT}{p(t)}\right) - D_k(t) \sum_{n=0}^p \alpha_n \sin\left(\frac{2\pi knT}{p(t)}\right) \\ B_k(t) &= C_k(t) \sum_{n=0}^p \alpha_n \sin\left(\frac{2\pi knT}{p(t)}\right) + D_k(t) \sum_{n=0}^p \alpha_n \cos\left(\frac{2\pi knT}{p(t)}\right) \end{aligned} \right\} \quad (2-21)$$

这里 $A_k(t)$ 和 $B_k(t)$ 分别为瞬时激励波形的余弦和正弦项系数。

对式 (2-21) 进行简单的转换, 可从瞬时激励波形中获得瞬时语音波形:

$$\left. \begin{aligned} C_k(t) &= \frac{A_k(t) \sum_{n=0}^p \alpha_n \cos\left(\frac{2\pi knT}{p(t)}\right) + B_k(t) \sum_{n=0}^p \alpha_n \sin\left(\frac{2\pi knT}{p(t)}\right)}{\left[\sum_{n=0}^p \alpha_n \cos\left(\frac{2\pi knT}{p(t)}\right)\right]^2 + \left[\sum_{n=0}^p \alpha_n \sin\left(\frac{2\pi knT}{p(t)}\right)\right]^2} \\ D_k(t) &= \frac{-A_k(t) \sum_{n=0}^p \alpha_n \sin\left(\frac{2\pi knT}{p(t)}\right) + B_k(t) \sum_{n=0}^p \alpha_n \cos\left(\frac{2\pi knT}{p(t)}\right)}{\left[\sum_{n=0}^p \alpha_n \cos\left(\frac{2\pi knT}{p(t)}\right)\right]^2 + \left[\sum_{n=0}^p \alpha_n \sin\left(\frac{2\pi knT}{p(t)}\right)\right]^2} \end{aligned} \right\} \quad (2-22)$$

这样就可以利用式 (2-21) 和式 (2-22) 对 PW 去除共振峰结构或加入共振峰结构。

2.3.3 对齐和 SCR

一般来说, 被提取的相继的原型激励波形未被适当地对齐, 即两个原型激励波形的类似特征没有出现在类似的相位 φ 处。但 PW 在提取时希望能在每个修正点都提供一个 PEW 的傅立叶级数表示式, 这就需要进行在量化和内插前对原型激励波形进行相位对齐。对齐需要 SCR 测度。为了防止对齐出现问题, 较短的 PW 应重复整数倍, 以便在对齐前匹配于较长的原型激励波形的长度。将原型激励波形 $s(t_{m+1}, \varphi)$ 相位移位 ξ , 使得 $s(t_{m+1}, \varphi)$ 和原来的原型激励波形 $s(t_m, \varphi)$ 间的类似性测度最大。可得

$$\begin{aligned} \xi &= \arg \max_{\xi'} \Xi[s(t_m, \varphi), s(t_{m+1}, \varphi - \xi')] \\ &= \arg \max_{\xi'} \sum_{k=0}^{\lfloor p(t) \rfloor} [A_k(t_m)A_k(t_{m+1}) + B_k(t_m)B_k(t_{m+1})]\cos(k\xi') + \\ &\quad [B_k(t_m)A_k(t_{m+1}) - A_k(t_m)B_k(t_{m+1})]\sin(k\xi') \end{aligned} \quad (2-23)$$

公式 (2-23) 中右边是以傅立叶级数系数表示的互相关。在时刻 t_{m+1} , 对齐的激励的余弦和正弦系数分别是

$$\begin{cases} \tilde{A}_k(t_{m+1}) = A_k(t_{m+1})\cos(k\xi) - B_k(t_{m+1})\sin(k\xi) \\ \tilde{B}_k(t_{m+1}) = A_k(t_{m+1})\sin(k\xi) + B_k(t_{m+1})\cos(k\xi) \end{cases} \quad (2-24)$$

从现在开始, 所有的傅立叶系数都假定是对齐的, 符号“~”将被略去。

2.3.4 原型激励波形的量化

由于在此这前的原型激励波形已被对齐, 相继的原型激励波形是非常类似的, 这样对原型激励波形进行差分编码是很有利的。在 PWI 的方法中, 码矢量是由傅立叶级数系数构成, 它表示了一些在时域中的带限脉冲。原理是可用原来量化的激励波形 $s_Q(t_m, \varphi)$ (以 20~30ms 的间隔传送) 和一个或多个码书来近似当前量化的原型激励波形 $s_Q(t_{m+1}, \varphi)$ 。设用 L 个码书矢量来激励波形时, 在时刻 t_{m+1} 处的量化激励可表示为

$$s_Q(t_{m+1}, \varphi) = g_0 s(t_m, \varphi) + \sum_{l=1}^L g_l C_{k_l}^{(l)}(\varphi) \quad (2-25)$$

这里的 $C_{k_l}^{(l)}$ 为第 l 个码书的第 k_l 个码矢量, g_l 是增益因子。对于具有 $\lfloor p(t)/2 \rfloor$ 个谐波的激励波形, 从每个存储的候选矢量中取出前 $\lfloor p(t)/2 \rfloor$ 个谐波, 使得 SNR 最大的增益和码书矢量则为所求, 即有

$$\{g_0, g_1, \dots, g_L, k_1, k_2, \dots, k_L\}$$

$$= \arg \max_{(g_0, \dots, g_L, k_1, \dots, k_L)} (SNR[H[u(t_{m+1}, \varphi)], H[g_0 u_Q(t_m, \varphi)] + \sum_{l=1}^L g_l C_{k_l}^{(l)}(\varphi)]) \quad (2-26)$$

假定优化的增益因子已经存在, 首先选择码本标号。为了在不降低性能的前提下对码书标号分层优化, 量化码书中的每个矢量要与前面码书得到的优化码矢量正交, 同时也要与原来量化的原型激励波形正交。最后再对增益因子量化。在增益因子 $\{g_i\}$ 量化前, 可以调整长时 SCR, 由于长时 SCR 用在量化的增益因子中, 所以不用传送额外的信息就能达到使长时 SCR 自适应的目的。

图 2-2 给出了原型激励波形 $s(t_{m+1}, \varphi)$ 的简单量化过程。

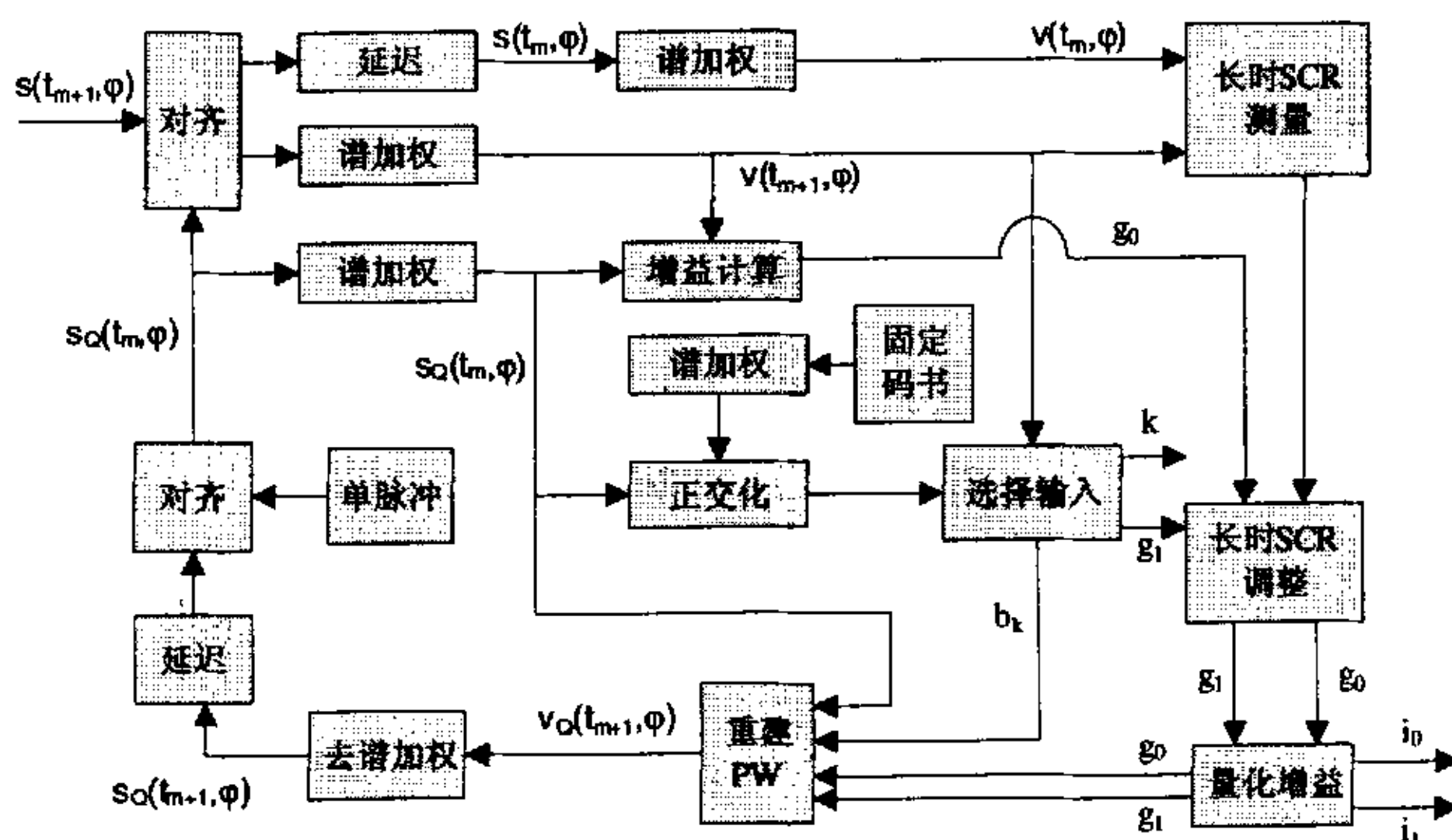


图 2-2 原型激励波形量化流程

Fig.2-2 PEW Quantization Flowchart

首先, 未量化的原型激励波形与原来量化的原型激励波形 $s_Q(t_m, \varphi)$ 对齐, 在当前和原来未量化的原型激励波形上测量长时 SCR。如果使用了谱加权, 那么在图 2-2 中原型波形应用 $v(\bullet, \bullet)$ 表示, 而不用 $s(\bullet, \bullet)$ 表示。在增益计算部分里, 将

$s(t_{m+1}, \varphi)$ 投射到 $s_Q(t_m, \varphi)$ 上可以计算出原来的原型激励波形的增益 g_0 。在与未量化的原型波形 $s(t_{m+1}, \varphi)$ 比较前，要使谱加权的码书矢量与 $v_Q(t_m, \varphi)$ 正交，确定最佳矢量 b_k ，传送该矢量的标号 k ，而最佳码书增益 g_1 用于调整长时 SCR。假定 $v_Q(t_m, \varphi)$ 的能量和 b_k 是归一化的，或者对于长时 SCR 调整部分是已知的，那么可根据公式 (2-13) 和公式 (2-14) 对增益进行调整，然后量化调整后的增益，传送它们对应的标号 i_0 和 i_1 。由 $v_Q(t_m, \varphi)$ ， b_k ， g_0 和 g_1 ，可以计算出当前谱加权的原型激励波形 $s(t_{m+1}, \varphi)$ ， $s(t_{m+1}, \varphi)$ 再与一个单脉冲原型波形对齐，以便保证基音脉冲位于原型波形的中心，单脉冲的符号必须与 $s(t_m, \varphi)$ 中的最大脉冲符号一致。对齐后的这个原型波形将用于下一个 PW 的量化。

而在接收端由符号 k ， i_0 和 i_1 ，以及过去的 PEW 重建当前原型激励波形的过程将在图 2-3 中给出。

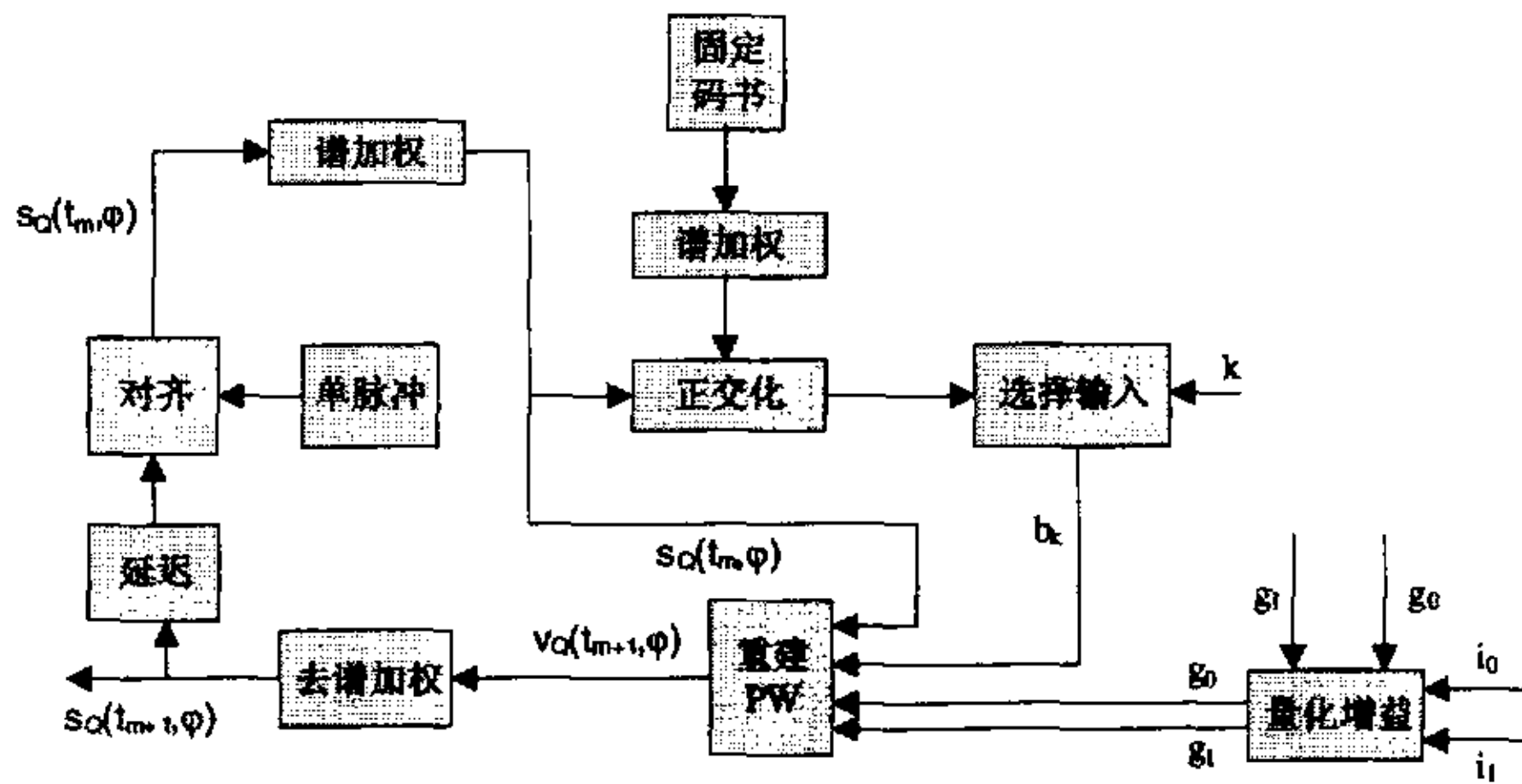


图 2-3 原型激励波形反量化框图

Fig.2-3 PEW Unquantization Flowchart

2.3.5 原型激励波形的内插

众所周知，傅立叶级数展开是线性的，从而 PEW 的线性内插等价于傅立叶级数系数的内插。由于基音周期在插值区间内变化使得 PEW 的谐波数不一定相等，为了方便内插，在较少谐波数的 PEW 中加入零幅度谐波。这与事实相符，即在抽样语音信号中，零幅度谐波信号通过防混迭滤波器的截止频率时可能出现，也可能不出现。

需要注意的是，当 2 个频谱长度不同时，可采用将比较长的谱截短或将较短

的谱后添零的方法。由于相邻帧的基音周期变化不大，且高频谱谐波具有相对低的能量，所以这种方法是可行的。

在 t_m 和 t_{m+1} 时刻提取 2 个已量化对齐的原型激励波形，设定两个波形的谐波数已被统一为 K 。可以按式 (2-5) 在区间 $t_m \leq t \leq t_{m+1}$ 上进行内插，得到用傅立叶级数表示的语音信号：

$$s(t) = \sum_{k=0}^K \left\{ [(1-\alpha(t))A_k(t_m) + \alpha(t)A_k(t_{m+1})] \cos[k\varphi(t_m)] + k \int_{t_m}^t \frac{2\pi dt'}{[1-\alpha(t')]p(t_m) + \alpha(t')p(t_{m+1})} + \right. \\ \left. [(1-\alpha(t))B_k(t_m) + \alpha(t)B_k(t_{m+1})] \sin[k\varphi(t_m)] + k \int_{t_m}^t \frac{2\pi dt'}{[1-\alpha(t')]p(t_m) + \alpha(t')p(t_{m+1})} \right\} \\ t_m \leq t \leq t_{m+1} \quad (2-27)$$

值得注意的是，瞬时激励波形的相位 $\varphi(t)$ 等于式 (2-27) 中 $k=1$ 时正弦和余弦函数的变量，而插值函数 $\alpha(t)$ 为相位 $\varphi(t)$ 的简单线性函数，即

$$\alpha(t) = \frac{\varphi(t) - \varphi(t_m)}{2\pi M} \quad (2-28)$$

这里的 M 是内插区间长度除以 2 个修正点处的周期的平均值，即

$$M = 2 \frac{t_{m+1} - t_m}{p(t_m) + p(t_{m+1})} \quad (2-29)$$

最后经变换得到

$$\varphi(t) = \left\{ \begin{array}{l} \varphi(t_m) + 2\pi \frac{-Mp(t_m) + \sqrt{M^2 p^2(t_m) + 2M[p(t_{m+1}) - p(t_m)](t-t_m)}}{p(t_{m+1}) - p(t_m)} \quad p(t_{m+1}) \neq p(t_m) \\ \varphi(t_m) + 2\pi \frac{t-t_m}{p(t_m)} \quad p(t_{m+1}) = p(t_m) \end{array} \right\} \quad (2-30)$$

