

基于奇异值分解的低速率波形内插语音编码算法

王贵平, 鲍长春, 张 鹏

(北京工业大学电子信息与控制工程学院, 北京 100022)

摘 要: 波形内插(WI)语音编码模型作为当今最具潜力的低速率语音编码方案之一, 因其良好的性能, 越来越受到人们的重视. 本文基于一种奇异值分解(SVD)的特征波形分解方法, 利用语音信号的感知特性, 将二维特征波形的幅度谱分成基本矩阵、过渡矩阵和补充矩阵, 并采用了不同的量化方法, 有效地降低了运算复杂度; 另外, 本文根据语音信号时变特性, 将三个矩阵分为三种组合模式表示特征波形幅度谱, 并引入周期因子和能量熵来衡量矩阵周期程度, 解决了奇异值分解后参数难于量化的问题, 提高了编码效率. 主观 A/B 测试表明, 本文提出的 2.4kbps SVDWI 编码器的重建语音质量略好于 21.4kbps MELP 编码器.

关键词: 语音编码; 波形内插; 特征波形; 奇异值分解; 周期因子

中图分类号: TN9121.3 **文献标识码:** A **文章编号:** 0372-2211(2006)01-0135-06

Low Bit Rates Waveform Interpolation Speech Coding Based on Singular Value Decomposition

WANG Gui-ping BAO Chang-chun ZHANG Peng

(School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100022, China)

Abstract As one of the current most potential low bit rates speech coding schemes waveform interpolation model with its high performance has been paid more and more attention. Based on a kind of Singular Value Decomposition method for decomposing Characteristic Waveform in WI speech coding the magnitude spectrum of CW is perceptually divided into basic matrix, transitional matrix and supplemental matrix to be quantized respectively which effectively reduce the computational complexity. In addition, there are three patterns to represent the magnitude spectrum of CW combined by the matrices above, and then Periodic Factor and Energy Entropy are introduced to indicate the periodicity of matrix, which solved the problems for quantization of parameters after SVD and improved the efficiency of coding. Subjective A/B listening tests indicated that the reconstructed speech quality of the 2.4kbps SVDWI codec is a little better than that of 21.4kbps MELP coder.

Key words speech coding waveform interpolation, characteristic waveform, singular value decomposition, periodic factor

1 引言

近年来, 4Kbps 及其以下低速率的高质量语音编码技术已经成为人们研究的焦点. 波形内插语音编码模型作为当今最具潜力的低速率语音编码方案之一, 以其良好的性能, 越来越受到人们的重视. 波形内插语音编码技术最先是瑞典皇家理工学院的 Klein 教授在 1991 年提出的, 并称为典型波形内插 (PWIP Prototype Waveform Interpolation) 语音编码算法^[1, 2]. PWI 以慢渐变的基音周期波形为基础, 从激励信号中每隔 20ms~30ms 提取一个基音周期段作为典型波形编码输出, 在合成端通过线性插值得到其他基音

周期段的激励信号, 合成的激励信号通过线性预测合成滤波器得到重建语音. PWI 在 (3.0~4.0)kbps 可获得好的重建语音质量, 但它仅适用于浊音语音, 清音语音还需用其他方法 (如 CELP) 来编码. 为了进一步降低编码速率, Klein 教授又提出了特征波形 (CW) 分解语音编码算法, 称为特征波形内插 (CWIC Characteristic Waveform Interpolation) 语音编码算法^[3, 4]. 该算法将语音信号表示为在残差域渐变的特征波形, 在浊音段, CW 慢渐变, 在清音段, CW 快渐变. 基于 CW 的这种特性, 用简单的非因果低通滤波器将 CW 分解为慢渐变波形 (SEW Slowly Evolving Waveform) 和快渐变波形 (REW Rapidly Evolving Waveform), 基于主观听觉, 对这两

收稿日期: 2004204225 修回日期: 2005211201

基金项目: 国家自然科学基金 (No 60372063); 北京市自然科学基金 (No 4042009)

种波形分别编码. 该方法避免了传统声码器清浊判决带来的影响, 适用于所有的语音段^[5,6].

