

基于 G. 729 的固定码书搜索的改进性研究

张艳萍¹ 朱忠进¹ 王文龙¹

摘要

在详细阐述 G. 729 协议的 CS-ACELP 算法工作原理的基础上,分析了该算法各功能模块的算法复杂度,针对其总体复杂度较高,对硬件实现条件要求较高的问题,对复杂度相对较高的固定码书搜索部分采用减少脉冲搜索位置和补充搜索相结合的方式进行了改进,加快了码书的搜索速度,降低了算法复杂度.通过对改进后的算法进行实验仿真,验证了其可行性和有效性.

关键词

固定码书搜索;共轭代数码激励线性预测(CS-ACELP);算法复杂度

中图分类号 TN912

文献标志码 A

0 引言

自 1996 年国际电信联盟 (ITU) 正式公布 G. 729 标准的 CS-ACELP (Conjugate Structure-Algebraic Code Excited Linear Prediction, 共轭代数码激励线性预测) 算法以来,因其在语音合成上的高质量表现而成为当前比较主流的编码方案之一,主要应用于语音网关,以软件形式实现数字化的语音数据收发^[1],然而该算法存在较高的运算复杂度,难以在低配置的硬件系统中实时实现.为降低其算法复杂度,相关研究人员提出了深度优先树^[2]、脉冲预选-替换、集中搜索等方法对算法复杂度较高的固定码书搜索部分提出了改进.本文在对 CS-ACELP 编解码算法原理分析的基础上,针对固定码书的搜索提出了削减脉冲位置搜索与补充搜索相结合的方法,降低了算法复杂度,使其能够满足中等语音质量要求的场合.

1 G. 729 编解码原理

1.1 编码

图 1 是 G. 729 标准的 CS-ACELP 算法语音编码原理.通信线路中的模拟语音信号经过滤波和 8 kHz 采样后,量化成 16 bit 的 PCM (Pulse Code Modulation, 脉冲编码调制) 数字信号进入编码器. CS-ACELP 算法在标准 PCM 或线性 PCM 的语音采样基础上,每 10 ms 生成一个 10 字节长的语音帧.对于每一个语音帧提取 CELP (Code Excited Linear Prediction, 码激励线性预测) 模型参数,这些参数在编码后以每帧 80 位进行信道传输.在解码端接收到的比特流通过激励码书解码成相应的 CELP 参数,再通过短时综合滤波器和长时综合滤波器重建语音帧,最后经过后向滤波进行语音增强^[3].

在预处理模块中对输入信号进行高通滤波和定标.预处理后的信号在所有的后续分析中作为输入信号.每 10 ms 一帧作 LP (Linear Prediction, 线性预测) 分析,以计算 LP 滤波器的系数.这些系数转换成线谱对 LSP (Linear Spectrum Pair),同时使用预测型两级矢量量化为 18 bit.激励信号采用合成-分析的搜索程序进行选择,在此过程中原始信号和重构信号之间的误差,按照知觉加权的失真测度使其最小化^[4].

激励参数每个子帧 (5 ms, 40 个样点) 确定一次.已量化和未量化的 LP 滤波系数用于第 2 子帧,而在第 1 子帧使用内插的 LP 系数,根

收稿日期 2010-11-25

资助项目 江苏省自然科学基金 (BK2009410)