下面将上面各式离散化表示：

首先将原型激励波形的时间标号离散化，将 2 个内插的原型激励波形记为 $s(n_0, m)$ 和 $s(n_1, m)$ ；相应的周期为 $p(n_0)$ 和 $p(n_1)$ ，用采样点数表示；一帧内的样点数记为 L 。式 (2-29) 变为

$$M = 2 \frac{L}{p(n_0) + p(n_1)} \quad (2-31)$$

插值函数为

$$\alpha(i) = \begin{cases} \frac{-Mp(n_0) + \sqrt{M^2 p^2(n_0) + 2M[p(n_1) - p(n_0)]}}{M[p(n_1) - p(n_0)]} & i = 0, 1, \dots, L-1: p(n_1) \neq p(n_0) \\ \frac{i}{Mp(n_0)} & i = 0, 1, \dots, L-1: p(n_1) = p(n_0) \end{cases} \quad (2-32)$$

$$\varphi(i) = 2\pi M \alpha(i) \quad i = 0, 1, \dots, L-1 \quad (2-33)$$

则式 (2-27) 变为

$$s(i) = \sum_{k=0}^K \{ [(1-\alpha(i))A_k(n_0) + \alpha(i)A_k(n_1)] \cos[k\varphi(i)] + [(1-\alpha(i))B_k(n_0) + \alpha(i)B_k(n_1)] \sin[k\varphi(i)] \} \quad i=0, 1, \dots, L-1 \quad (2-34)$$

如式 (2-34) 所示就是内插后得到的一帧信号。

内插后得到的激励信号通过 IIR 滤波器即可获得重建语音，在时间上要对 LP 滤波器系数内插，一般采用的是线谱频率参数的线性内插方法。同时由于重建语音 PCW 的相位是已知的，可以在 PCW 的倍数处修正滤波器。这种原型波形内插方法能够得到高质量的合成语音，利用此方法对提取的原型波形进行编码，并用内插重建未经编码的语音信号，将大大降低编码率^[25]。

2.3.6 PWI 算法的初始化

PWI 算法仅能用于产生浊音语音信号，而在清音段语音需要用到其他的方法进行编码。不同编码模式的使用要靠清浊判决。在浊音区的开始，过去估计的原型激励波形 $s_Q(t_m, \varphi)$ 不存在，此时可有三种方法估计过去的原型激励波形：

1. 利用 PW 的提取原理从帧分界点之前的重建语音信号中估计；此时在译码端要确定一个合适的开始相位 $\varphi(t_m)$ ，对于突然开始的浊音区， $\varphi(t_m)$ 并不严格。
2. 用单脉冲波形表示过去的原型激励波形，单脉冲波形的幅度由所传送的信息决定；

3. 用当前的原型激励波形 $s_Q(t_{m+1}, \varphi)$ 复制。这时可用一个位于中心的单脉冲作为过去的原型激励波形。

2.4 PWI 编码器性能分析

PWI 算法就是在一定的时间范围内以慢渐变的 PCW 为基础在浊音帧里提取和传送一段 PW，再用线性插值恢复未被传送的信号。它在语音编码中提供了一般的波形内插技术，同时考虑了浊音语音重建准确度的两个方面：波形形状准确度和周期性准确度，再结合了清音的 CELP 技术，使其成为一种具有巨大潜力的混合编码器。

Kleijn 博士给出了一种数码率是 3.5kb/s 的 PWI 语音编码方案，取帧长为 20ms，比特分配表如表 2-1 所示。

表 2-1 3.5kb/s PWI 编码器比特分配表^[20]

Table 2-1 3.5kb/s PWI Coder's Bit Allocation

参数	符号	比特分配
LP 系数	$\{\alpha_i\}$	30
基音周期	$p(t_{m+1})$	7
原来的 PW 增益	g_0	5
码书	$C_{(k_1)}^{(1)}(\varphi), C_{(k_2)}^{(2)}(\varphi)$	2×8
码书增益	g_1, g_2	2×6
合计		70

正式 MOS (Mean Opinion Scoring 平均评价得分) 测试结果并与其他语音编码算法比较的结果如表 2-2 所示。在 Kleijn 博士提出的编码方案中，清音语音用 CELP 算法进行编码，比特速率为 4.1kb/s。从 MOS 的分析结果来看，此 PWI/CELP 混合编码的语音质量接近 8kb/s 的 VSELP 编码算法。

表 2-2 3.5kb/s PWI 编码器 MOS 测试及比较结果^[20]

Table2-2 3.5kb/s PWI Coder's MOS Test and Comparison with Other Coders

编码器	比特率 (kb/s)	MOS
PCM	128	4.20
GSM	13.2	3.60
IS54	8.0	3.56
PWI	3.5/4.1	3.45

虽然 PWI 算法能产生高质量的浊音语音，但是它的局限性也是非常明显的，主要表现在：

1. PWI 仅适用于浊音语音，要想实现一个完整的编码系统，必须与其他的编码器-CELP，组合工作以处理清音语音。这使得编码器整体在过渡区易产生较明显的听觉失真，而且这种失真通常发生在 CELP 和 PWI 模式间转换处附近^[26]。

2. PWI 编码器中浊音语音线性变化这一基本假设，使得幅度的线性插值有时会产生过分的平滑，导致了不自然的重建语音。同时，基音周期的线性插值也引入了过强的周期性，有时会产生小的蜂鸣声^[27]。

3. PWI 编码器中的高复杂对齐过程，大大地提高了 PWI 算法的计算复杂度，使得其实时实现有很大的困难。

2.5 本章小结

本章详细讲述了波形内插语音编码算法中的基本算法—原型波形内插 PWI 算法的基本原理及其实现过程，并对 PWI/CELP 混合编码器的性能作了分析。

第三章 特征波形内插语音编码算法

3.1 基本思想

针对 PWI 算法的缺陷, Kleijn 博士提出了一种新的波形内插算法—特征波形内插算法 CWI (Characteristic Waveform Interpolation), 能够同时对浊音和清音语音编码而无需清浊判决。在 WI 方法中, 语音信号用渐变波形序列表示, 对于浊音语音信号来说, 这些波形都是简单的 PCW, 但是对于清音语音和背景噪声, 波长是变长的类噪声。当渐变波形不再对 PCW 限制时, 再用 PCW 或 PW 来描述渐变波形是不合适的。所以在 CWI 算法中, 改用特征波形来表示渐变波形。

CWI 算法将特征波形 CW (Characteristic Waveform) 分解为慢渐变波形 SEW (Slowly Evolving Waveform) 和快渐变波形 REW (Rapidly Evolving Waveform) 两个部分。SEW 波形表示的是语音信号的准周期成分, 而 REW 表示的是信号的非周期成分和噪声成分。这两种波形具有非常不同的感性需要, 可以分别对它们量化, 从而达到了增强编码效率的目的^[28]。

3.2 基本原理

CWI 算法将语音信号表示为一个二维表面 $u(t, \varphi)$, 沿相位轴 φ 显示按一定时间间隔提取的 PCW, 而沿时间轴 t 显示的是 PCW 的渐变波形。将沿 φ 轴显示的 PCW 称为特征波形 CW, 中间未被传送的 PCW 可通过 CW 的内插重建, 一旦求得各时刻的 PCW 及其相位函数, 即可完美重建原始语音信号。

CW 和两种类型的傅立叶变换相联系。一种是在给定的时刻 t 处, 沿 φ 轴的 $u(t, \varphi)$ 的傅立叶变换, 该变换类似于语音信号本身的短时傅立叶变换; 另一种是在给定的相位 φ 处, 沿 t 轴的 $u(t, \varphi)$ 的傅立叶变换, 该变换可以得到一个与 CW 的渐变速率相关的渐变频率谱, 整个渐变频率谱的带宽是基音周期的函数。要得到无失真的合成语音, 一般应至少每隔一个基音周期提取一个 CW, 若基音周期为 $p(t)$, 那么谱的渐变带宽最多为 $1/2p(t)$ ^[29]。

由于语音信号中, 准周期成分和噪声成分在感性上不同, CWI 算法可以相当高的提取率提取 CW, 再运用简单的 FIR 滤波器, 沿着时间轴 t 对语音信号 $u(t, \varphi)$ 分别进行低通和高通滤波。高通滤波部分得到一个快变波形 REW, 代表信号的类噪声部分; 低通滤波结果是一个缓变波形 SEW, 代表信号的准周期部分。REW 的幅度频谱可用低精确度的描述, 伴以一个较高的更新速率就可以保证得到好的

性能。与之相反，SEW 的幅度频谱则需要精确的描述，它的更新速率可由于小的渐变带宽而取得较低。

CWI 编码器的原理图如下：

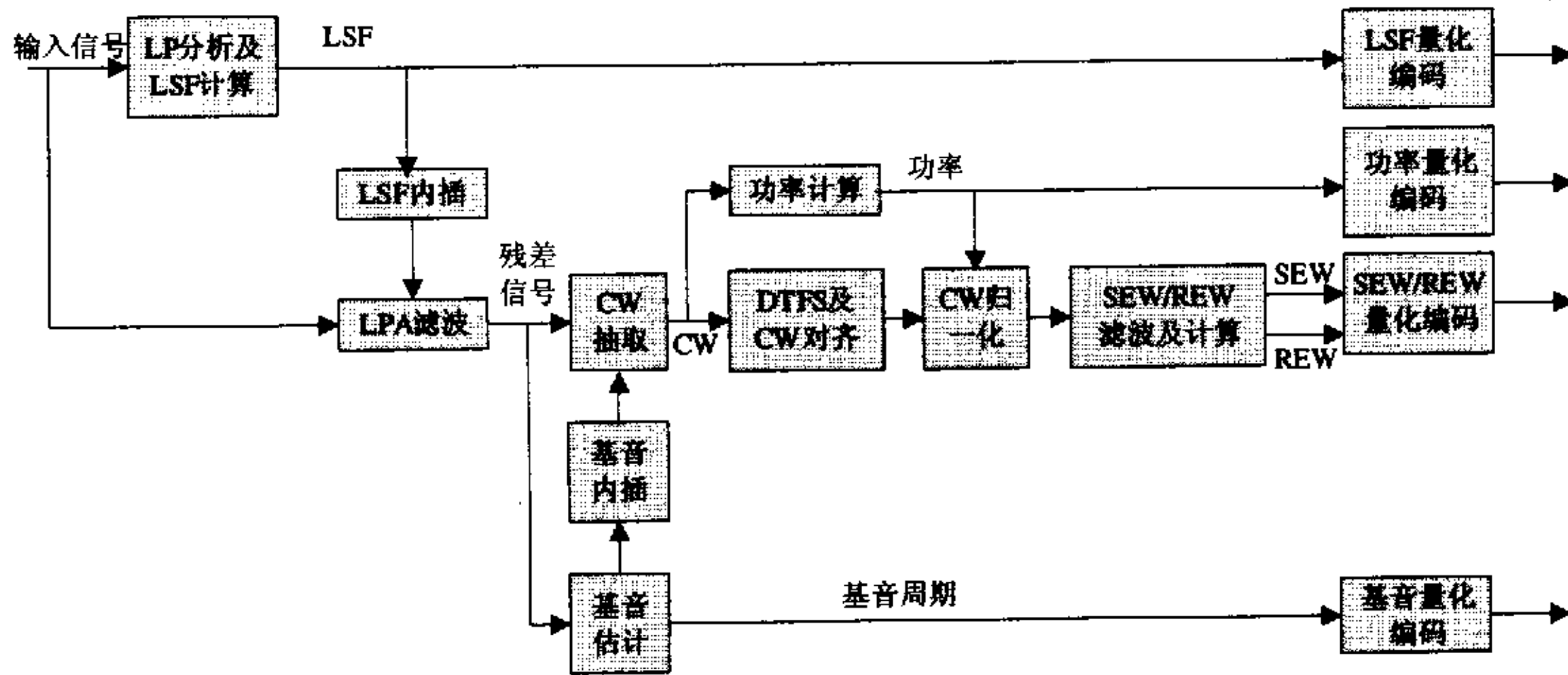


图 3-1 CWI 编码流程
Fig.3-1 CWI Encoder Flowchart

CWI 解码器的原理图如下：

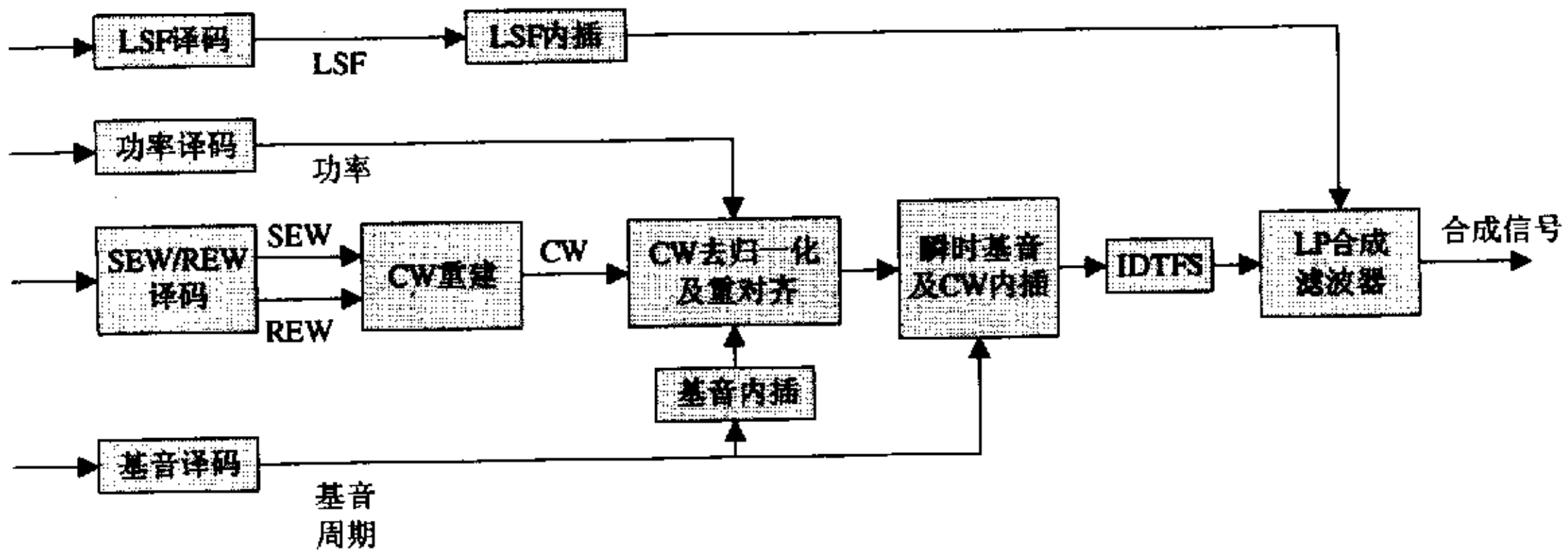


图 3-2 CWI 解码流程
Fig.3-2 CWI Decoder Flowchart

在编码器里，首先对输入信号进行 LP 分析，得到对应的残差信号。其次，由残差信号估计基音周期，再根据估计的基音周期提取 CW。为了使 CW 表达波形渐变的二维表面，提取的 CW 先被转换到 DTFS 域再进行对齐，并用功率对其

进行归一化。将已对齐归一化的 CW 滤波分解，得到慢渐变波形 SEW 和快渐变波形 REW，对 SEW 和 REW 分别量化编码，使得 CWI 算法能在低码率里产生高质量的合成语音。

解码是编码的逆过程。由译码得到的 SEW 和 REW 重建 CW，再由 CW 通过线性内插和 IDTFS 重建残差信号 (LPR)，将残差信号通过 LP 合成滤波器即可得到最终的重建语音信号。

3.3 实现过程

3.3.1 编码合成模块分析

CWI 语音编码器的功能是将语音信号分解为一系列的渐变 CW (二维表面) 和其他的正交参数，如 CW、LSF 系数、功率和基音周期等，CW 再被分解成 REW 和 SEW，并分别将它们量化后送出。

3.3.1.1 线性预测分析及线谱频率参数计算

1. 线性预测分析 LPA (Linear Prediction Analysis) 一般借助于线性预测误差滤波器来实现，线性预测误差滤波器的传递函数为

$$A(z) = 1 - \sum_{i=1}^p \alpha_i z^{-i} \quad (3-1)$$

其中， p 为预测器阶数， $\{\alpha_i\}_{i=1,2,\dots,p}$ 为线性预测器系数，实现框图如下：

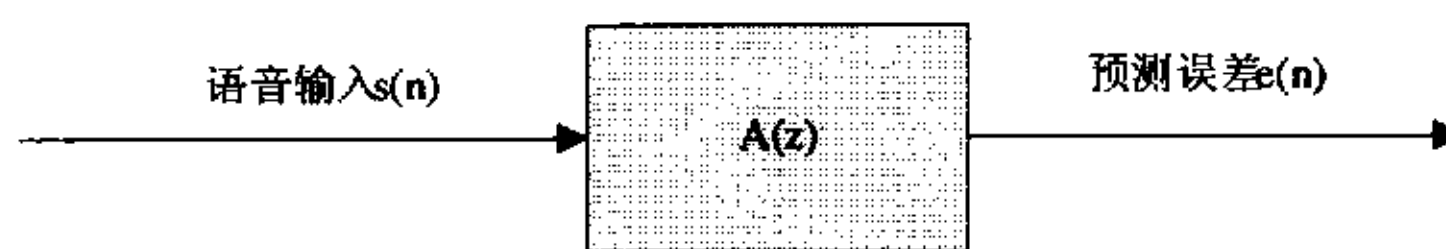


图 3-3 语音信号的线性预测模型

Fig.3-3 Linear Prediction Model of Speech Signal

输出 $e(n)$ 和输入 $s(n)$ 间满足如下关系：

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^p \alpha_i s(n-i) \quad (3-2)$$

其中， $\hat{s}(n)$ 称为 $s(n)$ 的预测值。

线性预测分析 LPA 实质上是设计一个预测误差滤波器 $A(z)$ ，即求解 $\{\alpha_i\}_{i=1,2,\dots,p}$ 使预测误差 $e(n)$ 在某个预定的准则下最小。理论上选择均方误差 $E[e^2(n)]$ 最小准则，按定义有

$$E[e^2(n)] = E\left[\left[s(n) - \sum_{i=1}^p \alpha_i s(n-i)\right]^2\right] \quad (3-3)$$