WI算法在较低的编码速率能取得较好的语音质量, 但不能实现高质量的语音编码. 在低速率语音编码中, 要得到高质量的合成语音, 就必须充分利用人类听觉系统的特性, 集中有限的编码资源表示语音信号的感知信息. 众所周知, 语音信号主要是由浊音语音和清音语音组成, 而从语音信号产生机理得知, 浊音和清音由于激励源的不同所产生的波形信号也完全不同, 造成了巨大的感知差异. 因此, 将浊音成分与清音成分分开, 并按各自不同的感知特性分别进行量化可以提高编码效率. 目前, 线性相位低通滤波是实现特征波形分解的最简单有效的方法之一, 即沿着时间轴, 将二维特征波形表面分解成 SEW 和 REW. SEW 表示了语音信号类周期成分(如浊音信号), REW 表示了语音信号类噪声成分(如清音信号). 为了进一步分解语音信号, Whisper实验室的 Chong等利用基音同步小波变换 (PSWT2Pitch Synchronous Wavelet Transform)^[7] 分解 CW. 这种方法得到了更多波形渐变的细节, 对不同感知特性的波形表面进行变维矢量量化, 但是编码器总延时比较大. 随后, 该实验室的 Lukasik等提出了将奇异值分解 (SVD) Singular Value Decomposition) 理论用于 CW 的分解^[8]. 此方法对当前 CW 表面进行分解不需要额外的延时, 并且根据编码比特的要求采用不同个数的奇异值可以在解码端重建不同精度的 CW. 因为它涉及到矩阵变换和计算奇异值、奇异矢量而具有非常高的运算复杂度, 并且分解后产生了大量的待量化参数, 需要比较多的比特数进行量化, 因此限制了其在低速率语音编码中的应用.

本文基于文献 [8], 提出了一种新颖的基于奇异值分解的特征波形的分解与量化模型^[9,10], 有效地解决了算法延时长、计算复杂和高比特量化参数的问题. 其中, 第二部分介绍了 WI 语音编码中基于奇异值分解的 CW 分解模型; 第三部分在前面分解模型的基础上介绍了特征波形的量化模型; 第四部分给出了 214kbpsSVD2WI 编码器的方案和性能评价; 第五部分中给出了全文的结论.

2 基于奇异值分解的特征波形分解模型

设 $A \in C_r^m \otimes (r > 0)$, 其中 m 和 n 分别表示矩阵 A 的行数和列数, r 表示矩阵 A 的秩, $A^H A$ 的特征值为 $K_1 \setminus K_2 \setminus \dots \setminus K_r \setminus K_{r+1} = \dots = K_n = 0$ 则称 $R_i = \sqrt{K_i} (i = 1, 2, \dots, n)$ 为 A 的奇异值, 存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V , 使得

$$U^H A V = \begin{bmatrix} 2_r & 0 \\ 0 & 0 \end{bmatrix} \text{ 其中, } 2_r = \text{diag}(R_1, R_2, \dots, R_r), \text{ 而 } R_i (i = 1, 2, \dots, r) \text{ 为 } A \text{ 的非零奇异值, 将上式改写为 } A = U \begin{bmatrix} 2_r & 0 \\ 0 & 0 \end{bmatrix} V^H \text{ 称之为 } A \text{ 的奇异值分解}^{[11]}.$$

奇异值分解方法计算复杂, 目前还没有完善的快速算法, 严重制约了它的实际应用, 本文提出将特征波形矩阵

作分块处理, 仅对其中一小块或两块根据语音感知特性作奇异值分解处理, 从而降低了复杂度. 具体过程如下: 首先, 提取的特征波形经过离散时间傅立叶级数 (DTFS) 表示后, 在频域完成对齐操作, 特征波形的对齐不仅使得各子帧的特征波形保持相位一致, 而且与 SVD 有着密切地联系. 尤其对于浊音语音, 对齐操作使得前后子帧相关性最大, 从而整个特征波形矩阵呈现出非常强的周期性. 对于这种周期较强的波形表面, 奇异值分解后得到一系列渐变波形, 但绝大部分信息和能量都集中在第一个二维渐变波形, 有利于量化编码. 对于清音语音, 各子帧 CW 经过对齐后同样增加了相关性, 也有利于下一步的处理. 实验表明人耳对一般的相位信息不是特别敏感, 因此当特征波形对齐以后, 将各子帧的 DTFS 系数利用极坐标形式去除其 DTFS 系数中的相位信息, 即

$$|CW_k(n)| = \sqrt{\text{DTFS}_k \cdot A_k^2(n) + \text{DTFS}_k \cdot B_k^2(n)} \quad k = 1, 2, \dots, P(n)/2 \quad (1)$$