作者简介

张艳萍,女,博士,教授,主要从事水声通信的研究. zypgjc@163.com

¹ 南京信息工程大学 电子与信息工程学院, 南京, 210044

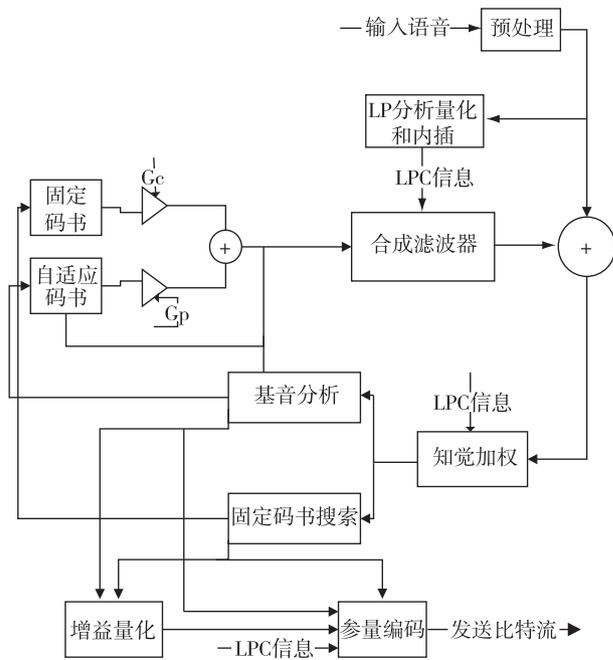


图1 CS-ACELP 算法编码原理

Fig. 1 CS-ACELP coding block diagram

据知觉加权语音信号每 10 ms 帧估计一次开环基音延时, 下面的操作每个子帧都重复进行, 目标信号是由通过加权合成滤波器滤过的 LP 残差计算得到的. 由 LP 残差和激励之间误差滤波修改这些滤波器的初始状态, 等效于从加权语音信号中减去加权合成滤波器的零输入响应. 计算加权合成滤波器的脉冲响应, 然后用目标信号和脉冲响应搜索开环基音延时附近的值作闭环基音分析(即寻找自适应码本的延时和增益), 分数基音延时分辨率 1/3 样点间隔, 在第 1 子帧用 8 bits 基音延时编码, 在第 2 子帧用 5 bits, 从目标信号中减去自适应码书的贡献. 新的目标信号用于固定码书搜索寻找最佳激励. 固定码书激励是使用 17 bit 的代数码书. 自适应码书和固定码书的增益用 7 bit 矢量量化. 最后, 滤波器存储器使用已经确定的激励信号来修正^[5].

1.2 解码

解码器的原理如图 2 所示. 首先, 在接收的比特流中提取参数序号, 对这些参数序号解码后, 得到对应于 10 ms 语音帧的编码参量. 这些参量包括 LSP 系数、2 个分数基音延迟、2 个固定码书矢量、2 组自适应码书和固定码书的增益. 每一个子帧的 LSP 系数要进行内插, 并转换为 LP 滤波器系数. 最后对每 5 ms 子帧完成以下运算^[6]:

1) 激励是由自适应码书矢量和固定码书矢量

按照各自的增益定标后相加而构成的;

2) 合成语音是由激励通过 LP 合成滤波器滤波而重构;

3) 重构语音信号后置处理, 包含了建立在长期和短期合成滤波器基础上的自适应后滤波器, 接着完成高通滤波和定标操作.

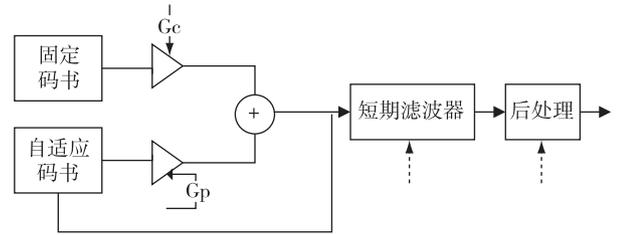


图2 CS-ACELP 算法解码原理

Fig. 2 CS-ACELP decoding block diagram

1.3 算法模块复杂度分析

CS-ACELP 算法主要包含以下几个模块: 语音信号的预处理、信号线性预测(LP)分析及 LP 系数的求解、LPC 系数和线谱对 LSP 的互换、语音信号知觉加权、基音分析、固定码书搜索、自适应码书搜索、自适应码书矢量及固定码书矢量增益求解及其量化、存储器更新、后置滤波处理等^[7].

与现有的一些编码方式进行比较, CS-ACELP 算法具有合成语音质量较高, 算法复杂度相对适中的优点, 因此成为当前比较新的一种语音压缩标准. 但是, 当这种算法应用于硬件条件苛刻, 特别是在要求硬件处理系统低成本化的情况时, 该算法的复杂度则显得较高.

表 1 是一帧实际语音处理过程中, 主要模块的运算量在该算法中所占的运算量比例.