为使 $E[e^2(n)]$ 最小，令

$$\frac{\partial E[e^2(n)]}{\partial \alpha_j} = -2E[e(n)s(n-j)] = 0 \quad (3-4)$$

将 $e(n)$ 按 (3-2) 式的表示代入 (3-4) 式中，得到

$$E[s(n)s(n-j) - \sum_{i=1}^p \alpha_i s(n-i)s(n-j)] = r(j) - \sum_{i=1}^p \alpha_i r(j-i) = 0 \quad 1 \leq j \leq p \quad (3-5)$$

其中， $r(j) = E[s(n)s(n-j)]$ 是 $s(n)$ 的自相关序列。

定义自相关矢量 \mathbf{r} 、自相关矩阵 \mathbf{R} 、和参数矢量 \mathbf{a} 分别为

$$\mathbf{r} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(p-1) \\ r(1) & r(0) & \cdots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$$

式 (3-5) 可写成矩阵形式如下：

$$\mathbf{r} - \mathbf{R}\mathbf{a} = \mathbf{0} \quad (3-6)$$

上式被称为 Yule-Walker 方程。

令最小均方误差为 E_p ，有

$$E_p = E[e^2(n)]_{\min} = E[e(n)\{s(n) - \sum_{i=1}^p \alpha_i s(n-i)\}] = E[e(n)s(n)] (\because E[e(n)s(n-j)] = 0, 1 \leq j \leq p)$$

$$= E[\{s(n) - \sum_{i=1}^p \alpha_i s(n-i)\}s(n)] = r(0) - \sum_{i=1}^p \alpha_i r(i) \quad (3-7)$$

组合式 (3-6) 和式 (3-7), 可得完整的、平稳信号的线性预测误差滤波器的求解方程式如下:

$$\begin{bmatrix} r(0) & r(1) & \cdots & r(p) \\ r(1) & r(0) & \cdots & r(p-1) \\ \vdots & \vdots & \vdots & \vdots \\ r(p) & r(p-1) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ -\alpha_1 \\ \vdots \\ -\alpha_p \end{bmatrix} = \begin{bmatrix} E_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3-8)$$

用自相关法和 Levinson-Durbin 算法, 可递推求出方程式 (3-8) 的解, 从而解得所求的 LP 系数 $\{\alpha_i\}_{i=1,2,\dots,p}$ 。

输入语音信号通过 LP 分析后得到 LP 系数 $\{\alpha_i\}_{i=1,2,\dots,p}$, 再将其转换成线谱频率参数 LSFs。为了保证谱包络的平滑过渡, 两个相邻帧的 LSF 参数要线性内插, 以得到每个子帧的 LSF 参数, 再将每组内插的 LSF 参数转换回 LPA 系数, 用于 LPA 滤波器, 即可得到当前帧的 LP 残差信号^[30]。

另外, 当前帧中内插的最后一组 LP 系数准备用于计算未来帧的残差样点, 这些未来的残差样点将用于基音估计和 CW 提取。

2. LP 参数与 LSF 参数之间的相互转换关系:

(1) LP 系数转换成 LSF 参数: 已知 LP 系数 $\{\alpha_i\}_{i=1,2,\dots,p}$, 可得到 LP 分析滤波器的系统函数 $A(z) = \sum_{i=0}^p \alpha_i z^{-i}, \alpha_0 = 1$ 。构造两个多项式 $P_A(z)$ 和 $Q_A(z)$ 分别为:

$$\begin{aligned} P_A(z) = A(z) - z^{-(p+1)} A(z^{-1}) = & 1 + p'_1 z^{-1} + p'_2 z^{-2} + \cdots + p'_{\frac{p}{2}} z^{-\frac{p}{2}} - \\ & p'_{\frac{p}{2}} z^{-\frac{(p+1)}{2}} - \cdots - p'_1 z^{-p} - z^{-(p+1)} \end{aligned} \quad (3-9)$$

$$\begin{aligned} Q_A(z) = A(z) + z^{-(p+1)} A(z^{-1}) = & 1 + q'_1 z^{-1} + q'_2 z^{-2} + \cdots + q'_{\frac{p}{2}} z^{-\frac{p}{2}} + \\ & q'_{\frac{p}{2}} z^{-\frac{(p+1)}{2}} + \cdots + q'_1 z^{-p} + z^{-(p+1)} \end{aligned} \quad (3-10)$$

$Q_A(z)$ 是对称多项式, $P_A(z)$ 是反对称多项式, 可通过如下变换得到两个对称多项式:

$$P'(z) = \frac{P(z)}{1-z^{-1}} = 1 + p''_1 z^{-1} + \dots + p''_{\frac{p}{2}} z^{-\frac{p}{2}} + \dots + p''_1 z^{-(p-1)} + z^{-p} \quad (3-11)$$

上式, $p''_1 = p'_1 + 1, p''_i = p'_i + p''_{i-1}, i = 2, 3, \dots, \frac{p}{2}$ 。

$$Q'(z) = \frac{Q(z)}{1+z^{-1}} = 1 + q''_1 z^{-1} + \dots + q''_{\frac{p}{2}} z^{-\frac{p}{2}} + \dots + q''_1 z^{-(m-1)} + z^{-p} \quad (3-12)$$

同理, 上式中 $q''_1 = q'_1 - 1, q''_i = q'_i - q''_{i-1}, i = 2, 3, \dots, \frac{p}{2}$ 。

由于 $P'(z)$ 、 $Q'(z)$ 的根在单位圆上, 令 $z = e^{j\omega}$, 则有 $z + z^{-1} = 2\cos(\omega)$, 令 $(z + z^{-1})/2|_{z=e^{j\omega}} = \cos \omega = x$, 则 $P'(z) = 0$ 、 $Q'(z) = 0$ 同是关于 x 的 $p/2$ 次代数方程。联立解此代数方程组求得 x , 再由 $\omega_i = \cos^{-1}(x_i)$ 即可求得 LSF 参数。

(2) LSF 参数转换成 LP 系数: 已知 LSF 参数为 $\{\omega_i\}_{i=1,2,\dots,p}$, 且定义

$$\begin{cases} x_{i/2} = \cos(\omega_i), & i = 2, 4, \dots, p \\ y_{(i+1)/2} = \cos(\omega_i), & i = 1, 3, \dots, p-1 \end{cases}$$

定理 1: 给定一个一元 n 次方程 $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x^1 + a_0 = 0$ ($a_n \neq 0$)

它的 n 个根为 x_1, x_2, \dots, x_n , 则

$$x_1 + x_2 + \dots + x_n = -\frac{a_{n-1}}{a_n}$$

$$x_1 x_2 + x_1 x_3 + \dots + x_{n-1} x_n = \frac{a_{n-2}}{a_n}$$

.....

$$x_1 x_2 \dots x_n = (-1)^n \frac{a_0}{a_n}$$

根据定理 1 可分别得到 p''_i 和 x_i 、 q''_i 和 $y_{\frac{i+1}{2}}$ 间的关系, 从而求出 $\{p''_i, q''_i\}_{i=1,2,\dots,\frac{p}{2}}$,

跟着由 $\{p''_i, q''_i\}_{i=1,2,\dots,\frac{p}{2}}$ 转换回 $\{p'_i, q'_i\}_{i=1,2,\dots,\frac{p}{2}}$:

$$\begin{aligned} p'_1 &= p''_1 - 1, p'_i = p''_i - p''_{i-1}, & i = 2, 3, \dots, \frac{p}{2} \\ q'_1 &= q''_1 + 1, q'_i = q''_i + q''_{i-1}, & i = 2, 3, \dots, \frac{p}{2} \end{aligned} \quad (3-13)$$

最后通过下式, 可从 $\{p'_i, q'_i\}_{i=1,2,\dots,\frac{p}{2}}$ 得到所求的 LP 系数:

$$\begin{cases} \alpha_i = \frac{1}{2}(p'_i + q'_i), & i = 1, 2, \dots, \frac{p}{2} \\ \alpha_{(p+1)-i} = \frac{1}{2}(p'_i - q'_i), & i = 1, 2, \dots, \frac{p}{2} \end{cases} \quad (3-14)$$

3.3.1.2 基音估计

基音估计的准确性对于 CWI 编码算法来说非常重要, 因为编码端的 CW 提取过程, 以及解码端的 CW 去归一化和内插过程都强烈依赖于基音估计。基音估计的方法很多, 根据 CWI 算法模型的特点, 选择在 LPA 残差域中估计基音周期。在残差域进行基音估计的有利之处在于共振峰的影响已经排除, 即残差谱接近是平坦的^[31]。

简要的基音估计算法描述如下:

基音估计器每帧估计一次基音周期, 对于每帧数据, 基音估计器在两个叠接窗上独立计算。第一个窗覆盖了整个当前帧, 而第二个窗覆盖了当前帧的半帧和将来帧的半帧。将来的半帧残差信号用当前帧最后内插的 LPA 系数 $\{\alpha_i\}_{i=1,2,\dots,p}$ 计算, 由此得到的基音估计器对编码器没有额外引入时延。

基音估计依赖于如下的归一化自相关函数:

$$\rho(\tau) = \max\{0, \min\left\{\frac{\sum_{i=0}^{L_f-1-\tau} r(i)r(i+\tau)}{\sqrt{\sum_{j=0}^{L_f-\tau-1} r^2(j) \sum_{k=0}^{L_f-\tau-1} r^2(k+\tau)}}, 1.0\right\}\} \quad P_{\min} \leq \tau \leq P_{\max} \quad (3-15)$$

式(3-15)中, τ 为整数基音延迟, $r(\bullet)$ 为LPA残差信号, 分母为归一化因子, 最大和最小函数的作用在于保持 $\rho(\tau)$ 介于0~1之间, P_{\min} 和 P_{\max} 为所允许的最小和最大基音周期^[32]。

选取使 $\rho(\tau)$ 最大的基音延迟为最优延迟, 即真正的基音周期。如果延迟值 τ 对应于信号的真正基音周期或真正基音周期的整数倍, 则对应的 $\rho(\tau)$ 接近于1。相反, 如果信号不具备周期性特征(如清音语音), 则 $\rho(\tau)$ 远小于1。

在求得每个窗的最优基音延迟后, 以一定的逻辑关系组合两个窗的最优延迟, 以得到在当前帧更可靠的基音延迟估计。设 $\{\tau_0, \rho(\tau_0)\}$ 和 $\{\tau_1, \rho(\tau_1)\}$ 分别为在第一个窗和第二个窗内得到的最优延迟和归一化自相关系数, $\rho(\tau_0)$ 和 $\rho(\tau_1)$ 同时分别作为延迟值 τ_0 和 τ_1 的可信度函数, 以供基音估计算法通过下面的逻辑组合求出最后的延迟估计 τ_{opt} :

```

if (  $\rho(\tau_0) > \rho(\tau_1) + 0.4$  )
{
    if (  $|\tau_0 - \tau_1| > 15$  )
         $\tau_{opt} = \tau_0$ 
    else
         $\tau_{opt} = [(\tau_0 + \tau_1) / 2]$ 
}
else
     $\tau_{opt} = \tau_1$ 
    
```

值得注意的是, 无论信号是否周期, 基音估计器均提供了一个基音估计值。

对于清音语音， $\rho(\tau)$ 的值相对很低，表明基音周期变化大，此时为了减小编码器的计算复杂度，可将基音估计值置为最小值 P_{\min} 。

3.3.1.3 基音周期内插

要完成 CW 的提取，首先要知道每个提取点的基音周期，这就需要对基音周期进行内插。在 CWI 编码器中采用了简单的线性内插技术来内插基音周期。令 $P(n_1)$ 和 $P(n_2)$ 为当前帧分界点处的基音周期， $n_1 < n_2$ ，则基音周期线性内插公式如下：

$$P(n) = \frac{(n_2 - n)P(n_1) + (n - n_1)P(n_2)}{n_2 - n_1} \quad n_1 \leq n \leq n_2 \quad (3-16)$$

这里 $n_2 - n_1 = L_f$ ， L_f 是样点数表示的帧长。

在语音信号中，尤其是在浊音段的开始和结尾处，基音周期有时会出现加倍或减半的情况。另外，基音估计器经常会出错，估计的基音周期常常是实际基音周期的整数倍^[33]。无论哪种情况都会使重建语音听起来有“喳喳”声，为了纠正这个问题，CWI 编码算法改进了基音周期的内插方法：

$$P(n) = \left\{ \begin{array}{l} \left\{ \begin{array}{l} \frac{C(n_2 - n)P(n_1) + (n - n_1)P(n_2)}{C(n_2 - n_1)}, n_1 \leq n < \frac{n_1 + n_2}{2} \\ \frac{C(n_2 - n)P(n_1) + (n - n_1)P(n_2)}{(n_2 - n_1)}, \frac{n_1 + n_2}{2} \leq n < n_2 \end{array} \right. \left. \begin{array}{l} \text{当 } P(n_1) < P(n_2) \text{ 时, } C = \left\lceil \frac{P(n_2)}{P(n_1)} \right\rceil \\ \text{当 } P(n_1) > P(n_2) \text{ 时, } C = \left\lceil \frac{P(n_1)}{P(n_2)} \right\rceil \end{array} \right. \end{array} \right. \quad (3-17)$$

因子 C 可作为基音是否加倍或减半的指示器，当 $C=1$ 时，说明基音没有加倍/减半发生，可直接用 (3-16) 式内插基音周期；当 C 大于 1 时就意味着估计的基音有加倍/减半情况发生，基音的内插公式改用 (3-17) 式，使得基音在中点不连续变化。

3.3.1.4 特征波形的提取

完成基音周期的估计及内插后，就可以在每帧上提取 CW 了。设每帧提取 M 个 CW， L_f 为每帧样点数。首先将当前帧划分为 M 个等长区间，定义每个区间的

末端点为 CW 的提取点，也就是相邻提取点间隔为 L_f/M 个样点数，定义编码器的子帧长度 $L_{sf} = L_f/M$ ，这就使得 CW 的提取点一直位于子帧的末端点。然后，以每个提取点为中心建立一个长度为内插基音周期的提取窗，窗所覆盖的 LP 残差信号段就是我们要提取的 CW。

在编码器提取完 CW 的后续步骤中，要对 CW 进行傅立叶变换。而在傅立叶变换中，CW 被周期延拓，所以在提取过程中如果不对 CW 的边界加以处理，周期延拓的 CW 会产生大的不连续性，引起听觉上的失真。为了解决这个问题，可以考虑在一定程度上放松每个提取点的位置，允许真正的提取点偏离理想提取点 ε 个样点，定义提取窗边界周围信号能量最小的 ε 为最佳偏离值。

需要注意的是，由于每个提取点允许有一定的偏移，两个相邻的 CW 可能分享同样的残差信号段，甚至有可能是一样的。对于浊音语音，每个提取的波形可以看成是一个 PCW，而对于清音语音，提取的波形为变长的类噪声段。理论上，当 $\varepsilon = 0$ 时，分析帧中的每个样点至少应该包含在一个 CW 中。

还有，提取窗有可能跨越分解点，完成当前帧的 CW 提取需要许多过去帧和未来帧的样点。如果 CW 的最大可能长度为 P_{max} ，则需要过去帧的样点数至少为 $P_{max}/2$ ，需要将来帧的样点数也至少为 $P_{max}/2$ 。这些未来的样点由 LPA 滤波器决定，不需要额外的算法延迟。

3.3.1.5 特征波形的傅立叶级数表示

在 CW 提取之后、CW 对齐之前，将 CW 转换到 DTFS 域，用离散时间的傅立叶级数表示 CW。单个一维 CW，其 DTFS 表示如下：

$$s(m) = \sum_{k=0}^{\lfloor P/2 \rfloor} [A_k \cos(\frac{2\pi km}{P}) + B_k \sin(\frac{2\pi km}{P})] \quad 0 \leq m < P \quad (3-18)$$

上式， P 为基音周期， $\{A_k\}$ 和 $\{B_k\}$ 是离散傅立叶级数系数， m 可为 $0 \leq m < P$ 之间的任意实值。

当 P 为偶数时， $\{A_k\}$ 和 $\{B_k\}$ 计算公式如下：

$$\left. \begin{aligned}
 A_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\
 B_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right]
 \end{aligned} \right\}, \quad k = 1, 2, \dots, \frac{P}{2} - 1$$

$$\left. \begin{aligned}
 A_k &= \frac{1}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\
 B_k &= \frac{1}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right]
 \end{aligned} \right\}, \quad k = 0, \quad k = \frac{P}{2}$$
(3-19)

当 P 为奇数时, $\{A_k\}$ 和 $\{B_k\}$ 计算公式如下:

$$\left. \begin{aligned}
 A_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\
 B_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right]
 \end{aligned} \right\}, \quad k = 1, 2, \dots, \frac{P-1}{2}$$

$$\left. \begin{aligned}
 A_k &= \frac{1}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\
 B_k &= \frac{1}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right]
 \end{aligned} \right\}, \quad k = 0$$
(3-20)

现在一单个的 CW 可用一组 DFTS 系数 $\{A_k\}$ 和 $\{B_k\}$ 描述。对式 (3-18) 中的所有参数附加一个离散时间标号即可得到 CW 序列的二维离散表示:

$$s(n, m) = \sum_{k=1}^{\lfloor P(n)/2 \rfloor} \left[A_k(n) \cos\left(\frac{2\pi km}{P(n)}\right) + B_k(n) \sin\left(\frac{2\pi km}{P(n)}\right) \right] \quad 0 \leq m < P(n) \quad (3-21)$$

上式中, $\{A_k(n)\}$ 、 $\{B_k(n)\}$ 以及 $P(n)$ 均是时变的。式中忽略了系数 A_0 和 B_0 , 因为在式 (3-19) 和式 (3-20) 中, B_0 的正弦项值 $\sin(0) = 0$; 而 A_0 表达的是信号的直流成分, 对语音的感性质没有影响。

现在, 式 (3-21) 是 CW 的二维表示式, 沿 m 轴显示的是单个的 CW, 而沿时间轴 n 显示的是 CW 的渐变形状。值得注意的是, CW 的长度取决于时变的基音周期 $P(n)$, 不同时刻的 CW 可能有不同的长度。