其中, $P(n)$ 为当前子帧的基音周期. 至此得到了一个 $P(n)$ 行、 $N (= 10)$ 列的矩阵, 表示了二维特征波形的幅度谱信息. 由于语音具有时变特性, $P(n)$ 通常认为在 202120 (样点) 间变化. 为了便于统一计算, 需要将 CW 作等维处理, 令 $P(n)$ 为最大值 120, 构造一个 60 @ 10 的矩阵表示 CW 幅度. 当各子帧的 CW 长度 (即 $P(n)$ 长度) 小于 120 时, 它的频域幅度矩阵需要补零来填满构造矩阵. 实验表明, 这种在频域添零, 使得较短的 CW 伸长从而化变维矢量为定维矢量的方法, 等价于时域上 CW 的带限内插. 然后, 为了降低复杂度, 将 $|CW|$ 构造矩阵分成三个矩阵: 基本矩阵、过渡矩阵和补充矩阵^[10]. 其中, 基本矩阵是由原构造矩阵的前 12 @ 10 组成的矩阵, 表示了 CWs 800Hz 以下的低频谱; 过渡矩阵是由原构造矩阵中间的 15 @ 10 组成的矩阵, 补充矩阵是原构造矩阵的剩余部分, 过渡矩阵和补充矩阵表示了 CWs 800Hz 以上的高频谱. 图 1 给出了 $|CW|$ 构造矩阵分块的过程, 其中 a) 图的 X 轴表示了样点数 (200 样点/帧, 分成 10 个子帧), Y 轴表示谐波个数, Z 轴表示幅度; b)、c) 和 d) 图的 X 轴表示子帧的标号, Y 轴表示各子帧的谐波个数, Z 轴表示幅度.

特征波形幅度矩阵之所以这样分成三个矩阵, 各自采用不同的量化方法, 除了基于降低运算复杂度的考虑, 还利用了 CW 幅度谱的感知特性. 实验表明, 线性预测残差的低频段包含重要的听觉信息, 必须采用比较精细的方法量化, 而高频段可以根据线性预测残差的平坦谱 (在理论上是平坦的) 特性在合成端恢复. 本文中 60 @ 10 的构造矩阵表示了残差信号的频率从 02400Hz 的幅度谱, 而基本矩阵 (12 @ 10) 正是表示了残差信号从 0280Hz 的低频信息, 可以采用比较精细的方法编码; 而构造矩阵后面的部分 (48 @ 10) 表示了残差信号从 8002400Hz 的高频信息, 为了减少对稀疏矩阵的大量运算而进一步细化, 结合谐波

个数时变的特点和编码比特的限制, 将这部分又分为过渡矩阵和补充矩阵, 下一部分将介绍三个矩阵的量化方法.

3 基于奇异值分解的 WI 编码器参数的量化

本文前面构造了一个定维矩阵 (60 @ 10) 表示特征波形幅度谱并将其分解成三个子矩阵: 基本矩阵、过渡矩阵和补充矩阵. 由于特征波形谐波数是时变的, 决定了三种模式: 基本矩阵单独表示特征波形. 基本矩阵和过渡矩阵表示特征波形、三个矩阵共同表示特征波形、下面将对这三种模式下的三个矩阵进行量化, 首先介绍三个矩阵本身不同的感知特性以及三种模式的统计分布. 图 2 是一段分别由基本矩阵、过渡矩阵和补充矩阵单独重建的语音信号.

由图 2 和实际重建语音的听觉效果可以得出: 基本矩阵包含了特征波形幅度谱基本的低频信息和主要的能量; 过渡矩阵既是幅度谱从低频成分到高频成分的过渡, 又是从基本矩阵到补充矩阵表示语音信号能量的过渡; 而补充矩阵表示谐波数较大时幅度谱的高频信息, 是对基本矩阵和过渡矩阵的高频成分的补充, 通常由于高频的衰落, 包含的能量较小.

另外, 实验表明^[10], 模式 2 和模式 3 所占的比例要比模式 1 大得多, 即特征波形用两个矩阵和三个矩阵表示的概率要比用一个矩阵表示的概率大, 这就决定了过渡矩阵和补充矩阵在编码中不容忽视.

综上所述, 基本矩阵包含了丰富的低频感知信息, 需要比较精确的量化; 过渡矩阵和补充矩阵表示了大部分的高频信息, 并且大部分的语音信号都需要过渡矩阵、补充矩阵与基本矩阵共同表示, 因此就决定了三个矩阵不同的量化方法.