表 1 主要模块运算量所占百分比

Table 1 Computational proportion of the main modules

模块	运算量/%
预处理	2
LP 分析	12
LSP 量化	13
码本搜索	62
后置滤波及后处理	6

由表 1 可以看出, 对于部分功能模块, 例如语音信号的预处理、信号加窗处理和相关性计算、LP 分析及 LP 系数的计算、增益计算等, 优化余地不大, 而对于占用声码器大量时间及资源的码书搜索(包括

自适应码书搜索和固定码书搜索)才是需要优化的重点. 本文主要完成固定码书搜索的分析和改进, 以降低算法的复杂度.

2 固定码书搜索分析及其改进

CS-ACELP 编码算法固定码书采用代数码书结构, 采用正负号脉冲交错设计方式. 每个 40 维(对应每个子帧的 40 个样点)的码矢仅有 4 个非零脉冲, 其脉冲幅度为 -1 或 $+1$, 且呈规律排列. 码矢 $c(n)$ 是由一个 40 维的零矢量在其确定的位置上安置单个脉冲并乘以其对应的幅度构成^[8]:

$$c(n) = S_0\delta(n - m_0) + S_1\delta(n - m_1) + S_2\delta(n - m_2) + S_3\delta(n - m_3), \\ n = 0, \dots, 39. \quad (1)$$

式(1)中: $\delta(0)$ 为单位脉冲; $m_i (i = 0, 1, 2, 3)$ 代表脉冲位置; $S_i (i = 0, 1, 2, 3)$ 代表脉冲幅度或符号. 固定码书矢量的搜索本质在于搜索各子帧的最佳码矢量 $c_k (k$ 为索引标记) 及其增益, 利用输入语音和加权重构语音之间的均方误差 E 最小化准则来搜索固定码书矢量:

$$E = \|\mathbf{x} - g\mathbf{H}\mathbf{c}_k\|^2. \quad (2)$$

式(2)中: \mathbf{x} 为自适应码书搜索目标信号减去自适应码书贡献所得的固定码书搜索目标信号; g 为自适应码书增益; \mathbf{H} 为下三角 Toeplitz 卷积矩阵, 主对角线元素为 $\mathbf{h}(0), \mathbf{h}(1), \dots, \mathbf{h}(39)$ ^[3]. 欲使 E 为最小, 令 $\partial E / \partial g = 0$, 可得最佳码书矢量的增益为

$$g = \frac{\mathbf{x}^T \mathbf{H} \mathbf{c}_k}{\mathbf{c}_k^T \mathbf{H}^T \mathbf{H} \mathbf{c}_k} = \frac{\mathbf{d}^T \mathbf{c}_k}{\mathbf{c}_k^T \boldsymbol{\varphi} \mathbf{c}_k}. \quad (3)$$

式中, \mathbf{d} 为固定码书搜索的目标信号与脉冲响应卷积所得的相关信号:

$$\mathbf{d}(n) = \sum_{i=n}^{39} \mathbf{x}(n) \mathbf{h}(i - n); n = 0, 1, 2, \dots, 39. \quad (4)$$

$\boldsymbol{\varphi}$ 为脉冲响应的自相关矩阵:

$$\varphi(i, j) = \sum_{n=j}^{39} \mathbf{h}(n - i) \mathbf{h}(n - j); \\ i = 0, 1, 2, \dots, 39; j = 0, 1, 2, \dots, 39. \quad (5)$$

将式(3)代入式(2), 使得下式最大的码书矢量即为所要寻求的矢量:

$$\frac{\mathbf{c}_k^2}{E} = \frac{(\mathbf{x}^T \mathbf{H} \mathbf{c}_k)^2}{\mathbf{c}_k^T \mathbf{H}^T \mathbf{H} \mathbf{c}_k} = \frac{(\mathbf{d}^T \mathbf{c}_k)^2}{\mathbf{c}_k^T \boldsymbol{\varphi} \mathbf{c}_k}. \quad (6)$$

满足式(6)的 \mathbf{c}_k 即为最佳码书矢量, 由关系式(3)可得码书增益. 由代数码书的结构特殊性, 式

(6)可转化为

$$\frac{\mathbf{c}_k^2}{E} = \frac{\sum_{i=0}^3 S_i \mathbf{d}(m_i)}{\sum_{i=0}^3 \varphi(m_i, m_i) + 2 \sum_{i=0}^2 \sum_{j=i+1}^3 S_i S_j \varphi(m_i, m_j)}. \quad (7)$$