为了方便, 可以对 CW 的长度进行归一化。通过如下的变换公式:

$$\varphi = \varphi(m) = \frac{2\pi m}{P(n)} \quad (3-22)$$

式 (3-21) 可变为

$$s(n, \varphi) = \sum_{k=1}^{\lfloor P(n)/2 \rfloor} [A_k(n) \cos(k\varphi) + B_k(n) \sin(k\varphi)] \quad 0 \leq \varphi(\bullet) < 2\pi \quad (3-23)$$

这样可使所有的 CW 有同样的长度 2π 。这个变换将在解码端的合成信号时用到。

3.3.1.6 特征波形的对齐

提取到的 CW 一般是不同相位的，也就是说这些特征波形的主特征在时间上并没有对齐。为了确保渐变 CW 的精确描述，必须进行 CW 的对齐。

在每个子帧的基础上，对当前的 CW 进行循环时间移位就可以完成当前 CW 与原来 CW 的对齐。原因是 DTFS 表达允许我们将 CW 看作是周期信号中的一个周期波形，循环时间移位等价于对 DTFS 系数加一个线性相位。

对齐模块图如下所示：

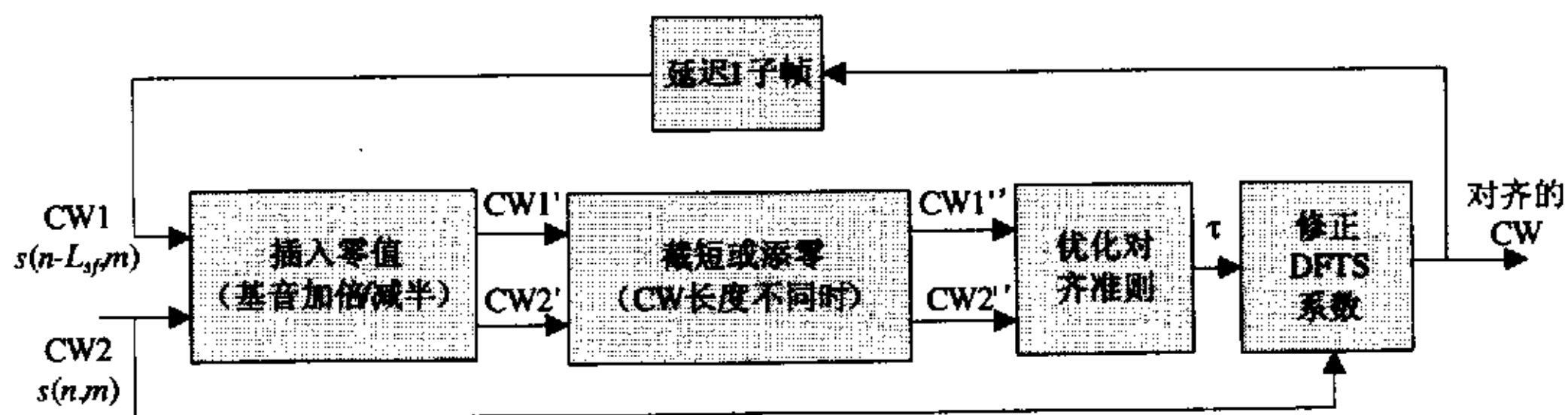


图 3-4 CW 对齐过程

Fig.3-4 CW Alignment Procedure

1. CW 长度相同时的对齐过程

当 CW 长度相同时，CW 对齐过程图里的插入零值模块和截短或添零模块均不起作用，输入信号 CW1 和 CW2 直接进入优化对齐准则模块。根据公式 (3-21)，原来的 CW 和当前的 CW 的 DTFS 表示分别是

$$\begin{aligned}
 s(n_0, m) &= \sum_{k=1}^M [A_k(n_0) \cos(\frac{2\pi km}{P}) + B_k(n_0) \sin(\frac{2\pi km}{P})] \\
 s(n_1, m) &= \sum_{k=1}^M [A_k(n_1) \cos(\frac{2\pi km}{P}) + B_k(n_1) \sin(\frac{2\pi km}{P})]
 \end{aligned} \tag{3-24}$$

这里的 n_0 和 n_1 对应为原来 CW 和当前 CW 的时间标号: $n_1 - n_0 = L_{sf}$ 。

$$\text{令} \quad \left. \begin{aligned} P &= P(n) = P(n-1) \\ M &= \lfloor P(n)/2 \rfloor = \lfloor P(n-1)/2 \rfloor \end{aligned} \right\} \tag{3-25}$$

其中, P 表示 CW 的长度, M 表示谱的谐波数。

假定当前的 CW 向右循环移位 T 个样点, 那么 $s(n_1, m)$ 变为

$$s(n_1, m-T) = \sum_{k=1}^M [A_k(n_1) \cos(\frac{2\pi k(m-T)}{P}) + B_k(n_1) \sin(\frac{2\pi k(m-T)}{P})] \tag{3-26}$$

通过使以 DTFS 形式表示的两个 CW 之间的互相关最大, 可确定 $s(n_1, m-T)$ 和 $s(n_0, m)$ 对齐所需要的最佳时间移位 T :

$$\begin{aligned}
 T = \arg \max_{0 \leq T' < P} \sum_{k=1}^M \{ [A_k(n_0)A_k(n_1) + B_k(n_0)B_k(n_1)] \cos(\frac{2\pi kT'}{P}) + \\
 [B_k(n_0)A_k(n_1) - A_k(n_0)B_k(n_1)] \sin(\frac{2\pi kT'}{P}) \}
 \end{aligned} \tag{3-27}$$

上式也可表示为归一化时间移位的形式, 将 $\tau = \frac{2\pi T'}{P}$ 代入上式, 可得

$$\begin{aligned}
 \tau = \arg \max_{0 \leq \tau' < 2\pi} \sum_{k=1}^M \{ [A_k(n_0)A_k(n_1) + B_k(n_0)B_k(n_1)] \cos(k\tau') + \\
 [B_k(n_0)A_k(n_1) - A_k(n_0)B_k(n_1)] \sin(k\tau') \}
 \end{aligned} \tag{3-28}$$

式 (3-28) 表示的方程被称为对齐准则, 是 CW 对齐过程的基础。

求得最佳移位后, 用三角恒等式展开式 (3-26), 可以得到对齐的 CW 的 DTFS 系数:

$$\left. \begin{aligned} A'_k(n_1) &= A_k(n_1)\cos\left(\frac{2\pi kT}{P}\right) - B_k(n_1)\sin\left(\frac{2\pi kT}{P}\right) \\ B'_k(n_1) &= A_k(n_1)\sin\left(\frac{2\pi kT}{P}\right) + B_k(n_1)\cos\left(\frac{2\pi kT}{P}\right) \end{aligned} \right\}, \quad k=1,2,\dots,M \quad (3-29)$$

若用归一化时间移位 τ 表示对齐的 CW 的 DTFS 系数, 式 (3-29) 变为

$$\left. \begin{aligned} A'_k(n_1) &= A_k(n_1)\cos(k\tau) - B_k(n_1)\sin(k\tau) \\ B'_k(n_1) &= A_k(n_1)\sin(k\tau) + B_k(n_1)\cos(k\tau) \end{aligned} \right\}, \quad k=1,2,\dots,M \quad (3-30)$$

对齐的 CW 表示式为:

$$s(n_1, m-T) = \sum_{k=1}^M [A'_k(n_1)\cos\left(\frac{2\pi km}{P}\right) + B'_k(n_1)\sin\left(\frac{2\pi km}{P}\right)] \quad (3-31)$$

2. 基音加倍/减半时的对齐过程

在自然语音中, 偶尔会发生基音的加倍或是减半现象, 使得提取的 CW 长度可能是实际 PCW 的若干倍或是若干分之一。如果提取的 CW 长度不同, 但是没有发生基音加倍/减半的情况, 则在 CW 的对齐过程中, 基音加倍/减半模块不起作用。当提取的 CW 存在基音加倍/减半的情况时, 为防止对齐出现问题, 较短的 CW 在对齐前必须周期延拓整数倍, 使其最大程度地匹配与较长的 CW。在频域内这样做等价于在较短的 CW 的原始谐波间插入零幅度谐波。

为了检测基音是否加倍/减半, 同样是用公式 (3-17) 中定义的指示因子 C 来判决。当 C 大于 1 时, 则发生了基音加倍/减半。

3. CW 长度不同但是没有基音加倍/减半时的对齐过程

当 CW 长度不同时, 就不能直接使用基于等维假设的对齐准则方程 (3-28) 式。可采用下列两种方法中的一种, 使得两个 CW 具有相同的长度, 再采用对齐准则方程令 CW 对齐。

- (1) 将较长的 CW 谱截断, 使其与短的 CW 具有相同的谐波数;
- (2) 将较短的 CW 谱添零, 使其与长的 CW 具有相同的谐波数。

第一种方法丢弃了高频谐波, 得到了暂时缩短的 CW, 虽然可能损失了一些

暂时细节，但是谱高端的谐波能量相对较低，使得截断的 CW 形状和原始 CW 的形状仍十分近似。第二种方法在频域添零，伸长了原来较短的 CW，使其匹配于较长的 CW，这种时间伸展实际上等价于时域上 CW 的带限内插。内插的本身并未对序列引入任何新的临时信息，却提高了精度。综合比较这两种方法，虽然得到的时间移位 τ 不同，在重建语音信号上却得到了相同的感性质量。而在考虑计算复杂度上，第一种方法占有优势，因为截断的序列具有较少的 DTFS 系数，计算复杂度下降。

3.3.1.7 特征波形的功率计算及归一化

完成 CW 的提取及对齐后，要对 CW 的功率进行归一化。CW 的功率被严格定义为一个 PCW 上每个样点的平均能量，将归一化的 CW 和未归一化的 CW 联系在一起。归一化的目的是将 CW 的功率和形状分离开来，以便对它们分别编码，提高编码效率。

前面已将 CW 转换到 DTFS 域，所以现在 CW 的功率计算和归一化都是在 DTFS 系数上完成。首先要导出 CW 的功率和它的 DTFS 系数间的关系。

定义时间 n 处的 CW 平均功率 $E(n)$ 为

$$E(n) = \frac{1}{P(n)} \sum_{m=0}^{P(n)-1} |s(n, m)|^2 \quad (3-32)$$

式中的 $P(n)$ 是 CW 的长度。结合式 (3-21) 和式 (3-32)，可得到

$$\begin{aligned} E(n) &= \frac{1}{P(n)} \sum_{m=0}^{P(n)-1} s(n, m) s^*(n, m) \\ &= \frac{1}{P(n)} \sum_{m=0}^{P(n)-1} s(n, m) \sum_{k=1}^{\lfloor P(n)/2 \rfloor} [A_k^*(n) \cos\left(\frac{2\pi km}{P(n)}\right) + B_k^*(n) \sin\left(\frac{2\pi km}{P(n)}\right)] \end{aligned} \quad (3-33)$$

由于语音和残差样点值都是实数，所以 $\{A_k(n)\}$ 和 $\{B_k(n)\}$ 一直是实数，有

$$\begin{aligned} A_k(n) &= A_k^*(n) \\ B_k(n) &= B_k^*(n) \end{aligned} \quad (3-34)$$

并且处理的是某一特定时刻的 CW，可将时间标号 n 从表达式中删去以简化表示，这时， $E(n)$ 可简化为

$$E = \frac{1}{P} \sum_{m=0}^{P-1} s(m) \sum_{k=1}^{\lfloor P/2 \rfloor} A_k \cos\left(\frac{2\pi km}{P}\right) + \frac{1}{P} \sum_{m=0}^{P-1} s(m) \sum_{k=1}^{\lfloor P/2 \rfloor} B_k \sin\left(\frac{2\pi km}{P}\right) \quad (3-35)$$

在式 (3-35) 中交换求和顺序, 再运用式 (3-19) 和式 (3-20), 得到

$$E = \begin{cases} \frac{1}{2} \sum_{k=1}^{P/2-1} (A_k^2 + B_k^2) + A_{P/2}^2 + B_{P/2}^2, & P \text{ 为偶数} \\ \frac{1}{2} \sum_{k=1}^{\lfloor P/2 \rfloor} (A_k^2 + B_k^2), & P \text{ 为奇数} \end{cases} \quad (3-36)$$

式 (3-36) 即为从 DTFS 确定 CW 功率的公式, 该式子说明了 CW 的平均功率粗略等于所有谐波成分的能量和。

将式 (3-36) 两边同时除以 E , 得到单位功率如下式:

$$1.0 = \begin{cases} \frac{1}{2E} \sum_{k=1}^{P/2-1} (A_k^2 + B_k^2) + \frac{A_{P/2}^2}{E} + \frac{B_{P/2}^2}{E}, & P \text{ 为偶数} \\ \frac{1}{2E} \sum_{k=1}^{\lfloor P/2 \rfloor} (A_k^2 + B_k^2), & P \text{ 为奇数} \end{cases} \quad (3-37)$$

将 E 合并到每个 DTFS 系数中去, 式 (3-37) 变为

$$1.0 = \begin{cases} \frac{1}{2} \sum_{k=1}^{P/2-1} \left[\left(\frac{A_k}{\sqrt{E}}\right)^2 + \left(\frac{B_k}{\sqrt{E}}\right)^2 \right] + \left(\frac{A_{P/2}}{\sqrt{E}}\right)^2 + \left(\frac{B_{P/2}}{\sqrt{E}}\right)^2, & P \text{ 为偶数} \\ \frac{1}{2} \sum_{k=1}^{\lfloor P/2 \rfloor} \left[\left(\frac{A_k}{\sqrt{E}}\right)^2 + \left(\frac{B_k}{\sqrt{E}}\right)^2 \right], & P \text{ 为奇数} \end{cases} \quad (3-38)$$

由上式可见, 归一化过程仅仅是用每个 DTFS 系数除以平均功率的方根 \sqrt{E} 。

总的说来, CW 的归一化过程可分为两步: 第一步是由式 (3-36) 计算 CW 的平均功率, 第二步是按式 (3-38) 得到归一化的 CW, 即具有单位功率的 CW 表示。

但是也要注意, 在 CW 归一化前或后进行对齐是一样的, 归一化只跟能量有关, 而对齐则只跟相位或是时域移位有关。另外, 当 CW 的功率太小时 (如安静语音段), 归一化过程必须省去, 否则会发生计算溢出。

3.3.1.8 特征波形的分解

CWI 算法利用人耳对浊音和清音感觉不同的特点, 改从感觉满意度的角度去改进低码率语音编码算法。CWI 不采用传统的浊音/清音判决方法, 而是将 CW 分解成为两种成分, 一种是表达准周期成分的慢渐变波形 SEW, 另一种是表示类噪声成分的快渐变波形 REW, 利用人耳对它们的感性差别对它们分别量化编码, 可以提高编码效率。

沿时间轴低通滤波 CW 渐变表面可以得到 SEW, 从 CW 中减去 SEW 可以得到 REW。对于浊音语音, SEW 和 REW 分别表示整形的类脉冲波形和噪声成分, 由于浊音语音段的周期性特征, SEW 的能量通常比 REW 高。对于清音语音, 信号变化十分迅速, 没有表现出明显的周期性, CW 分解后, 大部分能量分布到 REW 上^[34]。

因为 DTFS 运算为线性变换, 所以在时域低通滤波 CW 等价于在频域低通滤波 DTFS 系数。因此对 CW 的二维低通滤波可以直接对 $\{A_k(n)\}$ 和 $\{B_k(n)\}$ 系数进行。设在 CW 分解中采用有 N 个抽头的非因果 FIR 滤波器低通滤波器, 其脉冲响应为 $h_{CW}(m)$, 通过下式可以得到时刻 n 处的低通滤波的 CW:

$$\left. \begin{aligned} \tilde{A}_k(n) &= \sum_{i=-\lfloor N/2 \rfloor}^{\lfloor N/2 \rfloor} A_k(n - iL_{sf})h_{CW}(i) \\ \tilde{B}_k(n) &= \sum_{i=-\lfloor N/2 \rfloor}^{\lfloor N/2 \rfloor} B_k(n - iL_{sf})h_{CW}(i) \end{aligned} \right\}, \quad k = 1, 2, \dots, \left\lfloor \frac{P(n)}{2} \right\rfloor \quad (3-39)$$

上式的 $\tilde{A}_k(n)$ 和 $\tilde{B}_k(n)$ 就是时刻 n 处低通滤波的 CW, 也即是 SEW 的 DTFS 系数。

由于 CW 的维数是随基音周期而变化的, 滤波前可用 CW 长度不同的截断或添零技术来压缩或伸展 CW, 使得滤波窗内的所有 CW 与当前时刻 n 处的 CW 有相同的长度。

3.3.1.9 提取各参数的量化

截止目前, CWI 参数的提取分析已经完成, 跟着进入参数的量化阶段。CWI 待量化的参数有基音周期、LSF 参数、功率、CW(SEW 和 REW), 前两个参数每帧提取一次, 后面两个参数则每子帧提取一次。

对于 LSF 参数，一般采用次优矢量量化算法，常用的量化方案有分裂式矢量量化、多级矢量量化、预测式矢量量化和连接分裂矢量量化等^[35-37]。而最为常用的是分裂式矢量量化，LSF 参数比特流每帧传送一次。

基音周期是 20~120 之间的 101 个可能整数值（在抽样率为 8kHz 时），分配 7 个比特对其量化，每帧也是传送一次。

功率量化前需要额外的处理：输入功率要先转换到对数域，再经低通滤波减样。减样后的功率值分别用非自适应差分标量量化器量化。在接收端，信号功率被解码，在对数域借助域线性内插方法增样到原来的个数，最后经指数运算即可恢复信号功率。功率的编译码框图如下：