3.1.1 基本矩阵参数的量化

3.1.1.1 模式 1 基本矩阵单独表示 CW 幅度谱

基本矩阵经过奇异值分解后, 如果仅量化第一个奇异值及其相应的左右奇异矢量, 重建的语音周期性非常强.

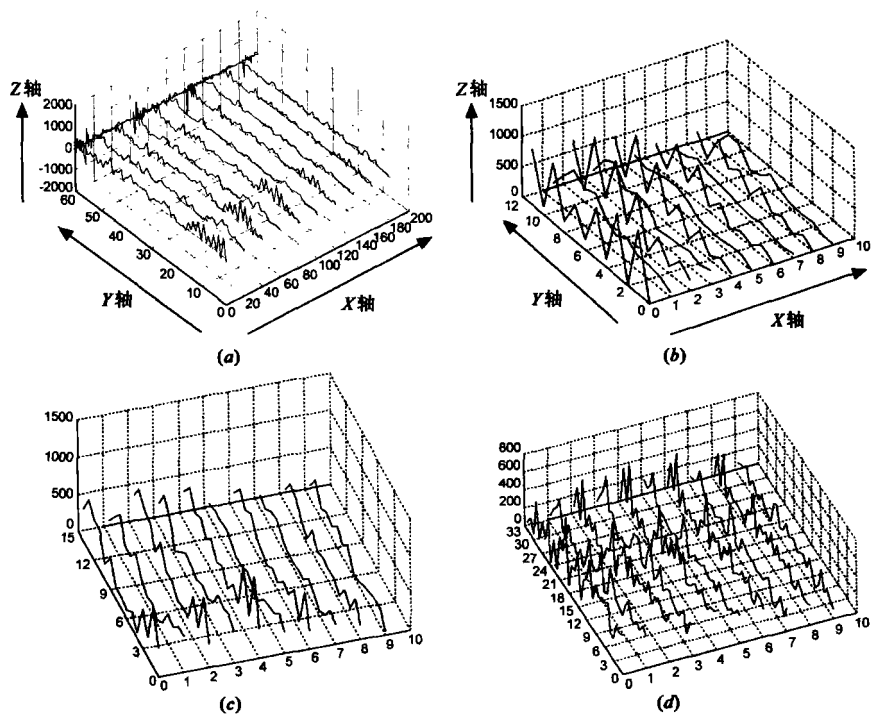


图 1 特征波形矩阵分块的过程
(a) 一段残差信号及其 CW 的幅度谱; (b) 12 @ 10 维的基本矩阵
(c) 15 @ 10 维的过渡矩阵; (d) 33 @ 10 维的补充矩阵

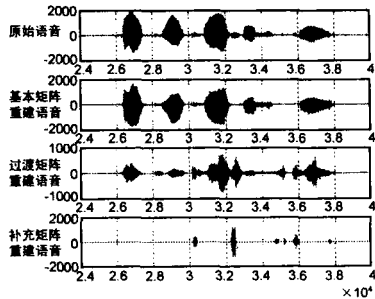


图 2 三个矩阵单独重建语音信号

对于浊音段效果很好, 但对于过渡段和清音段就会因周期性过强而产生蜂鸣声; 而另一方面由于存在冗余信息和编码比特的限制, 没有必要且不可能将 U₂ 和 V 全部量化. 因此, 在一定编码速率下, 需要量化的奇异值个数取决于原 CW 幅度谱的周期程度, 本文引入了周期因子和能量熵修正因子来描述原 CW 幅度谱的周期程度. 在说明周期因子定义之前, 先来分析奇异值大小分布与原 CW 幅度谱的周期性的关系.