简化该搜索过程: 选择适当的量化信号 $\mathbf{d}(n)$ 作脉冲幅度预判决, 即设置某脉冲幅度的符号等于 $\mathbf{d}(n)$ 在此位置上的符号, 即 $S_i = \text{sgn}[\mathbf{d}(m_i)]$, m_i 为第 i 个脉冲的位置. 则

$$\mathbf{c}_k = |\mathbf{d}(m_0)| + |\mathbf{d}(m_1)| + |\mathbf{d}(m_2)| + |\mathbf{d}(m_3)|, \\ E = 2 \left(\sum_{i=0}^3 \varphi'(m_i, m_i) + \sum_{i=0}^2 \sum_{j=i+1}^3 \varphi'(m_i, m_j) \right)$$

其中: $\varphi'(m_i, m_i) = \frac{1}{2} \varphi(m_i, m_i); i = 0, 1, 2, \dots, 39$.

$\varphi'(m_i, m_j) = \text{sgn}[\mathbf{d}(m_i)] \text{sgn}[\mathbf{d}(m_j)] \varphi(m_i, m_j); i = 0, 1, 2, \dots, 39; j = i + 1, \dots, 39$ ^[9].

理论上, 完成该算法下的全局搜索, 才可以找到最佳的固定码书矢量, 但是由于其计算复杂度较大, 很难在实际中应用, 不少研究人员提出了一些限制搜索范围的优化方法^[2,8]. 本文在相关研究结论的基础上, 提出了削减脉冲位置全搜索与补充搜索相结合的搜索方法.

最佳码矢量应满足式(6)为最大, 从式(6)可以看出相关信号 $|\mathbf{d}(n)|$ 和 $\frac{\mathbf{c}_k^2}{E}$ 值的大小有着重要的对应关系, 而相关信号在码书搜索前便可以计算. 对于码书脉冲的 40 个可能位置, 本文先只考虑单个脉冲的贡献度, 对于码书结构的每一层轨道用单一脉冲代入式(6), 可得:

$$\frac{\mathbf{c}_k^2}{E} = \frac{\mathbf{d}^2(m_i)}{\sum_{i=0}^3 \mathbf{h}^2(i)}. \quad (8)$$

将所得结果在同一轨道中与其他位置进行比较, 可以发现, 式(8)值的大小关系同其相关信号 $\mathbf{d}(m_i)$ 的大小成正比关系. 对于 4 个轨道中所有脉冲位置的相关信号 $\mathbf{d}(n)$, 按照 $|\mathbf{d}(n)|$ 值的大小关系从大到小降序排列, 重新排列组成新的码书结构, 如表 2 所示, 并以此顺序作为码书搜索的顺序. 码书搜索时, 为减少计算量, 不再进行全局最佳搜索, 而是按照 $|\mathbf{d}(n)|$ 大小, 削去低相关的脉冲位置, 搜索时只从 4 个非零脉冲的 8 或 16 个候选位置选出高相关部分, 具体要求可由处理器的性能和语音质量的要求来决定.

表 2 固定码书调整后的结构

Table 2 The schematic of fixed codebook after adjustment

脉冲	脉冲位置	相关信号
i_0	m_0	$ d(0) > d(1) > \dots > d(7) $
i_1	m_1	$ d(0) > d(1) > \dots > d(7) $
i_2	m_2	$ d(0) > d(1) > \dots > d(7) $
i_3	m_3	$ d(0) > d(1) > \dots > d(15) $

将经过相关计算而重排的代数结构码书按其相关性由高到低取出前 N 个脉冲位置, 本文 N 值取 5, 那么所有的可能搜索组合数为 $5 \times 5 \times 5 \times 10 = 1\,250$, 对每种可能的脉冲组合分别按照式(6)计算 $\frac{c_k^2}{E}$ 值, 使得其满足全局最大. 可见, 完成一次部分脉冲位置的全搜索过程需要 1 250 次运算, 当然, 在不明显影响语音质量的情况下可以通过对脉冲位置的设定来更改搜索次数.