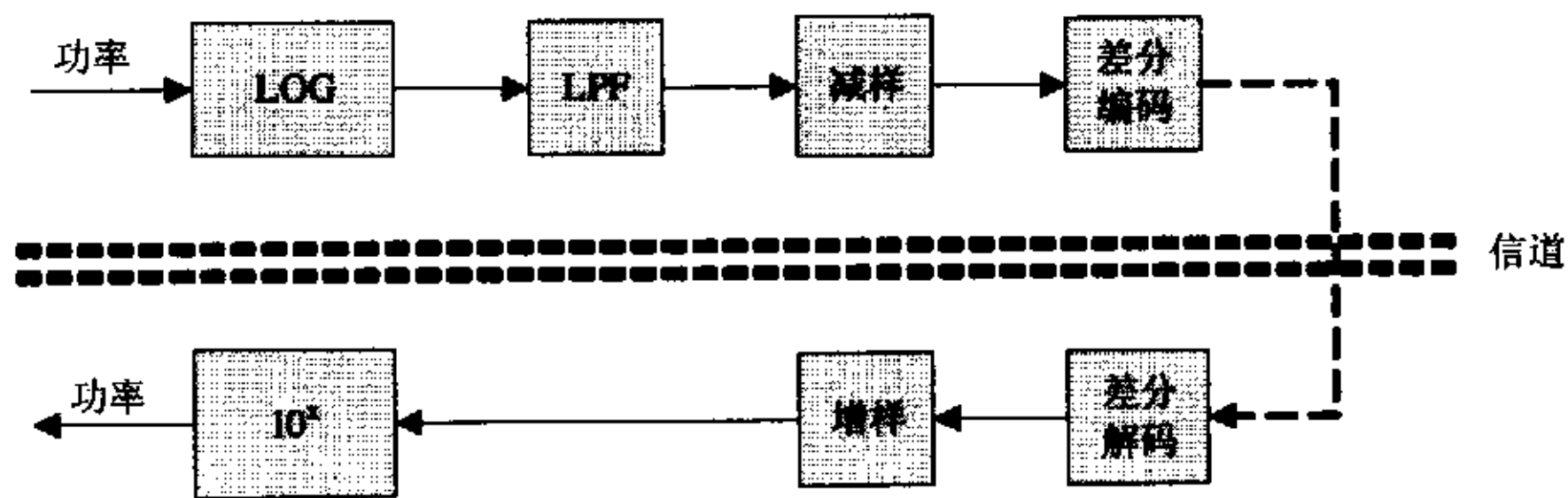


图 3-5 功率编译码流程

Fig.3-5 Quantization and Unquantization of Power

图 3-5 中，低通滤波器的作用是防止在减样过程中发生谱的混迭。

在本小节里，讨论的量化重点放在 REW 和 SEW 的量化上。

1. REW 的量化

REW 的量化过程是以感觉满意度为基础的，可由听力实验得出三个相关的有用结论：

(1) 用随机相位谱取代 REW 的相位谱时，语音质量降低很小；

(2) 用 1000Hz 的平方窗严格平滑每个 REW 的幅度谱时，察觉不出语音质量的降低；

(3) 在 5ms 间隔内取所有 REW 幅度谱的平均值时, 仅能察觉到非常小的听觉失真。

第一个结论表明 REW 的相位谱携带的感性信息不多, 在速率较低时可以不传送它的相位信息。后两个结论则指出 REW 幅度谱的时间分辨率比它的频率分辨率更为重要, 使得 REW 需要较高的更新速率, 但是可对其进行粗糙的量化就足以满足语音质量要求了。

在 REW 的编码端, 首先将输入的 REW 减样; 然后将每个减样的 REW 转换到极坐标形式, 这也意味着 REW 的相位谱被完全丢弃; 最后用一个小码书来量化 REW 的幅度谱。从上面的结论二可知这在 REW 方面来说, 已足够产生好的合成语音质量了。

由于 REW 幅度谱的维数与基音周期相关, 也即是说 REW 谱是变维的, 这就需要变维矢量量化 VDVQ (Variable Dimension VQ) 技术对 REW 谱量化。

2. SEW 的量化

对于非语音信号, SEW 相位是随机的, 波形是平坦的, 而对于语音信号, SEW 相位是从真实信号中抽取出来的典型脉冲相位频谱。在 800Hz 以上的 SEW 幅度, 它在感觉意义上不太重要, 可以近似为 $1 - |\text{REW}|$, 而在 800Hz 以下的, 一个 8 维的码本足以描述基本的频谱行为。这个比特分配与在编码器里采用的量化方案 (开环还是闭环) 无关。

在编码端, 为了补偿分解滤波器的不准确性, 首先将 SEW 减样; 跟着将减样的 SEW 转换到极坐标, 同样使相位谱被完全丢失; 最后将幅度谱分裂成三个不混叠的子带, 各个子带分别量化, 为了适应人耳对较低频率的分辨能力, 低频的子带应该用较多的比特量化, 而高频的子带则可用较少的比特量化。

同样地, SEW 的维数也是随着基音周期的变化而变化, 也需要 VDVQ 量化方法对其量化。首先将 SEW 幅度谱的训练数据变维到 SEW 的最大谐波数, 然后将每个训练矢量分裂为 3 个子带, 最后用 LBG 算法分别设计 3 个子带的码书^[38]。

3. 具有固定维数通用码书的 VDVQ 算法

(1) 变维公式

若当前帧的基音周期为 P , 谱的最大谐波数是 K_{max} , 则定义 K 值为 $K = \lfloor P/2 \rfloor$ 。

令 \mathbf{X} 表示一 K 维矢量, 其中第 i 个元素为 $x(i)$, \mathbf{Z} 表示一 K_{max} 维矢量, 其第 k 个元素为 $z(k)$, 则谱矢量可通过如下公式进行变维处理:

①从 K 维到 K_{max} 维变换:

$$z(k) = \begin{cases} x(i), & k = \lfloor 2 \times K_{max} \times i/P \rfloor, & 0 \leq i \leq K \\ 0.0, & k \text{ 为其他值}, & 0 \leq k \leq K_{max} \end{cases} \quad (3-40)$$

②从 K_{max} 维到 K 维变换:

$$x(i) = z(k), k = \lfloor 2 \times K_{max} \times i/P \rfloor, \quad 0 \leq i \leq K, 0 \leq k \leq K_{max} \quad (3-41)$$

(2) VDVQ 训练算法:

给定谱矢量的训练数据和体积为 N 、维数为 K_{max} 的初始码书, 以 LBG 算法迭代设计谱矢量码书。每次的训练均有两个关键迭代:

①用最近相邻准则 (NNR) 围绕码矢量群集训练矢量。依据式 (3-40) 将每个 K 维谱矢量变维到 K_{max} 维矢量, 在平方误差准则下, 如果满足

$$d(\mathbf{Z}_i, \mathbf{Y}_m) \leq d(\mathbf{Z}_i, \mathbf{Y}_j), \quad j = 1, 2, \dots, N \quad (3-42)$$

则分配 \mathbf{Z}_i 到群 C_m 中去, 这里的 \mathbf{Y}_j 是谱矢量码书中的第 j 个码字。

②依据均方失真测度将群聚质心替代旧的码字。对于每个群聚 $C_m, m = 1, 2, \dots, N$, 可求得一新的码矢量 \mathbf{Y}'_m , 质心由下式给出:

$$\mathbf{Y}'_m = \frac{1}{M_m} \sum_{\mathbf{Z} \in C_m} \mathbf{Z} \quad (3-43)$$

这里的 M_m 表示分配到群 C_m 中码矢量的个数。训练结束, 将最后的群聚质心作为谱矢量的码书矢量, 由此而得到的 K_{max} 维码书, 称之为通用码书。

(3) VDVQ 编码算法

一个维数为 K_{max} 、体积为 N 的通用码书在前已经确立, 其第 j 个码矢量表示

为 Y_j ，VDVQ 接收的输入矢量为 \mathbf{X} ，维数为 K ，编码过程如下：

根据式 (3-41)，将 K_{max} 维通用码书中的码矢量 $Y_j, j=1,2,\dots,N$ 转换为 K 维矢量 $\tilde{Y}_j, j=1,2,\dots,N$ ，使下式最小的作为要编码的码字序号：

$$E(j) = \frac{1}{K+1} \sum_{k=0}^K [x(k) - \tilde{y}_j(k)]^2, \quad 1 \leq j \leq N \quad (3-44)$$

式中 $x(k)$ 和 $\tilde{y}_j(k)$ 分别是矢量 \mathbf{X} 和矢量 \tilde{Y}_j 的第 k 个元素。如果谱矢量的通用码书体积为 N ，用 $\log_2 N$ 比特可完成对它的量化。

在接收端根据收到的基音译码值和通用码书的码字序号，利用式 (3-41)，将码字序号对应的 K_{max} 维码矢量转换为 K 维矢量，即可得到 K 维谱矢量，从而完成译码^[39]。

VDVQ 的实施原理图如下：

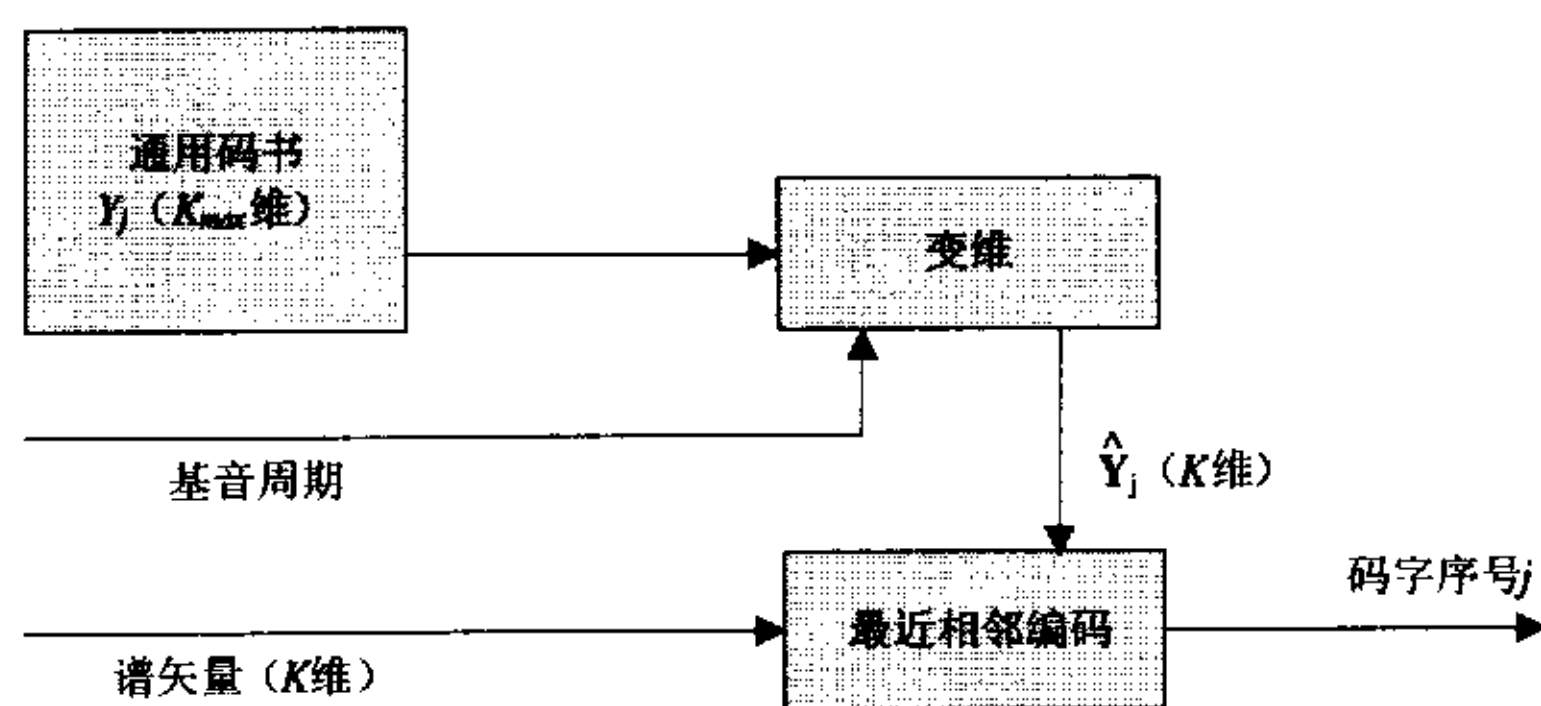


图 3-6 VDVQ 实施原理

Fig.3-6 VDVQ Execution Theory

至此，已完成 CWI 算法编码模块的分析，将在下一章节进行 CWI 算法解码合成模块的分析。相对于编码模块而言，解码合成模块要相对地简单。

3.3.2 解码合成模块分析

在合成模块里，运用 LSF 参数、基音周期、功率和归一化的 CW 可以重建语音。解码端接收到的是各个参数量化值，先经译码层反量化，再传送到合成模块。

正如在编码器里一样，合成模块中各部分的执行频率是不同的，有的每帧运行一次，有的则每子帧运行一次甚至更多。在 CWI 合成模块中的基音内插子模块和 LSF 内插子模块与 CWI 编码分析模块中的相同。

3.3.2.1 译码层

对 LSF 量化序号进行译码得到量化的 LSF，对得到的 LSF 进行内插，将通过内插得到的 LSF 转换成 LP 系数，得到 LP 滤波器。对基音周期译码，得到量化的基音周期。通过功率译码，得到量化的平均功率。

在 REW 的解码端，REW 的幅度谱被首先解码；然后被增样到编码端被减样前的频率，相当于在接收的谱中插入新的谱点，新的谱可通过相邻谱的线性内插或选择以前接收的谱来确定；最后，每个增样的 REW 幅度谱与随机相位谱组合返回到笛卡儿坐标系。值得注意的是，随机相位谱是在 $[0\sim 2\pi]$ 间均匀分布的独立随机变量值，并且随机相位谱是基于子帧的基础上组合到 REW 幅度谱上的。加入随机相位的目的是排除相继 REW 间的相关性，因为 REW 间过强的相关将导致重建语音带有蜂鸣声，尤其在清音语音段。同理，在 SEW 译码端，第一步是对 SEW 参数译码并组合各子带。为免谱的突发变化和明显的不连续将产生回声的重建语音，在各个子带的分界处使用线性内插技术平滑组合的谱。第二步对组合好的幅度谱附上固定的相位谱，转换到笛卡儿坐标。这里使用的固定相位谱来自于高基音周期的男性讲话者的浊音语音段，因为男性语音比女性语音提供了更多的基音谐波。最后一步，将 SEW 增样到和重建的 REW 一样的提取率。由于各个 SEW 的 DTFS 系数个数不同，SEW 的维数可能不同，可以采用合成小节中 CW 不等维情况下内插的方法来增样 SEW。SEW 和 REW 经过增样后，将两者简单相加即可得到重建的 CW。

3.3.2.2 过渡层

如 CWI 解码框图所示，CW 重建后，首先要对每个重建的 CW 进行功率去归一化运算。这个过程需要用到内插的基音周期以确定 CW 的长度。对式 (3-38) 进行逆运算，即用 \sqrt{E} 乘以每个 DTFS 系数来恢复 CW 功率。由于量化的参数有可能使得相继的 CW 不再保持准确的对齐，被恢复功率的 CW 必须进行重新对齐，对齐的过程仍然遵从对齐准则方程式 (3-28)。

3.3.2.3 合成层

在这里着重研究瞬时基音周期和 CW 的内插原理。在此之前，每个提取点处

重建的、对齐的 CW 及其基音周期长度经已确定，要重建一维残差信号，需要知道每个样点处的 CW 及其基音长度。增样 CW 可以选择简单的线性内插方法进行，当在两个等维的 CW 间完成增样时，可直接进行内插，但是当 CW 的长度不同或是基音长度发生加倍/减半情况时，就需要额外的处理以保证平滑内插。要注意的是，在 CW 内插时使用的瞬时基音的内插方法不同于编码分析端处的基音内插方法，不再使用方程式 (3-17)，而改用式 (3-16)。由这个内插产生的瞬时基音值总是对应于瞬时 CW 的长度。

瞬时基音及 CW 内插子模块的原理框图如下：

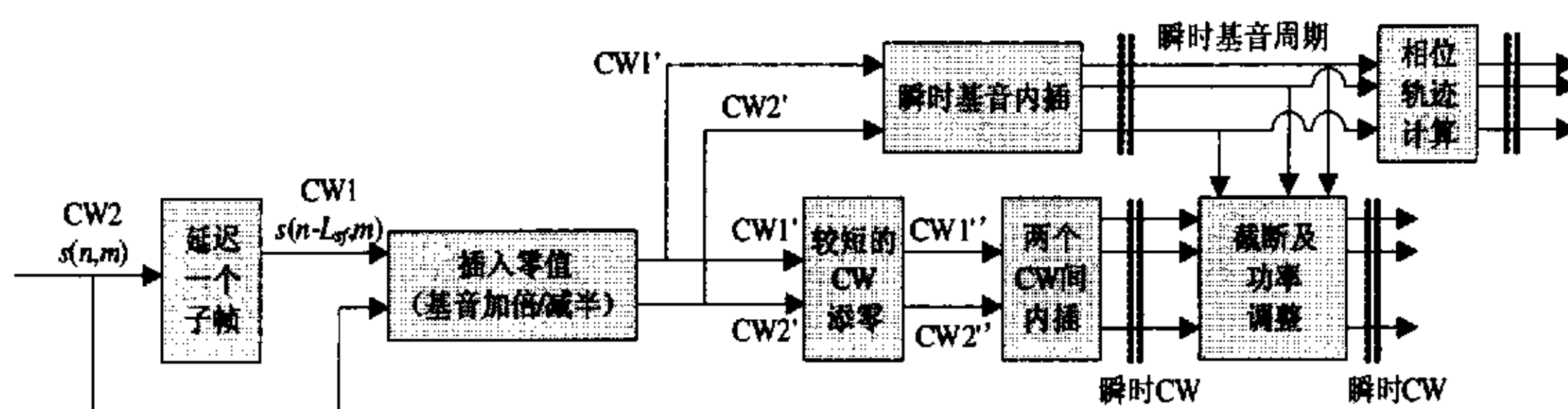


图 3-7 CW 内插过程

Fig.3-7 CW Interpolation Process

由上图可以看出，在处理瞬时基音和 CW 内插时会有三种不同的情况出现，分别讨论如下：

1. 等维情况下 CW 的内插

已知两个 CW 的长度相同，均为 P 。那么内插框图中的插入零值子模块、较短的 CW 添零子模块和截断及功率调整子模块都不执行。分别用 n_0 、 n_1 表示内插区间分解处的时刻，则通过 $s(n_0, m)$ 和 $s(n_1, m)$ 之间的内插，可以确定时刻 n 处的瞬时 CW— $s(n, m)$ 。时域中的内插表示为：

$$s(n, m) = \left(\frac{n_1 - n}{n_1 - n_0} \right) s(n_0, m) + \left(\frac{n - n_0}{n_1 - n_0} \right) s(n_1, m), \quad n_0 \leq n \leq n_1, 0 \leq m \leq P \quad (3-45)$$