图 3 给出了通常情况下奇异值在清浊语音段下的分布情况. 其中, 两条实线和四条虚线分别表示了两帧清音和四帧浊音 SVD 后的奇异值大小分布. 可以发现, 对于浊音语音, 奇异值大小衰减非常快, 尤其在第一个奇异值和第二个奇异值之间发生了陡降, 以至于浊音的四条曲线趋于重叠; 而对于清音语音, 奇异值大小分布整体呈线性缓慢下降. 实验表明, 当 CW 周期较强时, 第一个奇异值要远大于其他的奇异值, 其他的奇异值大小接近零值. 尤其当 CW 表现强周期时, 仅有第一个奇异值非零, 其他的奇异值都为零; 而当 CW 周期较弱时, 奇异值大小缓

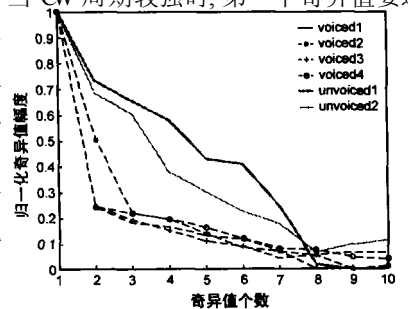


图 3 奇异值在清浊语音段的分布

慢下降. 尤其当 CW 各子帧互不相关, 呈现噪声特性时, 理论上各奇异值大小相等. 综上所述, CW 周期强弱与奇异值大小分布情况密切相关, 因此, 周期因子的定义也离不开奇异值大小的分布.

周期因子是从矩阵能量的角度引入的, 与奇异值大小分布有着密切关系. 若令原始 CW 幅度谱为矩阵 A , 经过奇异值分解后得到 $A=U2V^T$, 其中 $2 = \text{diag}\{R_1, R_2, \dots, R_{10}\}$, 定义 A 的近似矩阵为 $A = u_1 R_1 v_1^T$, 则原始矩阵 A 的能量为其 Frobenius 范数的平方

$$+A+^2_F = +uE^T v^2_F = +E^2_F = \sum_{i=1}^P R_i^2 \quad (2)$$

近似矩阵的能量为其 Frobenius 范数的平方, 即

$$+A+^2_F = +u_1 R_1 v_1^T +^2_F = R_1^2 \quad (3)$$

周期因子定义为

$$D_{peri} = \frac{+A+^2_F}{+A+^2_F} = R_1^2 / (\sum_{i=1}^P R_i^2), [1/p \ 1] \quad p = \text{rank}(A) \quad (4)$$

将周期因子归一化得

$$D_{peri} = \frac{D_{peri} - 1/p}{1 - 1/p} = \frac{1 - P D_{peri}}{1 - P}, [0 \ 1] \quad (5)$$

周期因子利用奇异值分布规律描述了近似矩阵从能量上逼近原始矩阵的程度, 当原始矩阵 A 为周期矩阵, 奇异值分解后 $R_1 > R_2 = R_3 = \dots = R_{10} = 0$ 则有 $A = A D_{peri} = 1$; 而当原始矩阵 A 为噪声矩阵时, 奇异值分解后理论上 $R_1 = R_2 = R_3 = \dots = R_{10} > 0$ $A = 10 @A$ $D_{peri} = 0$ 因此, 周期因子是对原始矩阵周期程度的描述.

由于谐波个数的时变, 尤其在过渡帧, 过渡矩阵中各列的幅值或能量并不一定相等, 因此这里采用各列的能量熵来修正周期因子^[12]. 设各列的能量分别为 E_1, E_2, \dots, E_n , 其中

$$E_i = \frac{1}{M} \sum_{j=1}^M |CW|_j^2 \quad (6)$$

对应各段的能量占比 e_1, e_2, \dots, e_n

$$e_i = E_i / (\sum_{j=1}^N E_j) \quad (7)$$

则能量熵定义为

$$ch = - \sum_{j=1}^N e_j \log e_j \quad (8)$$

归一化的能量熵为

$$chr = - \sum_{j=1}^N e_j \log e_j / \log N, [0 \ 1] \quad (9)$$

当各列的幅值越接近, 其熵越接近 1, 否则越小; 能量熵反映了原矩阵各列幅值的接近程度. 但当原矩阵各列接近零值时, 为了防止计算溢出, 则不用修正周期因子. 修正后的周期因子为:

$$D_{peri} = chr @D_{peri} \quad (10)$$

图 4 给出了 100 帧浊音和清音 (包括过渡音) 计算的周期因子修正前后的效果图, 其中横轴给出了帧数, 纵轴

表示周期因子的大小. 图 4 中的 a) 上面带星 (-*.) 的曲线表示在浊音帧计算的周期因子, 下面的带圆圈 (-o.) 曲线表示在过渡帧或清音帧计算的周期因子, 其中清浊判决是由归一化互相关函数基音检测算法的语音对数能量和周期性水平量共同确定的^[11]; 图 4(b) 是采用能量熵对图 4(a) 的周期因子的修正. 图 4(a) 与 (b) 进行比较可以得到: 修正前后, 浊音的周期因子变化不大, 过渡或清音的周期因子整体变小, 且在 0 到 019 间跳变; 修正前清浊音的周期因子发生混合, 难以区分. 修正后, 基本可以通过设定阈值分开周期强弱.