上述的搜索过程只搜索了前 N 个脉冲位置的情况, 舍去了后面的低相关部分, 尽管选中低相关值位置脉冲的可能性较小, 但是仍然有一定的可能性, 因此, 针对低相关值位置脉冲有必要做一个补充的粗略搜索. 对于低相关值位置脉冲, 作两脉冲组合和三脉冲组合的一个粗略搜索计算比较, 两脉冲组合中以 (i_0, i_1) 、 (i_0, i_2) 、 (i_0, i_3) 、 (i_1, i_2) 、 (i_1, i_3) 、 (i_2, i_3) , 三脉冲组合中以 (i_0, i_1, i_2) 、 (i_0, i_1, i_3) 、 (i_0, i_2, i_3) 、 (i_1, i_2, i_3) 来计算式(6)的 $\frac{c_k^2}{E}$. 这种补充的粗略搜索共计 $3^2 \times 3 + 3 \times 6 \times 3 + 3^3 + 3 \times 3^2 \times 6 = 270$ 次.

将完成 $N=5$ 的相关性码书结构的码书搜索与补充的码书搜索相结合, 取使得 $\frac{c_k^2}{E}$ 为最大的码书矢量为最终最佳码书矢量.

由上述的固定码书矢量的搜索算法可以看出, 该算法大大地缩小了搜索空间. G. 729 标准中, 如果按照全码书的全局最佳码书矢量作为最终矢量, 则要经过 $8 \times 8 \times 8 \times 16 = 8\,192$ 次的全局搜索计算, 本文的局部最优的搜索方案只需要 1 500 次左右, 与设立门限值限制第 4 环的搜索计算次数基本相当, 但搜索到的码书矢量质量有所提高.

3 实验结果及性能分析

本仿真设定语音采样频率为 8 kHz, 采样精度为 16 bit. 图 3 为一男声语音原始信号波形, 语音内容为“这里是语音信号处理实验”, 图 5 为其语谱. 将该

段语音信号通过仿真测试后得到合成语音信号, 图 4、6 分别为其波形和语谱.