将式 (3-21) 代入式 (3-45)，得到

$$\left. \begin{aligned} A_k(n) &= \left(\frac{n_1 - n}{n_1 - n_0} \right) A_k(n_0) + \left(\frac{n - n_0}{n_1 - n_0} \right) A_k(n_1) \\ B_k(n) &= \left(\frac{n_1 - n}{n_1 - n_0} \right) B_k(n_0) + \left(\frac{n - n_0}{n_1 - n_0} \right) B_k(n_1) \end{aligned} \right\}, \quad k = 1, 2, \dots, \lfloor P/2 \rfloor \quad (3-46)$$

上式表明：时域内两个 CW 的线性内插等价于它们 DTFS 系数的线性内插，内插是基于子帧的基础。由于边界点处的 CW 具有相同的长度，所以它们之间内插得到的 CW 也具有相同的长度，基音轮廓为常数。

2. 不等维情况下 CW 的内插

当边界点处的 CW 长度不同时（即系数 $\{A_k, B_k\}$ 的个数不同），基音周期在边界点之间的区间上是变动的。这样的话，CW 内插前要将短的 CW 长度调整到与长的 CW 相同，等价于对短的 CW 的 DTFS 系数添零。完成添零工作后，可用等维情况下的 CW 内插公式获得瞬时 CW。此时，内插过程框图中的插入零值子模块不工作。

在这过程中的瞬时基音是通过式 (3-16) 进行内插得到的，由于添零的原因，得到的基音周期值可能和内插得到的瞬时 CW 长度不同。为了避免这种不一致性，可将内插的 CW 长度截断到和内插的瞬时基音周期长度一致。DTFS 系数的截断可能导致 CW 功率的下降，为了保持截断前后的 CW 具有同样的功率，可用一个适当的因子标定 DTFS 系数。这些功能由截断及功率调整子模块完成。

3. 基音加倍/减半情况下 CW 的内插

若当前的 CW 比原来的 CW 长或是短很多，则很可能发生了基音加倍/减半的情况，这时插入零值子模块起作用。类似于 CW 的对齐过程，使用如式 (3-17) 中定义的同样的指示器 C 作为基音加倍/减半的判决准则。如果在子帧间基音确实发生了加倍/减半 ($C > 1$)，则将较短的 CW 重复整数倍，使其最大可能地与较长的 CW 匹配，相当于在较短 CW 原来的谐波间插入零幅度谐波。在较短的 CW 中插入零幅度谐波后，再进行不等维情况下的 CW 内插即可。

由于插入的零将保留在内插好的瞬时 CW 中，所以瞬时基音周期的内插在 CW 插入零之后进行。另外，为了增强相位轨迹的准确性，进入相位轨迹计算子模块时的瞬时基音周期不再取为整数。

需要注意的是，在帧分界处的 CW 内插是在前一帧的最后一个 CW 和当前帧

的第一个 CW 间进行的。

在相位轨迹计算子模块里，因为已经知道每个样点处的瞬时基音周期 $P(n)$ ，频率轨迹曲线 $F(n)$ 和 $P(n)$ 之间的关系为：

$$P(n) = \frac{1}{F(n)} \quad (3-47)$$

用 $\varphi(\bullet)$ 表示相位轨迹，则每个样点处的相位轮廓的逐点更新公式为：

$$\varphi(n) = \varphi(n-1) + \int_{n-1}^n \frac{2\pi}{P(n')} dn' \quad (3-48)$$

在上式， $\varphi(n)$ 和 $\varphi(n-1)$ 为当前的和原来的相位值；积分对应于变动相位在 $n-1$ 和 n 之间的增量区域。由瞬时基音周期的内插公式—公式 (3-16) 可知，基音周期在此积分区间线性变化，式 (3-48) 可变为：

$$\varphi(n) = \varphi(n-1) + \int_{n-1}^n \frac{2\pi}{(n-n')P(n-1) + (n'-n+1)P(n)} dn' \quad (3-49)$$

可直接估计该积分为：

$$\varphi(n) = \begin{cases} \varphi(n-1) + \frac{2\pi}{P(n) - P(n-1)} \ln \left[\frac{P(n)}{P(n-1)} \right], & P(n) \neq P(n-1) \\ \varphi(n-1) + \frac{2\pi}{P(n)}, & P(n) = P(n-1) \end{cases} \quad (3-50)$$

逐点执行式 (3-50) 即可将基音轨迹 $P(\bullet)$ 转换为相位轨迹 $\varphi(\bullet)$ 。但是要注意，当 $\varphi(n)$ 的值增大到比 2π 大时，必须从 $\varphi(n)$ 值中减去 2π ，满足 $0 \leq \varphi(\bullet) < 2\pi$ 。

在实时计算中，为了减少计算复杂性以减少时延，式 (3-48) 中的积分项可用下式近似

$$\frac{1}{2} \left(\frac{1}{P(n-1)} + \frac{1}{P(n)} \right) \quad (3-51)$$

上式取的是两个相继频率值的平均，则式 (3-48) 可简化为

$$\varphi(n) \approx \varphi(n-1) + \pi \left(\frac{1}{P(n-1)} + \frac{1}{P(n)} \right) \quad (3-52)$$

并且，在语音的开始处， $\varphi(0)$ 可取任意数，初始相位的偏移并不会影响重建语音的感性质量。一般取 $\varphi(0) = 0$ 。

一旦知道了每个样点的 CW 及其相位轨迹，则根据式 (3-23) 可实现已内插的 CW 的 DTFS 逆过程。得到一维残差信号 $r(n)$ 的表示式为：

$$r(n) = s(n, \varphi(n)) = \sum_{k=1}^{\lfloor P(n)/2 \rfloor} [A_k(n) \cos(k\varphi(n)) + B_k(n) \sin(k\varphi(n))] \quad 0 \leq \varphi(\bullet) < 2\pi \quad (3-53)$$

用重建的残差信号 $r(n)$ 去激励 LP 合成滤波器即可得到所求的合成语音。LP 合成滤波器的系统函数为 $H(z)$ ：

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p \alpha_i z^{-i}}$$

3.4 CWI 编码器性能分析

一般情况下，CWI 的合成语音与原来的语音不再保持同步，这主要是由基音估计的不准确性和波形内插过程中的瞬时基音和 CW 内插引起。正是因为 CWI 合成语音和原始语音的不同步使得 SNR 方法不再适用于测量编码器的性能，而改用主观 MOS 方法来评价 CWI 编码器的性能。

表 3-1 2.4kb/s CWI 编码器比特分配^[40]

Table3-1 2.4kb/s CWI Coder's Bit Allocation

参数	LSF	基音周期	功率	SEW	REW
比特分配	30	7	4	7	1 或 3
修正速率	40	40	80	40	160
比特率	1200	280	320	280	320

Kleijn 博士给出的是一种 2.4kb/s 的 CWI 语音编码方案。该编码方案的分析帧长为 25ms, 每帧提取 12 个 CW, 低于 800Hz 的 SEW 幅度谱用 7 比特矢量量化, 而高于 800Hz 的 SEW 幅度谱从 REW 幅度谱中推出。为了增加编码效率, 减样的 4 个 REW 每隔一个传送一次, 没传的 REW 用下一个或是原来的 REW 代替, 那么整个 REW 的编码速率为 $(1 \times 160) / 2 + (3 \times 160) / 2$ 比特。编码器的比特分配表如表 3-1 所示。

3.5 PWI 算法与 CWI 算法的比较

CWI 算法与 PWI 算法相比, CWI 算法无需清浊判决, 可以同时的清音信号和浊音信号用同一种算法进行编码, 避免了由不同编码模式转换而引起的失真。但是它的采样速率要求比 PWI 高很多, 而且在编解码端仍旧需要高复杂的对齐过程, 实时实现仍有困难。

本文将在下一章中提出一种基于 PWI 算法的改进方案, 在验证算法原理有效性的同时, 提高合成语音信号的质量和自然度, 降低编码器的计算复杂度。

3.6 本章小结

本章主要详述了特征波形内插算法的基本思想和实现过程, 对 CWI 算法在 2.4kb/s 数码率上构造的编码器方案的性能进行分析。在此基础上, 对两种主要的波形内插算法进行了比较, 引出了本文后面所要提出的改进 WI 算法。

第四章 改进的 WI 算法

4.1 算法改进的前提

我们在对 PWI 算法的 C 代码仿真的前提下，结合新的语音编码技术，提出了一种改进的 WI 编码算法流程，希望能在仿真实实现的过程中，证明这个改进算法对 WI 算法的发展有改良的贡献。

4.2 算法的基本原理框图

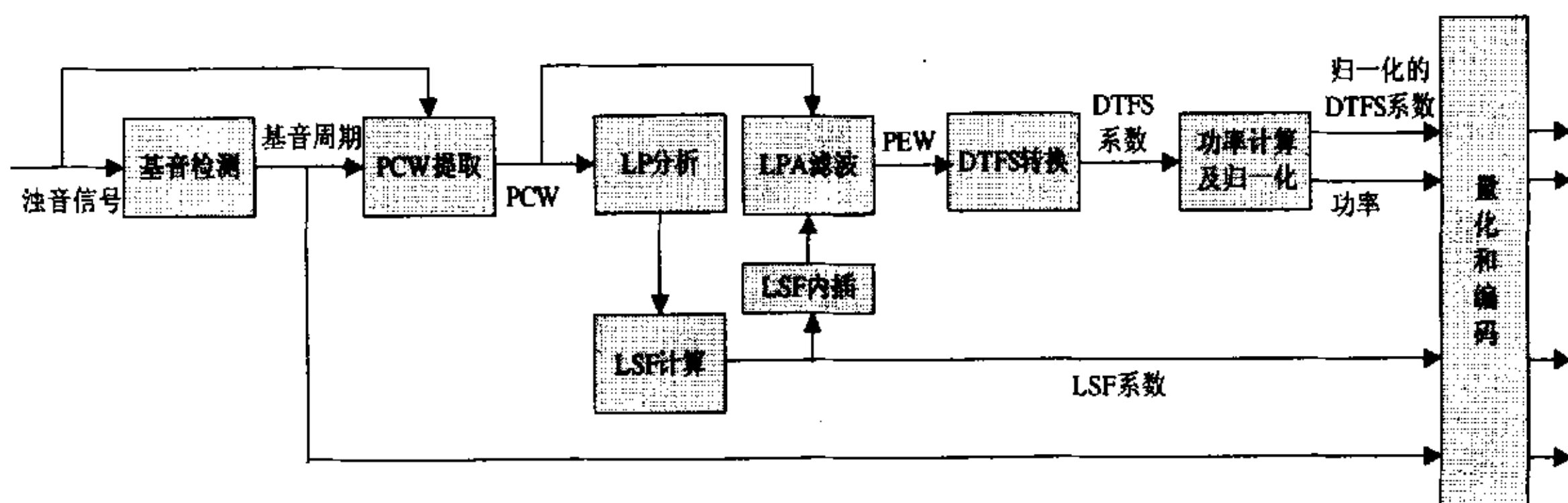


图 4-1 改进算法的分析流程

Fig.4-1 Modified Algorithm's Analytical Process

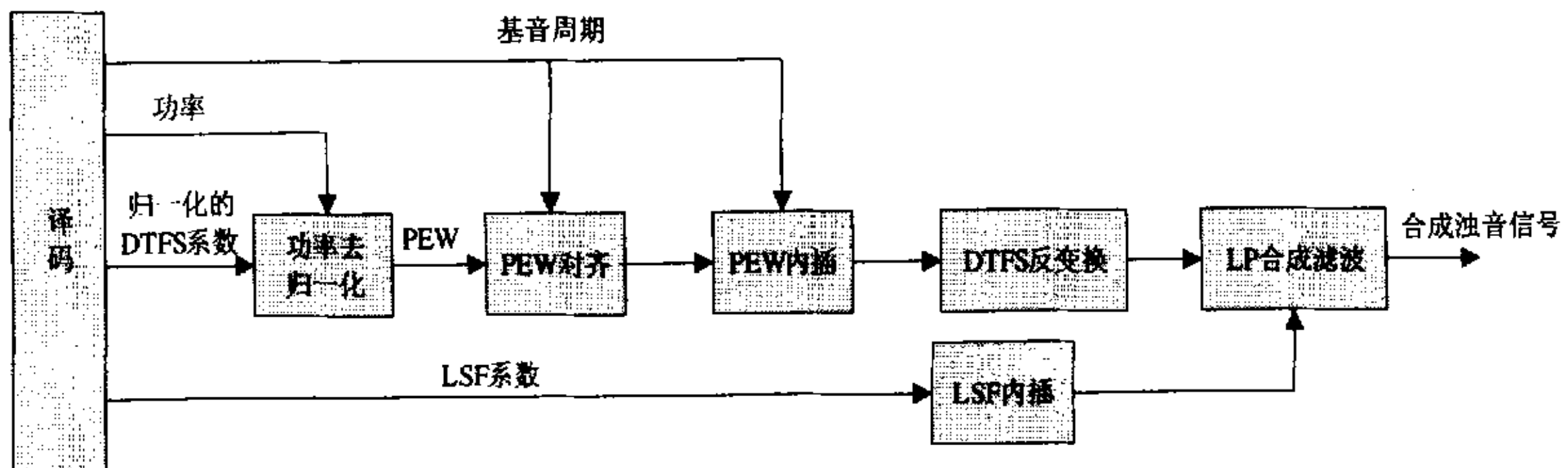


图 4-2 改进算法的合成流程

Fig.4-2 Modified Algorithm's Synthetic Process

4.3 算法实现流程

4.3.1 分析端

分析端使用的是 16kHz 抽样率的原始信号，分析帧长取 20ms，则每帧共有 320 个样点。为方便后面分析所用，这里定义帧长为 L_{sf} ，是样点意义上的长度，故有 $L_{sf} = 320$ 。

4.3.1.1 清浊判决及基音初步搜索

在本算法里，采用最大归一化自相关函数进行清浊判决，同时得到初步估计的基音周期。定义语音信号为 $u(n), n = 0, 1, \dots, L_{sf} - 1$ ，它的归一化自相关函数为

$$\rho(\tau) = \frac{\sum_{n=0}^{L_{sf}-1} u(n)u(n-\tau)}{\sum_{n=0}^{L_{sf}-1} u^2(n)} \quad (4-1)$$

在 $\tau \in [T_{\min}, T_{\max}]$ 范围内搜索 $\rho(\tau)$ 的最大值 ρ_{\max} 。其中 T_{\min} 和 T_{\max} 分别为浊音信号基音周期的可能最小和最大整数值。

设 uv_th 为判断清浊音的门限，若 $\rho_{\max} > uv_th$ ，则判断当前语音帧是浊音，此时对应于 ρ_{\max} 的 τ 值为当前语音帧的初步估计基音周期 P_0 ；若 $\rho_{\max} < uv_th$ 则判断当前语音帧是清音。

4.3.1.2 基音精细检测

目前的研究表明，当增加 CELP 算法中浊音语音的周期性时，尽管 SNR 下降，但感性质量增加了。在单脉冲激励编码方法中也可得到类似的结论。这些结果都表明，保持语音信号周期性的正确程度对感性质量非常重要。

基于要获得更高自然度的合成语音的考虑，就必须获得更接近原信号真实基音周期的基音周期估计值，所以有必要对已获得的基音周期再进行精细检测来提高基音周期的精度。

由于在浊音帧，基音粗估算法可以保证粗估基音偏离真正基音不超过 2~3 个点，所以我设计一个基音精细搜索算法来获得更为精确的基音周期。

设 $\varepsilon(P)$ 为基音周期 P 的函数, 表示基音周期 P 对应的预测合成信号 $u'(n)$ 与原始信号 $u(n)$ 之间的误差:

$$\varepsilon(P) = \sum_{n=0}^{2P-1} [u(n) - u'(n)]^2$$

P_i 为第 i 步搜索中 $\varepsilon(P)$ 最小值对应的基音值; Δ_i 为第 i 步的搜索步长且定义 $\Delta_1 = 1$, $i = 1, 2, \dots$ 。具体的基音精细搜索算法^[41]如下:

1. 计算 $P_0 - 3, P_0 - 2, P_0 - 1, P_0, P_0 + 1, P_0 + 2, P_0 + 3$ 共七个整数基音点的 $\varepsilon(P)$, 得到七点中 $\varepsilon(P)$ 最小值对应的基音值为 P_1 。
2. 令 $\Delta_i = \Delta_{i-1} / 2$, 计算 $P_{i-1} - \Delta_i, P_{i-1} + \Delta_i, i = 2, 3, \dots$ 的 $\varepsilon(P)$, 取已知点中 $\varepsilon(P)$ 最小值对应的基音值为 P_i 。
3. 重复步骤 2, 直至 Δ_i 等于所要求的精度为止, 则取最后的得到的 P_i 为精细搜索所求的基音周期。

4.3.1.3 PW 的提取

原型波形在原始的浊音帧上提取。以基音精细检测得到的精确的基音周期为长度, 每帧 16kHz 抽样的浊音语音上, 可提取出一个基音周期波形 PCW。提取方法仍旧采用原先 PWI 算法的最大预测增益准则及其实现方法。

首先, 对 PCW 定义一个起点, 一般在过零处或其附近, 可以使信号能量的变化对准则的影响最小。然后, 搜索长度近似等于基音周期的那段区间, 通过计算每个候选区间的带限离散傅立叶级数来计算最大预测增益。令 J 为以采样周期为单位的基音周期 (一般 J 不为整数), 对于一个带限为一半采样率的系统, 有 M 个谐波, 这里 $M \leq J/2 \leq M + 1$ 。用 $2M$ 个参数表示瞬时语音波形的离散傅立叶级数:

$$u(m, n) = \sum_{k=1}^M [C_k(n) \cos(\frac{2\pi km}{P(n)}) + D_k(n) \sin(\frac{2\pi km}{P(n)})] \quad (4-2)$$

如果长度为 J 的试探区间中心位于已有的样点处, 那么傅立叶级数能够准确地拟合语音样点。求得 M 个谐波的傅立叶级数后, 对所要求的延迟, 计算自相关函数:

$$r(m) = \frac{1}{2} \sum_{k=1}^M [C_k^2 + D_k^2] \cos\left(\frac{2\pi km}{J}\right) \quad (4-3)$$