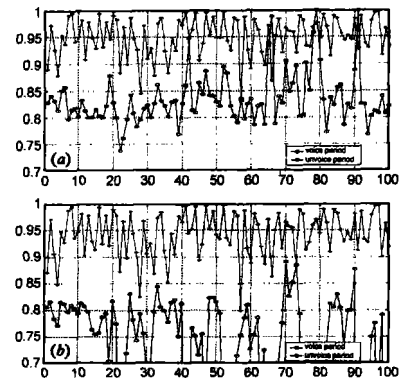


图 4 周期因子修正前后比较
(a)修正前的浊音和清音周期
(b)修正后的浊音和清音周期

基本矩阵, 当周期因子 $D_{peri} > 0190$ 时, 仅量化 u_1, R_1, v_1 , 其中 u_1 采用分裂式矢量量化方法 (SVQ), 分成两个 6 维的子矢量, 分别用 6 比特进行量化. v_1 先从 10 维下采样到 5 维, R_1 作为一个增益值与 v_1 共同用 5 比特矢量量化, 最后, 周期因子作为周期程度的判断标志用 1 比特量化.

当周期因子 $D_{peri} < 0190$ 时, 需要量化 $u_1, u_2; v_1, v_2; R_1, R_2$, 其中 u_1 分成两个 6 维的子矢量, 分别用 5 比特进行量化; u_2 也分成两个 6 维的子矢量, 分别用 4 比特进行量化. v_1, v_2 各下采样到 5 维, 与 R_1, R_2 分别用 5 比特和 4 比特量化. 周期因子量化同前.

31112 模式 2 和 3 基本矩阵表示 CW 的低频谱

基本矩阵包含了 CW 的低频信息, 实验表明, 当仅量化 u_1, R_1, v_1 时, 重建语音的主观听觉没有明显差别. 在模式 2 和 3 中, 因为只对 u_1, R_1, v_1 量化不需要计算周期因子. 其中 u_1 分成两个 6 维的子矢量, 分别用 6 比特和 5 比特进行量化; v_1 下采样到 5 维, 与 R_1 分别用 5 比特量化.

312 过渡矩阵参数的量化

过渡矩阵有着承上启下的作用, 包含了一定的能量和感知信息, 对重建语音也有着重要的影响. 下面分别介绍在模式 2 和模式 3 中的量化方法:

31211 模式 2 过渡矩阵表示 CW 的高频谱

模式 2 中, 过渡矩阵主要体现了 CW 幅度谱的高频成分, 根据人类听觉特性通常对细节成分采用比较粗糙的表示方法. 当过渡矩阵中的零值较多, 即各子帧谐波个数较小, 矩阵为稀疏矩阵时, 本文采用各列求均值并分成两个 5 维的矢量进行 5 比特量化; 当过渡矩阵中的非零值较多时, 每帧提取两个高频谱, 即分别提取第 5 子帧和第 10 子帧的高频谱, 进行 DCT 拟合逼近高频谱的包络, 将变维的

矢量转化为定维数的
的矢量,再对 DCT 系数
进行矢量化。本文对
DCT 的前 5 维系数分
别进行 5 比特矢量化,
图 5 给出了 DCT 拟
合过渡矩阵高频谱的
示例。其中横轴表示谱
波维数,纵轴表示谱的
幅度,实线表示 CW 幅
度谱,虚线表示 DCT
拟合的高频谱。

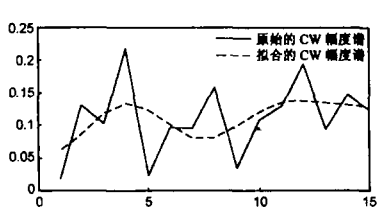


图 5 DCT 拟合过渡矩阵高频谱

31212 模式 3 过渡矩阵表示 CW 的中频谱

模式 3 中,过渡矩阵表现了 CW 的 80Hz 至 800Hz 的幅度谱成分,同时也是基本矩阵和补充矩阵能量的衔接。类似模式 2 当过渡矩阵中的零值较多时,本文对过渡矩阵各列分成上下两部分,分别计算平均值,再下采样到 5 维矢量进行 4 比特矢量化;当过渡矩阵中的非零值较多,即包含的信息比较丰富时,本文对过渡矩阵进行了奇异值分解,得到整个矩阵的主要能量,编码端仅量化 u_i, R_i, v_i 。对 15 维的 u_i 采用每 3 维做平均处理,得到 5 维子矢量并进行 4 比特矢量化,粗糙地表示了幅度谱随着频率的变化,而对 10 维的 v 下采样到 5 维子矢量进行 4 比特矢量化,表示了幅度谱沿着时间轴的渐变。