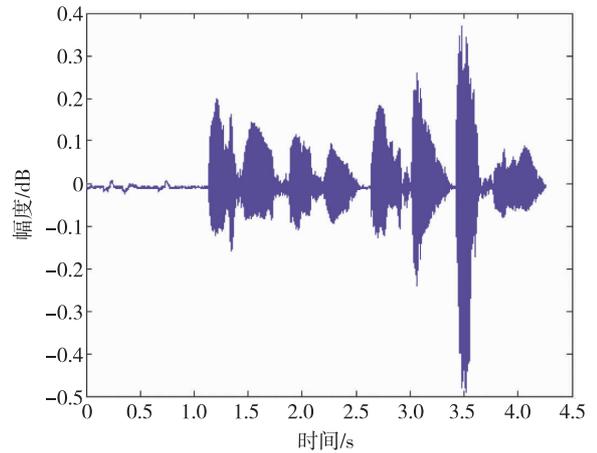


图 3 原始语音波形

Fig. 3 Original speech waveform

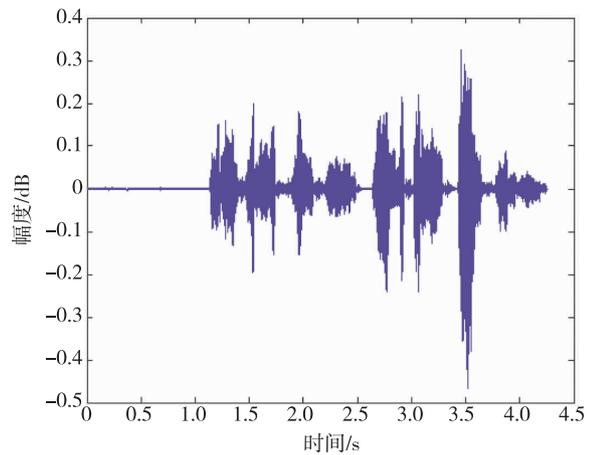


图 4 合成语音波形

Fig. 4 Synthesized speech waveform

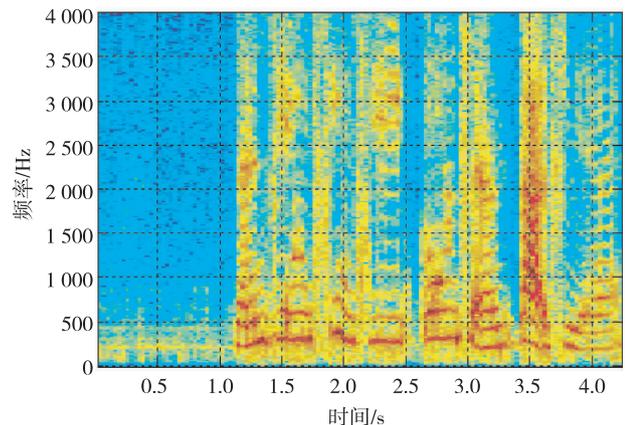


图 5 原始语音语谱

Fig. 5 Original speech spectrogram

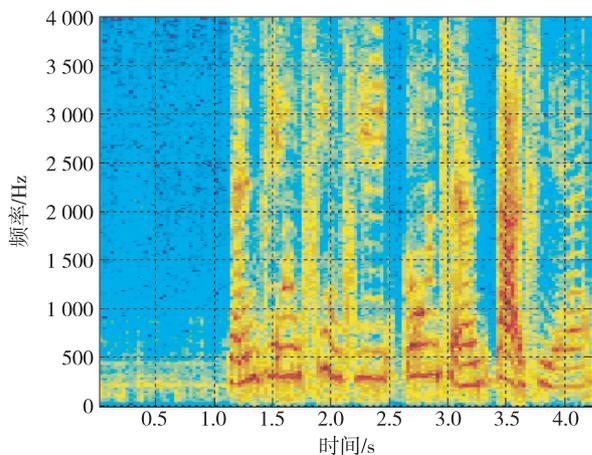


图6 合成语音语谱

Fig. 6 Synthesized speech spectrogram

将原始语音信号的波形、语谱图和重构语音的波形、语谱图进行比对,可以发现,原始语音和重构语音在波形特点、对应特定频率语音能量强弱上有较高的相似度.对重建语音信号进行主观试听,可以发现重建后的语音信号可懂度高,MOS 接近于 3.8.

和集中搜索法相比,本文采取的优化办法在计算次数上也不多.集中搜索法是采用限制第 4 层搜索次数的方法来改进码书的搜索速度,在码书搜索前确定一个阈值,在越过阈值时启动第 4 层的搜索,阈值 thr_3 定义为

$$\text{thr}_3 = av_3 + k(\max_3 - av_3),$$

其中 $\max_3 = \max |d(m_0)| + \max |d(m_1)| + \max |d(m_2)|$ 为前 3 个脉冲贡献的最大绝对相关; $av_3 = \frac{1}{8} \left(\sum_{n=0}^7 d(5n) + \sum_{n=0}^7 d(5n+1) + \sum_{n=0}^7 d(5n+2) \right)$; $k=0.4$. 在此条件下平均每子帧搜索从 512 次降为 90 次,每子帧搜索为 1 440 次,和本文的搜索次数相当.本文对 4 个脉冲位置采取了大概率命中可能的搜索效果更加准确.相关文献中提及的脉冲预选-替换方法,一般是在相关信号较大、脉冲命中率较高的前半区域选取脉冲位置进行置零替换,此时码书的搜索次数必然是降低了,仅为百余次,运算速度虽然能够得到显著的提高,但其对于低命中率的后半可能的脉冲位置采取了完全舍去,必然在很大程度上导致语音质量的下降,更重要的是这种算法为去除搜索上的冗余度,在设计上有较大的难度,在处理器已经能够处理的算法复杂度下,单纯追求速率而增加设计复杂度也是没必要的.

表 3 运算复杂度对比

Table 3 Comparison of computational complexity

算法	运算次数
CS-ACELP(理想状态)	8 192
集中搜索法	1 440
脉冲预选替换	百余次(视脉冲数而定)
本文算法	1 500

4 结语

基于 G. 729 标准的 CS-ACELP 算法可以完成高语音质量要求下的语音编解码过程,因此成为了现今比较主流的语音编码算法之一,但是,由于其运算量较大、复杂度较高,对硬件设备特别是处理器要求较高,很难在低端、低成本硬件要求下实时实现.本文采用了削减脉冲位置的局部全搜索与补充搜索相结合的新固定码书搜索方法,以牺牲较小语音质量为代价,在一定程度上减少了运算量,降低了复杂度,可以满足低配置、低成本处理器芯片的处理要求.