从上式得到一系列的自相关值后，得到长度为 J 的区间上产生的周期信号的线性预测增益 $G(J)$ 为：

$$G(J) = \frac{r(0)}{r(0) - \mathbf{a}^T \mathbf{R} \mathbf{a}} \quad (4-4)$$

这里的 $\mathbf{a} = [\alpha_1, \alpha_2, \dots, \alpha_p]^T$ 为短时预测器系数，可用自相关法对以原采样率的语音信号计算得到， p 为预测器阶数； \mathbf{R} 为自相关法矩阵，是对称的 Toeplitz 矩阵，它的第一行为 $r(0), r(T), r(2T), \dots, r[(p-1)T]$ ， T 为采样周期。

在 M 值的适当范围内重复式 (4-4)，具有最大预测增益的区间选择为所要提取的 PCW。

4.3.1.4 PCW 的 LP 分析

为了降低数码率，提高编码效能，不是直接传送 PCW 的参数，而是传送由 PCW 通过 LP 分析滤波的残差信号。PCW 首先进行 LP 分析，得到 LP 系数 $\{\alpha_i\}_{i=1,2,\dots,p}$ ， p 为线性预测的阶数。我们将得到的 LP 系数计算转换成线谱频率 LSF 参数。两个相邻帧的 LSF 要线性内插，再将每组内插的 LSF 参数转换回 LPA 参数，用于 LPA 滤波器，LPA 滤波器的系统函数为 $A(z)$ ， $A(z) = 1 - \sum_{i=1}^p \alpha_i z^{-i}$ 。当 PCW 通过 LP 分析滤波器后，即可得到当前帧的 LP 残差信号，沿用 PWI 算法中的称呼，记之为 PEW—原型激励波形 $s(m)$ 。

4.3.1.5 PEW 的 DTFS 变换

得到的一个 PEW 波形可用下式离散傅立叶级数表示：

$$s(m) = \sum_{k=1}^{\lfloor P/2 \rfloor} \left[A_k \cos\left(\frac{2\pi km}{P}\right) + B_k \sin\left(\frac{2\pi km}{P}\right) \right] \quad 0 \leq m < P \quad (4-5)$$

上式， P 为基音周期， $\{A_k\}$ 和 $\{B_k\}$ 是离散傅立叶级数系数。

由傅立叶级数展开原理可得到当前帧的 PEW 的傅立叶级数系数分别为 $\{A_k\}$ 、 $\{B_k\}$ 。当 P 为偶数时， $\{A_k\}$ 和 $\{B_k\}$ 计算公式如下：

$$\left. \begin{aligned} A_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\ B_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right] \end{aligned} \right\}, \quad k = 1, 2, \dots, \frac{P}{2} - 1 \\
 \left. \begin{aligned} A_k &= \frac{1}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\ B_k &= \frac{1}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right] \end{aligned} \right\}, \quad k = \frac{P}{2}$$
(4-6)

当 P 为奇数时, $\{A_k\}$ 和 $\{B_k\}$ 计算公式如下:

$$\left. \begin{aligned} A_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \cos\left(\frac{2\pi km}{P}\right) \right] \\ B_k &= \frac{2}{P} \sum_{m=0}^{P-1} \left[s(m) \sin\left(\frac{2\pi km}{P}\right) \right] \end{aligned} \right\}, \quad k = 1, 2, \dots, \frac{P-1}{2}$$
(4-7)

4.3.1.6 功率计算及归一化

已知 PEW 的傅立叶级数系数 $\{A_k\}$ 和 $\{B_k\}$, 可计算其功率 g_0 :

$$g_0 = \begin{cases} \frac{1}{2} \sum_{k=1}^{\lfloor P/2-1 \rfloor} (A_k^2 + B_k^2) + A_{P/2}^2 + B_{P/2}^2, & P \text{ 为偶数} \\ \frac{1}{2} \sum_{k=1}^{\lfloor P/2 \rfloor} (A_k^2 + B_k^2), & P \text{ 为奇数} \end{cases}$$
(4-8)

对该帧信号的 $\{A_k\}$ 和 $\{B_k\}$ 分别进行功率归一化, 即用每个 DTFS 系数除以功率的方根 $\sqrt{g_0}$:

$$\begin{cases} A'_k = \frac{A_k}{\sqrt{g_0}} \\ B'_k = \frac{B_k}{\sqrt{g_0}} \end{cases} \quad k = 1, 2, \dots, \lfloor P/2 \rfloor$$
(4-9)

4.3.2 合成端

在合成端可得到 PEW 的功率、LSF 系数、基音周期和归一化的 PEW 傅立叶级数系数 $\{A'_k\}$ 和 $\{B'_k\}$ 。

4.3.2.1 功率去归一化

功率去归一化是功率归一化的逆过程,即每个归一化的 DTFS 系数乘以功率的方根 $\sqrt{g_0}$ 便可恢复原来的 PEW 的 DTFS 系数:

$$\begin{cases} A_k = \sqrt{g_0} A'_k \\ B_k = \sqrt{g_0} B'_k \end{cases} \quad k = 1, 2, \dots, \lfloor P/2 \rfloor \quad (4-10)$$

4.3.2.2 PEW 对齐

一般地,我们译码后得到的 PEW 是不同相的,表明 PEW 的主特征在时间上并没有对齐。为了通过内插获得 PEW 波形的准确重建,在进行 PEW 内插前,必须要先对当前帧的 PEW 与前一帧的 PEW 进行对齐。

为方便对齐,先调整前后两帧 PEW 波形的谐波数。定义前一帧和当前帧的基音周期分别是 $P(n_1)$ 和 $P(n_2)$, 基音加倍/减半因子 C , 有

$$C = \left\lfloor \frac{P(n_2)}{P(n_1)} \right\rfloor \quad \text{当 } P(n_1) < P(n_2) \text{ 时}; \quad C = \left\lceil \frac{P(n_1)}{P(n_2)} \right\rceil \quad \text{当 } P(n_1) > P(n_2) \text{ 时}.$$

因子 C 作为基音是否加倍或减半的指示器,当 $C=1$ 时,说明基音没有加倍/减半发生;当 C 大于 1 时就意味着估计的基音有加倍/减半情况发生。当基音发生加倍或减半情况,即 $C = \theta$ (θ 为大于 1 的整数)时,将较短的 PEW 波形重复 $(\theta-1)$ 倍,使其与较长的 PEW 波形的长度最佳匹配。当没有发生基音加倍或减半情况时,往较短的 PEW 波形中加入零幅度谐波,使得前后两帧的 PEW 波形长度相同,此时有

$$P = P(n_1) = P(n_2)$$

$$\text{令} \quad M = \left\lfloor \frac{P(n_1)}{2} \right\rfloor = \left\lfloor \frac{P(n_2)}{2} \right\rfloor \quad (4-11)$$

对齐时,以前一帧 PEW 的相位为对齐基准,以帧为对齐基础。设当前 PEW 波形为 $s(n_2, m)$, 将其时域移位 T 个样点,即将当前的 PEW 波形循环时间移位,使得 $s(n_2, m)$ 和前一帧的 PEW $s(n_1, m)$ 间的类似性测度最大。因为 DTFS 的线性性质允许我们将信号的循环时间移位等价于对 DTFS 系数加一个线性相位。

通过使下式最大，可求得所需的时间移位 T ：

$$\begin{aligned}
 T &= \arg \max_{0 \leq T' < P} \Xi[s(n_1, m), s(n_2, m - T')] \\
 &= \arg \max_{0 \leq T' < P} \sum_{k=1}^M \{ [A_k(n_1)A_k(n_2) + B_k(n_1)B_k(n_2)] \cos\left(\frac{2\pi k T'}{P}\right) + \\
 &\quad [B_k(n_1)A_k(n_2) - A_k(n_1)B_k(n_2)] \sin\left(\frac{2\pi k T'}{P}\right) \} \quad (4-12)
 \end{aligned}$$

则在时刻 n_2 ，对齐的 PEW 的傅立叶级数系数 $\{\tilde{A}_k\}$ 和 $\{\tilde{B}_k\}$ 分别是

$$\left. \begin{aligned}
 \tilde{A}_k(n_2) &= A_k(n_2) \cos\left(\frac{2\pi k T}{P}\right) - B_k(n_2) \sin\left(\frac{2\pi k T}{P}\right) \\
 \tilde{B}_k(n_2) &= A_k(n_2) \sin\left(\frac{2\pi k T}{P}\right) + B_k(n_2) \cos\left(\frac{2\pi k T}{P}\right)
 \end{aligned} \right\} \quad (4-13)$$

对齐后的 PEW 可表示为下式：

$$s(n_2, m - T) = \sum_{k=1}^{\lfloor P(n_2)/2 \rfloor} \left[\tilde{A}_k(n_2) \cos\left(\frac{2\pi k m}{P(n_2)}\right) + \tilde{B}_k(n_2) \sin\left(\frac{2\pi k m}{P(n_2)}\right) \right] \quad (4-14)$$

从现在开始，所有的傅立叶系数都假定是对齐的，符号“ \sim ”将被略去。

4.3.2.3 PEW 内插

在前面的 PWI 算法中，使用了一个内插函数 $\alpha(t)$ 对 PEW 进行内插。考虑到 DTFS 的线性性质，我将内插函数简化为简单地按照一帧中所含的周期波形的个数进行线性内插。已知帧长为 L_f ，且 $L_f = 320$ 。而基音周期的长度，考虑到相继的原型激励波形 PEW 通常是非常类似的，可根据当前帧长与前一帧长的周期长度来估计。设前一帧的基音周期长度为 $P(n_1)$ ，当前帧的基音周期长度为 $P(n_2)$ ，前一帧所含的 PEW 个数为 N 。则有

$$\left. \begin{aligned}
 \text{当 } P(n_1) > P(n_2) \quad N &= \left\lfloor \frac{L_f}{P(n_1)} \right\rfloor \\
 \text{当 } P(n_1) < P(n_2) \quad N &= \left\lfloor \frac{L_f}{P(n_1)} \right\rfloor
 \end{aligned} \right\} \quad (4-15)$$

DTFS 系数内插公式为

$$\left. \begin{aligned} A_k(n) &= \left(\frac{N-n}{N}\right)A_k(n_1) + \left(\frac{n}{N}\right)A_k(n_2) \\ B_k(n) &= \left(\frac{N-n}{N}\right)B_k(n_1) + \left(\frac{n}{N}\right)B_k(n_2) \end{aligned} \right\} n = 0, 1, \dots, N-1; k = 1, \dots, \left\lfloor \frac{P(n)}{2} \right\rfloor \quad (4-16)$$

同理，基音周期的内插公式也为

$$P(n) = \left(\frac{N-n}{N}\right)P(n_1) + \left(\frac{n}{N}\right)P(n_2), \quad n = 0, 1, 2, \dots, N-1 \quad (4-17)$$

考虑到在内插时，由于每个内插的基音周期都在变动，使得 $\{A_k\}$ 和 $\{B_k\}$ 的维数也在变动，所以我在内插时判断，正在进行第 n 个周期波形内插的 $\{A_1, B_1\}$ 和 $\{A_2, B_2\}$ ，其计算所取的内插维数由当前被内插的 $\{A_k(n), B_k(n)\}_{n=0,1,\dots,N}$ 的维数决定，也就是由已内插得到的基音周期 $P(n)$ 决定，不足的维数部分视其值为零代入内插公式中，多余的维数部分则被舍去。

4.3.2.4 PEW 的 DTFS 反变换

PEW 的 DTFS 系数内插后，可得到内插后的 PEW 的傅立叶级数表示：

$$s(n, m) = \sum_{k=1}^{\left\lfloor \frac{P(n)}{2} \right\rfloor} \left[A_k(n) \cos\left(\frac{2\pi km}{P(n)}\right) + B_k(n) \sin\left(\frac{2\pi km}{P(n)}\right) \right], \quad 0 \leq m < P(n), 0 \leq n \leq N-1 \quad (4-18)$$

由于每帧截得的信号不会刚好是完整的整数个周期波形，为了避免相继帧的合成信号之间发生相位突变，在进行 PEW 的 DTFS 反变换的同时，我专门设定一个标记因子 η ，进行如下操作：

1. 每帧开始重建前都将 η 初始化为零，在开始内插后每计算一个信号样点 $s(n, m)$ ， η 累加一次。

2. 每帧内插时，当 η 已等于 L_f 而 n 或 m 还没有取到最大值时，无论此时的 n 或 m 取何值，均要停止内插，视作该帧的 PEW 信号已被重建完毕；当 n 或 m 已取到各自的最大值而 η 还没有达到 L_f 值时，取当前帧的 $\{A_2, B_2, P(n_2)\}$ 为参数继续重建样点，直到 $\eta = L_f$ 。

4.3.2.5 LP 合成滤波器

在译码后得到 LSF 系数。将相邻帧的 LSF 系数线性内插后再转换回 LPA 系数 $\{\alpha_i\}_{i=1,2,\dots,p}$ ，可组成 LP 合成滤波器。LP 合成滤波器的系统函数为 $H(z)$ ，

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p \alpha_i z^{-i}}$$

如式 (4-18) 表示的 PEW，通过 LP 合成滤波器后，被恢复成精确近似原浊音信号的合成浊音信号。

4.4 改进算法的仿真结果

在仿真中，取消浊音门限 uv_th 为 0.5；16kHz 抽样率时浊音信号基音周期的可能整数值为 40~240，所以 T_{\min} 和 T_{\max} 分别取值为 40 和 240；基音精细搜索的要求精度取为 $\frac{1}{8}$ ^[42]；LP 分析滤波器取 16 阶。

浊音语音的基音周期在 16kHz 时的可能整数值为 40~240，分配 8 个比特对其量化，每帧传送一次。归一化的 PEW 的傅立叶级数系数 $\{A_k^i\}$ 和 $\{B_k^i\}$ 的维数和基音周期有关，分配 7 个比特对它们进行量化，量化后所得序号每帧传送一次。分配 40 比特采用分裂矢量量化算法对 16 阶的 LSF 参数进行量化，即将 LSF 参数分裂为四、四、四、四的四组矢量，每组分配 10 个比特进行矢量量化。量化的 LSF 参数也是每帧传送一次。功率量化时，由于信号的对数功率比信号功率本身更富有感性意义，所以功率首先被转换到对数域，再用 4 个比特的差分标量量化器量化。

清音编码仍旧采用与原 PWI 编码器相同的 CELP 编码算法，使用了一个固定的码书，没采用自适应码书。

用本文的改进算法分别对元音“a”和一段语音—梦中水乡—“春天的黄昏，请你陪我到梦中的水乡”进行编码，得出的仿真结果如下：

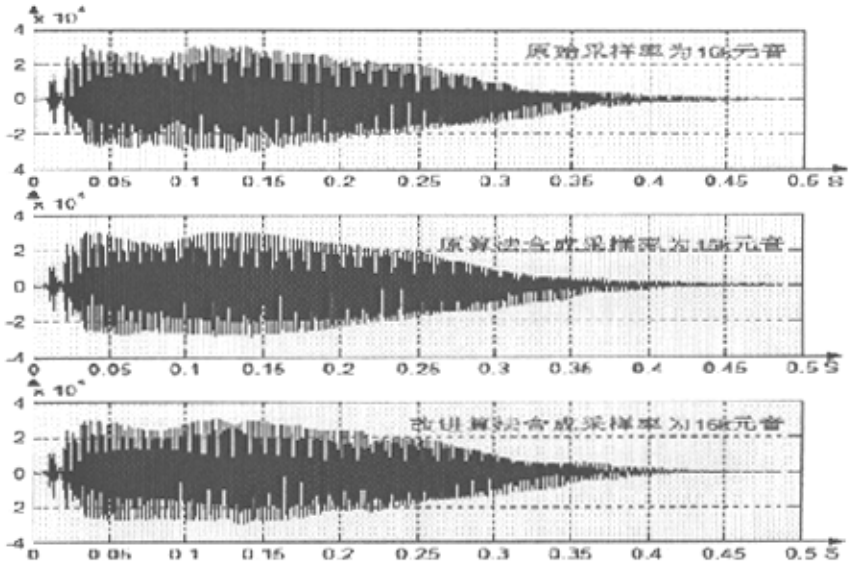


图 4-3 元音“a”的波形比较

Fig.4-3 Waveform Comparison of Vocal “a”

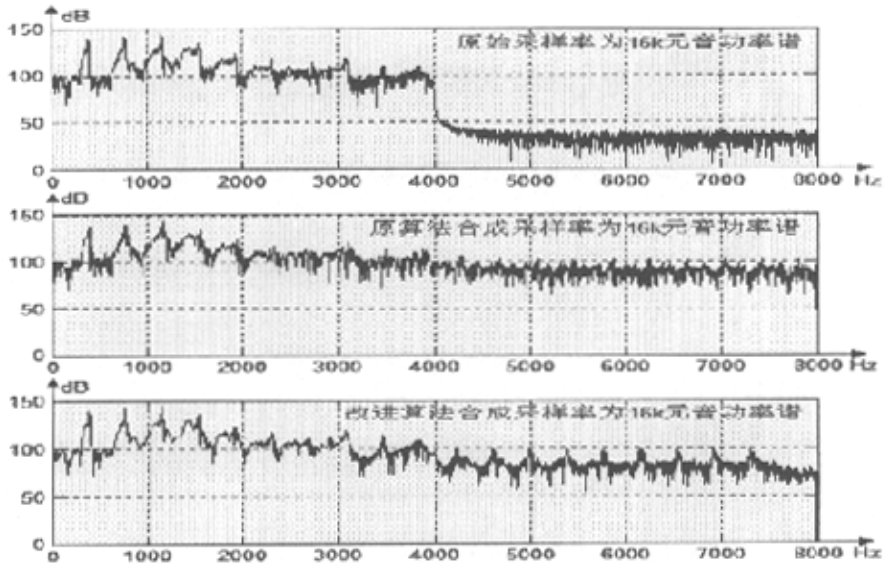


图 4-4 元音“a”的功率谱比较

Fig.4-4 Power Spectrum Comparison of Vocal “a”