313 补充矩阵参数的量化

补充矩阵仅在模式 3 中需要处理,它反映了 CW 的高频谱,即语音信号中的随机成分,由于人耳对语音随机成分的细节不敏感,所以可以对补充矩阵采用比较粗糙的表示形式。另外,理论上残差信号的高频谱趋于平坦谱,因此本文对补充矩阵的各列计算均值,然后下采样到 5 维矢量,用 4 比特进行矢量化。

4 SVD2W I 编码器的方案和性能评价

前面主要介绍了基于奇异值分解的特征波形的分解和量化,现在介绍编码器其他参数的量化方法:LSF 参数采用了预测分裂式矢量量化方法 (PSVQ)^[1] 进行量化,基音周期采用了 7 比特的均匀标量化;归一化 CW 功率的动态范围很宽,于是先转换为对数功率,然后采用插值量化的方法^[16]。图 6 给出了一个完整的 2.4 kbps SVD2W I 语音编码器的模型,并已经用 C 语言在计算机上进行了实现。其量化比特分配如表 1 所示。

表 1 2.4 kbps SVD2W I 语音编码器比特率分配

编码参数	LSF 参数	基音	CW	基本、过渡和补充阵
提取速率 (Hz)	40	40	40	40
比特分配	20	7	7	26
传输速率 (bits/s)	800	280	280	1040
总比特率 (bits/s)	800+ 280+ 280+ 1040= 2400			

在低速率语音编码中,由于相位信息已被忽略,波形无法匹配,所以以信噪比为基础的客观测度不再适用,通常对编码器进行主观评价。本文的 2.4 kbps SVD2W I 属于低速率语音编解码器,所以仍采用了主观语音质量 A/B 测试。

本文对 2.4 kbps SVD2W I 和 2.4 kbps MELP 两个编码器进行了非正式的语音质量主观对比测试 (A/B 测试。表 2 给出了 20 个语音短句的 A/B 测试结果。

表 2 SVD2W I 算法和 MELP 算法语音质量的测试结果比较

认为 2.4 kbps SVD2W I 编码器质量好	认为 2.4 kbps MELP 编码器质量好	认为二者没有差别
39.5%	37.7%	22.8%

5 结论

综上所述,本文基于文献 [8],提出了一种新颖的基于奇异值分解的特征波形的分解与量化模型,比传统的 WI 语音编码器不仅减少了一帧延时,而且提高了特征波形的表示精度,合成端可以根据不同的精度要求重建语音。另外,将特征波形幅度谱分为基本矩阵、过渡矩阵和补充矩阵,降低了编码器的运算复杂度,而且按照不同的感知特性分别编码,提高了量化效率。非正式的 A/B 测试表明, SVD2W I 编码器的质量要好于 2.4 kbps MELP 编码器。

参考文献:

[1] 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 233- 234

[2] Kleijn W B Continuous representation in linear predictive coding [A]. IEEE ICASSP 91 [C]. Toronto 1991 201- 204

[3] Kleijn W B Haagen J W waveform Interpolation for Coding and Synthesis Speech coding and Synthesis [M]. Elsevier Science 1995. 175- 207

[4] Kleijn W B Haagen J Transformation and decomposition of the speech signal for coding [J]. IEEE Signal Processing Letters 1994 1(9): 136- 139.