参考文献

References

- [1] 王浔. 基于 CS-ACELP 语音编码算法的研究及 DSP 实现[D]. 哈尔滨: 哈尔滨理工大学电气与电子工程学院, 2007
WANG Xun. Research based on CS-ACELP speech coding algorithm and implementaton on DSP [D]. Harbin: Harbin University of Science and Technology Institute of Electrical and Electronic Engineering, 2007
- [2] 李利, 陈刚. G. 729 中固定码本快速搜索算法研究[J]. 计算机工程与技术, 2010, 31(10): 2209-2215
LI Li, CHEN Gang. Study on fast fixed-codebook search algorithm in G. 729 speech coding [J]. Computer Engineering and Design, 2010, 31(10): 2209-2215
- [3] 王炳锡. 语音编码[M]. 西安: 西安电子科技大学出版社, 2002: 57-122
WANG Bingxi. Speech coding [M]. Xi'an: Xidian University Press, 2002: 57-122
- [4] 薛冰雷, 仇润鹤, 陈蕾. G. 729A 语音编解码器构成描述及其仿真[J]. 东华大学学报: 自然科学版, 2007, 33(1): 112-116
XUE Binglei, QIU Runhe, CHEN Lei. ITU-T G. 729A speech coding algorithm and implementation [J]. Journal of Donghua University: Natural Science Edition, 2007, 33(1): 112-116
- [5] 王帆, 周群彪, 蔡葵, 等. G. 729 语音编解码实现方案的研究及基于 DSP 的实现[J]. 四川大学学报: 自然科学版, 2007, 44(4): 785-790
WANG Fan, ZHOU Qunbiao, CAI Kui, et al. Implementation of G. 729 ADPCM on DSP [J]. Journal of Sichuan U-

- niversity: Natural Science Edition, 2007, 44(4): 785-790
- [6] 鲍长春. 数字语音编码原理[M]. 西安: 西安电子科技大学出版社, 2007: 49-67
BAO Changchun. Digital speech coding principle [M]. Xi'an: Xidian University Press, 2007: 49-67
- [7] 鲍长春. 低比特率数字语音编码基础[M]. 北京: 北京工业大学出版社, 2001: 48-72
BAO Changchun. Low bit rate digital speech coding basis [M]. Beijing: Beijing University of Technology Press, 2001: 48-72
- [8] 王虹, 陈锴. G. 729 声码器的码本搜索改进[J]. 电脑与信息技术, 2005, 13(3): 48-50
WANG Hong, CHEN Kai. A new codebook search method of G. 729 vocoder [J]. Computers and Information Technology, 2005, 13(3): 48-50
- [9] 张德民, 宋铁成, 张天骐, 等. 一种快速的代数码书搜索方法[J]. 重庆邮电大学学报: 自然科学版, 2010, 22(2): 183-187
ZHANG Demin, SONG Tiecheng, ZHANG Tianqi, et al. A fast algebraic codebook search method [J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2010, 22(2): 183-187
- [10] Parkhc, Chioyc, Leedy. Efficient codebook search method for ACELP speech coding[C] // Tsukuba, Ibaraki, Japan: IEEE, 2002: 17-19
- [11] Kyung Jin Byun, Hee Bum Jung, Minsoo Hahn, et al. A FAST ACELP codebook search method [C] // ICSP, 2002: 422-425

Fixed codebook improvement research based on G. 729

ZHANG Yanping¹ ZHU Zhongjin¹ WANG Wenlong¹

1 School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract Beginning with a detailed introduction of CS-ACELP algorithm based on G. 729 protocol, this paper does analysis of the algorithm's computational complexity of each functional modules, which shows high complexity and high demanding for the hardware system. Thus some improvements are made by reducing the impulse locations and adding general search in the low correlation impulse locations, which can accelerate the fixed codebook search and reduce the computational complexity. Experiment simulation is carried out to verify the feasibility and effectiveness of the improved algorithm.

Key words fixed codebook search; CS-ACELP; pcomplexity of the algorithm