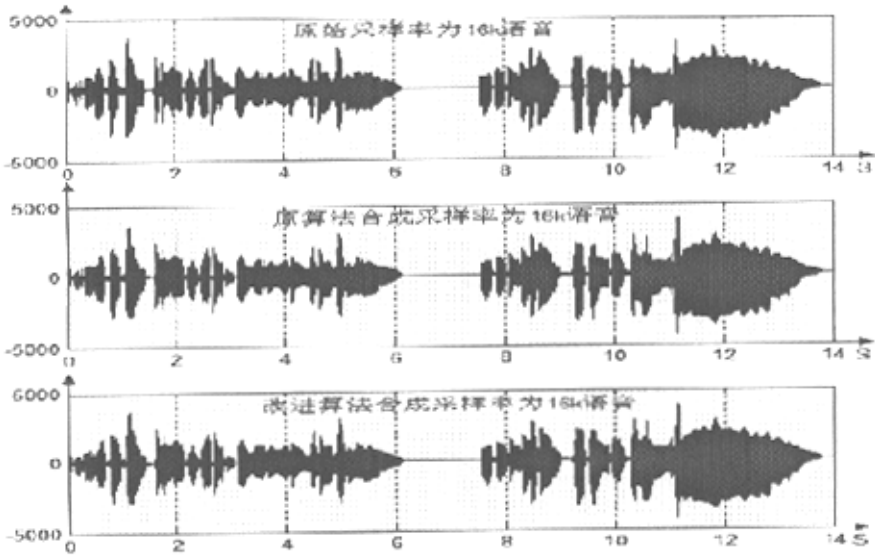


图 4-5 语音—梦里水乡的波形比较

Fig.4-5 Waveform Comparison of Speech-梦里水乡

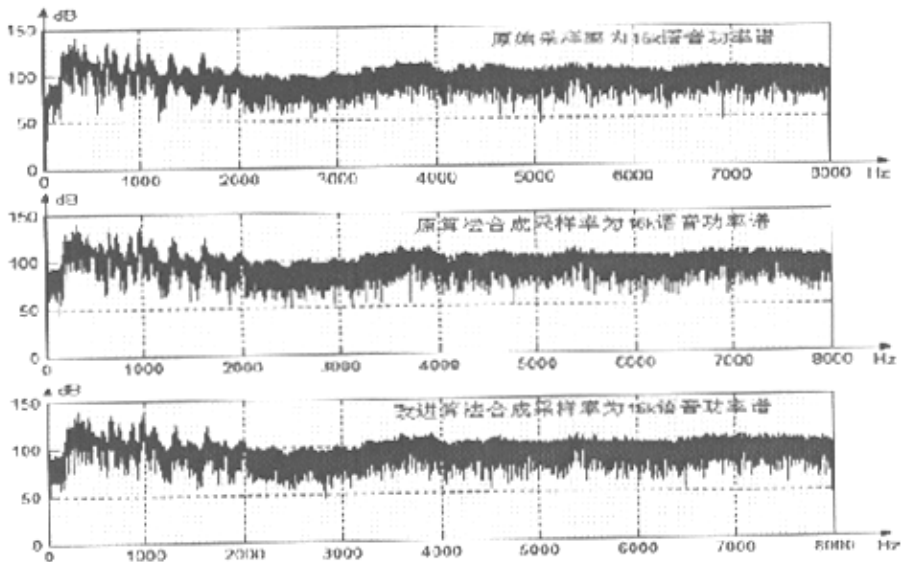


图 4-6 语音—梦里水乡的功率谱比较

Fig.4-6 Power Spectrum Comparison of Speech-梦里水乡

与原语音的接近度测试结果如表 4-1 所示：

表 4-1 合成语音的接近度测试

Table4-1 Aural Test of Synthetic Speech in Approximation

参与测试人数	合成语音相对原语音的接近度	
	元音“a”	梦里水乡语音
30 (人)	91%	80%

经过广泛的非正式试听，元音“a”和梦里水乡这段语音均被认为是声音清晰自然，试听时听者可完全放松，不需要注意力集中即可听清。其中，元音“a”的合成语音被认为与原音非常接近；而梦里水乡的合成语音的接近度则稍微下降，主要是由于清浊音过渡处的编码模式转换仍旧没有被处理好而引起的听觉失真。

由上面的波形图比较、功率谱图比较和输出合成语音的听力测试可得出结论：改进算法的合成语音质量自然度较高，语音质量非常接近于原语音信号。该改进算法是可行的。

4.5 改进算法的仿真分析

目前的语音编码算法研究表明：保持语音信号周期性的正确程度对于语音的感性质量非常重要。故而与原来的 PWI 算法相比，改进算法增加了基音精细检测步骤，使得每帧估计的基音周期更为接近于原浊音信号的真实基音周期，迎合浊音编码信号对基音周期的强烈依赖性。

同时，为避免相继帧的合成信号之间发生相位突变或是基音突变，从而在整体语音中造成变调、毛刺等感觉的情况出现，改进算法在波形内插前对波形进行了对齐；在信号合成时，采用了标记因子以保证合成信号的总样点数与原始浊音信号保持一致，即保证了相继合成浊音帧的相位连续。

改进 PWI 算法所获得的合成语音信号能够高度保持原语音信号的自然度和可懂度，语音质量明显比原始的 PWI 算法有较大的改善。但是由 PWI 算法的基本原理所局限，改进 PWI 算法仍旧需要配合清音编码器共同工作，使得清浊编码模式转换处附近的整体输出语音有听觉失真存在。针对这一点，可在 PWI 算法中追加清浊编码的合成处理：用 PWI 方法重建信号的最后相位 $\varphi(t)$ 对应于原残差信号中的某一时刻。该时刻一旦确定，就可以替换掉原语音信号的缓冲区，以便原

语音信号中的这个时刻对应于帧分界点，使得清音语音算法准确地开始于 PWI 编码的结束处，保证了重建语音信号的连续性^[22]。同时，随着语音编码技术的发展和新技术的出现，更多更好的编码技术可应用到 PWI 算法中去，有助于 PWI 算法进一步降低数码率的同时拥有很高的语音感觉质量。

4.6 本章小结

本章主要介绍了改进 PWI 算法的基本原理和实现流程，并对其仿真结果作出分析，指出改进的模块及由此而得到的结果，证明了 PWI 算法在语音压缩编码领域有着巨大的发展潜力。

结论

目前来说, 16kb/s 和 8kb/s 的语音编码技术已经标准化和产品化, 而 4~8kb/s 的语音编码技术已有区域性标准, 其国际化标准的制订和使用产品的研发正成为众多支持编码技术研究的商家的竞争焦点。而重要的研究焦点已逐步转向更低的码率, 尤其是以 MBE、STC、PWI 和 MELP 为基础的低速率编码器。

本文详细讲述了 WI 语音编码算法的原理, 并对 WI 语音编码的基本算法-PWI 算法, 结合现有的编码新技术, 提出了改进性的算法流程, 并进行了基于 C 代码的仿真。仿真结果证明了改进的 PWI 语音编码算法的有效性。改进算法在浊音语音的波形形状准确度和周期性准确度上均明显优于原来的 PWI 算法, 充分表明其在语音压缩编码方面仍旧有着巨大的发展潜力。若能与清音编码器结合, 并处理好清浊转换的过渡问题, PWI 算法必能成为具有巨大发展潜力的混合编码器。

语音信号处理是语言学和数字信号处理两个学科相结合的产物, 必然会随着这些学科的发展而发展, 波形内插语音编码算法也不例外。根据现在的语音编码技术的研究情况, 可以预料到它的发展方向主要在以下几个方面:

1. 简化现有算法: 现有的波形内插语音编码算法, 虽然处理得到的语音信号效果很好, 但是以算法复杂、实现速度较低为代价的。在不降低现有算法性能的前提下, 算法应尽量被简化, 提高运算速度, 增强算法的实用性。

2. 改进现有算法的语音质量: 波形内插语音编码算法由于十分地依赖基音周期, 有可能会在合成浊音信号中引入过强的周期性而引起失真。所以集中研究基音周期的统计分布, 并确定它的最佳内插函数, 是十分必要的。

3. 进一步降低数码率: 在保持相同语音质量的前提下, 尽可能地降低编码速率是每个语音压缩算法的目标。基于现有的量化技术及针对 PWI 算法中主要参数的变维特性, 变维矢量量化算法和 PWI 算法的结合是优先考虑的研究内容之一。

4. 寻找最佳的语音激励模型: 在波形内插语音编码算法中, PWI 算法仍旧需要清浊判决, 需要和别的清音编码算法结合才能组成完整的编码系统。那么合适的激励模型对于提高 PWI 编码器性能有着举足轻重的作用。

随着语音压缩编码技术的改进和新技术的出现, 波形内插语音编码将朝着高性能、低复杂度、实用化的方向发展, 从而更有力地推动低码率语音编码的发展。

参考文献

- [1] CCITT Fascicle III.3 Red Book, Recommendation G.711: Pulse Code Modulation of Voice Frequency, 1984
- [2] CCITT Fascicle III. 3 Red Book, Recommendation G.721: 32 kbit/s Adaptive Differential Pulse Code Modulation, 1984
- [3] CCITT Recommendation G.726.40-, 32-, 24-, and 16kb/s Adaptive Differential Pulse Code Modulation.
- [4] CCITT Report COM XVIII-R (26) C, Part C.2. Draft recommendation G72x: 7 kHz Audio Coding within 64kb/s, 1986
- [5] T. E. Tremain. The government standard linear predictive coding algorithm: LPC-10. *Speech Technol.*, Apr.1982:pp.40-49
- [6] Hellwig K.; Vary P.; Massaloux D.; Petit J.P.; Galand C. and Rosso M. Speech codec for the European mobile radio system. *Global Telecommunications Conference, 1989, and Exhibition. 'Communications Technology for the 1990s and Beyond'. GLOBECOM '89, IEEE, 27-30 Nov. 1989, vol.2, Pages: 1065 - 1069*
- [7] Fenichel R. Proposed Federal Standard 1016. National Communication System Office of Technology and Standards, Washington, DC 20305-2010, 1990
- [8] J. H. Chen, et al. A Low-Delay CELP Coder for the CCITT 16kb/s Speech Coding Standard. *IEEE Journal on Selected Areas in Communications*, 1992, 10(5), 830-849
- [9] Gerson I.A. and Jasiuk M.A. Vector sum excited linear prediction (VSELP) speech coding at 8 kbps. *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90, 1990 International Conference on, 3-6 April 1990, vol.1, Pages: 461 - 464*
- [10] Usai P.; Cosier G.; Pascal D.; Sotscheck J. and Kappelan M. Subjective performance evaluation of the GSM half-rate coding algorithm (with voice signals). *Comm.1995 IEEE International Conference, June 1995, Volume 1, Pages: 65-71*
- [11] Miki S.; Mano K.; Moriya Y.; Oguchi K. and Ohmuro H. A pitch synchronous innovation CELP (PSI-CELP) coder for 2-4 kbit/s. *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94, 1994 IEEE International Conference on, 19-22 April 1994, Volume II, Pages: II/113 - II/116*
- [12] ITU-T Recommendation G.729. Coding of Speech at 8kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP), 1996
- [13] ITU-T Recommendation G.723.1. Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, 1996

- [14] DeJaco A.; Gardner W.; Jacobs P. and Chong Lee. Qcelp: The North American Cdma Digital Cellular Variable Rate Speech Coding Standard. *Speech Coding for Telecommunications*, 1993. Proceedings, Pages: 5-6
- [15] 林奕琳, 李巧玲等. AMR-WB 语音编码算法及仿真. *计算机工程与应用*. 2003 年第 29 期第 39 卷:67-69
- [16] 李迅. 第三代移动通信系统中自适应多速率语音编码技术研究. 北京理工大学博士学位论文, 2002:32-53
- [17] Griffin D.W. *The Multi-band Excitation Vocoder*. Ph.D Thesis of Massachusetts Institute of Technology, 1987
- [18] McAulary R J and Quatieri T F. *Speech Analysis/Synthesis Based on a Sinusoidal Representation*. *IEEE Transaction on Acoustic Speech Signal Processing*, 1986, 34: 744-754
- [19] McCree A.V. and Barnwell III T.P. *A New Mixed Excitation of LPC Vocoder*. *Proc. IEEE ICASSP*, 1991:593-596
- [20] Kleijn W.B. *Encoding Speech Using Prototype Waveform*. *IEEE Transaction on Speech and Audio Processing*, 1993, 1(4):386-399
- [21] Kleijn W.B. and Haagen J.A. *Speech Coder Based on Decomposition of Characteristic Waveforms*. *Proc. IEEE ICASSP*, 1995:508-511
- [22] 鲍长春. *低比特率数字语音编码基础*. 北京:北京工业大学出版社, 2001: 218-257
- [23] 史学晶, 赵淑清. 一种浊音语音原型波形内插算法. *电声技术*, 2003(12):4-6
- [24] 鲍长春. 低比特率语音编码的若干问题研究. 西安电子科技大学博士后研究工作报告, 1997
- [25] 鲍长春. 一种有效的典型波形块内插方法. *电子科学学刊*, 1999, 21(6): 743-751
- [26] 江灏, 崔惠娟, 唐昆. 一种高质量的 2kb/s 语言编码算法 MWI. *清华大学学报*, 1998, 38(3):67-71
- [27] Ge Gao, Ching P.C. and Tan Lee. *A new approach to generating Pitch Cycle Waveform for Waveform Interpolation codec*. *Microprocessors and Microsystems of Elsevier Science*, 2002, 25(9-10):421-426
- [28] 张海. 基于 WI 的低速率语音编码算法研究与实现. 北京工业大学硕士学位论文, 2001:19-53
- [29] Kleijn W.B. and Paliwal K.K. *Speech Coding and Synthesis*. Elsevier Science B.V., 1995
- [30] 王田, 崔慧娟, 冯重熙. 改进型 LPC 声码器语音编码算法. *通信学报*, 1996,

17(5):1-8

[31] 徐金标,杜利民. 基音同步特征波形内插语音编码算法. 声学学报第 25 卷第 6 期:499-503

[32] 鲍长春,樊昌信. 基于归一化自相关函数的基音检测算法. 通信学报, 1998, 19(10):27-31

[33] 胡洪涛. 低速率语音编码的研究-2400 比特/秒帧内波形插值语音编码算法. 中国科学院声学研究所博士学位论文,2001:17-31

[34] Tanaka Y and Kimura H. Low-bit-rate Speech Coding Using a Two-dimensional Transform of Residual Signal and Waveform Interpolation. Proc. IEEE ICASSP, 1994: I173-I176

[35] 卓力,鲍长春. 一种高效、透明的线谱频率参数矢量量化器. 信号处理, 1999, 15(增刊):154-158

[36] 鲍长春,樊昌信. 线谱频率参数的分裂矢量量化. 电子科学学刊, 1998, 20(4): 508-514

[37] Kim M.Y., Ha N.K. and Kim S.R. Linked Split-Vector Quantization of LPC Parameters. Proc. IEEE ICASSP, 1995:741-744

[38] 鲍长春. 低码率声码器中残差谱的变维矢量量化. 电子学报, 1999, 27(1): 12-16

[39] Das A and Gersho A. Variable Dimension Vector Quantization of Speech Spectra for Low-rate Vocoders. IEEE Proc. Of the Data Compression Conference, 1994: 420-429

[40] Kleijn W.B., Yair Shoham, Deep Sen and Roar Hagen. A Low-Complexity Waveform Interpolation Coder. Proc. IEEE ICASSP, 1996:212-215

[41] 张军,肖自美,韦岗. 1.8~2.4kbps 多带激励语音编码器. 声学学报, 2002, 27(5): 398-404

[42] 杨行峻,迟惠生等. 语音信号数字处理. 北京:电子工业出版社,1995:276-278

攻读学位期间发表的学术论文

在学期间已发表（包括已接受待发表）的论文，以及已投稿、或已成文打算投稿、或拟成文投稿的论文情况（只填写与学位论文内容相关的部分）：

序号	作者（全体作者，按顺序排列）	题目	发表或投稿刊物名称、级别	发表的卷期、年月、页码	相当于学位论文的哪一部分（章、节）	被索引收录情况
01	林奕琳 李巧玲 李江源 韦岗	AMR-WB 语音编码算法及仿真	计算机工程与应用 中文核心期刊	2003 年 第 29 期 第 39 卷	第一章 第一节	
02	李巧玲 韦岗	两种自适应多码率语音编码算法仿真及性能比较	2003 年度华南理工大学电子与信息学科研究生学术研讨会	2003 年 P342-P348	第一章 第一节	

致谢

本论文自始至终是在导师韦岗教授的悉心指导和无微不至的关怀下完成的。韦岗教授不仅为我们提供了良好的科研环境，而且还在学习、生活等各个方面关心我们、培养我们，尤其是他那渊博的学识、严谨的治学态度、敏锐的洞察力以及分析问题的独到见解对我的研究生生活和今后的工作都起到了极大的促进和指引作用，使我受益终身，为此我衷心地向韦岗教授表示崇高的敬意和诚挚的谢意！

衷心感谢电信学院贺前华老师、林耀荣老师、陈芳炯老师、季飞老师、胡斌杰老师以及所有授课老师的谆谆教诲，感谢李伟刚书记、刘育辉书记、朱一成老师、聂文斐老师的关心帮助！正是他们的教导和关心才使我能顺利地完成论文，并在研究生阶段获得了很大的提高。

在这里，我特别感谢余华师兄、单志龙师兄，他们在我的三年研究生时间里，无论是学习上，或是生活上均无私地提供了帮助和指导，使我获益良多。同时，我也感谢 2001 级的同学林奕琳、宁更新、刘阳、杨志维、周蓓、方敏群，感谢他们三年来的朝夕相伴、互学互进，使我度过了一段珍贵的难忘时光，收获了深厚的友情。此外，我还要感谢各个实验室众多的师兄师姐、师弟师妹们的大力支持，尤其是秦伟师兄、黄德宏师弟、戚莹师弟、董浩亮师弟、宋雅师弟、曹燕师妹、李云洁师妹、以及蔡艺师妹。对以上的各位，我表示由衷的感谢和真诚的祝福！

最后，我要感谢家人对我的鼓励和培养，自始至终他们是我坚强而温暖的后盾，一直在关心我、帮助我，给我以勇气和信心去克服各种困难。借此机会，我要对他们表示我最衷心的感谢和呈上最美好的祝愿。

本文的工作得到了国家自然科学基金的资助，在此表示感谢。

李巧玲

2004 年 5 月