[5] 张海. 基于 WI 的低速率语音编码算法研究与实现 [D]. 北京: 北京工业大学电子信息与控制工程学院, 2001. 19- 72

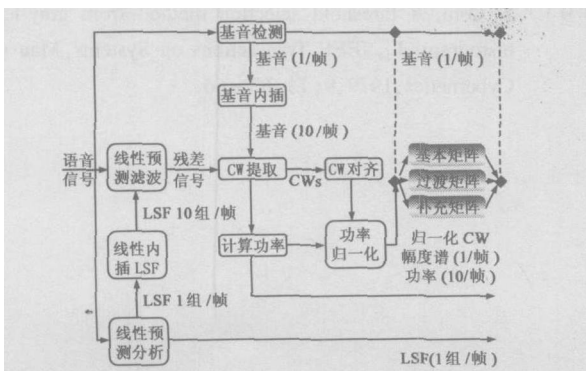
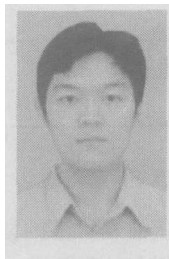


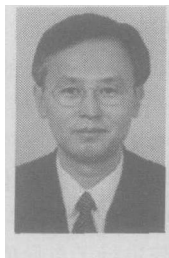
图 6 SVD2W I 编码器整体框图

- [6] 朱娜娜. 2kbps波形内插语音编码算法的研究 [D]. 北京: 北京工业大学电子信息与控制工程学院, 2003. 10 - 70
- [7] Chong NR, Burnett IS, Chicharo JF. A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition [J]. IEEE Transactions on speech and audio processing 2000, 8(3): 345- 348
- [8] Lukasiak J, Burnett IS. Scalable decomposition of speech waveforms [A]. 2002 IEEE Speech Coding Workshop Proceedings [C]. Tsukuba City, Ibaraki, Japan 2002. 135 - 137.
- [9] Guiping Wang, Changchun Bao. Low complexity decomposition for the characteristic waveform of speech signal [A]. ISCSLP2004 [C]. Hong Kong 2004. 145- 149.
- [10] 王贵平. 基于奇异值分解的低速率波形内插语音编码算法的研究 [D]. 北京: 北京工业大学电子信息与控制工程学院, 2005. 48- 78
- [11] 徐仲, 张凯院, 等. 矩阵论简明教程 [M]. 科学出版社. 2002. 90- 120.
- [12] 刘红星, 姜澄宇, 等. 基于矩阵奇异值分解的信号非周期性程度指标 [J]. 南京航空航天大学学报, 2000, 32 (1): 114- 118

作者简介:

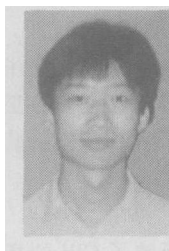


王贵平 男, 1980年出生于北京, 主要研究方向为低比特率数字语音编码, 2005年于北京工业大学获得工学硕士学位, 现就职于北阳电子.



鲍长春 男, 1965年6月出生于内蒙古赤峰市, 博士, 教授、博士生导师, 国际语音通信学会 (ISCA) 会员, 中国电子学会理事, 信号处理学会委员, 5通信学报与5信号处理学报编委, 主要研究领域为数字信号处理与语音编码.

E-mail: chdha@bjut.edu.cn



张鹏 男, 1980年出生于河北保定, 现为北京工业大学硕士研究生, 主要研究方向为低比特率数字语音编码.

(上接第 162 页)

- [3] Ho Ming Chen, Chung A C S, Yu S CH, et al. Multimodal image registration by minimizing Kullback-Leibler distance between expected and observed joint class histograms [A]. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Computer Society Press, 2003. 570- 576
- [4] 储颖, 唐超伦, 纪震, 等. 基于拉伸空间变换算法在 DSA 配准中的应用 [J]. 深圳大学学报理工版, 2005, 22(2): 127- 132
- [5] A Venot, V Leclerc. Automated correction of patient motion and gray values prior to subtraction in digitized angiography [J]. IEEE Transactions on Medical Imaging 1984, 3 (4): 179- 186
- [6] LV Tran, Jack Sklansky. Flexible mask subtraction for digital angiography [J]. IEEE Transactions on Medical Imaging 1992, 11(3): 407- 415
- [7] E H W Meijering, W J Niessen, M A Viergever. Retrospective motion correction in digital subtraction angiography: a review [J]. IEEE Transactions on Medical Imaging 1999, 18(1): 2- 21
- [8] E H W Meijering, K J Zuiderveld, M A Viergever. Image registration for digital subtraction angiography [J]. International Journal of Computer Vision, 1999, 31(2/3): 227- 246
- [9] N Ostu. A threshold selection method from gray level histogram [J]. IEEE Transactions on Systems, Man and Cybernetics 1979, 9(1): 62- 66